International Engineering Journal For Research & Development
Vol.6
Issue 6

# CLASSIFICATION & CLUSTER MINING WITH PARALLEL FUZZY ACO

**Dr.K.Sankar**
Anna University, India
**sankarkrish78@gmail.com**

---

## ABSTRACT

The main purpose of data mining is to extract the required knowledge from data. Data mining is considered as an inter-disciplinary field. Some of the data mining tasks include classification, regression, clustering, and dependence modeling and so on. The task discussed above is solved by using many of the data mining algorithms. The first and foremost step is to design the data mining algorithm for which task the algorithm is to be solved.

The parallel Fuzzy ACO technique is applied to the data mining tasks of both classification and clustering. The optimization algorithm to partition or creating clusters of data is more efficient. Depending on the nature of data when the level of uncertainty composition is higher in a given data set, parallel fuzzy ACO for classification and clustering gets complicated. Research works have been extensively conducted on MAX-MIN ANT for clustering and classification. The following chapter focuses on MAX-MIN ANT fuzzy for clustering and classification.The primary purpose of clustering algorithm is to segregate the data into self-similar clusters in such a way that the inter-cluster distance is maximized and the intra-cluster distance is minimized. Clustering algorithms can be divided into many types ranging from Hard, Fuzzy, Possibilistic to Probabilistic. Our work concentrates on clustering with parallel fuzzy ACO due to the reduced performance achieved using the sequential fuzzy ACO technique.

***Keywords: Ant Colony Optimization (ACO), MMAF (Max-Min Ant Fuzzy).***

## INTRODUCTION

Methods like ant-oriented clustering methods have profound great work in performing clustering techniques. Ant-based clustering techniques have received great influence from the researchers over the recent years. As the methods involve, to a greater extent, in exploring the analysis of data particularly suitable to perform exploratory data analysis, and also because many researchers have concentrated on improving performance of the overall system, stability of work involved, convergence and speed involved in real time applications.

The algorithms developed using the techniques of ant colony optimization have unique features such as distributed, flexible, and robustness in nature. Each individual ant is considered as a simple agent with limited power, but it exhibits the property of co-operation and stochastic iterative behavior. Clustering with parallel Fuzzy ant colony optimization (ACO) algorithm simulate the behavior of ants and at the same time interact with Fuzzy clusters. In the proposed algorithm, the property of ants is simulated to obtain good cluster centers.

### Behavior of Ant Colonies

In a colony which includes certain types of social insects such as ants, bees, wasps and termites the behavior of each insect is performed in such a way that the ants perform their tasks independently when compared to the other members of the colony. Subsequently the works performed by different insects are grouped with each other in such a way that the colony is capable of solving complicated tasks by way of cooperation.

It also necessitates that certain key problems related to the ant such as choosing and picking up materials, finding and accumulating the food which consists of involving complicated sophisticated planning are clarified by insect colonies without any necessity of supervisor or centralized controller. This cooperative behavior which develops from a group of social insects has been called "swarm intelligence" [1].

## Fuzzy Ant Parallel System

The proposed algorithm clustering with parallel fuzzy ACO is a work extended from the basic model of swarm intelligence techniques. The ant's co-ordinate in such a way so as to move cluster centers in feature space in a parallel manner in search of optimal cluster centers. The feature values are initialized between 0 and 1. Each ant is designated to a particular feature of a cluster in the partition. The feature space, cluster or the partition for a specific ant never gets changed as shown in Figure 1.1.

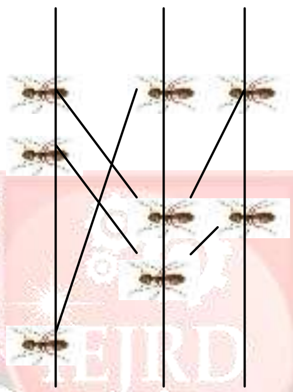

**Figure 1.1 Diagrammatical Representation of Parallel View**

The partition value is evaluated using the parallel reformulated criteria as shown in Equation (1.1). If the distance is denoted by $Dist_{jk}$ and the optimization is denoted by O then

$$\left.\begin{array}{l} O_{jk} = 0 \text{ if } Dist_{jk} \\ O_{jk} = 1 \text{ otherwise} \end{array}\right\}$$

$$P1(\gamma) = \Sigma \min (Dist_{1k}, Dist_{2k}, \ldots . Dist_{nk}) \quad (1.1)$$

$k=1,2,3,\ldots,m$

where $O_{jk}$ is the optimization factor for $P1(\gamma)$ in Equation (1.1).

The working of the Fuzzy Ant Parallel system is given as below: The work starts with the partitioning process as shown in Figure 1.2. Comparison of partition is performed between the current and previous partitions. If the current parallel partition is better than any of the previous sequential partitions in the ant's memory, then the ant remembers this parallel partition else the ant goes back to a better partition and this iteration continues.

In this way the ants makes sure that they do not remember the bad partition. There are two orders for the random movement of the ant. The positive parallel direction is when the ant is moving on the order from 0 to 1 and the negative parallel direction is when the ant is moving on the order from 1 to 0. The ants stop to perform after a fixed number of epochs.

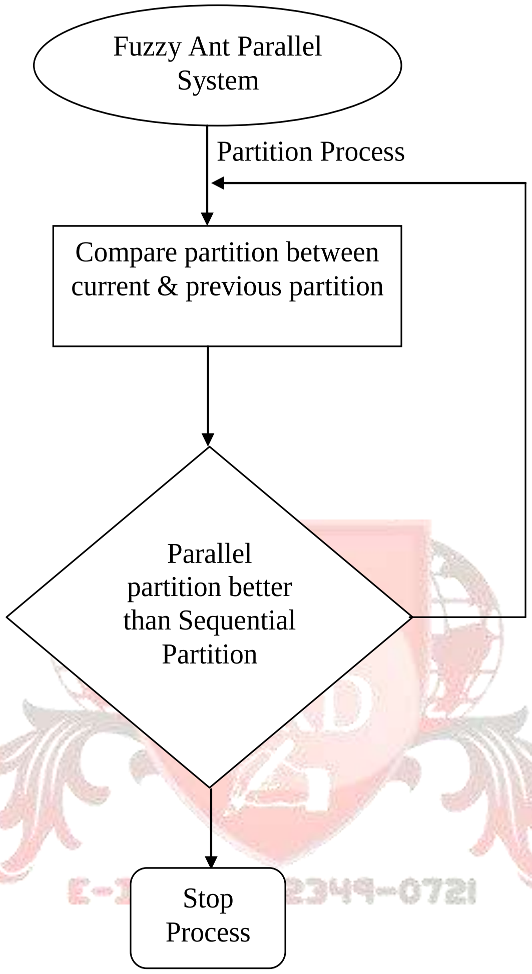


**Figure 1.2 Fuzzy Ant Parallel Partition process**

| Initialization | Begin |
| --- | --- |
| Select the feature values<br>in range between 0 and 1<br>Let the direction be<br>positive and negative,<br><br>positive denotes the movement<br>of ant in the feature space from 0 to 1<br>negative denotes the movement<br>of ant in the feature space from 1 to 0.<br><br>Let $P_r$ denote the probability<br>of the ant being at rest<br>Let $P_{cont}$ denotes the probability<br>of the ant being moving or continuing<br>Let Dist $j_k(y_k,\alpha_i)$ be the distance | i = 1<br>repeat [$P_r$]<br>Probability of ant being in rest<br>i = i + 1<br>until i = r<br>repeat [$P_{cont}$]<br>Probability of ant being continue<br>i = i + 1<br>until cont= r<br>end for<br><br>**Evaluate partition quality**<br>Evaluate the quality of partition for fuzziness and reformulation using the equation given below:<br>$R(\alpha) = \sum \min(Dist_{1k}, Dist_{2k}, \ldots Dist_{nk})$ where k = 1,2,..n<br>**Comparison of partition values**<br>**Let** j=1<br>**Repeat**<br>**If** ($Part_{new}$ better than $Part_{old}$) |

**Figure 1.3 Algorithm for Fuzzy Ant Clustering**

## EXPERIMENTAL EVALUATION

The system implementation of parallel fuzzy ant based parallel clustering algorithm for rule mining used three real data sets obtained from the UCI repository. The data sets were Iris Human Data Set, Wine Recognition Data Set and Glass Identification Data Set. The simulation conducted in MATLAB normalizes the feature values.

The minimum value of a dataset specific feature is mapped to 0 and the maximum value of the feature is mapped to 1. This work initializes the ants with random initial values with random directions. There are two types of directions, Left and Right. The Left direction denotes Negative and Right direction denotes Positive. The Positive (Right) direction indicates the ant is moving in the feature space from 0 to 1. The Negative (Left) direction indicates the ant is moving in the feature space from 1 to 0. Initially the memory is cleared. The ants are initially assigned to a particular feature within a particular cluster of a particular partition. The feature, cluster or the partition assigned to them is never changed by the ants.

Three data sets Glass Data Set, Wine Data Set, Iris Data Set were evaluated from a mixture of five Gaussians. The probability distribution across all the data sets is the same but the means and standard deviations of the Gaussian are different. Of the three data sets, two data sets had 350 instances each and the remaining one data set had 750 instances each. Each instance had two attributes. To visualize the Iris data set, the Principal Component Analysis (PCA) algorithm was used to project the data points into a 2D and 3D space.

## CONCLUSION

The clustering approach established in our work Parallel fuzzy ACO for clustering provides a framework for optimization of any intended purpose that can be expressed in terms of cluster centroids. The algorithm presents a high probability of skipping most poor native solutions resulting in an excellent separation of knowledge. The algorithm establishes here involves the amount of clusters be known and has a nominal understanding of parameter selections and results in clusters that are typically enhanced, optimized than those from previous parallel algorithms.

## References

1. Kanade, P.M. "Fuzzy ants as a clustering concept", Theses and Dissertations, University of South Florida, Scholar Commons, 2004.
2. Kanade, P.M. and Hall, L.O. "Fuzzy ants as a clustering concept", 22nd International Conference of the North American Fuzzy Information Processing Society, NAFIPS'03, pp. 227-232, 2003.
3. Kanade, P.M. and Hall, L.O. "Fuzzy ant clustering by Centroid Positioning", University of South Florida, Tampa FL 33620.
4. Abe, S. and Lan, M.S. "Fuzzy rules extraction directly from numerical data for function approximation", IEEE Trans. Syst. Man Cybern., Vol. 25, No. 1, pp. 119-129, 2005.
5. Song Mao, Chenglin Zhao, Zheng Zhou and Yabin Ye, "An improved fuzzy unequal clustering algorithm for wireless sensor network", Communications and Networking in China (CHINACOM), 6th International ICST Conference, pp.245-250, 2011.
6. Rajeswari, R. and Rajesh, R. "A modified ant colony optimization based approach for image edge detection", Image Information Processing (ICIIP), International Conference, pp.1-6, 2011.
7. Dr. K.Sankar Completed ME(CSE) in the year 2006 and Ph.D in the year 2013 in Anna University. He has 17 years of Teaching Experience. He has published 10 papers in International Journal and 25 papers presented in International & National conferences.