International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-3 Issue-3, February 2014

# An Effective Method for Clustering Using an Algorithm Stimulated By the Decision Making Processes of Bee Swarms

**M.Nandhini, K.Mohana Prasad**

***Abstract— Cluster is the task of grouping a set of objects into a different cluster. Objects in a cluster are more similar when compared with those object in the other in the cluster in some sense or the other Applications of clustering in biomedical research include gene expression data analysis, genomic sequence analysis, biomedical document mining and MRI image analysis . In data mining, clustering is used for scalability, interpretability and usability, insensitive to order of input records and high dimensionality. In this paper we proposed Correlative Artificial Colony where we consider more relationship between the employ bees and onlookers and extend the exploitation capacity of the ABC algorithm. The result shows that the proposed algorithm reduces computation time, increases the exploitation ability of ABC and produces a good solution in a limited amount of time.***

***Index Terms— artificial bee colony,clustering,correlative artificial bee colony .***

## I. INTRODUCTION

Data mining is a process of analyzing the data from different prospective and summarizing the result as useful information. Clustering is grouping of similar data into different clusters based on certain criteria in such a way that in some way the cluster members are similar. Clustering concept is used because of its simplicity, pattern detection and unsupervised learning process. There are four major clustering methods. They are Distance based, Hierarchical, Partitioning and Probabilistic. Artificial Bee colony algorithm is based on the behavior of honey bees. Best Food source is found by the movement of Onlookers, Employed Bees and Scout bees.

## II. RELATED ARTICLES

In [1] the clustering is done by using Ant Colony Optimization and k- Harmonic means algorithm and efficiency of both of the algorithms is compared. In ACOC, an artificial ant colony finds the shortest path by generating the pheromone trail following the path of real ant. K-harmonic means (KHM) algorithm minimizes the harmonic average from all points in N to all centers in the dataset, it uses the Harmonic Averages to find the distances from each data point. The standard deviation of KHM is larger than the ACOC.The ACOC has larger processing time compared to KHM.

In[2], The algorithm groups the full text articles based on the keywords that are extracted. So each cluster contains only those articles which contain that keyword as their part.

Then fuzzy C mean clustering is used to combine the clusters together if their similarity measure is greater than or equal to a specified threshold value. XML files containing the created clusters are obtained as the output. This works efficiently only when same data for text extraction is used. In[3], In order to implement cluster is not known a priori, ACPSO is image clustering algorithm based on particle swarm optimization algorithm has been proposed. ACPSO can separation image into compact and separated clusters without any information on the real number of clusters. The separation of each particle of swarm involves using evolving operators which aim to reduce energetically the number of cluster centers.It is very difficult to exercise control over a swarm.The complexity of a swarm system leads to unforeseeable results.

In[4], The mass of data points GCA is a clustering algorithm based on Gravitational Law and Newton's Second Motion Law. For an n-dimensional data set with n data points, each data point is considered as an object in the n-dimensional space with mass equal to 1.Not affected by noisy data. Dependent upon the number of iterations and the smallest distance value.

In[5], In order to overcome the local optimal solution of Fuzzy C means, here Fuzzy Particle Swarm Algorithm is integrated with it.PSO optimizes a problem by having a population of contestant solutions and moving the particles around in the search-space according to simple mathematical formulae over the particle's position and velocity. The method easily suffer from the partial optimism, which causes the less exact of speed and the direction.The problems of distribution and optimization are not met.The method can not work out in particles in the energy fields.

In[6], Genetic Bee Tabu K-Means Clustering Algorithm (GBTKC) Bees will be selected randomly and crossover operator will be used in selected bees to generate two off springs. Then mutation operator will be used on resulted off springs. If off springs is not in Tabu list, evaluate its fitness. This bee is selected for next generation and added to Tabu list. If the Tabu list is full, the last item will exit and the input item will be stated at top of the list. If off springs is not in Tabu list then off springs will be added to Tabu list, evaluated fitness of off springs, and assign off springs to next generation and this continues till stopping criterion. number of sites selected for neighbourhood searching out of n visited sites (m) and Long CPU time.

**Manuscript published on 28 February 2014.**
***** Correspondence Author (s)
**K.Mohana Prasad*,** Computer Science and Engineering, Sathyabama University, Chennai, India.
**M.Nandhini,** Computer Science and Engineering, Sathyabama University, Chennai, India.

## III. EXISTING SYSTEM

The honey bee in the ABC algorithm are categorized into three groups Employed,Onlooker,Scout bees. The colony is separated into equal numbers of employed bees and onlooker bees. The number of employed bees is equal to the number of food sources. Each solution in the search space consists of a set of parameters which represent a food source position. The quality of food sources value and is associated with its position. The employed bees are responsible for investigating their food sources and sharing the information about these food sources to employ the onlooker bees. The onlooker bees will chose a food source based on this information. A food sources that has higher quality will have a larger chance to be selected by onlooker bees than one of lower quality.An employed bee whose food source is rejected as low quality by employed and onlooker bees will change to a scout bee and then search randomly for new food sources. Thus, the exploitation is handled by employed and onlooker bees while the exploration will be maintained by scout bee.

## IV. PROPOSED WORK

In general, the ABC algorithm is used to find a best solution. In abc relationship is only between the employed bee and onlooker bee . it should be selected by roulette wheel selection and the one selected randomly.Therefore, it is not good enough to maximize the exploitation capacity.The Correlative Artificial Bee Colony algorithm is proposed based on the structure of ABC algorithm. Here We find a more relationships between the employed bees and the onlookers, to extend the exploitation capacity of the ABC algorithm. In Correlative Artificial Bee Colony (CABC), the mass in universal gravitational force equation is replaced by the parameter , which is a food source value represented by the position of the employed bee. The advantage is reduce the computation Time, increase Exploitation ability of ABC, produce a good solution in a limited amount of time.

The process of the Exclusive ABC can be described in 5 steps:

- Step 1: Initial Process
- Step 2: Move the Onlooker bees:
- Step 3: Move the Scout bees:
- Step 4: Update the Best Food Source Found So Far
- Step 5: Final Checking

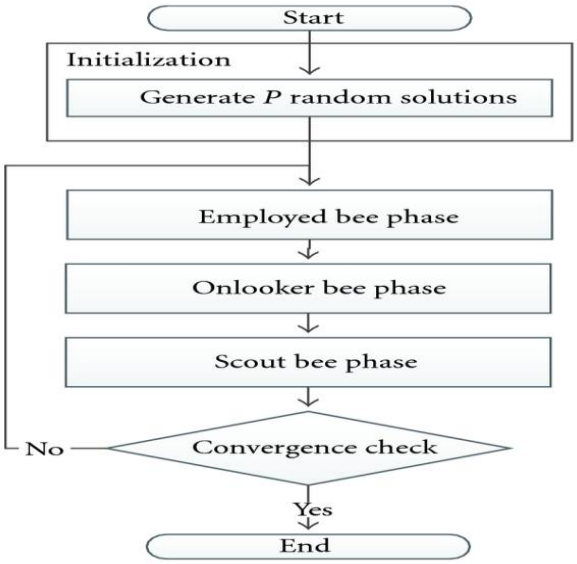


Fig 1. Flow chart diagram

## V. CONCLUSION

We have used Correlative Artificial Bee Colony Algorithm (CABC) which exploits the relationship between Onlooker and Employed bees more when compared to the conventional Bee colony Algorithms. In order to compare the results of CABC and ABC, benchmark functions have been used in the experiments. The experimental results have shown that the CABC presents stable and reliable capacity to find the near best solutions for the benchmark functions.

## REFERENCES

[1] M. Divyavani and T.Amudha "Comparing the Clustering Efficiency of ACO and K-Harmonic Means Techniques"International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.1, No.4, August 2011.

[2] Sumit Vashishta, Dr. Yogendra Kumar Jain ,"Efficient Retrieval of Text for Biomedical Domain using Data Mining Algorithm"(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 4, 2011

[3] Salima Ouadfel ,Mohamed Batouche and Abdlemalik Ahmed ,"Automatic image clustering using a swarm intelligence approach",IJCSI International Journal of Computer Science Issues,Vol.8,Issue 5,No 3,September2011.

[4] Tianwu Zhang, Hongshan Qu,"An Improved Clustering Algorithm",Proceedings of the Third International Symposium on Computer Science and Computational Technology(ISCSCT '10) August 2010 pp. 112-115

[5] Hesam Izakiana,, Ajith Abrahamb "Fuzzy C-means and fuzzy swarm for fuzzy clustering problem"Expert Systems with Applications 38 (2011) 1835–1838

[6] Mohammad Ali Shafia1 a, Mohammad Rahimi Moghaddam1, Rozita Tavakolian, "A Hybrid Algorithm for Data Clustering Using Honey Bee Algorithm, Genetic Algorithm and K-Means Method " Journal of Advanced Computer Science and Technology Research 1 (2011) 110-125


Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)