# Comparison of M-FPSO and C-FPSO Hybrid Methods in Clustering

Y. Ortakcı, C. Göloğlu

***Abstract*—Fuzzy C-Means (FCM) algorithm is one of the most common algorithms used in solving clustering problems. Nevertheless, FCM has a disadvantage of suffering from getting stuck on local optimum easily due to randomly selected starting points. FCM can be supported by Evolutionary Algorithms (EA) to overcome this disadvantage. In this paper, a hybridization method of FCM and Particle Swarm Optimization (PSO) is described in order to increase clustering performance and speed. In this hybrid method, two different methods, which are based on membership values (M-FPSO) and cluster center points (C-FPSO), are utilized in clustering process. It is seen from the computational results that M-FPSO is superior to C-FPSO in clustering performance, but C-FPSO shows fast convergence.**

## I. INTRODUCTION

CLUSTERING can be defined as a separation process of unlabeled data sets into different groups without any supervision. A group should contain the data on similar features; however the data in the other groups should be different as much as possible. Therefore, the aim of clustering is to separate a data set with *n* points to *k* clusters. While this differentiation is made, *k* clusters should be different from one another and each cluster should house similar data [1, 2].

Similarly, clustering algorithms separate data sets to different independent clusters by transforming clustering problems to a kind of minimization problems using some criteria like squared error function. From this point of view, a clustering problem can be handled as an optimization problem [3]. In clustering data sets two basic points are taken into consideration:

- A similarity measure, which will be used to cluster data sets, should be determined. This similarity measure should guarantee the data points in the same cluster that must be similar in one another, as well as the data points from different cluster must be different as much as possible. The most common similarity measure used is Euclidian distance measure.
- A method, which will cluster data set most quickly and right way, should be determined.

In clustering algorithms two different methods can be used based on the relationship of data points belonging to clusters: Crisp and Fuzzy clustering methods. In crisp clustering method, data belongs only to one cluster whereas in fuzzy clustering data can belong to more than one cluster with different membership values.

In solving clustering problems, many clustering algorithms are used. Among them, Fuzzy C-Means (FCM) local learning algorithm is one of the most commonly used clustering algorithm. However, FCM has a disadvantage such as getting stuck on local optimum easily [4]. FCM can be supported by evolutionary algorithms like Particle Swarm Optimization (PSO) to overcome this disadvantage. In this paper, the combination of FCM and PSO will be accomplished with two different ways and the results obtained will be compared.

## II. FUZZY C-MEANS ALGORITHM

FCM which is derived from K-Means clustering algorithm is firstly produced by Dunn and developed by Bezdek in 1989 for pattern recognition studies [5]. FCM algorithm uses fuzzy clusters and fuzzy membership values in fuzzy concept. Unlike a crisp clustering algorithm, a data point can be owned by different clusters with different membership values in FCM. It can be formulated as follows:

$A = \{ a_1 , a_2, \dots , a_i , \dots , a_n\}$ is a data set with $n$ elements and $a_i = \{a_{i,1}, a_{i,2}, \dots , a_{i,d}\}$ is a vector with $d$ dimension. $n,\ d,\ k \in N$ and $1 < k < n$ and $\mu_{ij}$ is used to cluster $A$ data set to $k$ cluster. $\mu_{ij}$ is membership value of $i$th data point to $j$th cluster. The constraints of $\mu_{ij}$ are as follows [6, 7]:

$$\mu_{ij} \in [0,1];\ \forall\ i = 1,2,\dots,n\ and\ \forall\ j = 1,2,\dots,k; \quad (1)$$

Each membership value must be in [0, 1] interval.

$$\sum_{j=1}^{k} \mu_{ij} = 1;\ \forall\ i = 1,2,\dots,n; \quad (2)$$

C. Göloğlu is with Karabük University, Faculty of Engineering, Department of Mechanical Engineering, 78050 Karabük, TURKEY (corresponding author to provide phone: +90-370-4338200; fax: +90-370-4333290; e-mail: cgologlu@karabuk.edu.tr).

Y. Ortakcı is with Karabük University, Faculty of Engineering, Department of Computer Engineering, 78050 Karabük, TURKEY (e-mail: yasinortakci@karabuk.edu.tr).

The sum of membership data points belongs to $k$ cluster must equal to1.

$$0 < \sum_{i=1}^{n} \mu_{ij} < n; \forall\, j = 1, 2, \ldots, k; \quad (3)$$

The sum of membership values belongs to a cluster must be in (0, n) interval.

The objective function of FCM is:

$$J_{FCM}(M, O; X) = \sum_{j=1}^{k} \sum_{i=1}^{n} {\mu_{ij}}^{m} {d_{ij}}^{2} \quad (4)$$

where $M$ is $(n \times k)$ dimensional $\mu_{ij}$ matrice. $O = \{o_1, o_2, \ldots, o_k\}$ is $(k \times d)$ dimensional matrice which represents center of a cluster. $d_{ij}$ is Euclidian distance between $i$th data point and $jth$ cluster center. $m$ is fuzziness parameter that is used to fuzzify membership values at $m > 1$. Here, the aim of FCM algorithm is to find the minimum objective function value of $J_{FCM}$.

The membership value of $i$th data point to $j$th cluster is calculated according to following equation;

$$\mu_{ij} = \frac{1}{\sum_{c=1}^{k} \left(\frac{d_{ij}}{d_{ic}}\right)^{\frac{2}{(m-1)}}} \quad (5)$$

and the center of a cluster is calculated according to following equation:

$$o_j = \frac{\sum_{i=1}^{n} ({\mu_{ij}}^{m}\, a_i)}{\sum_{i=1}^{n} {\mu_{ij}}^{m}} \quad (6)$$

The pseudo code of FCM can be written as follows [4]:

BEGIN
Define *m* and *c*
Initialize *O* matrice
    FOR *t*=0 TO $t_{max}$
    Calculate *M* matrice according to Eq. (5)
    Update *O* matrice according to Eq. (6)
    Calculate $J_{FCM}$ according to Eq. (4)
    END FOR
END

## III. PARTICLE SWARM OPTIMIZATION

PSO is an evolutionary algorithm which was introduced in 1995 by Kennedy and Eberhart [8]. PSO is a population based algorithm and searches solution space with a swarm started with random values. PSO is inspired by social behaviors of bird flock and fish school while they search food. PSO is formed with particles which each one represents a different solution. Every particle searches multi-dimensional solution space by exploiting its individual experience and the neighborhood particles' experiences. Unlike other evolutionary algorithms, particle in PSO has a velocity component, *V*, as well as position component, *X*. Dynamically updated velocity component enables the particle to change its position by searching solution space comprehensively. Velocity of a particle is updated according to particle's individual experience and neighborhood particles' experience. Particle tends to approach to a better solution by updating velocity. Random parameters are used in velocity update to prevent getting stuck on local optimum [9]. In a *D* dimensional search space, position vector of a particle is $X_i = (x_{i1}, x_{i2}, \ldots, x_{id})$, $g_{best}$ is the particle that has best solution, and $p_{best}$ is the best solution of every particle throughout the search period. Then, speed and position vector of an updated $i$th particle is formulated as follows:

$$V_{ij}^{t+1} = wV_{ij}^{t} + c_1 r_1 \left(P_{ij} - x_{ij}\right) + c_2 r_2 \left(P_{g_{best}} - x_{ij}\right) \quad (7)$$

$$X_{ij}^{t+1} = X_{ij}^{t} + V_{ij}^{t+1} \quad (8)$$

where $t$ is the iteration number; $j = 1, 2, \ldots, d$;

The new position vector $X_{ij}^{t+1}$ offers new solution. $r_1$ and $r_2$ have random values in interval of [0, 1] enabling the randomness of PSO. The pseudo code of PSO is given as follows:

Initialize population with *p* particles
REPEAT
    FOR *i*=1 TO *p*
    Calculate The Objective Function
    Update $p_{best}$
    Update Velocity according to Eq. (7)
    Update Position according to Eq. (8)
    END FOR
Update $g_{best}$
UNTIL Termination Criteria

## IV. PSO BASED FCM (FPSO)

Many clustering alternatives like $C = \{C_1, C_2, \ldots, C_k\}$ can be obtained when clustering a data set $A$ with $n$ data points to $k$ cluster. Process of selection of the best clustering alternative is an optimization problem. As it is mentioned before, the biggest disadvantage of FCM is to be dependant to its starting points. Choosing no right starting points may cause the algorithm to get stuck at local optima in FCM. Therefore, the performance of clustering can be increased by combining FCM and PSO algorithm [7]. Each particle in

PSO offers a clustering alternative. Every clustering alternative is evaluated according to objective function in Eq. (4). Throughout the iterations better clustering solutions are tried to be found among the alternatives.

In combination of FCM and PSO algorithm, two different methods can be employed [10]. In the first method, called as Center based FPSO (C-FPSO), the particles represent the values of cluster centers and are tried to find best cluster centers in FPSO. In the second method, called as Membership based FPSO (M-FPSO), the particles represent the membership values of every data point to every cluster and are tried to find best membership values.

### A. CENTER BASED FPSO (C-FPSO)

In C-FPSO, each particle represents $(k \times d)$ dimensional cluster center matrice, $O$:

$$O = \begin{bmatrix} o_{11} & \cdots & o_{1d} \\ \vdots & \ddots & \vdots \\ o_{k1} & \cdots & o_{kd} \end{bmatrix} \quad (9)$$

$\mu_{ij}$ values are calculated in each iteration according to Eq. (5) [4]. The objective function is calculated according to $\mu_{ij}$ values and $O$ matrice. The pseudo code of the proposed method is given as follows:

```
Initialize O matrice for p particle
FOR i=1 TO t_max
    FOR i=1 TO p
    Calculate M matrice according to Eq. (5)
    Calculate the objective function
    Update Velocity, V
    Update Position, X
    END FOR
END FOR
```

### B. MEMBERSHIP BASED FPSO (M-FPSO)

In M-FPSO, each particle represent $(n \times k)$ dimensional membership matrice $M$:

$$M = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1k} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu o_{nk} \end{bmatrix} \quad (10)$$

Center of clusters, $o_j$ values are calculated according to Eq. (6) [11]. The objective function is calculated according to $o_j$ values and $M$ matrice. The pseudo code of the proposed method is given as follows:

```
Initialize M matrice for p particle
FOR i=1 TO t_max
    FOR i=1 TO p
    Calculate O matrice according to Eq. (6)
    Calculate the objective function
    Update Velocity, V
    Update Position, X
    END FOR
END FOR
```

## V. EXPERIMENTAL RESULTS

In the paper, the performance of M-FPSO and C-FPSO are evaluated [12]. The codes of two methods are written in Microsoft Visual Studio 2008 C#. The codes are run in a notebook, which has Intel Core2 Duo 2.63 GHz processor and 4GB RAM on Windows 7 operating system. The application parameters used are acceleration coefficients ($c_1 = 2.0$ and $c_2 = 2.0$); inertia ($w = 0.75$); maximum iteration number ($t_{max} = 1000$); limit of velocity ($V_{max} = 1\%$); population size (50); and fuzziness parameter ($m = 2.0$).

In the experimental runs, iris dataset is used for classification. Iris flower dataset has four dimensional 150 data points. Every data point keeps the length and width of pedal and the length and width of sepal. The iris flower dataset comprises of three clusters. The name of cluster and the number of element is shown in Table 1.

TABLE 1
IRIS FLOWER CLUSTERS

| CLUSTER NAME | NUMBER OF ELEMENT |
|---|---|
| IRIS SETOSA | 50 |
| IRIS VIRGINICA | 50 |
| IRIS VERSICOLOR | 50 |

The codes for both M-FPSO and C-FPSO are run 20 times. The objective function results are given in Table 2.

TABLE 2
THE OBJECTIVE FUNCTION RESULTS

| DATA SET | METHODS | BEST | WORST | MEAN | $\frac{\text{Mean} - \text{Best}}{\text{Best}}$ | $\frac{\text{Mean} - \text{Worst}}{\text{Worst}}$ |
|---|---|---|---|---|---|---|
| IRIS | M-FPSO | 64.949 | 73.690 | 67.205 | 0.034 | -0.088 |
| | C-FPSO | 60.505 | 60.505 | 60.505 | 0.0 | 0.0 |

The clustering error rate (CE) is calculated according to Eq. (6) below. CE results are shown in Table 3. The results are taken from the best run results in 20 runs for both methods.

$$CE = \left(\frac{wrong\ clustered\ data}{total\ data\ numbers}\right) \quad (6)$$

TABLE 3
CLUSTERING RESULTS

| Methods | Cluster Name | 1st Cluster | 2nd Cluster | 3rd Cluster | CE |
|---|---|---|---|---|---|
| M-FPSO | Setosa | 50 | 0 | 0 | 6.00% |
| | Virginica | 0 | 3 | 47 | |
| | Versicolor | 0 | 44 | 6 | |
| C-FPSO | Setosa | 50 | 0 | 0 | 10.66% |
| | Virginica | 0 | 37 | 13 | |
| | Versicolor | 0 | 3 | 47 | |

According to Table 2, C-FPSO's standard deviation is 0.0. Therefore, C-FPSO is more stable than M-FPSO. C-FPSO has the same results in 20 runs and the objective function value in C-FPSO is lower than of M-FPSO. However, in Table 3, CE value of M-FPSO is lover than of C-FPSO. That shows the clustering scatter of M-FPSO is closer to the real clustering scatter. The scatter of iris dataset according to M-FPSO is shown in Fig. 1. The 3D figure was plotted according to three dimensions of iris data set.

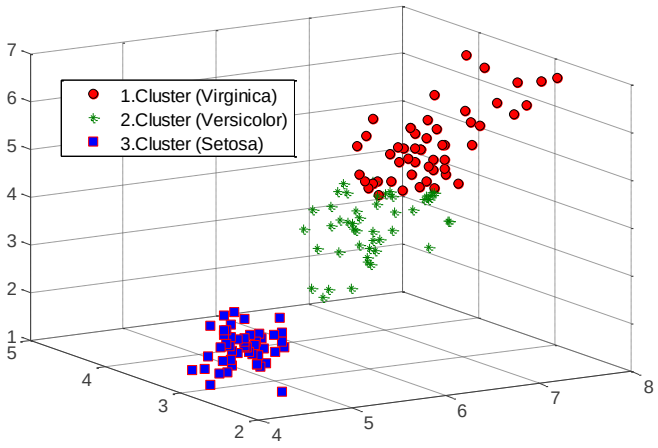


Fig. 1. The scatter of iris dataset

## VI. CONCLUSION

The application shows that clustering performance of M-FPSO is better than of C-FPSO for $t_{max} = 1000$. But C-FPSO is more stable with 0.0 standard deviation value than of M-FPSO. As it is shown in the Fig. 2, while C-FPSO convergences to minimum results at about the iteration of 100. M-FPSO convergences to the minimum result at about the iteration of 1000.

The different convergence performances are due to particle structure of M-FPSO and C-FPSO. In M-FPSO the particle represents $(n \times k)$ dimensional membership martice $(M)$. For the iris dataset, $M$ matrice is $(150 \times 4)$ dimensional. In C-FPSO the particle represents $(k \times d)$ dimensional cluster center matrice $(O)$. For the iris data set $O$ matrice is $(3 \times 4)$ dimensional. Therefore, C-FPSO convergences the optimum results quicker than M-FPSO for the iris data set used. So, the performance of clustering for both methods depends on dimension of particles. As result, if the number of elements in data set is larger than the cluster number, it is logical to use C-FPSO, otherwise it is logical to use M-FPSO. Meanwhile, the implementation on real engineering problems of the study is on the way.

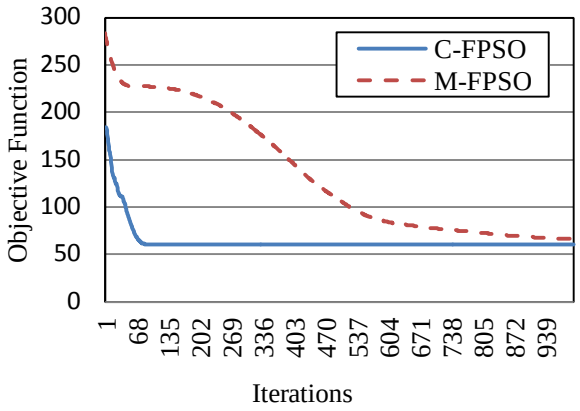


Fig. 2. The convergence curves of both methods

## REFERENCES


[1] S. Das, A. Abraham, A. Konar, "Automatic Clustering Using an Improved Differential Evolution Algorithm", *IEEE Trans., Man and Cybernetics, Part A: Systems and Humans*, 38 (1), pp. 218-237, 2008

[2] C. Y. Chen, H. M. Feng, F. Ye, "Automatic Particle Swarm Optimization Clustering Algorithm", *International Journal Of Electricial Engineering*, 13 (4), pp. 379-387, 2005

[3] M. G. H. Omran, A. P. Engelbrecht, A. Salman, "Dynamic Clustering using Particle Swarm Optimization with Application in Unsupervised Image Classification", *World Academy of Science*, pp. 199-204 2005

[4] H. Tang, B. Ding and W. Qi, "Research on traffic mode of elevator applied fuzzy C-mean clustering algorithm based on PSO", *International Conference on Measuring Technology and Mechatronics Automation*, pp. 582-585, 2009

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

[6] W. Wang, Y. Zhang, Y. Li and X. Zhang, "The Global Fuzzy C-Means Clustering Algorithm", *Proceedings of the 6th World Congress on Intelligent Control and Automation*, Dalian Chine, pp. 3604-3607, 2006

[7] H. Izakian, A. Abraham and V. Snášel, "Fuzzy Clustering Using Hybrid Fuzzy c-means and Fuzzy Particle Swarm", *World Congress on Nature & Biologically Inspired Computing*, pp. 1690-1694,2009

[8] J. Kennedy, R. Eberhart, "Particle Swarm Optimization" , *Proc. IEEE Int. Conf. on 4th Neural Networks*, Piscataway, NJ:IEEE Service Center, pp. 1942-1948, 1995

[9] R. Eberhart, J. Kennedy, "A New Optimizer Using Particle Swarm Theory", In *6th International Symposium on Micro Machine and Human Science*, Nagoya, pp. 39-43, 1995

[10] T. A. Runkler and C. Katz, "Fuzzy Clustering by Particle Swarm Optimization", *IEEE International Conference on Fuzzy Systems*, BC, Canada, pp. 601-608, 2006

[11] J. Yih, Y. Lin and H. Liu, "Clustering Analysis Method Based On Fuzzy C-Means Algorithm of PSO And PSO With Application in Real Data", *International Journal Of Geology*, 1, pp. 89-98, 2007.

[12] Y. Ortakcı, "Solving Engineering Problems with Particle Swarm Optimization", M.Sc. Thesis, Dept. Computer Engineering, Graduate School of Natural and Applied Sciences, Karabük University, Turkey, 2011.