Hybridization of Fuzzy C-Means with Bacterial Colony Optimization

K.Vijayakumari[#1], Dr. V. Baby Deepa[*2]

[#] *Department of Computer Science*
Government Arts College, affiliated to Bharathidasan university,
Karur, Tamilnadu, India
[1]k.vijikumari@gmail.com
[2]@deepamct@gmail.com

***Abstract*— Fuzzy clustering is a challenging task in data mining techniques. Fuzzy C-Means (FCM) clustering algorithm is applied to solve clustering problem. The present paper is proposed new hybrid fuzzy clustering algorithm which is combined with FCM and Fuzzy Bacterial Colony Optimization called as HFCA(Hybridization of Fuzzy clustering algorithm). The experimental results are shows that the proposed new hybrid fuzzy clustering algorithm obtained high clustering efficiency compared with FCM and FBCO.**

***Keywords*— Fuzzy Clustering, Fuzzy C-Means, Fuzzy Particle Swarm Optimization, Fuzzy Bacterial Foraging Optimization, Fuzzy Bacterial Colony Optimization.**

I. Introduction

Fuzzy clustering is a process of divided into small number groups from the given dataset based on their similarity [1]. In fuzzy clustering, Fuzzy C-Means (FCM) algorithm is a most fashionable partitioned, simple, easy to implemented fuzzy clustering algorithm [2]. However, it has many limitations including sensitive to initialization and fall into local optima problems. Many research works have been carried out in order to overwrite limitations of FCM. An eliminating the noise sensitivity fault of FCM has done with the help of Probabilistic FCM [3]. A spatial functions is used in FCM for extracting the spatial information of images [4]. However, these enhanced versions of FCM are not having ability to search the global solution in the search space.

To attempt these shortcomings, many swarm intelligent algorithms have been applied to solve local optima problem and achieves global optima. Some of the swarm algorithms are Genetic Algorithm (GA) [5], Ant Colony Optimization (ACO) [6], Particle Swarm Optimization (PSO) [7], Firefly Algorithm (FA) [8], Bacterial Foraging Optimization (BFO) [9], Bacterial Chemotaxis (BC) [10], Bacterial Colony Optimization (BCO) [11], Social Spider Optimization (SSO) [12], Efficient Stud Krill Herd (ESKH) [13]. The above mentioned optimization algorithms are easily applied to solve any complicated problem in order to achieve a global optimum solution from the given search space. Even though, different traditional optimization approaches are failed to achieve global optimum and lead to take more computation time.

The BCO algorithm is a well-known growing optimization method developed by Ben Niu et al. (2012) [11]. The BCO algorithm was achieved higher efficiency compared with GA, PSO, BFO, and some enhanced BFO when applied to some global optimization functions. The remarkable point of BCO is low convergence speed due to inner iteration process. In [11], the authors was mentioned the BCO could not faster compared with PSO.

In this research work, the FCM clustering algorithm is hybrid with FBCO in order to enhance fuzzy clustering quality. The strength of the proposed HFCA clustering method is evaluated using six well-known benchmark machine learning data sets such as Iris, Glass, Wisconsin Breast Cancer (WBC), Wine, Contraceptive Method Choice (CMC), and Vowel. The contributions of this research work are to enhance the accuracy of the clustering problem and finding optimum centroid values.

II. Related works

Developing an efficient fuzzy clustering model is a challenging task, which is another NP-complete problem. In the earlier stage, FCM clustering algorithm is applied to solve a fuzzy clustering problem[2]. The traditional FCM is failure to find a global optimum values due to its initialization of centroid values at beginning stage. Hence, many research works have been carried out in order to enhance the efficiency of FCM clustering algorithm. A hybridization of possibilistic c-means (PCM) and fuzzy c-means (FCM) is called PFCM in order to avoid various shortcomings of PCM and FCM [3].

The authors were introduced two new techniques for reducing the objective values of FCM by PSO. In which, PSO–V is represent each particle as component of a cluster center. PSO–U is represented each particle as an unnormalized and unscaled membership value [14]. Chaotic PSO was proposed in order to exploit the searching ability of FCM and escaping its major drawback of receiving stuck at locally optimal values. Similarly, gradient operator is implemented with PSO to accelerate convergence rate [15]. The fuzzy artificial BCO was applied to solve fuzzy clustering problem [16]. SSO is combined with fuzzy theory called fuzzy SSO to solve a fuzzy clustering problem with efficient manner[12].

The Artificial Bee Colony (ABC) algorithm was developed to solve fuzzy clustering problems. Hence, the performance of the FABC has achieved higher classification accuracy compared with FCM [16]. In [17], the author was developed a fuzzy based artificial bee colony optimization (FABC) algorithm for image segmentation. The proposed FABC is more significant comparing with other optimization techniques such as expectation maximum (EM), GA, and PSO. The FABC overcomes the drawbacks of FCM in terms of convergence, time complexity, and segmentation accuracy.

III.RESEARCH PROBLEMS

A. Fuzzy C-Means

The FCM is a well-known partitioned clustering algorithm developed by Bezdek et al. [2]. Given an unlabeled data set $O=\{O_1,O_2,......O_n\}$ in R^d dimensional area, the FCM is to partitions the data set into c (1 < c < n) fuzzy clusters with cluster centers $z=\{z_1,z_2,....,z_n\}$ [18].In fuzzy clustering, the data samples are described by a fuzzy matrix μ. It has with n rows and c columns. In which, the number of data samples is denoted by n and the number of classes is denoted by c. μ_{ij} is the element of the i^{th} row with j^{th} column. Them, μ is an indicates the degree of association of the i^{th} data samples with the j^{th} cluster. The distinctiveness of μ is defined as follows:

$$\forall_i \left[0,\ 1\right] \forall_i = 1,\ 2\ldots n\ \forall_j = 1,\ 2\ldots c \tag{1}$$

$$\sum_{j=1}^{c} \mu_{ij} = 1\ \forall_{i=1,2,....n} \tag{2}$$

$$o < \sum_{j=1}^{c} \mu_{ij} < n\ \forall_{j=1,2,....,c} \tag{3}$$

The minimization or maximization of the objective function is well-defined as follows,

$$J_m = \sum_{j=1}^{c} \sum_{i=1}^{n} \mu_{ij}^{m} d_{ij} \tag{4}$$

Algorithm 1: Fuzzy C-Means

Step 1: Initialize required parameters

Step 2: Compute the cluster centers

Step 3: Compute Euclidian distance

Step 4: Update the membership function

Where, $m(m>1)$ is a scalar values and is weighting exponents that regulating the fuzziness of the resultant clusters. d_{ij} is denote the Euclidian distance function from data object to the cluster center $z=\{z_1, z_2, ..., z_n\}$. The distance function d_{ij} is defined as follows,

$$d_{ij} = \left| o_i - z_j \right| \tag{5}$$

Where, the z_j centroids of the j^{th} cluster and i^{th} data objects which is defined by using below equation

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

If convergence criteria does not met after completing every iteration, then the membership values is updated as follows,

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

.

B. Fuzzy BCO

Compared with some conventional bacterial based optimization methods, the searching ability of the BCO is enhanced by newly created chemotaxis steps combined with communication process [11]. The communication mechanism is developed based on interactive communication scheme achieves optimum centroid values with least iterative process. Hence, this communication mechanism can reduce the computation cost and avoid premature convergence.

The BCO was developed by Niu and Wang in 2012 followed by the behaviour of E.Coli bacterial colony [19]. The behaviour of BCO algorithm is a combination of two bacterial foraging algorithms such as BFO[20] and BC algorithm [10]. The BCO algorithm has five steps including chemotaxis and communication, elimination and reproduction, migration process. The following subsection are briefly explained the steps of BCO [19].

a. Chemotaxis and Communication model

The communication is major activities of the bacterial colonies that are accompanied with two processes such as runs and tumbles. It has performed three kinds of information exchanges during the process of communication such as random direction, group information, and personal information. The bacterial colonies are used for an exchange the information and guide to them in the behaviours of movement and its guidelines. The process of the communication can be accomplished either run or tumble which is expressed as follows,

$$Position_i(T) = Position_i(T-1) + R_i.(Run_{info}) + R\Delta(i) \quad (12)$$

$$Position_i(T) = Position_i(T-1) + R_i.(Tumb_{info}) + R\Delta(i) \quad (13)$$

The relationship between the individual and the group of the bacterial colonies are considered by the position updating process. The information on the each bacterial colony is shared with the help of communication process.

b. Elimination and Reproduction method

Each bacterial colony is noticeable with a degree of energy level based on the ability of searching in the search space. The level of energy of each bacterial colony is defined by its probability. The energy level of each bacterial colony is sorted and analysed, then decide the qualification of bacterial colony used as a criterion. The assessment of the elimination and reproduction process can be defined as follows,

$$\begin{array}{ll} if\ L_i > L_{given}\ and\ i \in healthy, & then\ i \in Candidate_{repr} \\ if\ L_i < L_{given}\ and\ i \in healthy, & then\ i \in Candidate_{eli} \\ if\ i \in unhealthy, & then\ i \in Candidate_{eli} \end{array} \quad (14)$$

Where, L_i -is the energy level of i^{th} bacterial colony. All behavior of the bacterial colony was constrained surrounded by a reserved area. If bacterial colony moves left from reserving area, two schemes will be performed. Either, generate new individuals to replace the outer one or modify the forward way to retains them effectively. In this stage, the boundary control is very important. The outer bacterial colony is considered as an unhealthy and it's ready to eliminate from the populations, and then reproduce by the healthy bacterial colonies.

c. Migration Model

The migration is the process of constructing more nutrition and its performance is not followed by given probability. The bacterial colony would travel to a new random residence on the search space according to the given conditions satisfied. The BCO is performing the migration process in order to avoid the local minimum. It can be expressed as follows,

$$Position_i(T) = rand.(ub - lb) + lb \quad (15)$$

Where, lb and ub are the lower and upper borders of the colony positions respectively. $rand$ is a generate random number that range between 0 and 1.

c. Fuzzy BCO Clustering Algorithm

We concentrate the BCO for solving the problem of fuzzy clustering is proposed in this research works Some of modifications are made in the conventional BCO for solving fuzzy clustering problem. The major goal of the fuzzy clustering problem is to decrease the sum of distance values between samples which is defined as an objective function J_m as follows,

$$J_m = \sum_{j=1}^{c}\sum_{i=1}^{n}\mu_{ij}^{m} d_{ij} \qquad (16)$$

Where, K is determine the number of clusters, D is the number of dimension in the data sets of x. In the proposed FBCO, The fuzzy relation of the data sample X is defined as follows ,

$$X = \begin{bmatrix} \mu_{11} & \cdots & \mu_{1c} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \cdots & \mu_{nc} \end{bmatrix} \qquad (17)$$

Where, μ_{ij} is the membership function value for i^{th} data samples and j^{th} cluster. It should be satisfied with the equation (1) and (2). Furthermore, the complete lives of the bacterial colonies are involved, two processes in the chemotaxis such as running and tumbling. Therefore, the updating the position of the each bacterial colony as follows,

$$\begin{aligned} Position_i(T) = {} & Position_i(T-1) + C(i) \\ & *[f_i.(G_{best} - Position_i(T-1)) + (1-f_i) \\ & *(P_{best_i} - Position_i(T-1)) + turb_i] \end{aligned} \qquad (18)$$

$$\begin{aligned} Position_i(T) = {} & Position_i(T-1) + C(i) \\ & *[f_i.(G_{best} - Position_i(T-1)) + (1-f_i) \\ & *(P_{best_i} - Position_i(T-1))] \end{aligned} \qquad (19)$$

$$C(i) = C_{\min} + \left(\frac{Iter_{\max} - Iter_j}{Iter_{\max}}\right)(C_{\max} - C_{\min}) \qquad (20)$$

Algorithm 2: Fuzzy PSO

Step 1: Initialize the required parameters
Step 2: Create a swarm with P particles (X, pbest, gbestand V are n× c matrices).
Step 3: Initialize X, V, pbestfor each particle and gbestfor the swarm.
Step 4: Compute the cluster centers
Step 5: Compute the fitness value
Step 6: Compute pbestfor each particle.
Step 7: Compute gbestfor the swarm.
Step 8: Update the velocity matrix
Step 9: Update the position matrix
Step 10: If terminating condition 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

Where, $turb_i$ denotes the turbulent direction variance of the i^{th} bacterial colony, $C(i)$ - is the chemotaxis step size of the i^{th} bacterial colony, $f_i \in (0,1)$, G_{best} is the global best and P_{best} is the personal best of the i^{th} bacterial colony. $Iter_{\max}$, $Iter_j$ are the maximum iteration and current iteration respectively. In proposed FBCO, the n number of bacterial colonies of the position is defined as $Position$ and that is produced $D \times K$ cluster centroid. Hence, $n \times K$ position will be moved for achieving minimum distance that move to the global optimum values.

d. *Hybrid Fuzzy clustering Algorithm (HFCA)*

In the conventional FBCO, the G_{best} value is not changed during updating their position. Also, need more parameter tuning performs during process due to it has many inner loops. Hence, the FCM algorithm is faster than the FPSO, FBFO, FBCO algorithms due to fewer parameters evaluations, but it commonly falls into local optima and low clustering efficiency. The FBCO is obtained high clustering efficiency. Hence, in this research work, the FCM algorithm hybrid with FBCO algorithm to make a hybrid fuzzy clustering algorithm (HFCA) which keeps the advantages of both FCM and BCO algorithms. The output of FBCO is used as inputs to the FCM in order to accelerating the convergence rate. The proposed HFCA algorithm for fuzzy clustering problem can be indicated in Algorithm 3.

Experimental results analysis and discussions

The experimental results are implemented using MATLAB 2015b. The results of every clustering algorithm are obtained from the average values of 50 independent runs.

Dataset descriptions

The following six different kinds of machine learning datasets are obtained from UCI machine learning repository and it is used in order to evaluate the performance of the proposed method and its description are showed in Table 1. Fisher's iris consists of 3 different species of the iris flower, 150 samples and 4 features were composed in each species. Glass datasets have 214 data sample with six categories of glasses and it has 9 diverse features in each type. Wisconsin Breast Cancer (WBC) data set has 683 data objects with 2 classes and it considered by nine features. The wine dataset contains of 178 data samples with three diverse types of groups and considered by thirteen features. Contraceptive Method Choice (CMC) dataset contains 1473 data sample with 3 categories and nine features. Vowel data set has 871 Indian vowels with 3 features and 6 overlapping groups.

Algorithm 3: HFCA
Begin HFCA
Begin FBCO
Step 1: Initialize required parameters
Step 2: Each bacterial colony

Step 3: Chemotaxis and communication process in tumbling and swimming
Step 4: Calculate cluster center for each bacterial colony
Step 5: Find distance between data samples and cluster center
Step 6: Performing an interactive exchange
Step 6.1: If Individual Exchange
If Dynamic neighbour oriented
If found poorer fitness bacterial colony, then replace with random values
Else Random oriented
Replace poorer colony after finding poor fitness value of bacterial colony from the given data chooses by randomly
Step 6.2: Else if Group exchange
Replace the poorer after finding the best bacterial colony according to the fitness values
Step 7: Reproduction and elimination
Step 7.1: Search the healthy bacterial colony using its fitness values
Step 7.2: Perform the reproductions and elimination
Step 8: Migration process
Step 9: Update position for each bacterial colony
Step 10: If terminating situation is not met, then go to step 2.
End FBCO
Begin FCM
Step 11: Initialize required parameters
Step 12: Compute the cluster centers z_j
Step 13: Compute Euclidian distance
Step 14: Update the membership function μ_{ij}
End FCM
End HFCA

IV RESULTS AND DISCUSSION

The performance of the proposed clustering algorithm is compared with FBCO, FBFO, FPSO, and FCM. The Euclidean distance is used for finding the distance between the centroid and data samples. Table 2 illustrate the performance results of the objective functions for all compared algorithms.

In objective function, there are three types of values are calculated including worst, average, and best which is obtained from 50 independent runs. In which, worst value is represented maximum value of objective function. The best value is representing minimum values of the objective function values. The average value is calculated the average values of both minimum and maximum values. The best value is considered as a better performance and it is represented in bold letters in Table 2. Here, a low value of objective function is considered as a best performance because the major goal of this research work.

TABLE 1
DATASETS DESCRIPTIONS

Datasets	Instances	Features	Clusters
Iris	**150**	4	3
Glass	**214**	9	6
WBC	**683**	9	2

Wine	**178**	13	3
CMC	**1,473**	9	3
Vowel	**871**	3	6

TABLE 2
PERFORMANCE OF PROPOSED METHOD BASED ON INDEX VALUES

Datasets	**Methods**	**Objective Values**			**Median**	**Standard Deviations**
		Best	**Average**	**Worst**		
Iris	**FCM**	**71.65**	75.31	78.97	75.77	2.42
	FPSO	**65.16**	68.06	70.96	67.19	1.89
	FBFO	**62.03**	63.47	64.91	63.67	0.89
	FBCO	**58.02**	61.00	63.97	60.66	1.73
	HCFA	**56.06**	57.02	57.98	56.90	0.59
Glass	**FCM**	**76.22**	80.06	83.90	79.93	2.19
	FPSO	**69.10**	72.55	75.99	72.10	2.03
	FBFO	**65.01**	66.93	68.87	66.98	1.02
	FBCO	**63.03**	64.51	65.99	64.35	0.89
	HCFA	**60.01**	61.00	61.99	61.11	0.58
WBC	**FCM**	**2402.89**	2475.72	2547.89	2505.68	60.82
	FPSO	**2265.30**	2373.67	2482.05	2380.77	64.81
	FBFO	**1963.40**	2121.32	2279.25	2120.13	91.91
	FBCO	**1868.08**	1925.89	1983.70	1928.85	34.72
	HCFA	**1701.20**	1749.17	1797.13	1750.80	27.49
Wine	**FCM**	**12193.59**	12807.73	13420.73	12747.70	353.21
	FPSO	**11387.12**	11695.59	12004.59	11706.07	192.69
	FBFO	**11095.20**	11398.00	11700.81	11402.75	167.71
	FBCO	**10787.39**	10897.68	11007.97	10909.93	60.45
	HCFA	**10501.26**	10598.67	10696.08	10599.04	53.62
CMC	**FCM**	**3445.98**	3502.42	3558.86	3500.92	32.66
	FPSO	**3146.98**	3253.47	3359.95	3259.88	61.78
	FBFO	**2942.46**	3049.65	3156.83	3055.61	61.29
	FBCO	**2842.50**	2899.08	2955.66	2902.56	37.15
	HCFA	**2713.14**	2761.51	2809.88	2760.42	25.74
Vowel	**FCM**	**73472.47**	73912.95	74353.46	73932.02	264.18
	FPSO	**71477.02**	71912.01	72347.00	71907.80	255.20
	FBFO	**71070.09**	71505.51	71940.94	71493.56	260.72
	FBCO	**70975.65**	71261.60	71546.60	71243.61	170.14
	HCFA	**70578.27**	70899.83	70739.05	70717.10	98.31

V CONCLUSION AND FUTURE ENHANCEMENTS

Fuzzy clustering is a complex NP-Hard problem for solving with efficient manner, Hence HFCA clustering algorithm has proposed in this research work in order to solve clustering problem. The proposed method has obtained the high efficient clustering accuracy. The experiment results demonstrate that the proposed hybrid HFCA algorithm produced higher clustering efficiency. Furthermore, Type- 2 Fuzzy set is well-known fuzzy logic techniques in order to manage uncertainty data. Hence, future enhancement will focus on type-2 fuzzy set to enhance the performance clustering efficiency.

REFERENCES

[1] S. Alam, G. Dobbie, Y. S. Koh, P. Riddle, and S. U. Rehman, "Research on particle swarm optimization based clustering: a systematic review of literature and techniques," *Swarm and Evolutionary Computation,* vol. 17, pp. 1-13, 2014.
[2] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Computers & Geosciences,* vol. 10, pp. 191-203, 1984.
[3] N. R. Pal, K. Pal, J. M. Keller, and J. C. Bezdek, "A possibilistic fuzzy c-means clustering algorithm," *IEEE transactions on fuzzy systems,* vol. 13, pp. 517-530, 2005.
[4] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation," *computerized medical imaging and graphics,* vol. 30, pp. 9-15, 2006.
[5] U. Maulik and S. Bandyopadhyay, "Genetic algorithm-based clustering technique," *Pattern recognition,* vol. 33, pp. 1455-1465, 2000.
[6] T. Niknam and B. Amiri, "An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis," *Applied soft computing,* vol. 10, pp. 183-197, 2010.
[7] T. Cura, "A particle swarm optimization approach to clustering," *Expert Systems with Applications,* vol. 39, pp. 1582-1588, 2012.
[8] J. Senthilnath, S. Omkar, and V. Mani, "Clustering using firefly algorithm: performance study," *Swarm and Evolutionary Computation,* vol. 1, pp. 164-171, 2011.
[9] M. Wan, L. Li, J. Xiao, C. Wang, and Y. Yang, "Data clustering using bacterial foraging optimization," *Journal of Intelligent Information Systems,* vol. 38, pp. 321-341, 2012.
[10] S. D. Muller, J. Marchetto, S. Airaghi, and P. Kournoutsakos, "Optimization based on bacterial chemotaxis," *IEEE Transactions on Evolutionary Computation,* vol. 6, pp. 16-29, 2002.
[11] B. Niu and H. Wang, "Bacterial colony optimization," *Discrete Dynamics in Nature and Society*, vol. 2012, 2012.
[12] P. Padmavathi, V. Eswaramurthy, and J. Revathi, "Fuzzy Social Spider Optimization Algorithm for Fuzzy Clustering Analysis," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*, 2018, pp. 1-6.
[13] K. Baalamurugan and S. V. Bhanu, "An efficient clustering scheme for cloud computing problems using metaheuristic algorithms," *Cluster Computing*, pp. 1-11, 2018.
[14] T. A. Runkler and C. Katz, "Fuzzy clustering by particle swarm optimization," in *2006 IEEE international conference on fuzzy systems*, 2006, pp. 601-608.
[15] C. Li, J. Zhou, P. Kou, and J. Xiao, "A novel chaotic particle swarm optimization based fuzzy clustering algorithm," *Neurocomputing,* vol. 83, pp. 98-109, 2012.
[16] D. Karaboga and C. Ozturk, "Fuzzy clustering with artificial bee colony algorithm," *Scientific research and Essays,* vol. 5, pp. 1899-1902, 2010.
[17] A. Bose and K. Mali, "Fuzzy-based artificial bee colony optimization for gray image segmentation," *Signal, Image and Video Processing,* vol. 10, pp. 1089-1096, September 01 2016.
[18] K. Vijayakumari, M. Preetha, and K. Velusamy, "Performance Analysis of Clustering Based on Fuzzy System."
[19] B. Niu, H. J. D. D. i. N. Wang, and Society, "Bacterial colony optimization," vol. 2012, 2012.
[20] K. M. Passino, "Biomimicry of bacterial foraging for distributed optimization and control," *IEEE control systems magazine,* vol. 22, pp. 52-67, 2002.
[21] H. Chen, B. Niu, L. Ma, W. Su, and Y. Zhu, "Bacterial colony foraging optimization," *Neurocomputing,* vol. 137, pp. 268-284, 2014.
[22] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE transactions on pattern analysis and machine intelligence,* pp. 224-227, 1979.
[23] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," 1973.
[24] C. Borgelt, "Objective functions for fuzzy clustering," in *Computational Intelligence in Intelligent Data Analysis*, ed: Springer, 2013, pp. 3-16.