ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 2, Issue 3, March 2013

A modified Clustering algorithm based on Optimization method.

J.Mercy Geraldine[1], P.Kiruthiga[2]

***Abstract*— This study presents an effective optimization technique such as Ant Colony Optimization (ACO), Particle Swarm Optimization (PCO) and Fuzzy C-Means (FCM) for clustering the dataset. This technique is tested on several data and the performance of the modified algorithm, which is obtained to be better than the traditional algorithm.**

***Index Terms*—Ant Colony optimization, Particle swarm optimization, and Fuzzy C-Means.**

I. INTRODUCTION

Data mining is the process of extracting the data from the data warehouse. The data are extracted through many processes. The clustering of the data is done for grouping of the data through the centroids.

Clustering is processed through certain centroid. The centroid is mentioned and then the data are grouped together. The centroid is mentioned and data are grouped together as per the centroid.

According to their learning procedure, Pattern recognition methods can be divided into two groups. Supervised learning requires prior labeling of the training data to create a model of the given dataset. A supervised learning algorithm analysis the given training dataset and creates an output. This output is then compared with the desired output (label) and an error or feedback signal is created. Algorithm then updates itself according to this feedback signal in order to create a model of the given dataset.

. Fuzzy clustering methods that based on the objective function is the most studied in the literature and the most widely used in practice, such algorithm takes the clustering problem as a constrained optimization problem, by solving the optimization problem to determine the fuzzy partition and the clustering results in data set.

Fuzzy c-means clustering algorithm (FCM) is an effective algorithm based on the fuzzy concept. It is the most used method for clustering. The clustering effect of FCM is poor when the data set is in a higher dimension, and it will be difficult to find the global optimum [6].

Particle Swarm Optimization (PSO) is one of the heuristic algorithms nowadays used one. It is the evolutionary algorithms, and has proved to be very effective for solving global optimization, and gained lots of concentration in various engineering applications. It is not only a recently evolved high-performance optimizer which is easy to understand and implement, otherwise it also needs little computational secretarial. Usually only a few lines of code are used. It is a stochastic search technique with reduced memory condition, computationally efficient and it will be easy to implement and compared to other evolutionary algorithms.

Clustering problems can be recognized to optimization problems under certain conditions, Particle Swarm Optimization is an optimization algorithm which is based on the hypothesis of swarm intelligence, which could be implemented and it can be applied easily to solve various function optimization problems, otherwise the problems that can be transformed to function optimization problems.

PSO is easy to explain and implement, it also has a strong global search capability and a faster convergence.

ACO is the method of finding the shortest path through the convergence condition [1]. The ants or insects will find the shortest path to obtain the food. The shortest path among the other path is found through optimization technique.

Many iteration processes is done in the Ant Colony Optimization, whereas in each iteration ants will move from one state to another state. The intermediate state will be known as solution state. The Ant colony optimization is that process as follows

1. It visit each node exactly once;
2. A distant node has less chance of being chosen (the visibility);
3. The more intense the pheromone trail laid out on an edge between two nodes, the greater the probability that that edge will be chosen;
4. after completing its traverse, the ant deposits more pheromones on all edges it traversed.
5. After each iteration, trails of pheromones evaporate.

Many PSO and ACO-based Fuzzy Clustering Algorithms are proposed. However, in most of these algorithms the particle is encoded by cluster centers, less of these algorithms use the method that the particle is encoded by membership.

II. RELATED WORK

Cluster analysis problem is the data may fall into local minima and global minima [10]. The local minima problem is

***Manuscript received March, 2013*.**

***J.Mercy Geraldine**, HOD, Department of computer science, Anna University/ Srinivasan Engineering Colleg, Perambalur City ,India.*

***P.Kiruthiga**, Student, Department of Computer Science, Anna University/ Srinivasan Engineering College, Perambalur City, India.*

that the data may not have knowledge about another data in the cluster. Global minima are that occur between the local clusters. Local minimal is occur within the neighboring set of solution [4].

K-means clustering is that process clusters the data into k clusters. The data are clustered into k objects through the centroids. The k-means clustering will not process the data, if it is overlapped. The clustering of data is not possible in overlapped data [2-3].

III. FUZZY C-MEANS CLUSTERING ALGORITHM

It tries to partition a finite collection of n elements X={x1,…..,xn} into a collection of c fuzzy cluster with respect to some given principle. Given a finite set of data, the algorithm returns a list of c cluster center c={c1,…..,cm} and a partition matrix $U=U_{ij} \in [0,1]$, i=1,…n, j=1,…C. where each element U_{ij} says the degree to which element x_i belongs to cluster C_j.

In fuzzy clustering, each node has a degree of belonging to which cluster. But, it is not as fuzzy logic. Fuzzy logic is that belonging completely to just one cluster [9].

Fuzzy clustering will process on more number of clusters [11-12]. The centroid of cluster is mean of all points weighted by their degree of belonging to the cluster centers.

The specific steps for FCM:

Step 1: set the number of clusters c and fuzzy index, initializing the matrix of membership, set the maximum iterations n.

Step 2: then calculation of various cluster centers occur.

Step 3: matrix of membership are been calculated.

Step 4: repeat step 2 and step 3, until the conclusion of the maximum number of iterations.

It can also set a convergence accuracy as the condition for a loop terminates.

Ant colony optimization is method of finding the shortest path. This method is useful in salesman travelling problem. The process ACO is meant for finding the shortest path .The steps involved are

1. Each node is visited only once.
2. Distinct node has less chance of chosen.
3. Pheromone trail is secreted at each node (node visited).
4. Intensity of each pheromone is evaporated.

IV. FUZZY C-MEANS CLUSTERING ALONG WITH OPTIMIZATION METHOD.

Particle swarm optimization (PSO) is an optimization algorithm based on the theory of swarm intelligence, the support and competition among particles produced swarm intelligence to guide the optimization search. PSO is easy to describe and implement, it also has a strong global search capability and fast convergence [4-5]. But PSO also has defects, as in convergence condition, all particles are in the direction of the optimal, if optimal particle is not good enough, it can easily fall into local optimum .

Step 1: initializing the particle swarm occurs, by adding the population size, initial position and velocity of particles.

Step 2: calculate fitness for each particle, storage each particle best position and its fitness, and choose the particle that has the best fitness as ; best P best G

Step 3: then velocity and the position of each particle according their data are updated.

Step 4: calculate the fitness of each particle after update the position, compare the fitness of each particle with its best previous fitness , if better than it, then set the current position as ; best P best P

Step 5: compare the fitness of each particle with the group best previous fitness, if better than it, then set the current position as ; best G

Step 6: then to determine whether the results meet the conditions set by the end, if preconditions not met, then return to Step 3; then preconditions are met is verified, if it then stop iteration,

Step 7: output the optimal solution.

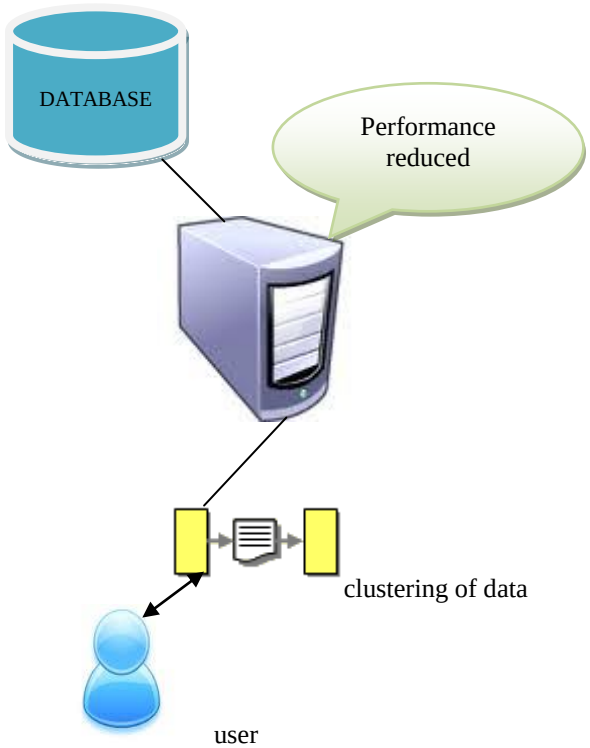

Figure 1. Reduction of performance during clustering.

The detailed steps of ACO based FCM is:

Step 1: initialize the optimization and the parameters involved in fuzzy clusteringof the data set are updated.

Step 2: for each particle, the corresponding clustering center according to the value of initial position are calculated.

Step 3: by updating the velocity and the position of each Particle in the cluster dimension the next step are processed.

Step 4: using condition approach to make the value of particle meet the constraints of fuzzy clustering. Then the corresponding clustering center according to the value of each particle are calculated.

Step 5: the fitness of each particle with the group best previous fitness $_{best}G$ are caculated.

Step 6:Determine whether the results has meet the conditions set by the end of the process, if preconditions not met, then

return to Step 3 then stop iteration, output the optimal solution.

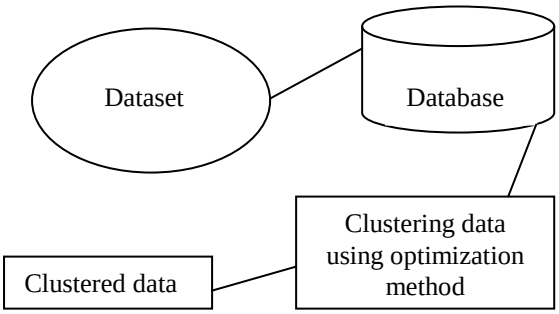

Figure 2. Process explanation

In FPSO algorithm the same as other evolving algorithms, a function is necessary for evaluating the generalized solutions and it is termed as fitness function. In this paper, Eq. (1) is used for evaluating the solutions.

$$f(x) = \frac{K}{Jm} \quad (1)$$

Whereas, the terms K is a constant and Jm is the objective function of FCM algorithm.

V. EXPERIMENTAL RESULT

In this section, we present an experimental setup for evaluating the glass, wine and vowel data set and evaluating the performance of the ACO-PSO-FCM algorithm. Then, these algorithms are presented and result obtained.

The algorithm uses the benchmark function that used in the data set. The data set will present the result in the appropriate function necessary for it.

We have used a somewhat different experimental methodology for each comparison in order to make the results obtained by ACO and PCO as comparable as possible to those obtained with the other methods.

It is important to highlight that unlike combinatorial optimization the comparison of algorithms for continuous optimization is usually not done. Usually each algorithm in case of combinatorial optimization, it is given the same amount of CPU time. The results obtained within that time are compared for the data set. Whereas, this makes the comparison of different algorithms more complicated, as the CPU time totally depends on the programming language used in this paper. The compiler, the skills of the programmer, and finally also on the machine used for running the several experiments. Hence, in case of combinatorial optimization it is strongly recommended to re-implement all the algorithms used in the comparison in order to make it fair. This still does not guarantee an entirely fair comparison, as it is difficult to ensure that the same amount of effort is put into optimization of the code of all the implemented algorithms.

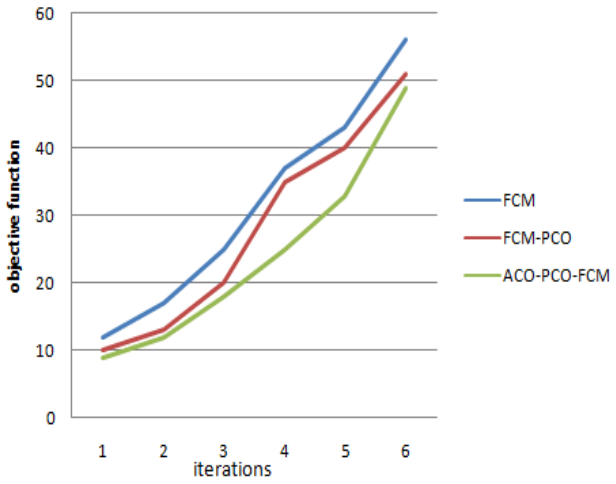

Figure 3. Comparison of algorithm.

For experimenting the algorithm three data set where considered and the process is done on it. The obtained results prove that the proposed algorithm is better when compared with the other algorithm. The data set considered are.

- Glass, where it consists of 214 objects and 6 different types of glasses in it. Whereas, each type consist of 9 features;
- Wine, where it consists of 178 objects and 3 different types of it is characterized by 13 features of data set.
- Vowel data set, where it consists of 871 Indian Telugu vowel sounds, it has three features in it along with six overlapping clusters.

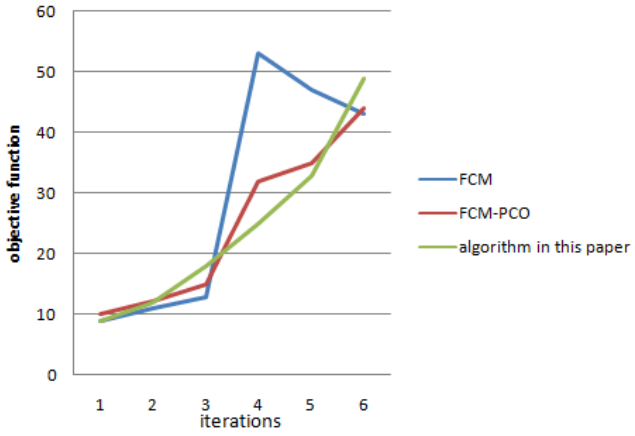

Figure 4. Comparison of algorithm in wine data set

The comparison between the different dataset present that classification of the data set is an effective method. The algorithm classify the wine, glass and vowel exactly. These method will classify the data set with probability. Overlapping of dataset cluster are performed accurately and appropriately.

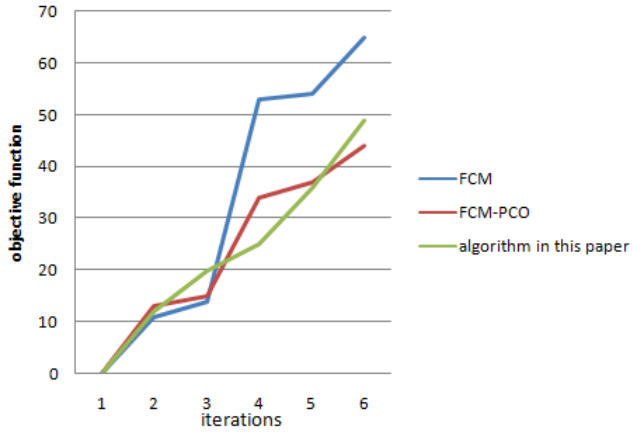

Figure 5. Comparison of algorithm in Vowel data set

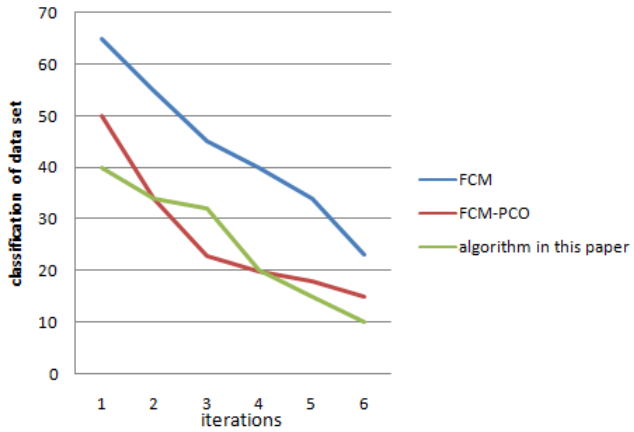

Figure 6. classification of data set in vowel data set.

These data sets are processed in data of low, medium dimensions. The dimension is obtained and theses cover data in all dimensions. The classification of the data set is formulated and where the data set are divided according to their class. The process is done at dimension of data set.

The combo of this algorithm will be superior to the other algorithm. The process will reduce the local minima and global minima problem. The local minima will reduce and appropriate result is obtained.

VI. CONCLUSION

It is aimed to study that the proposed algorithm outperforms the traditional algorithm. The algorithm fuzzy c-means can be easily fascinated in local optima. In addition, the particle swarm algorithm and ant colony algorithm is a global tool and it can be implemented. It is applied easily to solve various function optimization problems, otherwise the problems that can be transformed to function optimization problems. In order to overcome the drawbacks of the fuzzy c-means in this paper, a integrated algorithm fuzzy particle swarm algorithm and ant colony optimization is used. The results of Experiments over three well-known data sets such as, Glass, and Wine, shows that the proposed method is efficient.

REFERENCES

[1] Dorigo, M., Di Caro, G., Gambardella, L.M., 1999. Ant algorithms for discrete optimization. Artificial Life 5 (2), 137–172..
[2] Cai WL, Chen SC, Zhang DQ. Fast and robust fuzzy cmeans clustering algorithms incorporating local information for image segmentation. Pattern Recognition, 2007, 40(3): 825-833.
[3] Jiayin Kang, Lequan Min, Qingxian Luan, Xiao Li and Jinzhu Liu. Novel modified fuzzy c-means algorithm with applications. Digital Signal Processing. 2009, (19): 309- 319.
[4] Riccardo Poli, James Kennedy and Tim Blackwell. Particle Swarm Optimization. Swarm Intell. 2007, (1): 33-57.
[5] Yang GQ, ZHU CM. Particle swarm optimization algorithm based fuzzy kernel clustering method. Journal of Shanghai Jiao Tong University, 2009, 43(6): 935-939.
[6] Thomas A. Runkler, Christina Katz. Fuzzy clustering by particle swarm optimization [C]. 2006 IEEE International Conference on Fuzzy Systems. 2006: 601-608.
[7] Riccardo Poli, James Kennedy and Tim Blackwell. Particle Swarm Optimization. Swarm Intell. 2007, (1): 33-57.
[8] Wang JW, LI HN. Summary of particle swarm optimization algorithm. Modern Computer, 2009, 301: 22- 27.
[9] N.A. Mohamed, “Modified fuzzy C-mean algorithm for medical image segmentation,” M.Sc. thesis, Elect. Eng. Dept., Univ. Louisville, KY, 1999.
[10] R. Krishnapuram and J.M Keller, ”A possibilistic approach to clustering”. IEEE Transactions on Fuzzy Systems, 1(2), pp.98-110, 1993.
[11] Nikhil R. Pal, Kuhu Pal, James M. Keller, and James C. Bezdek, ”A possibilistic fuzzy c-means clustering algorithm”, IEEE Transactions on Fuzzy Systems, 13(4), pp.517–530, 1997.
[12] J.C. Bezdek, J. Hathaway, M.J. Sabin, and W. T. Tucker, ”Convergence Theory for Fuzzy C-Means: Counterexamples and Repairs”, IEEE Trans, 17, pp. 873- 877, Sept./Oct. 1987