Comparative Study of Fuzzy PSO (FPSO) Clustering Algorithm and Fuzzy C-Means (FCM) Clustering Algorithm

Phyo Phyo Oo, Ei Chaw Htoon
University of Computer Studies, Yangon
phyoo695@gmail.com, eichawhtoon@uit.edu.mm

Abstract

Swarm intelligence that mimic the natural collective intelligence to solve the computational problem has emerged and widely used in data mining. Particle swarm optimization (PSO) is a kind of swarm intelligence algorithm. Fuzzy clustering is an important research in several real-world applications. Fuzzy particle swarm optimization (FPSO) is a fuzzy clustering algorithm that can be optimized with the use of PSO algorithm to get global optima. Fuzzy c-means (FCM) is one of the most popular fuzzy clustering techniques. In this paper, FPSO and FCM clustering algorithms will be implemented. These two methods were compared in term of execution times and fuzzy objective function (J_m) by using datasets namely Iris Plants, Breast Cancer and Wine from UCI (University of California) Machine Learning repository.

Keywords: *Fuzzy Clustering, Particle Swarm Optimization, Fuzzy Particle Swarm Optimization (FPSO), Fuzzy C-means (FCM)*

1. Introduction

Clustering is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. Clustering techniques are applied in many application areas such as pattern recognition, data mining, machine learning, etc. Clustering algorithms can be broadly classified as Hard, Fuzzy, Possibilistic and Probabilistic [3].

Fuzzy clustering is the process of grouping a set of physical or abstract objects into classes of similar objects. It is partition methods that can be used to assign objects of the data set to their clusters. In fuzzy clustering, data elements can belong to more than one cluster, and associated with each element is a set of membership levels. These indicate the strength of the association between that data element and a particular cluster.

Swarm Intelligence draws inspiration from the collective intelligence emerging from the behavior of a group of social insects such as bees, ants and bird flock.

PSO is a population-based optimization method, which could be implemented and applied easily to solve various function optimization problems [8]. FPSO is a version of particle swarm optimization for TSP (Travelling Salesman Person Problem) [9]. The most popular fuzzy clustering algorithm is fuzzy c-mean (FCM) which introduced by Bezdek (1974) and now it is widely used. It is easily trapped in local optima [1].

In this paper, FPSO and FCM clustering algorithms will be implemented. These two methods will be compared in term of execution time and fuzzy objective function *(J_m)* by using datasets namely Iris plant, Breast Cancer and Wine.

The rest of this paper is organized as follows. Section 2 describes the related work. In Section 3, we discuss the background theories that are used. In Section 4, the system design and architecture are described. Section 5 describes the implementation of FPSO and FCM clustering algorithms using datasets namely Iris plants, Wine and Breast Cancer. Section 6 describes the experimental results and conclusion is given in Section 7.

2. Related Works

A hybrid fuzzy clustering method based on FCM and FPSO is proposed which make use of the merits of both algorithms. The proposed method, in order to overcome the shortcomings of the fuzzy c-means we integrate it with fuzzy particle swarm algorithm. Experimental results over six well-known data sets, Iris, Glass, Cancer, and Wine, show that the proposed hybrid method is efficient and can reveal very encouraging results in term of quality of solution found [4].

A hybrid algorithm, namely FPSO, combining the FCM and the particle swarm optimization (PSO) algorithms in order to cluster supplier in fuzzy environments. These method is to present a new approach for a PSO algorithm to clustering suppliers under fuzzy environments into manageable smaller groups with similar characteristics. Their numerical analysis [2] indicates that the PSO improves the performance of the FCM algorithm.

3. Background Theory

3.1 Clustering techniques

Clustering is a dynamic field of research in data mining. Clustering techniques can be categorized as follows: (1) Partition methods (K-means, FPSO, FCM, etc.), (2) Hierarchical methods (agglomerative (bottom-up) merging, divisive (top-down) splitting), (3) Density-based methods (Density-based Spatial Clustering of Applications with Noise (DBSCAN), etc.), (4) Model-based methods (Expectation-Maximization, Conceptual Clustering, Neural Network Approach), etc.

3.2 PSO - Swarm intelligence

Swarm intelligence (SI) draw inspiration from the collective intelligence emerging. Two popular swarm intelligence techniques are Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO). PSO is a global optimization algorithm for the continuous search space mainly useful for hard problem. ACO is swarm based intelligence technique that mimics the behaviors of ant colonies and it can be used for finding shortest path, and data clustering.

3.3 FPSO Clustering

FPSO [9] is a modified particle swarm optimization for Travelling Salesman Person Problem (TSP). It allows one piece of data to belong to two or more clusters. These indicates the strength of the association between that data element and a particular cluster.

In FPSO algorithm *X*, the position of the particle, shows the fuzzy relation from set of data objects, $X = \{x_1, x_2,..., x_n\}$, to set of cluster centers, $c= \{c_1, c_2,..., c_C\}$. *X* can be expressed as follows:

$$X = \begin{bmatrix} u_{11} & \cdots & u_{1C} \\ \vdots & \ddots & \vdots \\ u_{N1} & \cdots & u_{NC} \end{bmatrix}$$

Eq. (1)

In which u_{ij} is the membership function of the *i*th object with the *j*th cluster with constraints stated in (6) and (7). Therefore, the position matrix of each particle is the same as fuzzy matrix *u* 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 [-1, 1]. Eqs. (2) and (3) for updating the positions and velocities of the particles based on matrix operations.

$$V(t+1) = w*V(t)+(c_1 * r_1)*(pbest(t)-X(t)) + (c_2 * r_2) * (gbest(t)-X(t)) \quad \text{Eq. (2)}$$

$$X (t+1) = X(t)+ V(t+1) \quad \text{Eq. (3)}$$

Where *X* and *V* are position and velocity of particle respectively. Initial weight $w=0.9$. w is inertia weight, c_1 and c_2 are positive constants, called acceleration coefficients which control the influence of *pbest* and *gbest* on search process, *P* is the number of particles in the swarm, r_1 and r_2 are random values in range [0, 1].

After updating the position matrix, it may violate the constraints stated in (5) and (6). So, it is necessary to normalize the position matrix. First, 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 [0, 1] and then the matrix undergoes the following transformation without violating the constraints:

$$Xnormal = \begin{bmatrix} u_{11}/\sum_{j=1}^{C} u_{ij} & \cdots & u_{1C}/\sum_{j=1}^{C} u_{1j} \\ \vdots & \ddots & \vdots \\ u_{N1}/\sum_{j=1}^{C} u_{N1} & \cdots & u_{NC}/\sum_{j=1}^{C} u_{Nj} \end{bmatrix} \quad \text{Eq. (4)}$$

In FPSO algorithm the same as other evolutionary algorithms, we need a function for evaluating the generalized solutions called fitness function. In this paper, Eq. (5) is used for evaluating the solutions.

$$f(x) = \frac{K}{J_m} \quad \text{Eq. (5)}$$

where *K* is a constant and J_m is the objective function of FCM algorithm (Eq. (9)). The smaller is J_m , the better is the clustering effect and the higher is the individual fitness *f(x)*. The FPSO algorithm for fuzzy clustering problem can be stated as follows:

Algorithm1.FuzzyPSO (FPSO) clustering:

1. Initialize each particle with *k* random cluster centers (centroids).
2. Repeat for iteration-count =1 to Maximum-iterations
   (a) Repeat for each particle *i*
      (i) Calculate membership matrix
      (ii) Calculate fitness
   (b) Find the personal best position and global best position of each particle.

(c) Update the cluster centroids according to velocity updating and coordinate updating formula of PSO.

3.3.1 Particle swarm optimization (PSO)

Particle swarm optimization (PSO) is a population-based stochastic optimization technique inspired by bird flocking and fish schooling which is based on iterations/generations [7].

PSO Algorithm

Randomly initialized position and velocity of particles.
While ***termination condition*** is not reached
for i =1 to number of particles do
Evaluate the fitness f($X_i(t)$)
Update P_{best} and P_{gbest}
Update velocity
Update the position of the particle.
end for
end while

The termination condition in proposed method is the maximum number of iterations or no change values in *gbest* in a number of iteration.

3.4 FCM Clustering

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters [5]. It is one of the most widely used fuzzy clustering algorithms. This method is frequently used in pattern recognition [6]. Fuzzy c-means partitions set of *n* objects $X = \{x_1, x_2,..., x_N\}$ in R^d dimensional space into *C (1< C < N)* fuzzy clusters with cluster centers or centroids. The fuzzy clustering of objects is described by a fuzzy matrix *u* with *N* rows and *C* columns in which *N* is the number of data objects and *C* is the number of clusters. The characters of *u* are as follows:

$$u_{ij} \in [0,1] \quad \forall_i = 1,2,\dots,N; \quad \forall_j = 1,2,\dots,C \qquad \text{Eq. (6)}$$

$$\sum_{j=1}^{C} u_{ij} = 1 \quad \forall_i = 1,2,\dots,N \qquad \text{Eq. (7)}$$

$$0 < \sum_{i=1}^{N} u_{ij} < N \quad \forall_j = 1,2,\dots,C \qquad \text{Eq. (8)}$$

The objective function of FCM algorithm is to minimize the Eq. (9):

$$J_m = \sum_{j=1}^{C} \sum_{i=1}^{N} {u_{ij}}^m d_{ij} \qquad \text{Eq. (9)}$$

Where $d_{ij} = \|x_i - c_j\|$ Eq. (10)

Fuzzy partitioning is carried out through an iterative optimization of the objective function show above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{1}{m-1}}} \qquad \text{Eq. (11)}$$

$$c_j = \frac{\sum_{i=1}^{N} {u_{ij}}^m x_i}{\sum_{i=1}^{N} {u_{ij}}^m} \qquad \text{Eq. (12)}$$

In which *m* is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster *j*, x_i is the *i*th of d-dimensional measured data, c_j is the d-dimension center of the cluster, and $\|*\|$ is any norm expressing the similarity between any measured data and the center.

The FCM algorithm is iterative and can be stated as follows:

Algorithm 2. Fuzzy C-Means (FCM):

1. Select m (m>1) ; initialize the membership function values u_{ij}, *i= 1, 2, ...,N ; j=1,2,...,C.*
2. Compute the cluster centers c_j, *j=1, 2, ..., C,* according to Eq.(12).
3. Compute Euclidian distance d_{ij}, *i=1, 2, ... , N; j= 1, 2, ..., C.*
4. Update the membership function u_{ij}, *i=1, 2, ..., N*; *j= 1, 2, ..., C* according to Eq. (11) .
5. If not converged, go to step 2.

Several stopping rules can be used. One is to terminate the algorithm when the relative change in the centroid values becomes small or when the objective function, Eq. (9), cannot be minimized more. The FCM algorithm is sensitive to initial values and it is likely to fall into local optima [4].

3.5 Datasets for two methods comparison

For the purpose of comparing the performance of FPSO and FCM clustering algorithms, Iris plants datasets, Breast Cancer dataset and Wine datasets are used because these datasets cover data of low and medium dimensions. By using these datasets, this system can obtain superior results than other datasets and it can escape from local optima. These datasets are obtained from the University of California Irvine (UCI) machine learning repository [10]. The features of each dataset are expressed in Table 1.

Table 1. The features of dataset for the System

	Datasets	Instances	No: of attributes	Name of attributes	Name of ClassLabel
1	Iris plants	150	4	Sepal Length Sepal Width Petal Length Petal Width	Iris-Setosa Iris-Versicolor Iris-virginica
2	Breast Cancer	699	10	1.Sample code number 2.Clump thickness 3.Uniformity of cell size 4.Uniformity of cell shape 5.Marginal adhesion 6.Single epithelial cell size 7.Bare nuclei 8.Bland chromatin 9.Normal nucleoli 10.Mitoses	Benign Malignant
3	Wine	178	13	1.Alcohol 2.Malic acid 3.Ash 4.Alcalinity of ash 5.Magnesium 6.Total phenols 7.Flavanoids 8.NonFlavanoids phenols 9.Proanthocyanins 10.Color intensity 11.Hue 12.OD280/OD314 of diluted wine 13.Proline	Types of wine 1. 2. 3.

4. System Design and Architecture

Figure 1 shows the overview design of the system. In this system, FPSO clustering and FCM clustering methods are used to cluster Iris plant, Breast Cancer, and Wine datasets respectively. Moreover, the performances of each method on each dataset are also compared and analyzed.

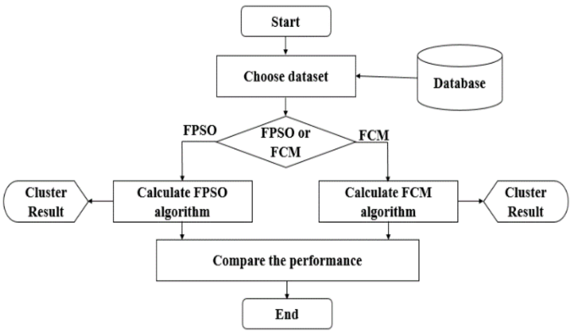

Figure 1. Overview Design of the System

5. Implementation

The implementation of the proposed system is presented in this section. These two methods were implemented based on java programming language with Microsoft Access 2013 database.

5.1 FPSO

Firstly, user enters iterations and cluster number to process the system with random cluster centers (centroids) from databases and calculate Euclidean distance from data element and a particular cluster according to equation 10. Then in FPSO, fuzzy objective function (J_m) of FCM can be used as fitness function. To calculate fitness, membership matrix in equation 11 must be first computed. Then, the fuzzy objective function (J_m) is calculated in equation 9, the fitness can be taken by $1/J_m$. And find P_{best} and P_{gbest} . Then, update the cluster center according to velocity updating in equation 2 and coordinate updating in equation 3. The system program continues until stopping criterion is complete. When the stopping criterion is complete, cluster output results are displayed. Figure 2 shows the flow of FPSO clustering method.

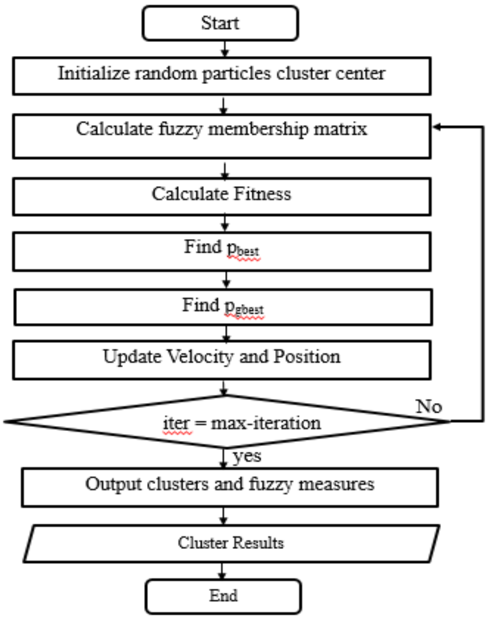

Figure 2. Flow of FPSO clustering method.

5.2 FCM

In the FCM, random cluster centers and membership matrix are initialized. And calculate cluster center according to equation 12 and update membership matrix in equation 11. Then the system calculates fuzzy objective function according to equation 9. And the system continues until stopping criterion is complete. When the stopping criterion is complete, the cluster result are outputted. The detailed flow of FCM method is illustrated in Figure 3.

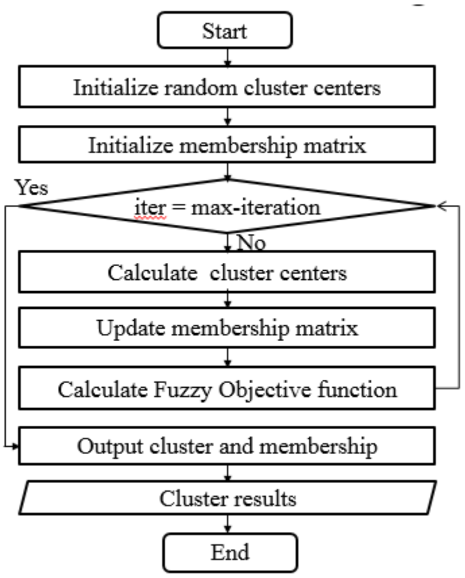

Figure 3. Flow of FCM clustering method

6. Experimental Result

To evaluate the performance of the system, experimental result of the system is shown in Tables. Table 2, 3, 4 show the compare the result of fuzzy objective function and execution time of FPSO and FCM clustering methods for each dataset depend on iteration with particle 30. These system is optimized with particle 30, iteration 100 for 150 records of Iris plants dataset, particle 30, iteration 30 for 699 records of Breast Cancer dataset and particle 30, iteration 10 for 178 records of Wine dataset respectively. Table 5, 6, 7 show the compare the result of fuzzy objective function and execution time of FPSO and FCM for datasets depend on particles with iteration 50. These system is optimized with iteration 50, particle 30 for 150 records of Iris plants dataset, iteration 50, particle 40 for 699 records of Breast Cancer dataset and iteration 50, particle 30 for 178 record of Wine dataset.

Table 2.

Iris Plants dataset					
Particle		30			
Iteration		10	30	50	100
FPSO	J-measure	67.68	66.98	67.56	66.66
	execution time (seconds)	0.04	0.25	0.34	1.07
FCM	J-measure	69.82	70.76	70.48	70.59
	execution time (seconds)	1.105	1.125	1.14	1.2

Table 3.

Breast Cancer dataset					
Particle		30			
Iteration		10	30	50	100
FPSO	J-measure	1505.7	1480.9	1496.3	1508.2
	execution time(seconds)	0.13	0.59	0.705	0.78
FCM	J-measure	1554.3	1554.6	1551.55	1554.1
	execution time(seconds)	1.125	1.255	1.235	1.303

Table 4.

Wine dataset					
Particle		30			
Iteration		10	30	50	100
FPSO	J-measure	12165.6	12257.6	12379.6	12213.2
	execution time(seconds)	0.16	0.344	0.735	1.125
FCM	J-measure	12472.5	12436.4	12455.4	12433.8
	execution time(seconds)	1.175	1.162	1.243	1.335

Table 5.

Iris Plants dataset				
Iteration		50		
Particle		30	40	50
FPSO	J-measure	66.93	67.79	68.03
	execution time(seconds)	0.53	0.566	0.66
FCM	J-measure	70.03	70.37	70.27
	execution time(seconds)	1.155	1.15	1.15

Table 6.

Breast Cancer dataset				
Iteration		50		
Particle		30	40	50
FPSO	J-measure	1516.4	1487.26	1497.7
	execution time(seconds)	0.69	0.937	1.016
FCM	J-measure	1556.7	1552.6	1552.1
	execution time(seconds)	1.22	1.225	1.24

Table 7.

Wine dataset				
Iteration		50		
Particle		30	40	50
FPSO	J-measure	12190.9	12226.9	12378.9
	execution time(seconds)	0.84	0.84	1.11
FCM	J-measure	12351.1	12479.5	12517.1
	execution time(seconds)	1.36	1.271	1.35

7. Conclusion

The objective of this study is to create an effective tool for building fuzzy PSO clustering algorithm (FPSO) and fuzzy C-means (FCM) clustering algorithm to help us making a proper clustering of various classes of datasets. PSO is global stochastic tool which could be implemented and applied easily to solve various function optimization problems. FPSO algorithm could be used for finding global optimal solutions. FCM is one of the most popular fuzzy clustering techniques and is easily trapped in local optima. A detailed comparison between FPSO and FCM clustering algorithms showed that the model constructed from FPSO clustering algorithm is much more efficient than other model based on the FCM clustering algorithm in terms of the execution time and fuzzy objective function.

REFERENCES

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

[2] Esmaeil Mehdizadeh. ISSN 1750-9653, England, UK. *A fuzzy clustering PSO algorithm for supplier base management*. International Journal of Management Science and Engineering Management. Vol.4 (2009) No.4, pp. 311-320

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

[4] Hesam Izakian, Ajith Abraham. *Expert Systems with Applications* 38(2011) 1835-1838

[5] J.C. Bezdek (1981): "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York

[6] J.C. Dunn (1973): "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *Journal of Cybernetics 3:32-57*

[7] Kennedy,J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of the IEEE international conference on neural networks* (pp. 1942-1948)

[8] Kennedy,J., & Eberhart, R., "Swarm intelligence", *Morgan Kaufmann Publishers, Inc.*, 2001, San Francisco, CA.

[9] Pang, W., Wang, K., Zhou, C., & Dong, L. (2004). Fuzzy discrete particle swarm optimization for solving traveling salesman problem. *In Proceedings of the fourth international conference on computer and information technology* (pp. 796-800). IEEE CS Press.

[10] UCI Machine Learning Repository. [http://archive.ics.uci.edu/ml/]. Irvine, CA: University of California, Center for Machine Learning and Intelligent Systems.