WALIA journal 31(S3): 245-252, 2015
ISSN 1026-3861

A new method of fuzzy clustering by using the combination of the firefly algorithm and the particle swarm optimization algorithm

Marzieh Haghdoost [1,*], Ali Haroun Abadi [2], Seyed Javad Mir Abedini [2]

[1]*Department of Computer Software, Bushehr branch, Islamic Azad University, Bushehr, Iran*
[2]*Department of Computer Software, Central Tehran Branch, Islamic Azad University, Tehran, Iran*

Abstract: Fuzzy clustering algorithm is one of the data mining methods that is applied in different fields. According to the fuzzy clustering algorithm, each object is allocated to the clusters regarding its percentage of belonging to each of the clusters. Finding the cluster centers is one of the main objectives of the clustering and it is possible to apply the swarm intelligence methods in order to accurately find the cluster centers. The swarm intelligence algorithms have separately been applied for clustering and they received the optimized solution. In order to solve the problems of the fuzzy clustering algorithm, the combined method based on the firefly algorithm and the particle swarm optimization algorithm is applied. In order to determine the validity, the suggested method is tested on four standard data collections received from the valid site of UCI. The simulation results shows that the combination of two algorithms do the clustering more accurately than applying each of them separately.

Key words: *Firefly algorithm; Particle swarm algorithm; Data clustering; Fuzzy c-means*

Introduction

Clustering is one of the data mining methods used to analyze data. Regarding its significance, some algorithms are presented to improve the clustering. The issue that most matters is determining the clusters centers. In order to obtain correct and optimized clustering, many algorithms have been applied, however, each of them have had some weaknesses, though, it is been attempted to improve them. For insistence, an algorithm based on the optimization of ant colony is represented and it attempts to find the optimal cluster center. Moreover, a method is represented to get rid of the local minimums (1). furthermore, the bee algorithm is applied for clustering, and then it is compared to several other intelligent algorithms such as the genetic algorithm to indicate its superiority over them as a result of the quality of its solution and the less processing time needed for the clustering(2). The combination of firefly algorithm with the k-harmonic-means algorithm are applied to obtain better numbers of the cluster and getting rid of the local optima (3). In another attempt, the firefly algorithm is combined with the k-means algorithm in order to find the correct and more accurate centers for the clusters in an optimized way. moreover, a new method was represented for diagnosing diabetes by combining the artificial intelligent methods such as the genetic algorithm for decreasing the quality and fuzzy systems, making instantaneous decisions, and FCM(5). Similarly, a new method was represented by combining the artificial bee colony algorithm and the particle swarm optimization algorithm in order to cluster data (6). The firefly algorithm is applied to decrease the problems related to clustering, and finally it is compared with bee algorithm and the artificial bee colony algorithm (7). Moreover, the combination of the ant algorithm and the particle swarm algorithm is represented for clustering and the simulated results shows that the suggested algorithm is much better than the ant algorithm and the particle swarm algorithm each applied separately.

Fuzzy clustering

The main purpose of clustering is to discover the structure of collected data. The clustering analysis intends to organize a collection of data in a series of the clusters so that the data in each of the clusters could have the highest degree of similarity and the data related to different clusters would have the maximum degree of dissimilarity. In fact, clustering is a process that receives a set of input data and categorizes them into several subsets. In clustering, dividing N data (pattern) into C(clusters) provides the possibility of creating many parts. Therefore, some of the optimization techniques which are called clustering methods should be applied. Generally, the clustering methods could be categorizes into three main groups: the hierarchical clustering, the model based clustering, and the categorization based clustering which is also called the objective function clustering.

The purpose of fuzzy c-means clustering algorithm is to cluster data so that all of the data could have the minimum distance to the centers (9). In other words, during the clustering process, the centers are changed so that the data could have the

* Corresponding Author.

minimum distance to the centers. The algorithm is based on uncertainty and each of the data in the different clusters has the degree of membership varying from zero to one. The clustering minimizes the distance between data and centers and it defines the cost function as follows:

$$Q = \sum_{i=1} \sum_{k=1} u_{ik}^{m} \|x_k - v_i\|^2 \quad (1)$$

Where xk stands for the k data in X matrix , vi stands for the i center in V matrix, uik stands for the degree of membership in K data in the i cluster, and m stands for the fuzzy coefficient. This function affects the distance of the point to the center of cluster on the cost regarding the degree of membership of the data in each cluster. Moreover, it considers the sum of it for all the data as the cost for clustering. In relation 1, symbol II stands for the distance function, it calculates the distance of data K to the center of cluster I, therefore, it is defined as:

$$\|x_k - v_i\|^2 = \sum_{j=1}^{\dim} (x_{kj} - v_{ij})^2 \quad (2)$$

Dim stands for the dimension of the input data, vij stands for the j dimension of the i cluster in V matrix, and Xkj stands for the kj dimension which is related to matrix X. Categorization (classification) matrix should have two following conditions:

$$0 < \sum_{k=1}^{N} u_{ik} < N, \qquad i = 1, 2, \ldots, c$$

$$\sum_{i=1}^{c} u_{ik} = 1, \qquad k = 1, 2, \ldots, N$$

It means that the sum of degrees of membership of each point in the whole cluster should be equal to one and the sum of degrees of membership of all data in each cluster must be between zero and the number of data. For optimizing, the objective function 1 should calculate its derivatives regarding U and V. the derivative of cost function regarding the cluster centers is defined as:

$$v_{st} = \frac{\sum_{k=1}^{N} u_{ik}^{m} x_{kt}}{\sum_{k=1}^{N} u_{ik}^{m}} \quad (3)$$

Before calculating, the categorization matrix with limitation is added to the cost function and the new function of its derivative is calculated as follows:

$$V = \sum_{i=1}^{c} u_{ik}^{m} d_{ik}^{2} + \lambda \left(\sum_{i=1}^{c} u_{ik} - 1 \right) \quad (4)$$

$$u_{st} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{st}^2}{d_{jt}^2} \right)^{\frac{1}{m-1}}} \quad (5)$$

The fuzzy c-means clustering algorithm could minimize the cost function in a repeated way. This algorithm starts from an initial value for the categorization matrix, and then it updates the matrix of clusters centers, and it repeats the same process. The pseudo code of the algorithm is demonstrated as follows:

procedure FCM-CLUSTERING ($\mathbf{x}$) **returns** prototypes and partition matrix
input: data $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$
local: fuzzification parameter: m
threshold: ε
norm: $\|\cdot\|$
INITIALIZE-PARTITION-MATRIX
$t \leftarrow 0$
repeat
for $i = 1 : c$ **do**

$$\mathbf{v}_i(t) \leftarrow \frac{\sum_{k=1}^{N} u_{ik}^{m}(t)\mathbf{x}_k}{\sum_{k=1}^{N} u_{ik}^{m}(t)} \text{ (compute prototypes)}$$

for $i = 1 : c$ **do**
for $k = 1 : N$ **do**
update partition matrix

$$u_{ik}(t+1) = \frac{1}{\sum_{j=1}^{c} \left(\frac{\| \mathbf{x}_k - \mathbf{v}_i(t) \|}{\| \mathbf{x}_k - \mathbf{v}_j(t) \|} \right)^{2/(m-1)}} \text{ (update partition) matrix}$$

$t \leftarrow t + 1$
until $\| U(t+1) - U(t) \| \leq \varepsilon$
return U, V

Fig. 1: pseudo code of fuzzy c-means algorithm

Though the fuzzy c-means successfully solves many of the clustering problems, it cannot solve other problems because of its ultimate result dependence on the initial value of the centers and getting trapped in the local optima.

Particle swarm optimization algorithm (PSO)

The particle swarm optimization which is known as the birds' algorithm is one of the powerful and favored algorithms for optimization and it is mostly used because of its relatively high convergence speed. Though it has a short life cycle in comparison with the older algorithms such as the genetic algorithm, it has superiority in the applied realms and it is considered as the first choice (13). PSO algorithm was first described by Kennedy and Eberhart in 1975(11). The techniques has had a considerable growth and its main version of the algorithm is recognizable among the present versions. A person's social effectiveness and social learning have made him able to stabilize his knowledge.

Humans solve their problems with the help of others and interacting with their beliefs, tendencies, and changing behaviors. The changes could be imagined as the individuals' movement toward each other through a social awareness space. PSO is an evolutionary computational algorithm inspired by the nature and the iteration. The source of the inspiration has come from the social behavior of animals such as the movement of birds flock or the fish school. Therefore, PSO begins with an initial randomly population matrix and it is similar to many of the evolutionary algorithms such as the imperialist competitive algorithm or the continuous genetic algorithm. In PSO, the particles fly in the super space (a space with more than three dimensions) and it has two essential abilities:

a memory to reserve the best position within itself awareness of the best position in its neighborhood or in the whole space of responses of the swarm members in order to transfer good places to one another via communication and also to match its position and speed with the good places.

Each particle needs the following information in order to impose the proper change on its place and velocity:

The best generality known for every body and whenever each particle recognizes a new place it immediately updates that related information for other particles.

The best neighborhood which is obtained by the particle via communication with the subsets of the group.

The place which is the best experience the particle has ever had.

All of the particles are influenced by the best generality to get close to it. The particles roam in the search space near the "best generality" and they do not explore the rest of the space and it is called convergence. If the inertia coefficient of velocity is considered small, all of the particles will decrease their speed to get close to the velocity of zero in the best generality. One of the ways to get out of the convergent condition (improper) is to give the initial value to the position of particles (after the occurrence of the convergence).

The structure of PSO

Each particle follows the optimized particles in the current position and it continues its movement in the problem space. A group of particles are formed randomly at the beginning of the process, and then the PSO updates the generations to find the optimized solution. In each step, the particle applies two the best ones to update the value. The first best one refers to the best position the particle has so far experienced. The mentioned position is known and protected. The other best value is applied by the algorithm. The pbest refers to the position obtained by the particle swarm and it I demonstrated by the gbest (10).

After finding the best values, the speed and place of each particle are updated by the following equations:

$$v_i^{k+1} = w*v_i^k + c1*r1*(pbest_i - x_i^k) + c2*r2*(gbest - x_i^k)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1}$$

c1 and c2 are the accelerating constants with the positive values and rand is a random number between 0 and 1. v stands for the velocity and x determines the place of particles at each moment. The right side of the relation consists of three parts. The first part refers to the current speed of the particle, and the second and the third parts are responsible for the speed change of the particle and its rotation toward the best personal experience and the best group experience. If the first part is not considered in this equation, the speed of particles will only be determined with regard to the best experience of the particle and the best experience of the all particles. Therefore, the best particle out of all of them will be constant in its place, while the others move toward that particle.

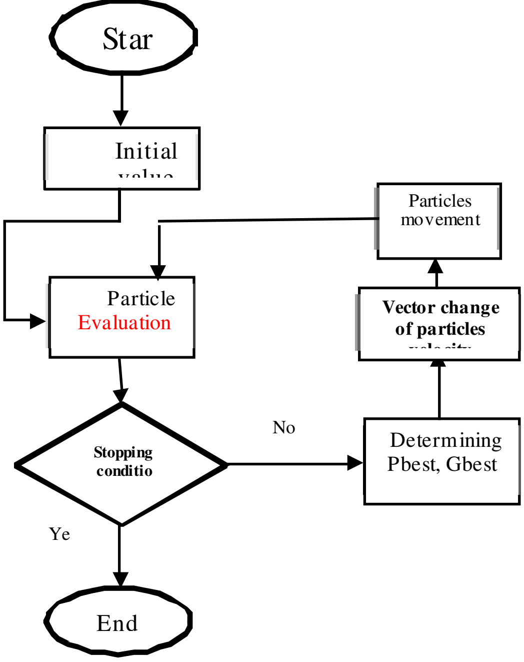


Fig. 2: PSO algorithm

In fact, the particle swarm movement without the first part of the above relation will be a process through which the search space will gradually be minimized. Moreover, the local search is formed around the best particle. However, if the first part of the above relation is considered, the particles will pass their normal way to reach the limitation wall and do the total search.

The pseudo code of PSO is represented as follows:

For each particle
Initialize particle
End For
Do
For each particle
Calculate fitness value of the particle fp
*/*updating particle's best fitness value so far)*/*
If fp is better than pBest
Set current value as the new pBest End For
*/*updating population's best fitness value so far)*/*
Set gBest to the best fitness value of all particles
For each particle
Calculate particle velocity according equation (1)
Update particle position according equation (2)
End For While maximum iterations OR minimum error criteria is not attained

Firefly algorithm

Firefly algorithm (FA) is based on the behavior of the firefly's brightness. The firefly algorithm is based on the three main stages:

1. All the fireflies are unisexual; therefore, a firefly attracts all other fireflies.
2. The attraction depends on the brightness. If one firefly has less brightness, it will be attracted to the firefly with more brightness. If the distance between them increases, the brightness of both of them will also decrease. If none of the fireflies have brightness, they will move randomly.
3. The brightness of the fireflies will be determined or influenced by the objective function.

The firefly algorithm is a population based algorithm for determining the total optimum of the objective function. In the firefly algorithm, the fireflies are randomly distributed to the search space. A firefly attracts other neighbors via the brightness density. The attraction of fireflies depends on the brightness of fireflies. The brightness density is calculated by the objective function. The density has a reversed relation with the distance r between the fireflies $I \propto 1/r2$. Any firefly indicates a candidate solution so that in each iteration, the candidates move toward the best solution, the firefly with the best solution has the most brightness.

FA algorithm includes two important parts (6): changing the brightness density and formulating the attraction. Attracting each firefly is determined by using the brightness density. The objective function in this state is defined by the relation 1.

Two phases of the FA are determined as follows:

1. Changing the brightness density: the values of the objective function are applied to find the brightness density. assume that there is a swarm of n fireflies and xi is the indicator of a solution for the firefly i, considering f (xi) as the evaluation value, the change of brightness density will be defined based on the following equation:

$$I_i = f(x_i),\ 1 \le i \le n.$$

2. Moving toward the attracting firefly: attracting an appropriate firefly occurs based on the observed brightness density related to the neighbors of the fireflies. Each firefly has a level of attraction of β that explains the power of a firefly for attracting other fireflies. However, attracting β depends on the distance rij between the fireflies i and j by considering the xi and xj.

$$r_{ij} = ||x_i - x_j||$$

The function of attracting β(r) is determined as follows:

$$\beta(r) = \beta_0 e^{-\gamma r^2}$$

β0 stands for the attracting when r=0 andγ stands for attracting brightness. Moving from the firefly i in the attracted place of xi toward a more attractive firefly j in the place of xj is achieved as follows:

$$x_i(t+1)=x_i(t)+\beta_0 e^{-\gamma r^2}\left(x_i - x_j\right) + \alpha\,(\text{rand} - 0.5)$$

The pseudo code of FA algorithm is determined as follows:

```
Objective function f(x), x = (x1, ..., xd)T
Generate an initial population of fireflies  xik, i = 1, 2, . . . , n and k = 1, 2, . . . , d
where d=number of dimensions
Maxgen: Maximum no of generations
Evaluate the light intensity of the population Iik which is directly proportional to f (xik)
Initialize algorithm's parameters
While(t <Maxgen)
  For i= 1 :n
    For j= 1: n
      If (Ij<Ii)
        Move firefly i toward j in d- dimension using Eq. (9)
      End if
        Attractiveness varies with distance r viaexp[−r2]
    Evaluate new solutions and update light intensity using Eq. (6)
    End for j
  End for i
Rank the fireflies and find the current best
End while
Post process results and visualization
```

Proposed clustering method

The combination of PSO and FA

Totally, there are two methods for combination (12):

1- The transitional technique: figure 3 indicates the main idea for combining PSO and FA. PSO algorithm would be performed for a while, and then it will be transited to the FA and the FA will be performed. Later, it will return to the PSO and this process will be repeated up to the time that it reaches the stopping condition.

2- Parallel technique: according to this technique, the population is divided into two parts and each part is involved in one technique. The algorithm performs two techniques and a certain number of the best members of the population will be selected for each iteration. The members with the most value of fitness will mostly be selected.

The pseudo code of the transitional technique is mentioned as follows:

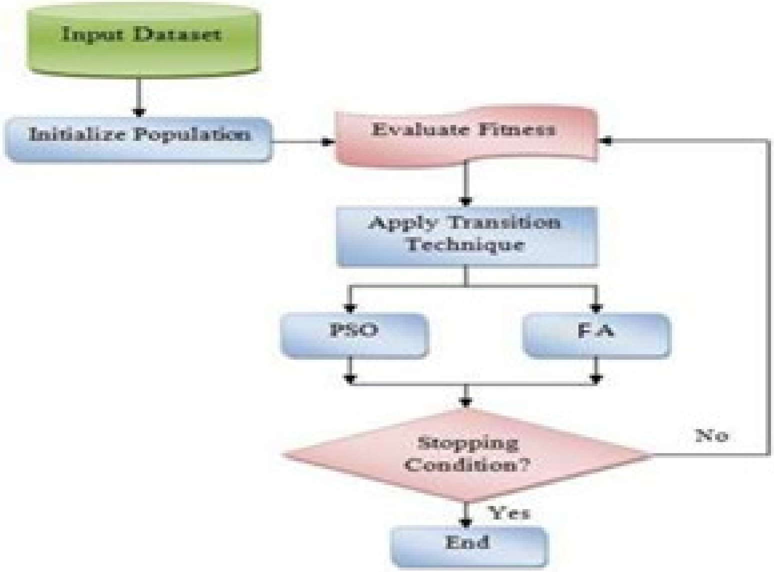


Fig. 3: flowchart of the transitional technique

Step 1: Select a population of size n randomly and initialize the population.

Step 2: Start with number of iterations equal to zero.

Step 3: Perform either F.A operations or P.S.O operations

Step 4: Generate output of the algorithms

Step 5: Evaluate fitness for each individual

Step6: Perform these transitions until termination condition is reached.

Pseudo code of the parallel code will be determined as follows:

Step 1: Select a population of size n randomly and initialize the population

Step 2: Evaluate fitness for each individual

Step 3: If termination criteria is not met, split the population to do selective reproduction and velocity updating. Depending on the algorithm employed

Step 4: If algorithm used is G.A perform crossover and mutation operations. Else perform personal best calculation

Step 5: Repeat this process until final solution is reached.

The transitional technique is applied in this paper. At first, PSO algorithm is performed for a while, and then it is transited to the FA algorithm and it also will be performed for a while, and it will return to PSO and this process will be repeated up to reaching the stopping condition.

Simulation results

The present paper studies the fuzzy clustering of the data and it intends to investigate the behavior of the combined algorithm of FA and PSO. In fact, the accuracy of clustering and the optimization of the algorithm are tested. In order to reach accurate investigation, four famous data sets are used. The data sets are described as follows:

Table 1: the description of data sets

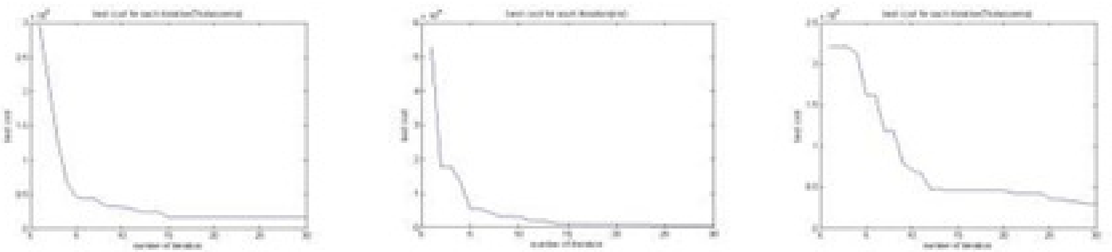
Dataset	Instances	Features	Classes
Glass	214	9	7
Breast Canser	699	9	2
Iris	150	4	3
Ecoli	336	8	7

For the suggested algorithm, the FA algorithm and the PSO algorithm, the population size equals 50, the iteration numbers equal 40, and the clusters numbers equal 14. Moreover, the parameter of $C2=2$ ، $C1=2$ ، $\gamma=1$ ، $\beta 0=2$ ، $\alpha=0.2$ are considered. The following tables indicate the results of performing clustering algorithms of FCM via firefly, clustering algorithm of FCM via PSO, and clustering algorithm of FCM via a combination of FA and PSO on the four datasets. Comparing different algorithms has been done by using two criteria of the mean distance between the clusters centers and the cost function of the fuzzy clustering. The more the distance between the clusters, the better the clustering would be. Moreover, the second criteria which is the cost function of the fuzzy clustering is considered as the objective function in the evolutionary algorithms. The less the objective function value, the better the clustering would be. Figures 1-4 show the convergence of the cost function of each algorithm in association with the different data sets. As it is observed, the clustering algorithm of FCM will sooner reach the convergent solution by combining the PSO and FA on the four data sets.

The collected data set of Glass:

Table 2: the values of the distance mean between the clusters center and the cost value of the clustering on the data set of Glass

Cost value of clustering	Distance mean between the clusters center	Clustering algorithm
673.814033	24.8456	FCM-Firefly
658.460061	27.9182	FCM-PSO
657.152653	29.748	FCM-PSO-Firefly



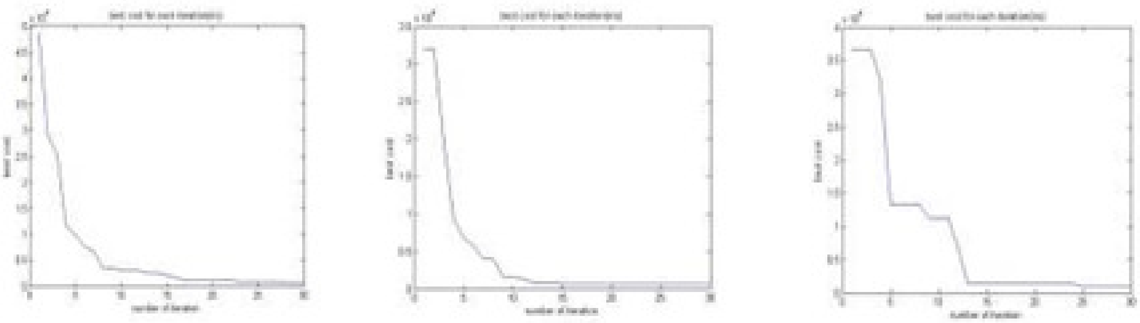
FCM PSO Firefly FCM PSO FCM Firefly

Fig. 4: convergence figure of the objective function on the data set of Glass

Data set of iris:

Table 3: the values of the distance mean between the clusters center and the cost value of the clustering on the data set of iris

Cost value of clustering	Distance mean between the clusters center	Clustering algorithm
9.201613	2.4905	FCM-Firefly
9.529928	2.3672	FCM-PSO
9.193547	2.5834	FCM-PSO-Firefly



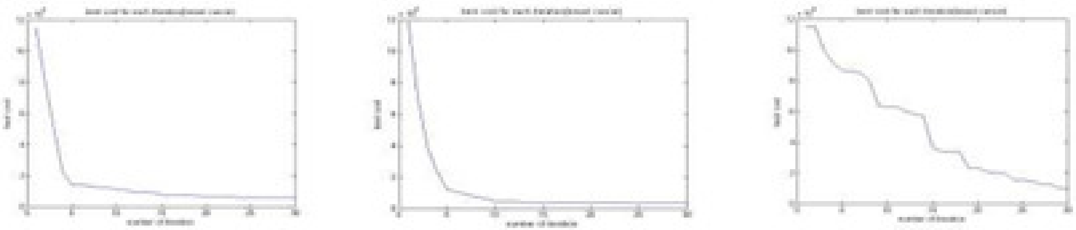
FCM PSO Firefly FCM PSO FCM Firefly

Fig. 5: convergence figure of the objective function on the data set of iris

Data set of breast cancer:

Table 4: the values of the distance mean between the clusters center and the cost value of the clustering on the data set of breast cancer

Cost value of clustering	Distance mean between the clusters center	Clustering algorithm
2025.173299	7.7798	FCM-Firefly
2025.204955	7.7944	FCM-PSO
2020.767385	7.9232	FCM-PSO-Firefly



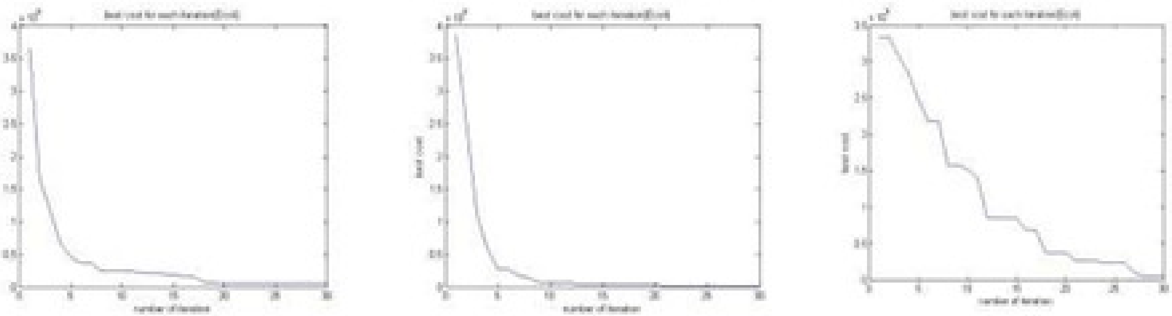
FCM PSO Firefly   FCM PSO   FCM Firefly

Fig. 6: convergence figure of the objective function on the data set of breast cancer

Data set of Ecoli:

Table 5: the values of the distance mean between the clusters center and the cost value of the clustering on the data set of Ecoli

Cost value of clustering	Distance mean between the clusters center	Clustering algorithm
3.015976	0.45316	FCM-Firefly
3.011383	0.45202	FCM-PSO
3.006375	0.46993	FCM-PSO-Firefly



FCM PSO Firefly   FCM PSO   FCM Firefly

Fig. 7: convergence figure of the objective function on the data set of Ecoli

Conclusion

In this paper, the firefly algorithm and the particle swarm optimization algorithm have been applied and a solution for solving the fuzzy clustering problem has been represented. In the suggested combined algorithm, at first, the maximum distance between the clusters centers, and then the cost of the clustering have been calculated and it is considered as the objective function. The obtained results are directly compared with the firefly algorithm and the particle swarm optimization algorithm. The results show that the suggested combined algorithm calculates the clustering more accurately than the firefly algorithm or the particle swarm optimization algorithm. Moreover, the cost of the clustering of the suggested algorithm is considerably less than the mentioned algorithms.

References

[1] Dorigo, M., St¨utzle, T. (2009, may). Ant Colony Optimization. *Institut de Recherches Interdisciplinaires et de Développements en Intelligence Artificielle*, 15-25.

[2] Changsheng.Z, Dantong.O, & jjaxn.N, (2010, julay). An artificial bee colony approach for clustering. *Expert system with Applications, 37*(7), 4761-4767.

[3] Amin abshori, A., & Bakhtiari, A. (2012). A New Clustering Method Based on Firefly and KHM. Journal of Communication and Computer 9, 387-391.

[4] Senthilnath, J., Omkar, S. N., & Mani , V. (2011, September). Clustering using firefly algorithm: Performance study. *Swarm and Evolutionary Computation, 1*(3), 164-171.

[5] Hadadnia, J., Vahidi, J., Gharakhani, A., & Fioozi, M. (2012). Fuzzy diagnosis of diabetes is

based on the laws and properties of the optimal combination of artificial intelligent systems for data mining . International Conference on Nonlinear Modeling & Optimization(ICNMO), 1-7.

[6] Dhote, C. A., Thakare, A. D., & Chaudhari, S. M. (2013). Data Clustering Using Particle Swarm Optimization and Bee Algorithm. *International Conference on Communication Systems and Network TechnologY*, 1-5.

[7] Senthilnath, J., Omkar, S. N., & Mani , V. (2011, September). Clustering using firefly algorithm: Performance study. *Swarm and Evolutionary Computation, 1*(3), 164-171.

[8] Fazeli Hasan Abadi, M., Hamidi, A., Piri, S., & Tezerez, M. (2014). Survey data clustering using hybridization strategy of continuous particle swarm optimization algorithm based on Ant Colony Optimization. *International Journal of Science and Applied Research. Volume 3*, 36-42.

[9] De Oliveira, J Valente,Pedrycz, Witold: 'Advances in fuzzy clustering and its applications' (Wiley Online Library, 2007. 2007)

[10] Dhote, CA, Thakare, Anuradha D,Chaudhari, Shruti M: 'Data clustering using particle swarm optimization and bee algorithm ,'in Editor (Ed.)^(Eds.): 'Book Data clustering using particle swarm optimization and bee algorithm' (IEEE, 2013, edn.), pp. 1-5

[11] Eberhart, Russ C,Kennedy, James: 'A new optimizer using particle swarm theory', in Editor (Ed.)^(Eds.): 'Book A new optimizer using particle swarm theory' (New York, NY, 1995, edn.), pp. 39-43

[12] Hyma, J, Jhansi, Y,Anuradha, S: 'A new hybridized approach of PSO & GA for document clustering', International Journal of Engineering Science and Technology, 2010, 2, (5), pp. 1221-12.26

[13] Kennedy, James, Kennedy, James F,Eberhart, Russell C: 'Swarm intelligence' (Morgan Kaufmann, 2001. 2001)

[14] Nayak, Janmenjoy, Nanda, Matrupallab, Nayak, Kamlesh, Naik, Bighnaraj,Behera, Himansu Sekhar: 'An improved firefly fuzzy c-means (FAFCM) Algorithm for clustering real world data sets': 'Advanced Computing, Networking and Informatics-Volume 1' (Springer, 2014), pp. 339-348.