

# Agglomerative Ants for Data 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

Clustering is a data mining technique for the analysis of data in various areas such as pattern recognition, image processing, information science, bioinformatics etc. Hierarchical clustering techniques form the clusters based on top-down and bottom-up approaches. Hierarchical agglomerative clustering is a bottom-up clustering method. Ant based clustering methods form clusters by picking and dropping the objects according to surroundings. This paper proposes an agglomerative clustering algorithm, AGG\_ANTS based on ant colonies. AGG\_ANTS clusters the objects by moving ants on the grid and merging their loads according to similarity resulting in bigger clusters. It avoids the calculation of similarity in the surrounding and pick/drop of objects again and again resulting in a more efficient algorithm.

**Keywords:** Clustering, Hierarchical, Agglomerative, Ant colony

## 1. INTRODUCTION

Clustering is the task of assigning a set of objects into clusters so that the objects in the same cluster are more similar to each other than to those in other clusters. The four main classes of clustering algorithms available are partitioning methods, hierarchical methods, density based methods and grid based methods [8]. Hierarchical algorithms build clusters gradually. Hierarchical clustering is further subdivided into agglomerative and divisive. An agglomerative approach begins with each pattern in a distinct cluster, and successively merges clusters together until a stopping criterion is satisfied. A divisive method begins with all patterns in a single cluster and performs splitting until a stopping criterion meets. Most hierarchical clustering algorithms are variants of the single-link and complete link algorithms. These two algorithms differ in the way they characterize the similarity between a pair of clusters. In the single-link method, the distance between two clusters is the minimum of the distances between all pairs of patterns drawn from the two clusters (one pattern from the first cluster, the other from the second). In the complete-link algorithm, the distance between two clusters is the maximum of all pair wise distances between patterns in the two clusters. In either case, two clusters merge to form a larger cluster based on minimum distance criteria.

Nowadays ant based algorithms [1] are becoming popular. Ant based Clustering is inspired by brood sorting in ant colonies. In these algorithms, artificial ants make heaps of objects just as real ants make heaps of dead bodies. Ants move in the workspace randomly. If an unloaded ant encounters an object, it picks up the object if it is surrounded by dissimilar objects. It continues walking and drops the object where it finds objects similar to the object it is carrying. The pick and drop actions result in clusters of similar objects without any initial knowledge of number of clusters.

Here an agglomerative clustering method is proposed based on ant colonies. The ants will agglomerate if there will be some similarity in their load, otherwise keep on walking on the workspace. This paper is organized as follows: Section 2 explains the hierarchical agglomerative clustering algorithm. Ant based clustering research is discussed in section 3. Proposed method is introduced in the section 4. Section 5 and 6 discusses the experimental results and conclusion.

## 2. HIERARCHICAL AGGLOMERATIVE CLUSTERING

Hierarchical agglomerative clustering is a bottom-up clustering method. It starts with every single object in a single cluster. Then, in each successive iteration, it agglomerates the closest pair of clusters by satisfying some similarity criteria, until all of the data is in one cluster. The procedure is as follows:

1. Initially each item  $x_1, x_2, x_3 \dots x_n$  is in its own cluster  $C_1, C_2, C_3 \dots C_n$ .
2. Repeat until there is only one cluster left.
3. Merge the nearest clusters  $C_i$  and  $C_j$ . The concept of nearest clusters may be based on different linkage variations as follows:

- Single-linkage:  
$$d(C_i, C_j) = \min_{x_i \in C_i, x_j \in C_j} d(x_i, x_j)$$
- Complete-linkage:  
$$d(C_i, C_j) = \max_{x_i \in C_i, x_j \in C_j} d(x_i, x_j)$$
- Average-linkage:  
$$d(C_i, C_j) = \frac{\sum_{x_i \in C_i, x_j \in C_j} d(x_i, x_j)}{|C_i||C_j|}$$

## 3. ANT BASED CLUSTERING

Firstly Deneubourg et. al [2] proposed the basic ant model for clustering. He focused on clustering objects by using a group of real-world robots. In his 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.

Gutowitz [4] improved this model by giving the ants the capacity to sense the complexity of their neighborhood. The ants would not try to pick or drop anything in areas with low complexity. These complexity-seeking ants were able to avoid actions that did not contribute to the clustering process, performing their task more efficiently.

Lumer and Faieta [3] extended the basic model with a concept of continuous similarity measure. Each ant can only sense the similarity of the objects in their immediate region. The probability of picking or dropping an object was a function of the measure of similarity.

Monmarche [5] proposed an algorithm where several objects are allowed to be on the same cell of the workspace grid. Each cell with one or more objects together corresponds to a cluster. Each ant is also capable of carrying more than one object at a time.

Ngenkaew [6] proposed two multiple pheromone concepts in ant based clustering with ant nest algorithm and with ant memory algorithm. Artificial Trailing pheromone and Foraging pheromone help ants to decide which direction to go or where to pick up or drop the item of food.

Handl [9] presented a comparative study of the performance of ant-based clustering against traditional methods. Study realized that ant clustering is very fast algorithm for high-dimensionality data and its interesting features like no assumption on the shape of the clusters, automatic identification of the number of the clusters and robust behavior to the effects of outliers.

Brown et. al [10] proposed hierarchical clustering by making a movement zone around each cluster. Worker ants move only in this zone. Then the resulting clusters are merged by the queen ant through a connection..

Li et. al [11] proposed a new for calculating the distance based on reachability paths between two objects, called distance with connection.

Jiang et. al [7] redefined the behavior of ant and colony similarity. Groups of objects are activated one by one to be artificial ants. This method does not search for the sample objects as ants are the objects themselves. It can reduce the number of iterations.

Feghi et. al [12] presented a new Adaptive Ant-based Clustering Algorithm for clustering data sets. The algorithm takes into account the properties of aggregation pheromone and perception of the environment.

#### **4. AGGLOMERATIVE ANTS CLUSTERING METHOD**

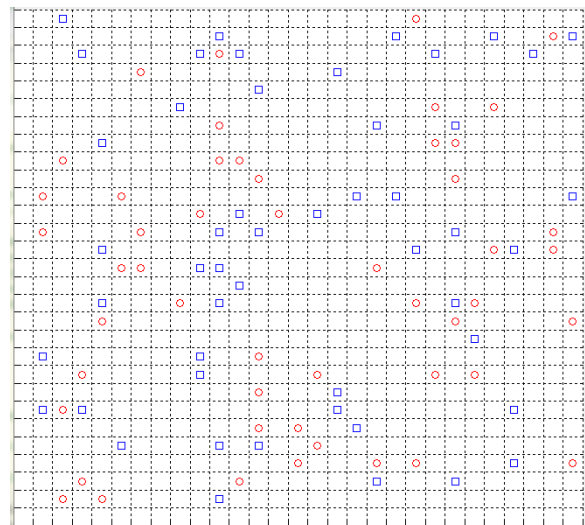
A new method for clustering using ants based on hierarchical agglomerative clustering is proposed. Every object is modeled as an ant. The workspace is considered as 2-D grid. Ants will move randomly on the workspace. Now the ants will form the clusters based on the similarity among the objects. If an ant meets another ant while walking, one ant will hand over its object to other one if their load is similar, so, one ant will hold both the objects resulting in a cluster of two objects. The clusters will go on merging until no other merge is possible. The handing over of cluster to another ant on a way according to similarity criteria is referred as agglomeration. If an ant agglomerates with some other ant, it will stop working. We are modeling the agglomerated ants as dead ants. So number of ants will decrease with successive iterations. Finally the number of ants that survive will give us the number of clusters automatically.

#### **Algorithm AGG\_ANTS:**

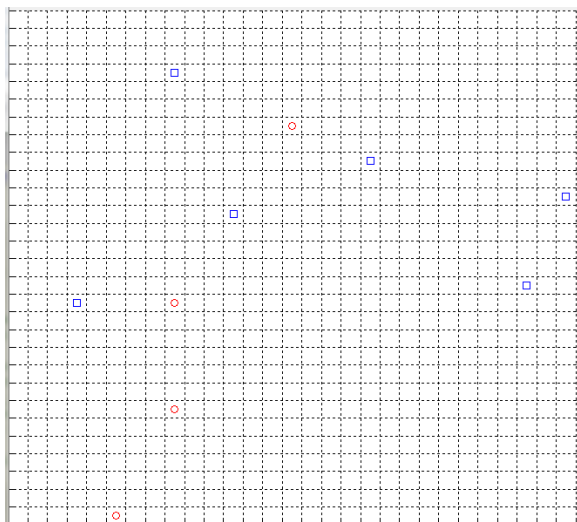
1. Initialize the ants on the grid locations with objects. Every ant is forming a cluster.
2. Ants move randomly.
3. If two ants meet, they agglomerate according to similarity criteria.
4. The agglomeration continues until no more merge is possible.

### **5. EXPERIMENTAL RESULTS**

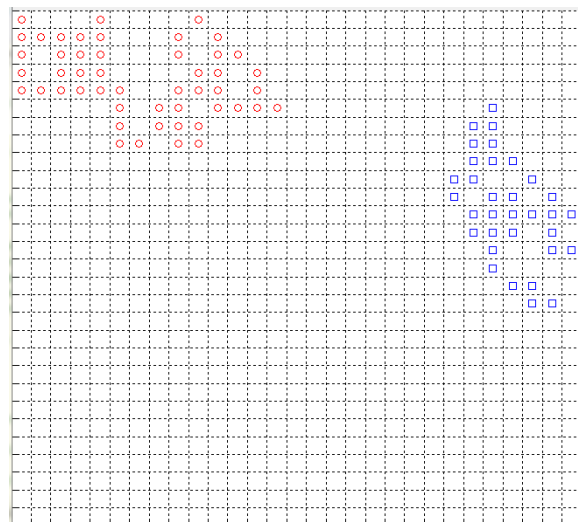
Ants have been simulated in MATLAB. The simulations have shown very good and fast results. The objects have been distributed randomly on the grid. Every object is modeled as an ant. So, initially the number of ants is same as the number of objects. As algorithm runs, the number of ants starts decreasing. All the experiments are done on 30×30 grid. In one experiment, we have taken 100 objects of two different types, 50 objects of 'o' type and 50 objects of '□' type. Total 100 ants start with clustering process and finally two ants survive. During the last iteration, the survived ants drop their load in neighborhood of their current location. The random distribution of objects, intermediate results and final clustering results are shown in Fig. 1, Fig. 2 and Fig. 3 respectively.



**Fig. 1: Initial random distribution of objects**



**Fig. 2: Intermediate results (current iteration = 218)**



**Fig. 3: Final clusters dropped by survived ants**

One experiment is done on a small set of numeric values. We took the set {10, 4, 6, 23, 78, 65, 45, 5, 66, 72, 99, 18, 22, 77, 39, 56, 80, 52, 88, 40} of numeric values of cardinality 20. So experiment starts with 20 ants. Objects less than mean of the values are considered similar to each other and dissimilar with those that are greater than mean. Results are shown in Table I.

## 6. CONCLUSION

A method AGG\_ANTs for data clustering has been proposed and implemented. The method is based on the hierarchical agglomerative clustering algorithm. Ants carry objects and merge their load until they get the final clustering solutions. Experiments have shown good and fast results. The method is also applied to a set of numeric values to check the applicability of AGG\_ANTs to numeric data.

**Table 1: Results studied during clustering process**

No of Iteration	No of Ants	No of clusters	Clusters
1	20	20	$C_1=10; C_2=4; C_3=6; C_4=23; C_5=78; C_6=65; C_7=45; C_8=5; C_9=66; C_{10}=72; C_{11}=99; C_{12}=18; C_{13}=22; C_{14}=77; C_{15}=39; C_{16}=56; C_{17}=80; C_{18}=52; C_{19}=88; C_{20}=40;$
500	3	3	$C_1=(10, 30, 39, 4, 6)$ $C_4=(23, 22, 45, 5, 18)$ $C_{11}=(99, 88, 72, 77, 66, 65, 56, 80, 78, 52)$
1000	2	2	$C_4=(23, 22, 45, 5, 18, 10, 30, 39, 4, 6)$ $C_{11}=(99, 88, 72, 77, 66, 65, 56, 80, 78, 52)$

## 7. REFERENCES

- [1] Marco Dorigo, V. Maniezzo and A. Colorni, 'Ant System: Optimization by a colony of cooperating agents', IEEE Transactions on Systems, Man, and Cybernetics-Part B, Vol. 26, No.1, pp. 1-13, 1996.
- [2] J.L. Deneubourg, S. Gross, N. R. Franks, A. Sendova-Franks, C. Detrain and L. Chretien, 'The dynamics of collective sorting: Robot-like ants and ant-like robots', Simulation of Adaptive Behavior: From Animals to Animats, pp. 356-363, 1991.
- [3] E.D. Lumer and B. Faieta, 'Diversity and adaptation in populations of clustering ants', Proc. of the Third International Conference on The Simulation of Adaptive Behavior: From Animals to Animats 3, pp. 449-508. MIT Press, 1994.
- [4] H. Gutowitz, 'Complexity-seeking ants', In Proc. of the Third European Conference on Artificial Life, 1993.
- [5] N. Monmarche, M. Slimane and G. Venturini, 'On improving clustering in numerical databases with artificial ants', Advances in Artificial Life, pp. 626-635, 1999.
- [6] W. Ngenkaew, Satoshi Ono and S. Nakayama, 'Pheromone- Based Concept in Ant Clustering', In Proc. of 3<sup>rd</sup> International conf. on Intelligent System and Knowledge Engineering, pp. 308-312, 2008.
- [7] Hong Jiang, Qingsong Yu and Yu Gong, 'An Improved Ant Colony Clustering Algorithm', 3<sup>rd</sup> International Conference on Biomedical Engineering and Informatics, IEEE 978-1-4244-6498-2/10, pp. 2368-2372, 2010.
- [8] A.K. Jain, M.N.Murty and P.J.Flynn. 'Data Clustering: A Review', ACM Computing Surveys, Vol. 31, No. 3, pp. 264-323, September 1999.
- [9] J. Handl, J. Knowles and M. Dorigo, 'Ant Based clustering: a comparative study of its relative performance with respect to k-means, average link and I-D-som', Technical Report No. TR/IRIDIA/2003-24, Universite Libre de Bruxelles, Belgium, 2003.
- [10] J.B. Brown and M. Huber, 'Pseudo-hierarchical ant-based clustering using Automatic Boundary Formation and a Heterogeneous Agent Hierarchy to Improve Ant-Based Clustering Performance', 2010 IEEE international conference on SMC, pp. 2016-2024, 2010.
- [11] Shanfei Li, Wei Huang, Kewei Yang, Yuejin Tan, 'An Improved Ant-Colony Clustering Algorithm Based On the Innovational Distance Calculation Formula', 2010 Third International Conference on Knowledge Discovery and Data Mining, pp. 342-346, 2010.
- [12] I. El-Feghi, M. Errateeb, M. Ahmadi and M. A. Sid-Ahmed, 'An Adaptive Ant-Based Clustering Algorithm with Improved Environment Perception' International Conference on Systems, Man, and Cybernetics, San Antonio, TX, USA - October 2009 published in IEEE 978-1-4244-2794-9/09, pp. 1431-1438, 2009.