Performance of the Fuzzy Clustering using Pattern Reduction Technique

Saranya T[1], Gayathridevi P[2]
[1]*Research Scholar*
[2]*Asst.Professor*
[1, 2] *PG and Research Department of Computer Science*
[1,2] *Sengunthar College of Arts and Science, Tiruchengodu,*
[1,2] *Namakkal, Tamilnadu, India*
(E-mail: saranya88mail@gmail.com)

***Abstract*—**the conventional fuzzy clustering algorithms are taking more computation time to complete their process dure to the extremely large datasets. The conventional clustering algorithm fails to scale up well when treatment the large datasets. Therefore, to reduce the computation expensive of the traditional clustering algorithms, a pattern reduction technique is used to reduce the computation cost of the fuzzy clustering algorithm. In this research work, the fuzzy particle swarm optimization (FPSO) algorithm is an integrated with pattern reduction algorithm in order to accelerate computation performance of the FPSO. The experimental results indicate that the proposed method can significantly reduce the computation time of the PR+FPSO clustering algorithm and the proposed fuzzy algorithm is to contrast with Fuzzy C-Means (FCM) and FPSO.

Keywords*—Fuzzy Clustering; Fuzzy C-Means; Fuzzy Particle Swarm Optimization; Pattern Reduction; Computational Cost*

INTRODUCTION

Clustering is the process of grouping the objects from the given datasets based on the similarities between the objects. A high similarity data object belongs to similar group and others in dissimilar groups. The clustering problem is a NP- Hard problem to solve efficiently [1]. There are two types of clustering methods such as hard clustering and fuzzy clustering. The hard clustering assign the data objects into single cluster.

In fuzzy clustering, the data objects are assigned to different group with the help of their membership value. Fuzzy clustering is a well-known method applied to solve many real world problems. In which, Fuzzy C-Means is a most popular unsupervised clustering algorithm developed by Bezdek and it is easy to implement, simple [2]. However, The FCM algorithm is falling into local optima due to an initialization of cluster center randomly.

To solve the drawbacks of convention clustering algorithms, many optimization algorithms have been employed to solve clustering problem such as Simulated Annealing (SA), Differential Evolution (DE), Genetic algorithms (GA), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), and Artificial Bee Colony (ABC) Clustering algorithm. Among them, the fuzzy PSO clustering techniques has been developed for solving clustering problem in efficient manner [3].

In this paper, concentrates on fuzzy particle swarm optimization clustering algorithm is well-famous optimization methods which is applied for solving clustering problem. Conversely, in the clustering problem, the machine learning algorithms can lead to a high computational rate, poorer convergence rate and lessening the performance due to the high dimensionality and huge of data samples in datasets[6].

Hence, in this paper, we concentrates on pattern reduction methods in order improve the generalization performance of the fuzzy clustering algorithm. The concept of the pattern reduction is process of removing the data sample which was located in sample group at more than two or three iteration. The pattern reduction is a significant way to reduce the computation time of the clustering algorithm.

The pattern reduction method [4] is an incorporated with fuzzy PSO is proposed in this paper and it is used to reduce the computational efficiency. The objective of this paper is to remove the static patterns to decrease the computation complexity with maintain the quality. The motivation of the research work is to improve the computation efficiency of the FPSO with the help of pattern reduction techniques.

The remainder of the paper is structured as follows: the details of the methods are described in Section 2, experimental results presented and discussed in Section 3. We conclude the paper in Section 4.

METHODOLOGY

Fuzzy C-Means

First The Fuzzy C-Means algorithm is an iterative and well famous partitioning clustering algorithm that find clusters in data and it is use the concept of fuzzy membership rather than assigning a pixel to a single cluster, each pixel will have diverse membership values on each cluster [1]. Given an unlabeled data set $O = \{O_1, O_2, ...O_n\}$ in $\mathbb{R}^d$ dimensional space, the fuzzy c-means is the partitions data set into *c (*$1 < c < n$*) and* $z = \{z_1, z_2, ..., z_n\}$ cluster centers. The fuzzy clustering of objects is illustrates by a fuzzy matrix μ with n rows and c columns. Here, n is the number of data points and c is the number of clusters. μ_{ij} is the element in the i^{th} row and j^{th} column in μ.

The distinctiveness of μ is as follows:

$$\forall_i \ [0, 1] \ \ \forall_i = 1, 2 \ldots n \ \ \forall_j = 1, 2 \ldots c \quad (1)$$

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

$$o < \sum_{j=1}^{c} \mu_{ij} < n \quad \forall_j = 1, 2 \ldots c \quad (3)$$

In the FCM, The objective function is to minimize the error between the data samples by using following equations,

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij} d_{ij} \quad (4)$$

Where,

$$d_{ij} = \left| o_i - z_j \right| \quad (5)$$

Form the above equation, d_{ij} is the Euclidian distance from object $O = \{O_1, O_2, \ldots\ldots O_n\}$ to the cluster center z_j. The $z = \{z_1, z_2, \ldots., z_n\}$, centroids of the j^{th} cluster is achieved using below equation

$$z_j = \frac{\sum_{i=1}^{n} \mu_{ij}^{m} o_i}{\sum_{i=1}^{n} \mu_{ij}^{m}} \quad (6)$$

Algorithm 1: Fuzzy C- Means

Step 1: Initialization process

Step 2: Compute the cluster centers by using (6)

Step 3: Compute Euclidian distance by using (5)

Step 4: Update the membership function by using the following equations,

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{m-1}}} \quad (7)$$

Step 5: If not meet convergence, go to step 2.

Fuzzy Particle Swarm Optimizations

The A modified particle swarm optimization for TSP is called fuzzy particle swarm optimization (FPSO)[2]. In FPSO method, the position and velocity of particles is redefined to signify the fuzzy relation between variables.

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

$$X = \begin{pmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{pmatrix} \quad (8)$$

Here, μ_{ij} is the membership function of the i^{th} object with the j^{th} cluster with limitations stated as Eq. (1) and Eq. (2). Consequently, the velocity of each particle is stated using a matrix with the size of n rows and c columns. Those are in range between -1 and 1.

The updating the positions and velocities of the particles are based on the following equation,

$$V(t+1) = w \otimes V(t) \oplus c_1 r_1 \otimes pbest(t) \Theta c_2 r_2 \otimes gbest(t) \Theta X(t) \quad (9)$$

$$X(t+1) = X(t) \oplus V(t+1) \quad (10)$$

After updating the position matrix, it may break the constraints given as Eq. (1) and Eq. (2). Hence, it is required to normalize the position matrix. First, the matrix is needed to make all the negative elements as a zero.

If all elements in a row of the matrix are zero, they require evaluating with the help of series of random numbers that ranges between 0 and 1. Then, the matrix is to undertake the following transformation without violating the constraints as follows,

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

In FPSO, fitness function is requiring a function for estimating the generalized solutions. The fitness function is defined for fuzzy clustering as follows,

$$f(X) = \frac{K}{J_m} \quad (12)$$

Here, K is a constant and J_m is the objective function as defined in the FCM algorithm (Eq. 4). The least value of objective function is considered as the better accuracy and the highest value of the objective function is defined as the worst.

Algorithm 3: Fuzzy PSO for fuzzy clustering:

Step 1: Initialize the parameters including population size P, c1, c2, w, and the maximum iterative count.

Step 2: Create a P particles with swarm (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: Calculate the cluster centers for each particle using by Eq. (6).

Step 5: Calculate the fitness value of each particle using by Eq. (12).

Step 6: Calculate pbest for each particle.

Step 7: Calculate gbest for the swarm.

Step 8: Update the velocity matrix for each particle using by Eq. (10).

Step 9: Update the position matrix for each particle using by Eq. (11).

Step 10: If not met the termination conditions, 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.

Pattern Reduction based Fuzzy Particle Swarm Optimizations

The most of the data set has large data samples that may take more computation time to completing their learning process. In this circumstance, paying concentration to reduce the computation cost of the machine learning algorithm is an essential task.

In this research work, concentrates the pattern reduction techniques to reduce the computation time of the Fuzzy PSO clustering algorithm. The goal of the pattern reduction algorithm is to avoid the redundant or repeat calculation in the machine learning algorithm[3].

Generally, the pattern reduction is performed as two tasks,

- Pattern compression
- Pattern removal

The pattern reduction compresses and removes the patterns by the choosing of the pattern to be removed. Then, the pattern $D[r_i^l]$, where $r_i^l \in \square_i^l$ as representative pattern and setting the values its average of all patterns are removed,

$$D[r_i^l] = \frac{\sum_{j=1}^{\left|\square_i^l\right|} D[r_{ij}^l]}{\square_i^l} \tag{13}$$

Since, the computation cost of the machine learning algorithms will be the high due to the repetition calculation. Then, the each and every pattern removed, $z \in \square_i^l$ and the values $M\left[z\right]$ is set to zero and it is defined as follows,

$$M\left[z\right] = \begin{cases} \square_i^l & if\ z \in \square_i^l\ and\ z = r_i^l \\ 0 & if\ z \in \square_i^l\ and\ z \neq r_i^l \end{cases} \tag{14}$$

The essential of the pattern reduction techniques is to keep tract the average values and the removed patterns. Finally, cluster center updates as follows,

$$C_i^l = \frac{\sum_{j=1}^{\left|C_i^l\right|} M[s_{ij}^l] \times D[s_{ij}^l]}{\sum_{j=1}^{\left|C_i^l\right|} M[s_{ij}^l]} \tag{15}$$

The calculating the objective functions of fuzzy clustering algorithm as follow,

$$J_m = \sum_{i=1}^{k} \sum_{j=1}^{\left|C_i^l\right|} \left\| D\left[s_{ij}^l\right] - C_i^l \right\|^2 \times M[s_{ij}^l] \tag{16}$$

Algorithm 4: Fuzzy PR+PSO for fuzzy clustering:

Step 1: Initialize the parameters including population 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: Calculate the cluster centers for each particle using by Eq. (6).

Step 5: Calculate the fitness value of each particle using by Eq. (12).

Step 6: Detect the set of static patterns using Eq. (13)

Step 6: Compress the patterns and remove pattern form data sets Eq. (15)

Step 6: Calculate pbest for each particle.

Step 7: Calculate gbest for the swarm.

Step 8: Update the velocity matrix for each particle using by Eq. (9).

Step 9: Update the position matrix for each particle using by Eq. (10).

Step 10: If convergence 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.

EXPERIMENTAL RESULTS AND DISCUSSIONS

The experimental results conducted different independent run according to the clustering algorithm, for the FCM, 100 independent runs, 10 independent runs for FPSO and PR based FPSO. The simulation results are summarized in Table 1-3.

The results of clustering algorithm have calculated by using objective function as followed by Equation (4) and the minimum value of the objective function is considered as a best optimal solution which is highlighted in bold letters.

The proposed Pattern Reduction based Fuzzy Particle Swarm Optimization clustering algorithm obtained superior results compared with FPSO and FCM clustering algorithm in all data sets and it can ability to find both local and global minima. Experimental results are conducted using MATLAB 2015.

Details of the Datasets

To evaluate the performance of the proposed clustering algorithms, an investigation carried out on the five type of well-known data sets which is obtained from UCI repository of machine learning database [4].

- The iris dataset has three different class of iris flower and 150 data samples were collected in each class with four features.
- The Glass datasets consists of 214 data objects with 6 different types of class. Each class has nine features.

Results

In order to evaluate the performance of the proposed model, the comparison made against FCM, FPSO, PR based FPSO clustering algorithm. The following tables are discussed about the proposed models and comparisons algorithms such as FCM, FPSO, and proposed PR+FPSO.

PERFORMANCE COMPARISON OF THE PROPOSED METHOD FOR IRIS DATASETS

Algorithms	**Iris Dataset**		
	Worst	***Average***	***Best***
FCM	72.58	70.70	68.82
FPSO	68.21	67.81	67.42
PR + FPSO	**66.32**	**66.11**	**65.91**

PERFORMANCE COMPARISON OF THE PROPOSED MODEL FOR GLASS DATASETS

Algorithms	**Glass Dataset**		
	Worst	***Average***	***Best***
FCM	75.42	74.39	73.37
FPSO	73.22	74.39	72.57
PR+ FPSO	**71.44**	**70.66**	**69.89**

From the TABLE I, it shows the performance of the proposed model for Iris dataset. The performance of the proposed model is produced 6.2 % of enhanced results compared to FCM and 1.8 % of enhanced results compare to FPSO. the graphical representation shows in Figure 1.

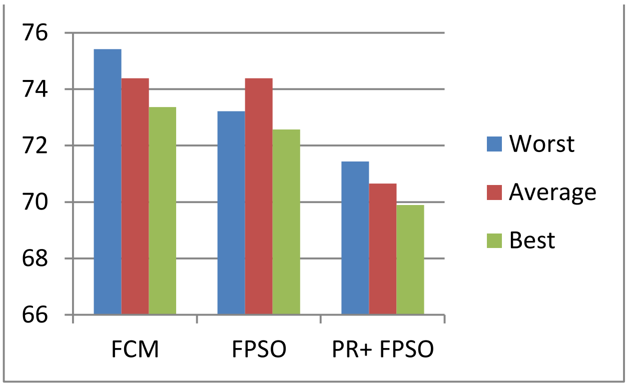

Figure 1: Performance Comparison of the proposed model for iris dataset

From the TABLE II, it shows the performance of the proposed model for Glass Datasets. The performance of the proposed model is produced 3.78 % of enhanced results compared to FCM and 1.78 % of enhanced results compare to FPSO. Figure 2 illustrates the performance of the proposed model for glass datasets.

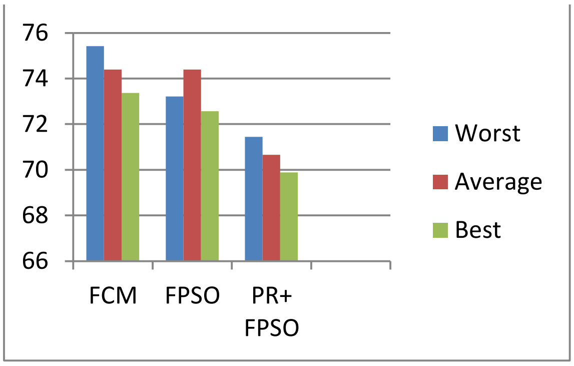

Figure 2: Performance Comparison of the proposed model for Glass dataset

PERFORMANCE COMPARISON OF THE PROPOSED MODEL BASED ON COMPUTATION COST

Datasets	**Glass Dataset**		
	FCM	***FPSO***	***PR+FPSO***
Iris	364.59	281.92	**246.03**
Glass	428.71	392.74	**324.24**

REFERENCES

[1] S. C. Cohen and L. N. de Castro, "Data clustering with particle swarms," in Evolutionary Computation, 2006. CEC 2006. IEEE Congress on, 2006, pp. 1792-1798: IEEE.

[2] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," Computers & Geosciences, vol. 10, no. 2-3, pp. 191-203, 1984.

[3] T. A. Runkler and C. Katz, "Fuzzy clustering by particle swarm optimization," in Fuzzy Systems, 2006 IEEE International Conference on, 2006, pp. 601-608: IEEE.

[4] K. Alsabti, S. Ranka, and V. Singh, "An efficient k-means clustering algorithm," 1997.

[5] H. Izakian and A. Abraham, "Fuzzy C-means and fuzzy swarm for fuzzy clustering problem," Expert Systems with Applications, vol. 38, no. 3, pp. 1835-1838, 2011.

[6] M.-C. Chiang, C.-W. Tsai, and C.-S. Yang, "A time-efficient pattern reduction algorithm for k-means clustering," Information Sciences, vol. 181, no. 4, pp. 716-731, 2011.

[7] S. Hettich, C. L. Blake, and C. J. Merz, "UCI repository of machine learning databases," ed, 1998.