Vol. 9(31), Jan. 2019, PP. 4098-4110

Contents list available at IJMEC

International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)

Journal Homepage: www.aeuso.org

PISSN: 2411-6173 - EISSN: 2305-0543

A New Clustering Method Using Evolutionary Algorithms for Determining Initial States, and Diverse Pairwise Distances for Clustering

Seyed Muhammad Hossain Mousavi

Department of Computer Engineering Bu Ali Sina University Hamadan, Iran

Phone Number: *+98-09332892726*

****Corresponding Author's E-mail:*** *mosavi.a.i.buali@gmail.com*

Abstract

In recent decades, various clustering methods and systems have been proposed, but a lot of them had problem in selecting initial clusters' centers. In this paper a new clustering method based on Differential Evolution (DE) and Particle Swarm Optimization (PSO) algorithms for fixing initial clusters' centers problem and different pairwise distances like (Euclidean, City-block and Chebyshev) for setting final clustering output, has been proposed. After choosing initial clusters' centers using DE and PSO algorithms, a kind of process to changing clusters' centers will be applied, which is based on finding nth farthest samples from each cluster and then calculating the mean for each nth farthest samples for each cluster, and finally inverting the calculated mean for each nth farthest samples. With this approach, problem in choosing initial clusters will fixes. After choosing best initial clusters, the main process of finding and biasing clusters' centers and allocating samples to closest cluster take place. This system has been validated with some of benchmark clustering datasets such as: Fisher-iris, Ionosphere, User-Knowledge-Modeling, Blood transfusion ,Breast cancer and compared with famous clustering methods like K-Means, Fuzzy C-means (FCM), Gaussian Mixture Model (GMM), Self-Organizing Maps (SOM) and returned promising results.

Keywords: *Initial Clusters' Centers, Pairwise Distances, Differential Evolution Algorithm (DE), Particle Swarm Optimization (PSO), Clustering, Clustering Datasets.*

1. Introduction

The task of assigning set of objects into groups (based on similarity between objects) is called clustering in which every group is called a cluster. Unsupervised learning is a machine learning task for inferring the output from unlabeled data. Clustering is one of the approaches to unsupervised learning for unsupervised classification. Clustering is a fundamental data analysis method. It is widely used for pattern recognition [17], feature extraction, vector quantization (VQ), image segmentation, function approximation, and data mining [1], machine learning, image analysis, information retrieval, bioinformatics, data compression, and computer graphics. Cluster analysis is an important statistical methodology used in wide variety of fields including artificial intelligence, business, biology, psychology and medicine [2]. Clustering techniques were originally conceived by Aristotle and Theophrastos in the fourth century B.C. and in the 18th century by Linnaeus [3], but it was not until 1939 when one of the first comprehensive foundations of these methods was published [4]. These

Article History:
Received Date: Oct. 10, 2018
Accepted Date: Nov. 16, 2018
Available Online: Jan. 03, 2019

techniques have traditionally been applied to classification problems [5], where the task to solve is how to organize observed data into meaningful structures [6].

1.1. Fuzzy Logic

Based on Prof Lotf-Ali Asgarzade, founder of fuzzy logic and fuzzy sets [18], as complexity increases, exact sentences lose their meaning, and meaningful sentences their precision. Following this rule, there is a need for converting the absolute results to the results possessing range. The fuzzy hypothesis is also based on this. Considering the similarity of fuzzy concept and probabilities, it can never be possible to say that these two concepts are one. For instance, in crisp problems, we say: "an accident occurred", but in a similar fuzzy case, we say: "how much the damage was", and for instance the intensity of it is shown as between a range of 0 and 1 [19], [20]. Fuzzy c-means or FCM [9] [10] is clustering in fuzzy mode, which will be used in this paper.

1.2. Evolutionary Computing

Evolutionary computing is a research area within computer science. As the name suggests, it is a special flavor of computing, which draws inspiration from the process of natural evolution. It is not surprising that some computer scientists have chosen natural evolution as a source of inspiration: the power of evolution in nature is evident in the diverse species that make up our world, each tailored to survive well in its own niche. The fundamental metaphor of evolutionary computing relates this powerful natural evolution to a particular style of problem solving – that of trial-and-error [21]. In artificial intelligence, an evolutionary algorithm (EA) is a subset of evolutionary computation, a generic population-based metaheuristic optimization algorithm. An EA uses mechanisms inspired by biological evolution, such as reproduction, mutation, recombination, and selection. Candidate solutions to the optimization problem play the role of individuals in a population, and the fitness function determines the quality of the solutions. Differential evolution algorithm is one of the popular evolutionary algorithms, which we will use it to find best initial cluster centers. Differential Evolution (DE) is a relatively recent heuristic (it was created in the mid-1990s) proposed by Kenneth Price and Rainer Storn [22] [23] [24] which was designed to optimize problems over continuous domains. This approach originated from Kenneth's Price attempts to solve the Tchebycheff Polynomial fitting Problem that had been posed to him by Rainer Storn [25].

2. Prior works

Some of the famous data clustering methods which is used in this paper and also some of other significant methods are introduced. K-means (MacQueen, 1967) [7] [8] is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a cunning way because different locations cause different results. So, better choice is to place them as far as possible, far away from each other. The next step is to take each point belonging to a given data set, and associate it to the nearest centroid. When no point is pending, the first step is completed. At this point it is needed to re-calculate k new centroids as barycenters of the clusters resulting from the previous step. After achieving these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop, it may be noticed that k centroids change their location step by steps until no more changes are done. In other words, centroids do not move any more.

Fuzzy clustering (also referred to as soft clustering) is a form of clustering in which each data point can belong to more than one cluster [11]. Fuzzy c-means (FCM) clustering was developed by J.C. Dunn in 1973 [9] and improved by J.C. Bezdek in 1981[10].

Gaussian mixture models (GMM) are often used for data clustering. Usually, fitted GMMs cluster by assigning query data points to the multivariate normal components that maximize the component

posterior probability given the data. That is, given a fitted GMM. Cluster assigns query data to the component yielding the highest posterior probability. This method of assigning a data point to exactly one cluster is called hard clustering [12]. For an example showing how to fit a GMM to data, cluster using the fitted model, and estimate component posterior probabilities [13]. Moreover, GMM clustering can accommodate clusters that have different sizes and correlation structures within them. Because of this, GMM clustering can be more appropriate to use than, e.g., k-means clustering. In 2010, K.-L. Du gave a comprehensive overview of competitive learning based clustering methods which the main goal was clustering based on neural network approach [14]. Also, Gabrys, Bogdan, and Andrzej Bargiela made a system based on fuzzy Min-Max neural network for clustering and classification [15]. In 2008 Das, Swagatam, Ajith Abraham, and Amit Konar used DE algorithm for clustering, which had so robust and satisfactory results [16].

Self-Organizing Maps (SOM) [33] is another clustering method, which is going to use in validation section for comparison purposes.

3. Proposed method

In this paper a new clustering system based on Differential Evolution (DE) [22] [23] and Particle Swarm Optimization (PSO) algorithm [34] for fixing initial cluster centers problem and different pairwise distances like (Euclidean [26], city-block [26] and chebyshev [27]) for setting final clustering output, is proposed. After choosing initial cluster centers by differential evolution algorithm, a kind of process to changing cluster centers will be applied, which is based on finding nth farthest samples from each cluster and then calculating the mean for each nth farthest samples for each cluster, and finally inversing the calculated mean for each nth farthest samples cluster. With this approach, problem in choosing initial clusters will be fixed. After choosing best initial clusters, the main process of finding and biasing cluster centers and allocating samples to closest cluster takes place. The main process will happen between 3 to 5 iteration, based on number of pairwise distances which default is 3 kind of distances (Euclidean [26], city-block [26] and chebyshev [27]). But it is flexible according to the datasets distribution form (by changing the pairwise distances types). This loop repeats to the end of the iterations. It is mentionable that, evolutionary clustering for DE clustering is based on [35] and for PSO clustering is based on [36].

Fig 1 shows the flowchart of DE algorithm. Also Fig 2 represents PSO algorithm's flowchart. The flowchart of the proposed clustering system is represented in Fig 3. Also Fig 4, shows the workflow of different pairwise distances which is employed in this paper. Pairwise distances methods have been formulated in (1), (2) and (3). Fig 5, represents the proposed method with an example. Table 1 represents the parameters value for DE and PSO algorithms used for finding initial cluster centers.

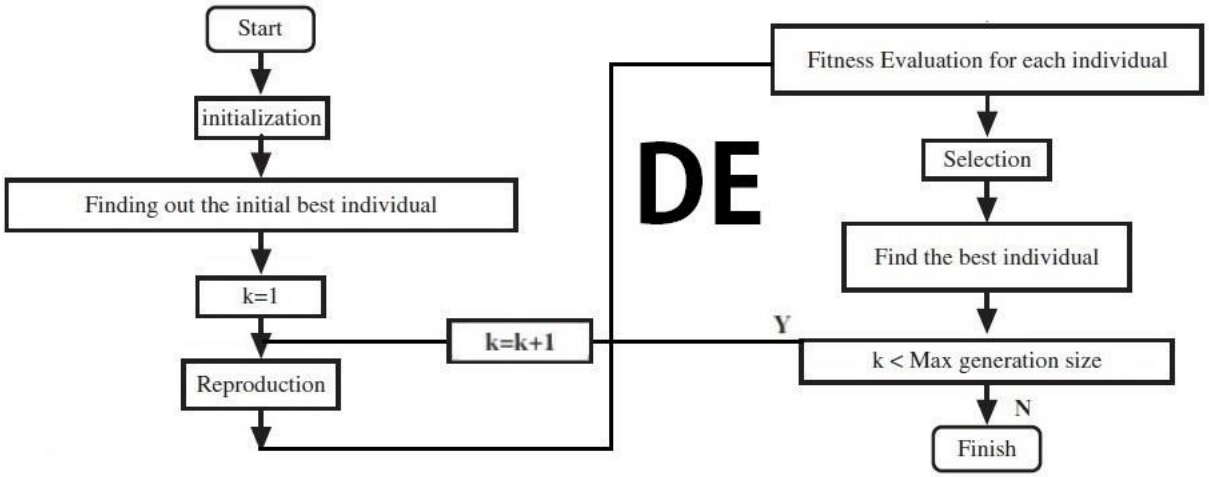

Figure 1: Flowchart of DE algorithm

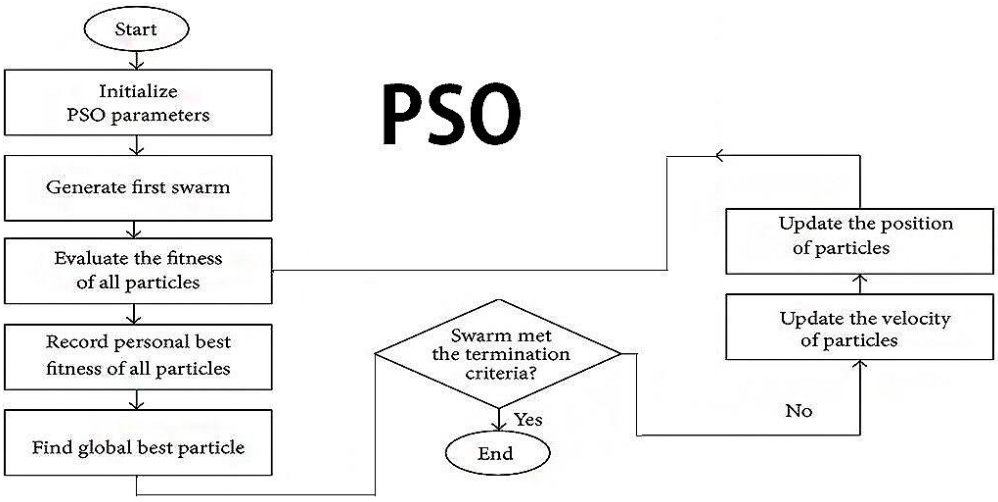

Figure 2: Flowchart of PSO algorithm

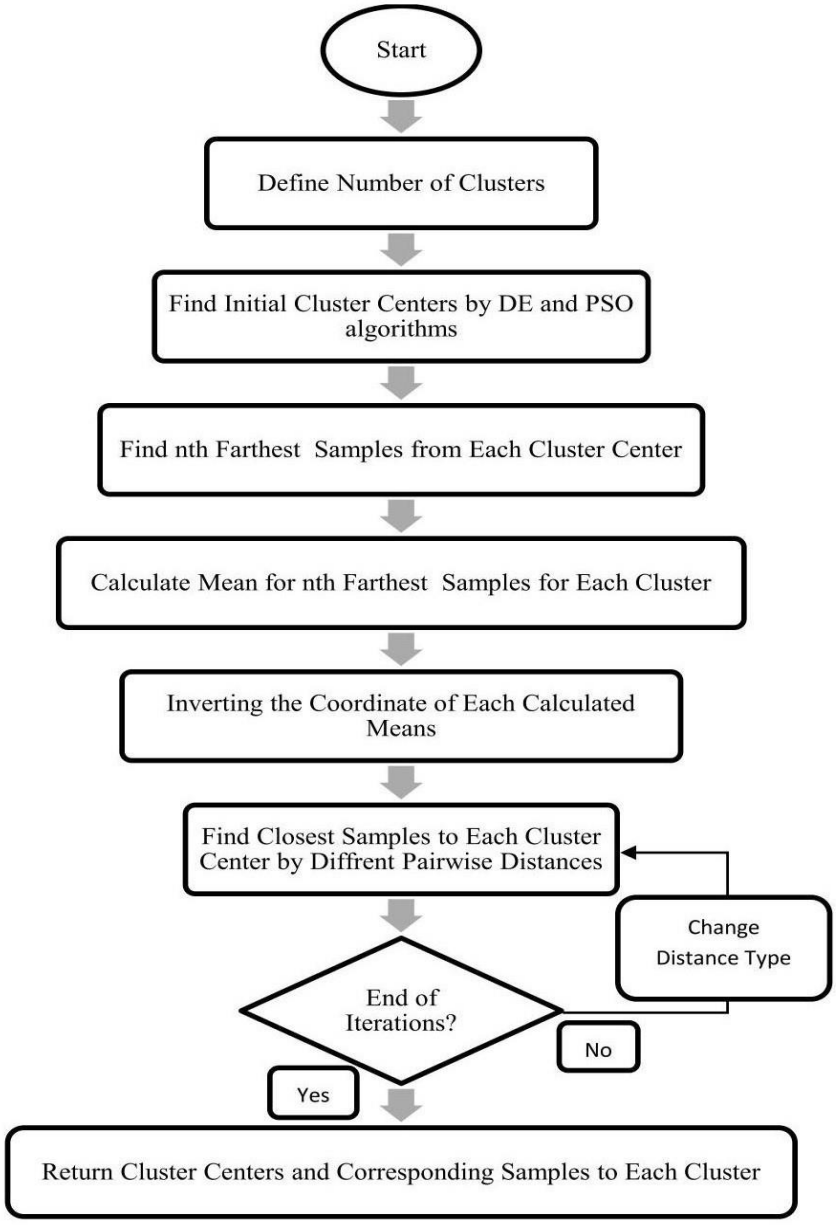

Figure 3: Proposed Method's Flowchart

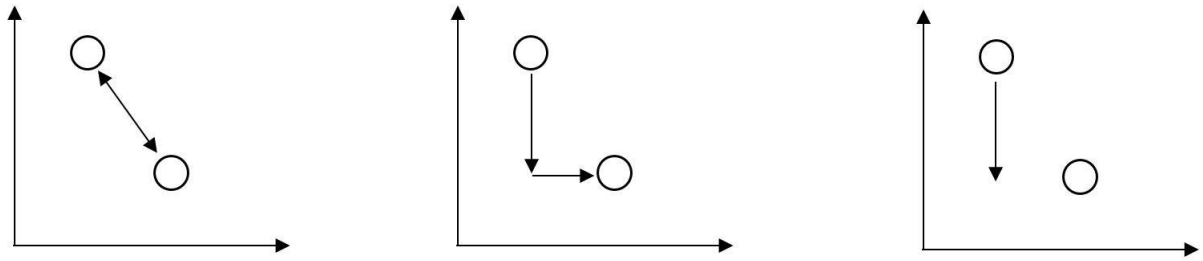

Figure 4: Different pairwise distances workflow between two points and two dimensions. From left to right (Euclidean, city-block, chebyshev or chessboard)

Given an m-by-n data matrix X, which is treated as m (1-by-n) row vectors x1, x2, ..., xm, the various distances between the vector x_s and x_t are defined as follows:

$$d_{st}^2 = (x_s - x_t) - (x_s - x_t)' \quad (1)$$

$$d_{st}^2 = \sum_{j=1}^{n} |x_{sj} - x_{tj}| \quad (2)$$

$$d_s = max_j\{|x_{sj} - x_{tj}|\} \quad (3)$$

Table 1: Parameters value for DE and PSO algorithms used for finding initial cluster centers

DE							
Lower Bound	Upper Bound	Iterations	Population Size	Lower Bound of Scaling Factor	Upper Bound of Scaling Factor	Crossover Probability	Mutation Probability
[0 0]	[1.4 1.8]	200	60	0.2	0.8	0.2	0.2
PSO							
Lower Bound	Upper Bound	Iterations	Swarm Size	Inertia Weight	Cognitive Coefficient	Crossover Probability	Mutation Probability
[0 0]	[1.4 1.8]	300	80	0.8	1.5	0.2	0.2

As an example in Fig 5, there are 8 step by step parts which demonstrate the whole the procedure. Part 1: differential evolution algorithm performance in 200 iterations with optimized value of 73.80 based on Table 1. Part 2: differential evolution algorithm result for 2 clusters (finding initial cluster centers). Part 3 and 4: green stars are initial cluster centers founded by DE algorithm. Red circles are, nth farthest samples from each cluster, which n is 5 in this example. Yellow circles are, mean of nth farthest samples from each cluster.

4. Validation and results

This system has been validated with some of the benchmark classification datasets (without label) and also clustering datasets like Fisher-iris [28], Ionosphere [29], User-Knowledge-Modeling [30] [31], Breast cancer [37] and Blood transfusion [38], and compared with famous clustering methods like K-means [7] [8], Fuzzy c-means (FCM) [9] [10], Self-Organizing Maps (SOM) [33] [39] and Gaussian mixture model (GMM) [13], and returned satisfactory results. Table 2 represents used datasets information in the research. Also Table 3 shows validation results for proposed method versus other four famous methods on 5 benchmark datasets (based on our test), using DE algorithm for setting initial clusters. Table 4 shows same target but using PSO algorithm for setting initial clusters.

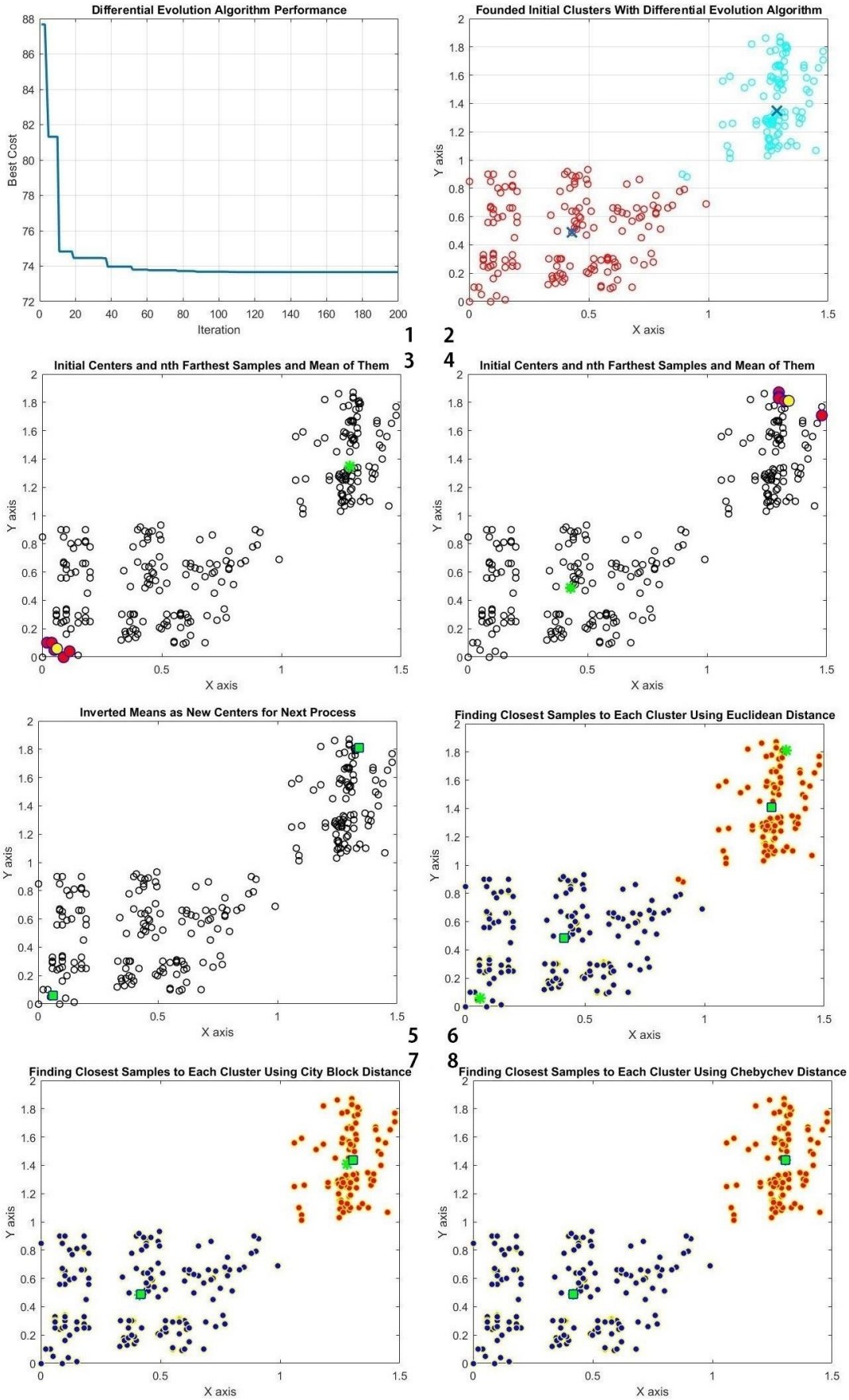

Figure 5: Proposed method with an example (3 distance iteration)

4.1. Datasets

Here we have five benchmark datasets for validation and comparison purposes. Classification datasets, consider as unlabeled data to be in unsupervised form.

Table 2: Datasets information

Dataset	Associated Tasks	Number of Instances	Number of Attributes	Number of Classes	Year
Fisher-iris	Classification	150	4	3	1936
Ionosphere	Classification	351	34	2	1989
User-Knowledge-Modeling	Classification, Clustering	403	5	4	2009
Breast cancer	Classification	286	9	2	1986
Blood transfusion	Classification	748	5	2	2008

4.2. Famous clustering methods

In following there are three famous clustering algorithms' formulas, which is formulated in (4) for K-means and (5) for Fuzzy c-means.

$$J = \sum_{j=1}^{k} \sum_{i=1}^{x} ||X_i^{(j)} - C_j||^2 \tag{4}$$

Where $||X_i^{(j)} - C_j||^2$ is a chosen distance measure between a data point $X_i^{(j)}$ and the cluster center C_j, is an indicator of the distance of the n data points from their respective cluster centers.

The FCM algorithm attempts to partition a finite collection of n elements X={x1,….xn} into a collection of c fuzzy clusters with respect to some given criteria. Given a finite set of data, the algorithm returns a list of c cluster centres C={c1,…,cc} and a partition matrix. W=w i,j E [0,1], i=1,…,n, j=1,…,c, where each element, w i,j , tells the degree to which element, xi , belongs to cluster cj. The FCM aims to minimize an objective function:

$$arg\ min\ c \sum_{i=2}^{n} \sum_{j=1}^{c} w_{ij}^{m} ||x_i - c_j||^2$$

Where:

$$w_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\left|\left|x_i - c_j\right|\right|}{\left|\left|x_i - c_k\right|\right|}\right)^{\frac{2}{m-1}}} \tag{5}$$

The algorithm which is used in practice to find the mixture of Gaussians that can model the data set is called EM (Expectation-Maximization). The formula is detailed, but for more information about it refer to [32].

Self-organizing maps learn to cluster data based on similarity, topology, with a preference (but no guarantee) of assigning the same number of instances to each class. Self-organizing maps are used both to cluster data and to reduce the dimensionality of data. They are inspired by the sensory and motor mappings in the mammal brain, which also appear to automatically organizing information topologically. Self-organizing feature maps (SOFM) learn to classify input vectors according to how they are grouped in the input space. They differ from competitive layers in that neighboring neurons in the self-organizing map learn to recognize neighboring sections of the input space. Thus, self-organizing maps learn both the distribution (as do competitive layers) and topology of the input vectors they are trained on [33] [39].

In sum, learning occurs in several steps and over many iterations. :

1. Each node's weights are initialized.
2. A vector is chosen at random from the set of training data.
3. Every node is examined to calculate which one's weights are most like the input vector. The winning node is commonly known as the Best Matching Unit (BMU).
4. Then the neighborhood of the BMU is calculated. The amount of neighbors decreases over time.
5. The winning weight is rewarded with becoming more like the sample vector. The neighbors also become more like the sample vector. The closer a node is to the BMU, the more its weights get altered and the farther away the neighbor is from the BMU, the less it learns.
6. Repeat step 2 for N iterations.

For Tables 3 and 4, there are following items for validation: the intra-cluster distances, i.e. the distance between data vectors within a cluster, where the objective is to minimize the intra-cluster distances. The inter-cluster distances, i.e. the distance between the centroids of the clusters, where the objective is to maximize the distance between clusters. Quantization, in mathematics and digital signal processing, is the process of mapping input values from a large set to output values in a smaller set. The difference between an input value and its quantized value is referred to as quantization error. More info on these items in [40].

Fig 6, 7 and 8 represent results of different clustering methods on ionosphere dataset (dimensions 33 and 34), Fisher-Iris dataset (dimensions 2 and 4) and User-Knowledge-Modeling dataset (dimensions 1 and 5) vs proposed method result (using DE algorithm at first step). Fig 9 and fig 10 represent quantization error for comparing the performance of each method (include our method) on 5 benchmark datasets using DE and PSO algorithms as initial starter.

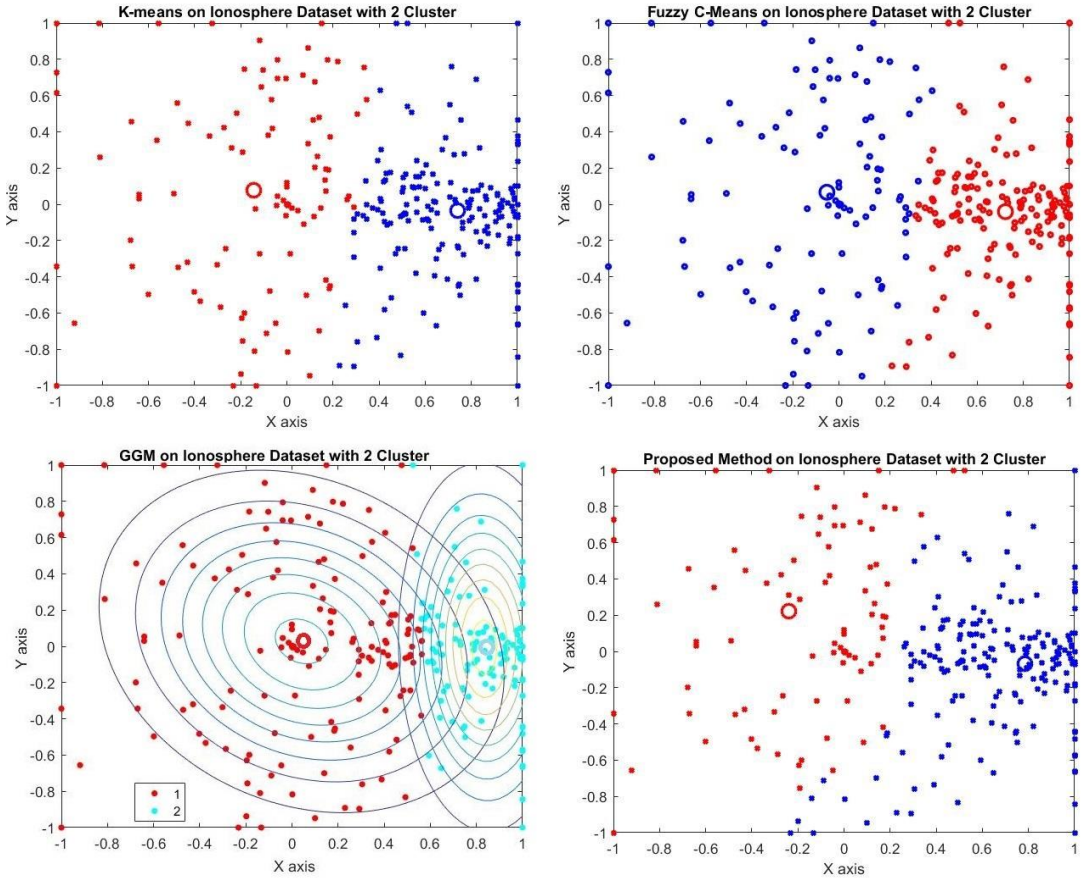

Figure 6: Results of different clustering methods on ionosphere dataset (dimensions 33 and 34) vs proposed method result

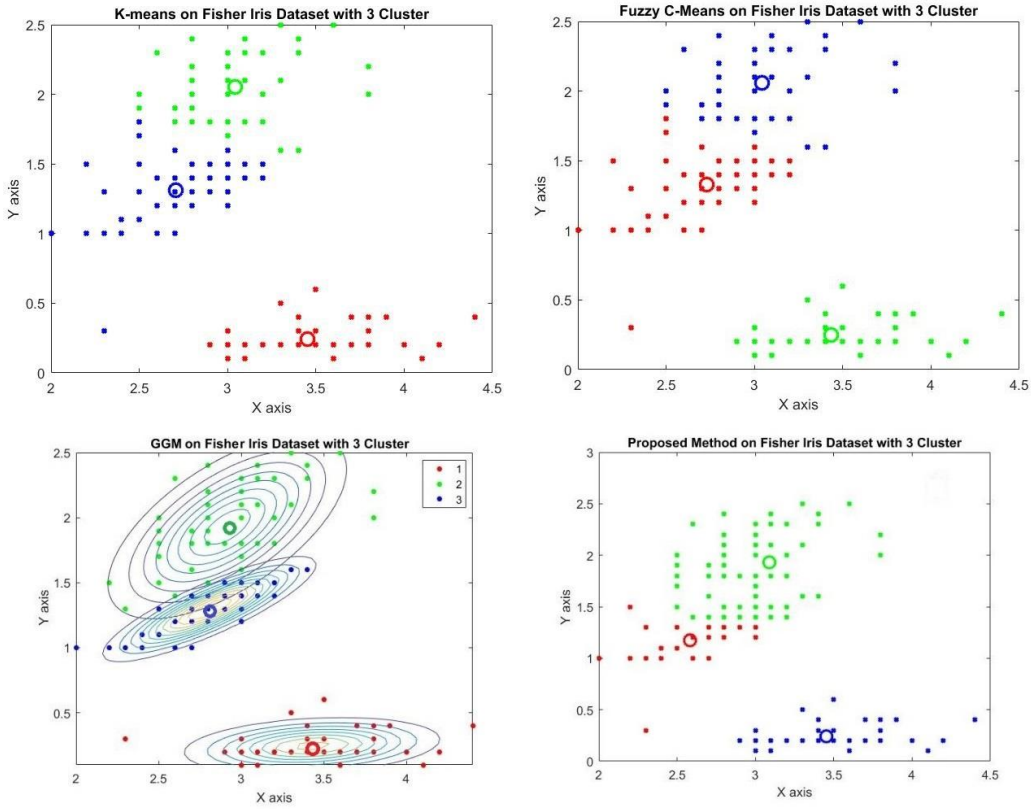

Figure 7: Results of different clustering methods on Fisher-Iris dataset (dimensions 2 and 4) vs proposed method result

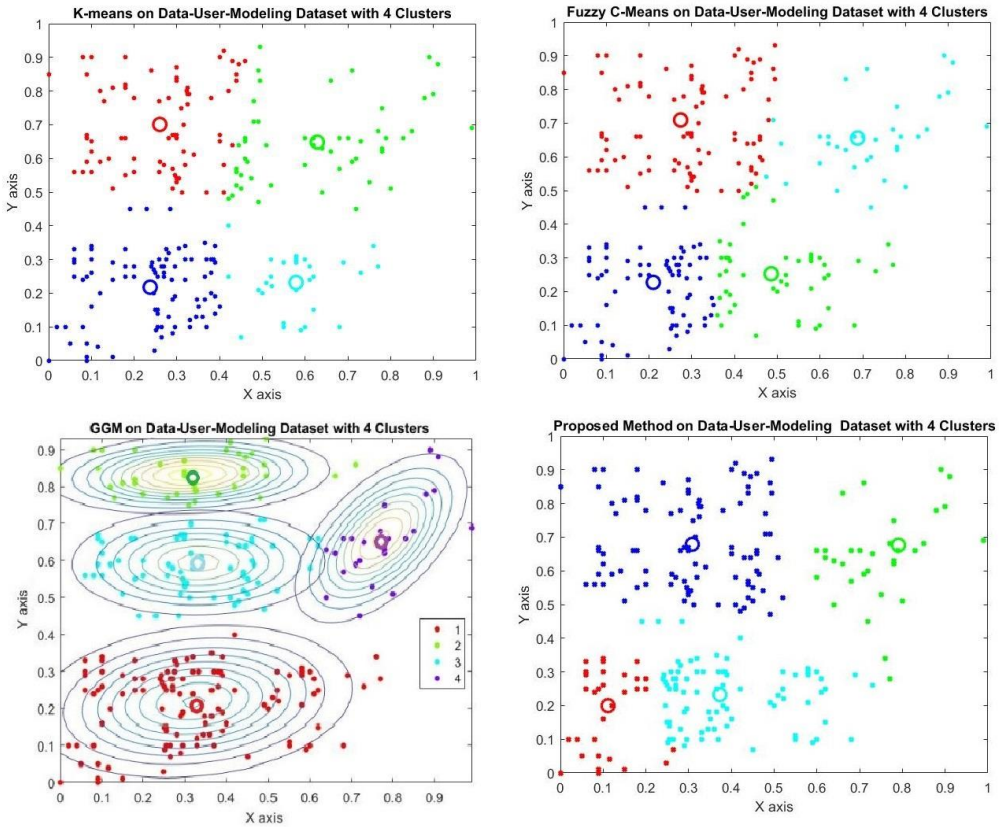

Figure 8: Results of different clustering methods on Data-User-Modeling dataset (dimensions 1 and 5) vs proposed method result

Table 3: Validation results for proposed method versus other four famous methods on 5 benchmark datasets using DE algorithm for setting initial clusters

Dataset (problem)	Algorithm	Quantization error	Intra-cluster distances	Inter-cluster distances
Fisher-iris	K-means	0.649±0.146	3.374±0.245	0.887±0.091
	FCM	0.677±0.242	3.778±0.890	0.917±0.079
	GMM	0.730±0.106	3.114±0.142	0.832±0.001
	SOM	0.650±0.344	4.471±0.200	0.845±0.092
	Proposed	**0.611±0.003**	3.281±0.995	0.813±0.028
Ionosphere	K-means	1.898±0.158	4.859±0.387	1.224±0.040
	FCM	2.179±0.288	5.249±0.240	3.528±0.072
	GMM	3.003±0.011	5.432±0.757	2.421±0.313
	SOM	**1.887±0.144**	4.991±0.301	1.628±0.110
	Proposed	1.970±0.000	4.443±0.370	1.514±0.101
User-Knowledge-Modeling	K-means	3.256±0.313	3.121±0.005	3.001±0.014
	FCM	3.666±0.072	7.499±0.012	6.555±0.013
	GMM	4.909±0.012	4.405±0.003	2.727±0.991
	SOM	3.411±0.040	5.000±0.666	4.312±0.100
	Proposed	**3.100±0.020**	7.113±0.124	5.500±0.009
Breast cancer	K-means	1.999±0.054	6.599±0.332	1.824±0.251
	FCM	2.179±0.257	7.119±0.300	2.822±0.159
	GMM	2.143±0.009	7.495±0.887	3.821±0.222
	SOM	1.980±0.151	6.521±0.312	1.324±0.881
	Proposed	**1.935±0.880**	6.633±0.319	2.814±0.208
Blood transfusion	K-means	0.993±0.100	12.356±5.002	10.400±2.011
	FCM	3.122±0.221	11.412±4.244	11.756±2.001
	GMM	3.103±0.931	12.792±4.751	09.511±1.400
	SOM	1.856±0.166	15.330±1.101	08.002±4.109
	Proposed	**0.912±0.097**	11.999±2.319	11.998±2.997

Table 4: Validation results for proposed method versus other four famous methods on 5 benchmark datasets using PSO algorithm for setting initial clusters

Dataset (problem)	Algorithm	Quantization error	Intra-cluster distances	Inter-cluster distances
Fisher-iris	K-means	0.612±0.100	4.172±0.288	0.833±0.056
	FCM	0.627±0.002	3.648±0.325	0.900±0.037
	GMM	0.732±0.100	3.174±0.272	0.821±1.987
	SOM	0.757±0.399	4.241±0.224	0.795±0.000
	Proposed	**0.611±0.297**	3.280±1.090	0.919±1.818
Ionosphere	K-means	1.445±0.158	3.900±0.007	1.821±0.087
	FCM	1.899±0.288	3.749±0.200	1.008±1.212
	GMM	2.012±0.011	5.455±0.447	3.436±0.933
	SOM	1.667±0.144	3.561±1.398	1.621±0.136
	Proposed	**0.988±0.000**	2.144±0.333	1.901±0.909
User-Knowledge-Modeling	K-means	3.400±0.003	3.521±1.405	3.000±0.000
	FCM	3.920±0.011	6.411±0.066	6.144±0.119
	GMM	5.900±1.095	6.755±2.099	2.640±0.898
	SOM	3.444±0.540	5.256±0.001	4.300±0.128
	Proposed	**3.353±0.128**	6.173±0.422	5.303±1.033
Breast cancer	K-means	1.801±0.194	6.119±0.002	1.694±0.318
	FCM	2.166±0.821	8.997±1.312	2.892±0.109
	GMM	2.003±0.014	75.005±1.217	2.751±0.007
	SOM	1.883±0.001	6.540±0.014	1.311±0.814
	Proposed	**1.701±0.010**	4.007±0.008	2.800±0.007
Blood transfusion	**K-means**	**0.879±0.140**	07.512±5.072	12.410±2.922
	FCM	2.111±0.244	14.919±4.244	11.600±2.056
	GMM	3.196±0.950	10.700±4.241	09.923±1.111
	SOM	0.944±0.996	14.000±1.163	09.082±4.800
	Proposed	0.901±0.002	12.639±2.306	10.361±2.689

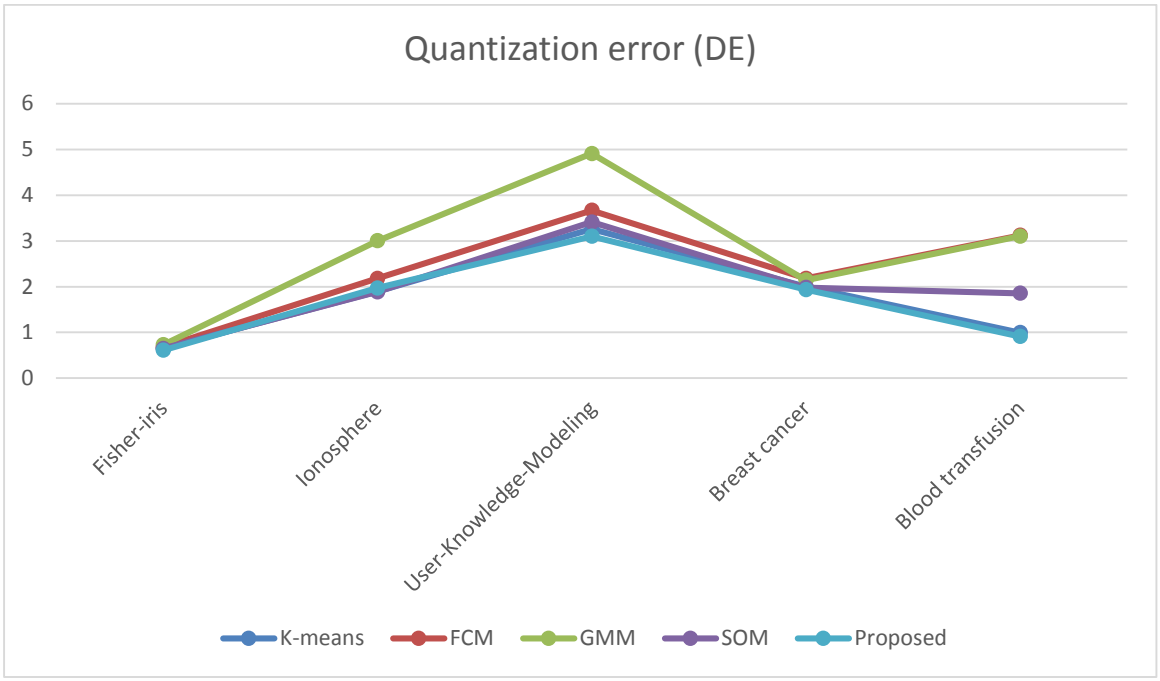

Figure 9: Quantization error for comparing the performance of each method on 5 datasets using DE algorithm as initial starter

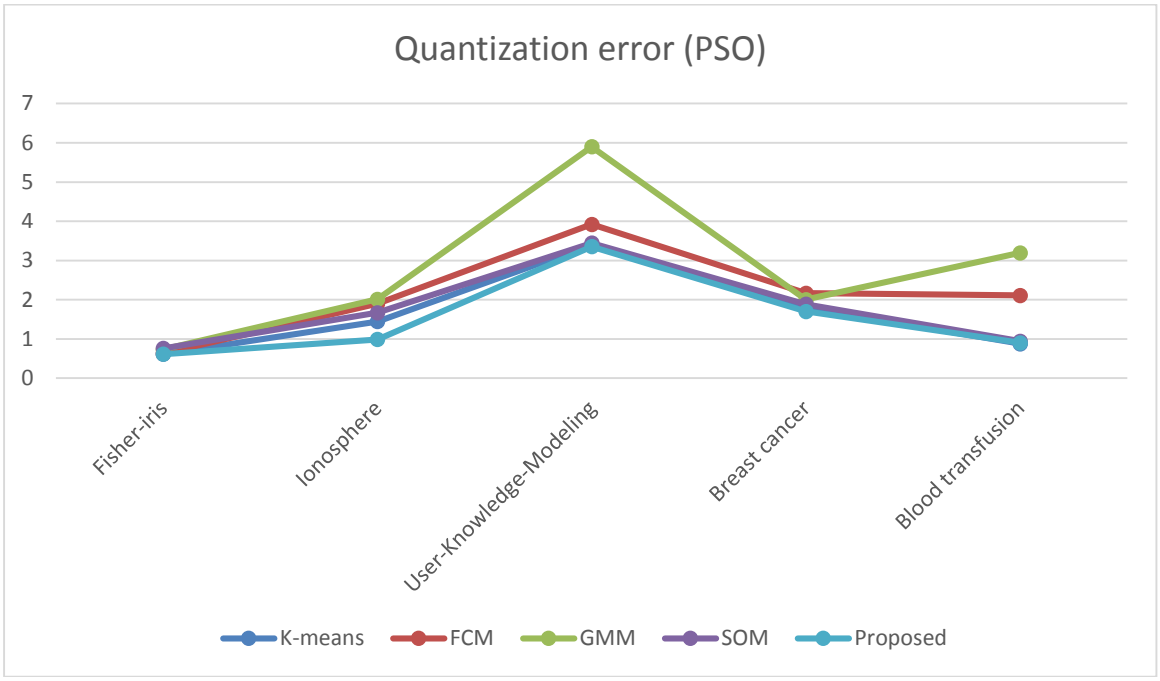

Figure 10: Quantization error for comparing the performance of each method on 5 datasets using PSO algorithm as initial starter

4.3. Clustering result

According to Tables 3 and 4 and based on Figures 9 and 10, lowest quantization error is belongs to proposed clustering method except for Ionosphere dataset in Table 3 which is belongs to SOM with 1.887 error percentage and Blood transfusion dataset with 0.879 error percentage which is belongs to K-means clustering method in Table 4. Other than that, all the lowest error percentages are belong to proposed method.

Conclusion and suggestion

Combining evolutionary clustering and using different pairwise distances and some other pre-processing methods, can make suitable clustering system. As it is clear in sections 3 and 4, proposed system can compete with famous clustering methods and return promising results. With this approach, the problem of initial cluster center state is fixed. Applying proposed method on some other benchmark datasets and comparing with some of other clustering methods like GA, PSO and DE clustering is of future works.

References

[1] Du, K-L. "Clustering: A neural network approach." Neural networks 23.1 (2010): 89-107.

[2] Mangiameli, Paul, Shaw K. Chen, and David West. "A comparison of SOM neural network and hierarchical clustering methods." European Journal of Operational Research 93.2 (1996): 402-417.

[3] C. Linnaeus, Clavis Classium in Systemate Phytologorum in Bibliotheca Botanica. Amsterdam, The Netherlands: Biblioteca Botanica, 1736.

[4] R. C. Tryon, Cluster Analysis. Ann Arbor, MI: Edward Brothers, 1939.

[5] J. A. Hartigan, Clustering Algorithms. New York: Wiley, 1975.

[6] González, Jesús, et al. "A new clustering technique for function approximation." IEEE Transactions on Neural Networks 13.1 (2002): 132-142.

[7] MacQueen, James. "Some methods for classification and analysis of multivariate observations." Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. Vol. 1. No. 14. 1967.

[8] Lloyd, Stuart. "Least squares quantization in PCM." IEEE transactions on information theory 28.2 (1982): 129-137.

[9] Dunn, Joseph C. "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters." (1973): 32-57.

[10] Bezdek, James C. Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media, 2013.

[11] Suganya, R., and R. Shanthi. "Fuzzy c-means algorithm-a review." International Journal of Scientific and Research Publications 2.11 (2012): 1.

[12] Theodoridis, S., and K. Koutroumbas. "Pattern recognition. 2009."

[13] McLachlan, Geoffrey, and David Peel. Finite mixture models. John Wiley & Sons, 2004.

[14] Du, K-L. "Clustering: A neural network approach." Neural networks 23.1 (2010): 89-107.

[15] Gabrys, Bogdan, and Andrzej Bargiela. "General fuzzy min-max neural network for clustering and classification." IEEE transactions on neural networks 11.3 (2000): 769-783.

[16] MLA Das, Swagatam, Ajith Abraham, and Amit Konar. "Automatic clustering using an improved differential evolution algorithm." IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans 38.1 (2008): 218- 237.

[17] Duda, Richard O., Peter E. Hart, and David G. Stork. Pattern classification. John Wiley & Sons, 2012.

[18] zimmermann h.j. “fuzzy set theory and its applications”, 2nd edition,kluwer academic pub,(1991).

[19] Klir, G. J., & Folger, T .A., “Fuzzy sets,Uncertainty,and Information”,Prentice-Hall,(1988).

[20] Mosqueira-Rey, E., Moret-Bonillo, V., & Fernández-Leal, Á. (2008). “An expert system to achieve fuzzy interpretations of validation data. Expert Systems with Applications”, 35(4), 2089–2106.

[21] Eiben, Agoston E., and James E. Smith. Introduction to evolutionary computing. Vol. 53. Heidelberg: springer, 2003.

[22] R. Storn and K. Price. Differential Evolution: A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces. Technical Report TR-95-012, International Computer Science Institute, Berkeley, California, March 1995.

[23] R. Storn and K. Price. Differential Evolution - A Fast and Efficient Heuristic for Global Optimization over Continuous Spaces. Journal of Global Optimization, 11:341–359, 1997.

[24] K. V. Price, R. M. Storn, and J. A. Lampinen. Differential Evolution. A Practical Approach to Global Optimization. Springer, Berlin, 2005. ISBN 3-540-20950-6.

[25] Coello, Carlos A. Coello, Gary B. Lamont, and David A. Van Veldhuizen. Evolutionary algorithms for solving multi-objective problems. Vol. 5. New York: Springer, 2007.

[26] Deza, Elena; Deza, Michel Marie (2009). Encyclopedia of Distances. Springer. p. 94

[27] Cyrus. D. Cantrell (2000). Modern Mathematical Methods for Physicists and Engineers. Cambridge University Press. ISBN 0-521-59827-3.

[28] Fisher,R.A. "The use of multiple measurements in taxonomic problems" Annual Eugenics, 7, Part II, 179-188 (1936); also in "Contributions to Mathematical Statistics" (John Wiley, NY, 1950).

[29] Sigillito, V. G., Wing, S. P., Hutton, L. V., \& Baker, K. B. (1989). Classification of radar returns from the ionosphere using neural networks. Johns Hopkins APL Technical Digest, 10, 262-266.

[30] H. T. Kahraman, Sagiroglu, S., Colak, I., Developing intuitive knowledge classifier and modeling of users' domain dependent data in web, Knowledge Based Systems, vol. 37, pp. 283-295, 2013.

[31] Kahraman, H. T. (2009). Designing and Application of Web-Based Adaptive Intelligent Education System. Gazi University Ph. D. Thesis, Turkey, 1-156.

[32] A.P. Dempster, N.M. Laird, and D.B. Rubin (1977): "Maximum Likelihood from Incomplete Data via theEM algorithm", Journal of the Royal Statistical Society, Series B, vol. 39, 1:1-38

[33] Vesanto, Juha, and Esa Alhoniemi. "Clustering of the self-organizing map." IEEE Transactions on neural networks 11.3 (2000): 586-600.

[34] Kennedy, James. "Particle swarm optimization." Encyclopedia of machine learning. Springer, Boston, MA, 2011. 760-766.

[35] Das, Swagatam, Ajith Abraham, and Amit Konar. "Automatic clustering using an improved differential evolution algorithm." IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans 38.1 (2008): 218-237.

[36] Van der Merwe, D. W., and Andries Petrus Engelbrecht. "Data clustering using particle swarm optimization." Evolutionary Computation, 2003. CEC'03. The 2003 Congress on. Vol. 1. IEEE, 2003.

[37] Wolberg, W.H., & Mangasarian, O.L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. In Proceedings of the National Academy of Sciences, 87, 9193—9196.

[38] Yeh, I-Cheng, King-Jang Yang, and Tao-Ming Ting. "Knowledge discovery on RFM model using Bernoulli sequence." Expert Systems with Applications 36.3 (2009): 5866-5871.

[39] Map, Self-Organizing, and Teuvo Kohonen. "Self-organizing map." Proceedings of the IEEE 78 (1990): 1464-1480.

[40] Van der Merwe, D. W., and Andries Petrus Engelbrecht. "Data clustering using particle swarm optimization." Evolutionary Computation, 2003. CEC'03. The 2003 Congress on. Vol. 1. IEEE, 2003.

Authors

Seyed Muhammad Hussain Mousavi received his MS.c degree from Bu-Ali Sina University Hamadan-Iran. He received his MS.c degree in Artifactual Intelligence in 2017. His research interests are Evolutionary algorithms, Pattern recognition, Image Processing, Fuzzy Logic, Human computer interactions, Classifications and clustering, Artificial intelligence, RGB_D data, Expert systems, Kinect, Data mining, Facial expression recognition, Face recognition, Age estimation and Gender recognition.

Email: mosavi.a.i.buali@gmail.com

Phone Number: +98-09332892726