Fuzzy Clustering using Fuzzy Particle Swarm Optimization

Dr. C. GOMATHI MCA., M.Phil., Ph.D.,
Assistant Professor,
Department of Computer Science,
Sri Vasavi College, Erode.

K. VELUSAMY M.Sc., M.Phil. (Ph.D)
Research Scholar
Department of Computer Science
Sri Vasavi College, Erode

Abstract –Clustering is crucial for scalable enterprise applications, as you can improve performance by simply adding more nodes to the cluster. K-means is very popular stiff clustering algorithms which panels data objects into k clusters. Fuzzy C-Means clustering algorithm is one of the most trendy clustering method is more efficient, straightforward, easy to implement and perceptive to initialization. A disadvantage of K-Means and Fuzzy C-Means together algorithms is easily falling in narrow optima. Fuzzy Particle swarm optimization algorithm helps to work out narrow optima. Experimental results prove that fuzzy PSO advance giving extremely ready for action results for fuzzy clustering and FPSO and objective functions values go one better than is better-worth the presentation of other to be had algorithms.

Keywords- K-Means, Fuzzy C Means, Fuzzy Particle Swarm Optimization

I INTRODUCTION

Cluster is the method of transmission data stuff to a set of put out of joint groups call Clustering [5]. Clustering methods are practical in many submission areas such as pattern recognition, machine learning etc. In outgoing, clustering is an unsupervised knowledge mission as very little or no previous data is given apart from input data sets. The main benefit of a clustered solution is mechanical recovery from failure, that is, recovery without user intervention. Disadvantages of clustering are difficulty and lack of ability to recover from database corruption. The fuzzy C means algorithm introduced by Jim Bezdek m [7]. It is first and nearly all stylish fuzzy clustering algorithms. Fuzzy C means straightforward says FCM. Fuzzy C Means clustering algorithm is narrow hunt algorithm because it's aware to original assessment successfully.

Particle swarm optimization (PSO) is an inhabitants-based optimization device, which could be implemented and sensible with no trouble to make clear a range of task optimization problems, or the problems that can be altered to task optimization problems (Kennedy & Eberhart) [10]. Pang, Wang, Zhou, and Dong, proposed a version of particle swarm optimization for TSP called fuzzy particle swarm optimization (FPSO). In this paper, a fuzzy clustering algorithm based on FCM and FPSO called FCM–FPSO is projected. The new outcomes over six real-life data sets indicate the FCM–FPSO algorithm is superior to the FCM algorithm and FPSO algorithm [13]. On the next step shift, particle swarm optimization algorithm is denoted by PSO. A made to order particle swarm optimization is also known as fuzzy particle swarm optimization (FPSO) projected by peng. We thrash out in one of the mainly accepted clustering algorithm which objects k clustering. Then FCM balance fuzzy PSO to get improved product which is best. We converse about iris dataset. Several papers are describes panel clustering algorithms by squared metric for single-handed key many facts. Iris in sequence include batch of inputs to produce algorithm for K means, Fuzzy C means evaluate with extra two techniques Particle Swarm Optimization is best. Same time we calculate objective values, to find best, worst and average.

II K MEANS

K-means clustering generate an exact number of disjoint, flat (non-hierarchical) clusters. Future it well suited to generating globular clusters. Following method is numerical, unsupervised, non-deterministic and iterative. K-means clustering algorithm is the largest part accepted paneling cluster algorithms. In earlier days, several papers are describes paneling clustering algorithms using squared inaccuracy metric for isolated key many details hectic cubes. The plaza inaccuracy for every cluster is defined as the computation of square detachment between each part in the class and the center, called intra-detachment. Hence the peak task in the square inaccuracy clustering algorithm is to reduce the sum of all intra- detachment. Established sequential algorithm starts with original set of K clusters, and move every patterns (or data) to a cluster if it minimizes the square inaccuracy.

K-Means Algorithm

K-Means is one of the basic clustering techniques, which is make each cluster is represented by the mean ideals of the entity in the cluster. The algorithm has as an input a predefined number of clusters that is the K from its name. The K-Means stands for a regular, an average place of all the members of an exacting cluster. The K-means Algorithm is an iterative procedure whereby concluding results depend on the standards selected for original centroid.

A Centroid is a simulated position in the hole of records which represents a standard place of the exacting cluster. The coordinates of this point are averages of attribute values of all examples that belong to the cluster. Resolution functional model in huge datasets has involved large importance value recently and of the most extensively read out problems in this area is the classification of clusters or tightly occupied regions, in a multi-dimensional dataset. The most important problem with the K-mean algorithms is that of choice of original centroids.

Finally, this problem solving procedure aims at minimizing a point task, in this case a squared inaccuracy function. The point function

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^j - c_j \right\|^2 \qquad (2.1)$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a selected distance compute among a data point $x_i^{(j)}$ and the cluster centre c_j, is an pointer of the distance of the *n* facts points from their relevant cluster centre.

ALGORITHM 1: K MEANS

Step 1: The position as a cluster center and select arbitrarily
k points
Step 2: Assign each one point to bring together centroid
Step 3: Re-computational the location of centroid for all the
point have been passing on
Step 4: With the exception of the centroid not modify go to
step 2

III FUZZY C MEANS

Fuzzy C Means is a method of centroid which allows one set of data it does belong to two or more clusters. This method developed by Dunn in 1973, improved by Bezdek in 1981 frequently used in pattern recognition.

The Fuzzy C-Means algorithm (often truncated to FCM) is an iterative algorithm that discover clusters in data and which uses the assumption of fuzzy connection as an alternative of conveying a pixel to a single cluster, each pixel will have alternate connection values on each cluster. Fuzzy C-means panels set of n stuff $O = \{o_1, o_2, ...o_n\}$ in R^d dimensional space into $c(1 < c < n)$ fuzzy clusters with $z = \{z_1, z_2, ... z_n\}$ cluster centers or centroids. The fuzzy clustering of objects is described by a fuzzy environment μ with n rows and c columns in which n is the number of data stuff and c is the number of clusters. μ_{ij}, the component in the ith row and jth column in μ, indicates the degree of association or membership function of the ith entity with the jth cluster. The characters of μ are as follows:

$$\forall_i [0,1] \forall_i = 1,2...n \forall_j = 1,2...c$$

$$\sum_{j=1}^{c} \mu_{ij} = 1 \quad \forall_i = 1, 2 \ldots n \qquad (2.2)$$

$$o < \sum_{j=1}^{c} \mu_{ij} < n \quad \forall_j = 1, 2 \ldots c \qquad (2.3)$$

The point purpose of FCM algorithm is to decrease the Eq. (4):

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij} d_{ij} \qquad (2.4)$$

Where $$d_{ij} = \left| o_i - z_j \right| \qquad (2.5)$$

In which, m (m>1) is a scalar termed the weighting model and pedals the fuzziness of the ensuing clusters and d_{ij} is the Euclidian distance from object $z = \{z_1, z_2, ..., z_n\}$ to the cluster center z_j. The z_j, centroids of the j$_{th}$ cluster, is obtained using under an equation

$$z_j \frac{\sum_{i=1}^{n} \mu_{ij}^{m} o_i}{\sum_{i=1}^{n} \mu_{ij}^{m}} \quad (2.6)$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}}}$$

(2.7)

The Fuzzy C-Means algorithm (FCM) is an iterative algorithm that finds clusters in data and which uses the concept of fuzzy membership, [8] instead of assigning a pixel to a single cluster, each pixel will have different membership values on each cluster. It panels set of n objects $O = \{o_1, o_2, ...o_n\}$ in R^d dimensional [14] space into $c(1< c <12\{,,...\}n)$ fuzzy clusters with $z = \{z_1, z_2, ...z_n\}$ cluster centers. The fuzzy clustering of stuff is described by a fuzzy matrix μ with n rows and c columns in which n is the number of data stuff and c is the number of clusters. In component ith row and j[th] column μ, point out the degree of connection or membership task of the i[th] point with the j[th] cluster. The characters of μ are as follows:

ALGORITHM 2: FUZZY C MEANS

Step 1: Choose m (m>1) μ_{ij} i=1,2,…n, j=1,2…c Initialize
the membership task values to be
Step 2: The Eq.(2.6) to work out the cluster centers z_j ,
j = 1,2,..., c
Step 3: Euclidian distance to work out d_{ij}, i = 1,2,..., n;
j = 1,2,..., c by using Eq. (2.5).
Step 4: Renew the membership task μ_{ij},
i = 1, 2... n;
j = 1,2,..., c by using

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}}} \quad (2.7)$$

Step 5: If not solve, go to step 2.

IV PARTICLE SWARM OPTIMIZATION

Particle swarm optimization (PSO) is a population based stochastic optimization technique, introduced by Kennedy and Eberhart, [10] inspired by social behavior of bird flocking or fish schooling.

Each particle keeps track of its coordinates in the problem space which are associated with the best solution (fitness) it has achieved so far. (The fitness value is also stored.) This value is called *pbest*. Another "best" value that is tracked by the particle swarm optimizer is the best value, obtained so far by any particle in the neighbours of the particle. This location is called *lbest*. When a particle takes all the population as its topological neighbours, the best value is a global best and is called *gbest*. After finding the two best values, the particle updates its velocity and positions with following equation (2.8) and (2.9).
The PSO starts with a population of particles whose positions represent the potential solutions for the studied problem and some human categories are two types male and female in those velocities are randomly initialized in the search space.

$$v(t+1) = w.v(t).c_1 r_1 (pbest(t) - X(t)) + c_2 r_2 (gbest(t) - X(t))$$
k=1,2…P (2.8)

$$X(t+1) = X(t) + v(t+1) \quad (2.9)$$

X and V are position and velocity of particle correspondingly, w is inertia weight, c1 and c2 are positive constants, called acceleration coefficients which control the manipulate of pbest(t) and gbest(t) on the look for process, P is the number of particles in the group, r1 and r2 are learning factors. Usually random values range between 0 and 1.

A. Fuzzy Particle Swarm Optimization

A particle swarm optimization with fuzzy set theory is called fuzzy particle swarm optimization (FPSO), which is proposed by Peng [9] et al.,. Using fuzzy relation between variables, FPSO redefines the position and velocity of particles. And it also applied for clustering problem. In this method X is the position of particle, the fuzzy relation for the set of data objects, $O = \{o_1, o_2, ...o_n\}$ to set of cluster centers, $z = \{z_1, z_2, ...z_n\}$, Can be expressed as follows

ALGORITHM 3: FUZZY PARTICLE SWARM OPTIMIZATION (FPSO)

Input: Dataset
Output: target Values
Step 1. Initialize the parameters including hand or leg size
P, c1, c2, w and then maximum iterative count.
Step 2. Swarm matrices to generate P particles (X, pbest,
gbest and V are n × c matrices).
Step 3. Initialize X, V, pbest for each particle and gbest for
the swarm.
Step 4. Work out the cluster centers for each particle using
by (2.6)
Step 5. To find the condition value of each particle using

by $f(x) = \frac{K}{J_m}$

Step 6. pbest compute for each particle.
Step 7. gbest calculate for the swarm.
Step 8. Update the velocity matrix for each particle using
by (11)
Step 9. Update the position matrix for each particle using
by (12)
Step 10.If finish state is not met, return back step 4

B. Fuzzy particle swarm optimization for fuzzy clustering

A customized particle swarm optimization are called fuzzy particle swarm optimization (FPSO) proposed by Peng et al. In which technique the position and velocity of particles redefined to represent the fuzzy relation between variables. In this part we describe the system for fuzzy clustering problem.

In FPSO algorithm X, the position of particle, shows the fuzzy relation from set of data objects, $o = \{o_1, o_2, ..., o_n\}$ to set of cluster centers, $z = \{z_1, z_2, ..., z_n\}$, X Can be expressed as follows:

$$X = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{pmatrix} \quad (2.10)$$

In which μ_{ij} is the membership function of the i_{th} object with the j_{th} cluster with constraints stated Eq. (2.1) and Eq. (2.2) therefore we can see that the position matrix of each particle is the same as fuzzy matrix μ in FCM algorithm. Also the velocity of each particle is stated using a matrix with the size *n* rows and *c* columns the elements of which are in range between -1 and 1. We get the Eq. (2.11) and Eq. (2.12) for update the positions and velocities of the particles base of the matrix.

$$v(t+1) = w \otimes v(t) \oplus c_1 r_1 \otimes pbest(t) \Theta X(t) \oplus c_2 r_2 \otimes (gbest(t) - \Theta X(t))$$

k=1,2…P
(2.11)

$$X(t+1) = x(t) \oplus v(t+1) \quad (2.12)$$

After update the arrangement matrix, it may go against the constraint given Eq. (2.1) and Eq. (2.2). So it is necessary to normalize the position matrix. First we make all the negative elements in matrix to become zero. If all elements in a row of the matrix are zero, they need to be re-evaluated using series of random numbers within the interval between 0 and 1, and then the matrix undergoes the following transformation without violating the constraints:

$$Xnormal = \begin{pmatrix} \mu_{11} / \sum_{j=1}^{c} \mu_{ij} & \cdots & \mu_{1c} / \sum_{j=1}^{c} \mu_{1j} \\ \vdots & \ddots & \vdots \\ \mu_{n1} / \sum_{j=1}^{c} \mu_{nj} & \cdots & \mu_{nc} / \sum_{j=1}^{c} \mu_{nj} \end{pmatrix}$$

(5.13)

In FPSO algorithm the same as other evolutionary algorithms, we need a function for evaluating the generalized solutions called fitness function. In below equation is used for evaluating the solutions.

$$f(x) = \frac{K}{J_m} \quad (2.14)$$

K is a stable and J_m is the point function of FCM algorithm (Eq. (5.14)). The minor is J_m, the better is the clustering result and the upper is the individual fitness
f (X). The FPSO algorithm for fuzzy clustering problem can be declared as follows:

ALGORITHM 4: FUZZY PSO FOR FUZZY CLUSTERING

Step 1 : Choose the initialize parameter together with human
part size p, c1, c2, w and the maximum iterative
count.
Step 2: Generate a swarm with P particles (X, pbest, gbest
and V are n× c matrices).
Step 3: Initialize X, V, pbest for each particle and gbest for
the swarm.
Step 4: Work out the cluster centers for each particle using
by Eq. (2.6).
Step 5: Calculate the fitness value of each particle using by
Eq. (2.14).
Step 6: Compute pbest for each particle.
Step 7: Calculate gbest for the swarm.
Step 8: Update the velocity matrix for each particle using
by Eq. (2.11).
Step 9: Update the position matrix for each particle using
by Eq. (2.12).
Step10: If terminate state is not met, go to step 4.

The termination condition in proposed method is the maximum number of iterations or no improvement in gbest in a number of iterations

V EXPERIMENTAL RESULTS AND DISCUSSION

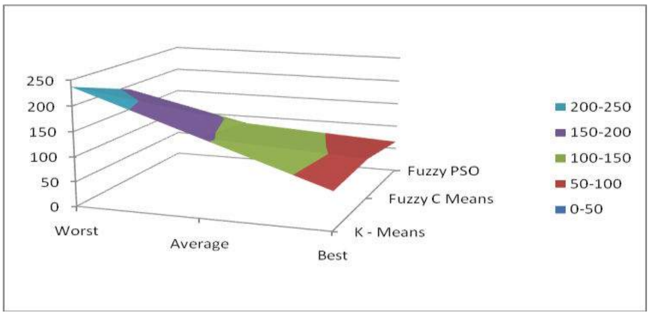
The performance of the proposed algorithm is evaluated under five different job datasets and the results are compared against the conventional algorithms such as K –means, FCM, FPSO. The results that are obtained under different experiments are given in following

TABLE I: Show the object values of the method is best

Methods	Worst	Average	Best
K - Means	237.1	156.7	**76.3**
Fuzzy C Means	192.7	137	**81.3**
Fuzzy PSO	65.4	65.1	**64.9**

Table I we compared to K- Means, Fuzzy C Means algorithms compare these two algorithm Fuzzy PSO best. Above the Table I to find take objective values and calculate the total number of iteration. Finally Fuzzy Particle Swarm Optimization is best values are (76.3,81.3,64.9)

Fig 1. Analysis of objective values of algorithms



Performance objective values of K means, Fuzzy C means and Fuzzy PSO with their iteration values got from objective matrix with implementation of MATLAB 6.0. In this chart includes our implementation algorithm values. Here very best one Fuzzy PSO objective values is best.

VI CONCLUSION AND FUTURE WORK

The PSO is a global optimizer for continuous variable problems with efficient. Small number of parameters to deal and the large number of processing elements is the advantages of the PSO. Enable by FPSO to fly around the solution space effectively.

Research this work the problem of premature convergence of the Fuzzy C means clustering algorithm based on PSO. It's very sensitive to noise and less effective when handling the data set. Based on the better Particle Swarm Optimization algorithm is existed. The experimental results, the proposed method significantly improves the clustering effect of the fuzzy PSO based compared those three to find out the objective values.

This research seeks at efficient clustering the traditional paneling clustering techniques K-Means, Fuzzy-C Means and FPSO. FPSO is giving the better results as compared to all other algorithms on iris datasets. The future work includes, Ant Colony Optimization (ACO), Bee Colony Optimization (BCO). Etc.,

REFERENCES

[1] Bezdek, J. (1974). Fuzzy mathematics in pattern Classification. Ph.D. thesis. Ithaca, NY:Cornell University.

[2] Chun-Wei Tsai, Kai-Cheng Hu, Ming-Chao Chiang Chu-Sing Yang, "Ant colony optimization with dual pheromone tables for clustering" Fuzzy Systems (FUZZ), 2011 IEEE International Conference',978-1- 4244-7315-1.

[3] Chen Yanyum, Qiu Jianlin, Gu Xiang, Chen Jianping, Ji Dan Chen Li,"Advances in Research of Fuzzy C- Means Clustering Algorithm", IEEE, 10.1109/NCIS.2011.104

[4] Eduardo Silva Palmeira,"A Comparison between K- Means, FCM and CKmeans Algorithms, 2011 IEEE,978-0 -7695-4628-5/

[5] Gan, G., Wu, J., & Yang, Z. (2009). A genetic fuzzy k- modes algorithm for clustering categorical data. Expert Systems with Applications (36), 1615–1620.

[6] Han and Kamber, Data Mining Concepts and Techniques, Morgan Kauffman Publishers

[7] Hathway, R. J., & Bezdek, J. (1995). Optimization of clustering criteria by reformulation. IEEE transactions on Fuzzy Systems, 241–245.

[8] H. Izakian, A. Abraham / Expert Systems with Applications (2011) Fuzzy C-means and fuzzy swarm for fuzzy clustering problem1835–1838

[9] Hesam Izakain a, Ajith Abraham, "Fuzzy C-mean and

fuzzy swarm for fuzzy clustering problem," 0957- 4171 2010 Elsevier Ltd., doi:10.1016/j.eswa.2010.07.112

[10] Kennedy,J., & Eberhart R."Swarm Intellignece. Morgan Kaufmann.

[11] K.Sankar, Dr.K.Krishnamoorthy Cluster Mining with Ant Colony Optimizer using Fuzzy inferences. International Conference on Computing, Communication and Networking Technologies

[12] V. Ramesh, K. Ramar, S. Babu (2013) Parallel K-Means Algorithm on Agricultural Databases, IJCSI International Journal of Computer Science.

[13] Stuti Karol, Veenu Mangat Evaluation of a Text Document Clustering Approach based on Particle Swarm Optimization, IJCSNS International Journal of Computer Science and Network Security, VOL.13 No.7, July 2013 130.

[14] Semantically Enhanced Document Clustering Based on PSO Algorithm Sridevi.U. K., Nagaveni.N., European Journal of Scientific Research, ISSN 1450- 216X Vol.57 No.3 (2011), pp.485-493

[15] Sanya,Hainan China,"SVM Combined with FCM and PSO for Fuzzy Clustering, 7th International Conference on Computational Intelligence and Security" IEEE, 978-0-7695-4584-4.

[16] K.Velusamy, R.Manavalan Performance Analysis of Unsupervised Classification Based on Optimization *International Journal of Computer* Applications (0975 – 8887) Volume 42– No.19, March 2012