An improved method of fuzzy c-means clustering by using feature selection and weighting

Amirhadi Jahanbakhsh Pourjabari[†], **Mojtaba Seyedzadegan**[††*]

[†] Department of Computer, Buinzahra Branch , Islamic Azad University, Buinzahra , Iran
[††] Department of Computer and Electrical Engineering , Buein Zahra Technical University, Buein Zahra, Qazvin , Iran
***** Corresponding Author: zadegan@bzte.ac.ir; Tel.: +989203186280***

Summary

Fuzzy C-means has been utilized successfully in a wide range of applications, extending from the clustering capability of the K-means to datasets that are uncertain, vague and otherwise are hard to be clustered. In cluster analysis, certain features of a given data set may exhibit higher relevance in comparison to others. To address this issue, Feature-Weighted Fuzzy C-Means approaches have emerged in recent years. However, there are certain deficiencies in the existing methods, e.g., the elements in a feature-weight vector cannot be adaptively adjusted during the training phase, and the update formulas of a feature-weight vector cannot be derived analytically. In this study, an Improved Feature-Weighted Fuzzy C-Means is proposed to overcome to these shortcomings. A novel initialization scheme for the fuzzy c-means algorithm was proposed. Finally, the proposed method was applied into data clustering. The experimental results showed that the proposed method can be considered as a promising tool for data clustering.

Key words:
Data mining, Clustering, Fuzzy c-means clustering (FCM), Feature-weight vector.

1. Introduction

Machine learning is an important branch of artificial intelligence whose main goal is to study and propose those methods which are able to learn from data. In many problems, machine learning may be applied where the process of extracting information from data is complex [1]. These methods have been employed with both supervised and unsupervised learning, where in the case of supervised learning, class labels are used to guide the machine learning algorithm. For several problems, it is not feasible to obtain data labels, therefore approaches for unsupervised learning like clustering are more suitable to solve these problems.

Clustering algorithms play an important role in discovering useful knowledge from large databases. The goal is that the objects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups. Clustering is a mathematical tool that attempts to discover structures or certain patterns in a dataset, where the objects have a certain degree of similarity inside each cluster. It can be achieved by various algorithms. Cluster analysis is a repetitive process of knowledge discovery. It will require modifying parameter and preprocessing until the results achieve the desired properties.

Clustering is a useful tool for understanding and visualizing available structures in data. Fuzzy C-means is one of the commonly used and efficient objective function-which is based on clustering techniques. Data clustering or cluster analysis is an important field in pattern recognition, machine intelligence and computer vision community, that has had numerous applications in the last three decades. Generally, clustering term is known as grouping a set of N samples into C clusters whose members are similar in some sense. This similarity between different samples is either a suitable distance based on numeric attributes, or directly in the form of pair-wise similarity or dissimilarity measurements. With a clustering technique, a collection of objects or feature vectors is partitioned into clusters. In the past few decades, many clustering algorithms have been developed, which mainly contain hierarchical clustering (such as Single Link and Complete Link), partitional clustering (such as k-means, fuzzy C-means, Gaussian Mixture and Density Estimation) and spectral clustering. Moreover, in the last few years, fuzzy C-means (FCM) clustering and spectral clustering algorithms are research focus [2].

Fuzzy clustering introduces the concept of membership into data partition, for this reason that membership can indicate the degree to which an object belongs to the clusters definitely, and actually represents the data partition more clearly. The fuzzy set theory was introduced by Zadeh [3], and it was successfully applied in image segmentation. The fuzzy c-means algorithm was proposed by Bezdek [4], based on fuzzy theory, it is the most widely studied and used algorithm in data clustering for its simplicity and ability to retain more information from images.

For most of the previous works of fuzzy c-means clustering, clustering data was used to demonstrate that the performance of fuzzy c-means clustering is affected by different feature weight, but if the feature weights are not properly chosen for fuzzy c-means clustering, the algorithm performs poorly. Therefore, it is important to

Manuscript received October 5, 2016
Manuscript revised October 20, 2016

select suitable feature weights to ensure the proper performance of fuzzy c-means clustering. Previous works proposed a feature-weight learning approach based on a defined similarity measure and an evaluation function to improve the performance of fuzzy c-means clustering. However, the defined similarity measure and evaluation function are complicated and difficult to interpret. In this paper, we propose a simple approach for fuzzy c-means weighting. In this paper, the features with higher relevance are more important to form the optimal clustering result than those with lower relevance.

The rest of the paper is organized as follows: In Section 2, the previous works is reviewed. In Section 3, the modified fuzzy c-means clustering algorithm is proposed. The experimental results and the comparison with a set of algorithms from the literature are presented in Section 4. Finally, in Section 5, we draw a conclusion and discuss the developing prospects of this work.

2. Related Work

Clustering plays an important role in pattern recognition, image processing, and computer vision. By a clustering technique, a collection of objects or feature vectors are partitioned into clusters. In the past few decades, many clustering algorithms have been developed, which mainly contain hierarchical clustering (such as Single Link and Complete Link), partitional clustering (such as k-means, fuzzy C-means, Gaussian Mixture and Density Estimation) and spectral clustering. Moreover, in the last few years, fuzzy C-means (FCM) clustering and spectral clustering algorithms are researching focuses.

Unlike the hard clustering techniques (each object is assigned to one and only one cluster), fuzzy C-means clustering allows an object to belong to a cluster with a grade of membership. Moreover, when there is not enough information about the structure of the data, fuzzy C-means clustering algorithm can handle this uncertainty better, and has been widely applied to the data clustering area. However, there are still some open problems in FCM algorithm [5, 6].

In Clustering, one of the most widely used algorithms is fuzzy clustering algorithms. Fuzzy set theory was first proposed by Zadeh in 1965 [3] & it gave an idea of uncertain belonging which was described by a membership function. For each clusters, the data points are assigned for membership values and fuzzy clustering algorithm allow the clusters to grow into their natural shapes. The fuzzy clustering algorithms can be divided into two types: one is Classical fuzzy clustering algorithms and the other is Shape based fuzzy clustering algorithms.

Fuzzy clustering is a powerful unsupervised method for the analysis of data and construction of models. In many cases, fuzzy clustering is more natural than hard clustering. Objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned for membership degrees between 0 and 1 indicating their partial membership. Fuzzy c-means algorithm is the most widely used algorithm in this regard.

As a partitioning clustering method, Fuzzy C-Means (FCM) has been widely studied in many fields [7, 8]. FCM was first proposed by Dunn in 1974 [9] and then it was developed by Bezdek [10].

In [11] a fuzzy clustering approach for time series based on cepstral coefficients, i.e., based on the fuzzy logic is proposed. Time series were classified in the frequency domain by considering their cepstral representations.

In [12] an improved method for image segmentation using the fuzzy c-means clustering algorithm (FCM) is proposed. They suggested these results for further improvement by acting at three different levels. This algorithm is widely experimented in the field of image segmentation, revealing very successful results.

In recent years, many improved versions of FCM have been proposed [13-16]. To eliminate the shortcoming caused by the random selection of the initial centers, [13] a new approach to the initialization of the FCM algorithm is proposed. The choice of initialization scheme is of importance because the optima and partitions identified by the FCM algorithm may vary depending on the selected initial cluster centroids. This method is based on the idea that the dominant colors in an image are likely to belong to separate clusters.

The [17] effectiveness of introducing the K-means++ initialization scheme into the context of Fuzzy C-means is investigated empirically. This method improves the way in which Fuzzy C-means initializes its clusters and has several advantages over the discussed methods. This method achieves superior clustering (in terms of validity indexes) compared to a used random initialization like the standard and fewer iterations.

Graves and Pedrycz [18] presented a comprehensive comparative analysis of kernel-based fuzzy clustering and fuzzy clustering. The kernel-based clustering algorithms can cluster specific non-spherical clusters such as the ring cluster quite well, performing FCM and GK for the same number of clusters.

The blind application of the conventional FCM algorithm to image segmentation often performs badly because: (i) FCM is very sensitive to noise and imaging artifacts since segmentation is decided only by pixel intensities, i.e. no spatial information in the image context is considered; (ii) The efficiency of FCM highly depends on the initialization step, because the iterative process easily falls into a locally optimal solution; (iii) The FCM algorithm is based on the Euclidean metric distance, so only it can be used to detect the data classes with the same super spherical shapes.

For the FCM and its improved versions, it is assumed that all the features of the samples in a given data set make equal contribution while constructing the optimal clusters. However, for certain real-world data sets, some of the features can exhibit higher relevance in the clustering information than others. Thus, the features with higher relevance are more important to form the optimal clustering results than those with lower relevance. Therefore, it is desirable to revisit an FCM method in which different features possess different weights. Recently, several Feature-Weighted Fuzzy C-Means (FWFCM) approaches [19-21] have been proposed to address the mentioned problem. These variants exhibit two separate stages. At the first stage, the feature-weight vector is determined. Then, at the second stage, the FWFCM is trained by the samples with their features weighted by the obtained feature-weight vector. Unfortunately, the elements of the feature-weight vector are fixed in this second stage, which might not fully reflect their relevance in the clustering process. Therefore, much effort has been devoted to adjust the feature-weight vector during the training course of the FCMs[22, 23]. However, these approaches assign different feature weights for different features of the clusters rather than for different features of the entire data set. The fuzzy c-means algorithm (FCM) is one of the most popular fuzzy clustering algorithms where the membership degrees of the data are obtained through iterative minimization of a cost function, subject to the constraint that the sum of membership degrees over the clusters for each data are equal with 1. The FCM algorithm suffers from several drawbacks: it also tries to minimize the intra-cluster variance as well, and has the same problems like k-means algorithm; the minimum is a local minimum, and the results depend greatly on the initializations In addition, the FCM algorithm is very sensitive to the presence of noise. The membership of noise points might be significantly high. The FCM algorithm cannot distinguish between equally highly likely and equally highly unlikely, and it is sensitive to the selection of distance metric.

3. Proposed Method

In order to improve the performance of clustering algorithm, many researchers have proposed different remarkable feature selection methods for clustering. Feature weighting can be considered as the generalization of feature selection since it assigns a proper weight value (which is on the closed interval [1, 0]) for each feature instead of giving either one, to retained features, or zero, to eliminated features.

The feature selection and weighting processes aim to improve the performance of the data mining tasks. Because no data reduction is performed, the main target is a better result of the data mining process, e.g., a better clustering or a more accurate classifier. Feature selection has been a fertile field of research and development since 1970s in statistical pattern recognition, machine learning and data mining

At present, feature weighted fuzzy clustering has become a very active area of research, and numerous approaches that develop weighted feature have been combined into fuzzy clustering.

With further and continuous research on weighted fuzzy clustering algorithms, many attributed weighting approaches have been incorporated into clustering. Feature weighting is considered to be a successful method for assessing the quality of attributes due to its simplicity and effectiveness.

In this paper, an improved version of feature-weighted fuzzy C-Means is proposed. In this version, the algorithmic framework and convergence properties of the proposed method are recapitulated. Proposed method dynamically updates feature-weight vectors in its training phase rather than utilizing a fixed feature-weight vector. Because the feature-weight vector of the traditional fuzzy c-means algorithm remains fixed during the clustering procedure, the significance of certain features to the changing cluster information cannot be appropriately manifested.

Let the sample to set D = {Xj} j=1N, with $\mathrm{X_j} = (\mathrm{x_{j1}, x_{j2}, \dots, x_{jd}}) \in \mathcal{R}^{\mathrm{d}}$, where N is the number of elements in the sample set, and D is the dimension of the feature space. The FCM clustering method minimizes the following objective function:

$$J(U,V;D) = \sum_{i=1}^{C}\sum_{j=1}^{N} \mu_{ij}^{m}\left[d_{ij}^{(w)}\right]^{2} \qquad (1)$$

where $\mathrm{U} = (\mu_{\mathrm{ij}})_{\mathrm{C\times N}}$ is a fuzzy partition matrix in which its element μ_{ij} denotes the membership of the jth sample, $\mathrm{X_j}$ belongs to the ith cluster, $\mathrm{V} = (\mathrm{V_1, V_2, \dots, V_C})^{\mathrm{T}} = (\mathrm{v_{iq}})_{\mathrm{C\times d}}$ is the center matrix that is composed of C cluster centers, m > 1 is the fuzzification exponent, and $\| . \|$ is the Euclidean norm. It should be noted that the membership μ_{ij} should satisfy the constraints $\mu_{\mathrm{ij}} \epsilon [0, 1]$ an$\sum_{\mathrm{i=1}}^{\mathrm{C}} \mu_{\mathrm{ij}} = 1$. Furthermore, $\mathrm{d_{ij}^{(w)}} = \|\mathrm{diag(w)(X_j - V_i)}\|$ with $\mathrm{w} = (\mathrm{w_1, w_2, \dots, w_d})$ is a feature-weight vector, $\mathrm{diag(w)} = \begin{pmatrix} \mathrm{w_1} & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mathrm{w_d} \end{pmatrix}$ while the elements $\mathrm{w_q}$ in the feature-weight vector w satisfy $\sum_{\mathrm{q=1}}^{\mathrm{d}} \mathrm{w_q} = 1$. Thereafter, we obtain the following update equations:

$$\mu_{ij=}\frac{1}{\sum_{k=1}^{C}\left(\frac{\left[d_{ij}^{(w)}\right]}{\left[d_{kj}^{(w)}\right]}\right)^{\frac{1}{m-1}}} \qquad (2)$$

$$V_i = \frac{\sum_{j=1}^{N} \mu_{ij}^{m} X_j}{\sum_{j=1}^{N} \mu_{ij}^{m}} \tag{3}$$

$$w_q^t = \frac{1}{\sum_{l=1}^{d} \left(\frac{\sum_{i=1}^{C} \sum_{j=1}^{N} \left[\mu_{ij}^{t}\right]^{m} \left(x_{jq} - v_{iq}\right)^{2}}{\sum_{i=1}^{C} \sum_{j=1}^{N} \left[\mu_{ij}^{t}\right]^{m} \left(x_{jl} - v_{il}\right)^{2}} \right)} \tag{4}$$

In this paper, in order to, initialize feature-weight vector, the term variance is used.

Term variance is the simplest univariate evaluation of the features and indicates that the features with larger values of variance contain valuable information. Term variance of feature can be calculated as follows:

$$TV(F_i) = \frac{1}{|S|} \sum_{j=1}^{|s|} (A_{ij} - \overline{A_i})^2 \tag{5}$$

where A_{ij} indicates the value of feature F_i for the pattern j, and $|S|$ is the total number of patterns

Algorithm. Feature-Weighted Fuzzy C-Means

```
Input       Dataset
Output      Final fuzzy partition matrix:
            Final center matrix:
            Final feature-weight vector:
  Begin algorithm
    Initialize number of clusters C , fuzzification exponent m and fuzzy partition matrix
    Initialize feature-weight vector using normalized Term Variance:
    while (not achieve termination condition)
      Update the cluster centers
      Calculate the distances
      Update the fuzzy partition matrix:
      Update the elements in the feature-weight vector :
    End while
  End algorithm
```

4. Experimental results

In order to evaluate the high performance of proposed method, performance test is illustrated for the proposed fuzzy clustering algorithm, and the proposed clustering algorithm results are compared with that of other state-of-the-art approaches.

The basic characteristics of these eight datasets are summarized in Table 1.

Table 1. The collection of data used

	#Samples	*#Features*	*#Clusters*
Iris	150	4	3
Crude oil	56	5	3
Bupa	345	6	2

The feature weights, the generated clustering error using the three different methods, i.e., the FWFCM was proposed by Wang et al., the FWFCM was proposed by Hung et al. and these proposed methods on the Iris data set, are illustrated in Table 2.

For the Iris data set, the two feature weighted fuzzy c-means and proposed method outperform the traditional FCM approach because their clustering error are all less than that of FCM. In comparison with two feature weighted fuzzy c-means methods, the proposed approach is the most efficient one because the learning time is the shortest and its error rates is less than other methods. From Table 1, we can see that, the error rate for FCM is 12. The performance of FCM clustering without feature weighting is the worst case. On the other hand, error rate for Wang et al and hung et al and Xing and Ming, are 6, 4 and 3.33 respectively.

The results for the different methods on the crude oil and Bupa data sets are included in Tables 3 and 4. Similar to Table 2, also it can be seen from the results that the proposed methods achieved the lowest error compared to the other methods.

From Tables 2–4, it can be easily observed that the proposed method achieves superior performance on all of the three data sets in comparison to other three methods.

Table 2. Clustering results obtained for Iris data set

Paper	*Feature-weight vector*	*Error*
FCM	-	12
Wang et al [24]	(0.0005, 0.0000,1.9829,0.1355)	6
Hung et al [22]	(0.1020,0.1022,0.3377,0.4580)	4
Xing and Ming [21]	(0.11, 0,11, 0.43, 0.33)	3.33
Our method	(0.1194,0.1134,0.4346,0.3327)	3.95

Table 3. Clustering results obtained for Crude_oil data set

Paper	*Feature-weight vector*	*Error*
FCM	-	37.5
[24]	(0.0542,1.6552, 0.8471,0.1593,0.0163)	37.5
[22]	(0.1570,0.1732,0.3638,0.1064,0.1996)	39.29
[21]	(0.1134,0.1007, 0.1023,0.6228,0.0608)	30.36
Our method	(0.1143,0.1071, 0.1123,0.6281,0.0382)	30.12

Table 4. Clustering results obtained for Bupa data set

Paper	*Feature-weight vector*	*Error*
FCM	-	48.41
[24]	(0.6240,0.5724,0.4655,0.7079,1.5158,0.8796)	47.25
[22]	(0.5365,0.0458,0.5099,0.4653,0.4813,0.0341)	49.86
[21]	(0.1536,0.0803, 0.2337,0.2361,0.2107,0.0856)	45.80
Our method	(0.1563,0.0831, 0.2117,0.2361,0.2107,0.1021)	45.72

To get an intuitive view, the clustering results of the above four approaches on the three data sets are depicted in Fig 1 to 3. These figures demonstrate the performances of the proposed methods on iris, Crude_oil and Bupa datasets, respectively. It can be observed from these figures that the samples which are scaled by an appropriate feature-weight vector can improve the performance of FCM, while an inappropriate vector can deteriorate the performance of FCM.

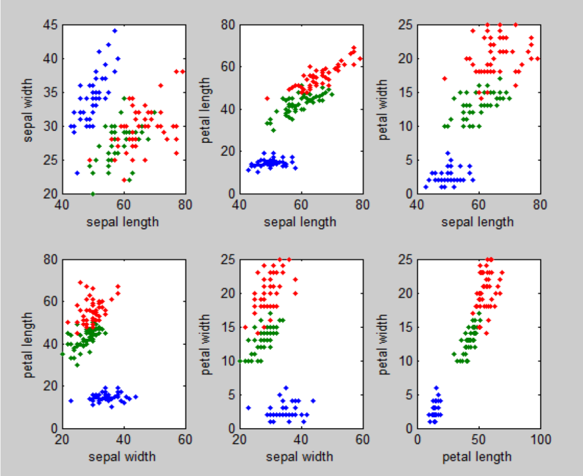


Fig1. The clustering results for iris data set

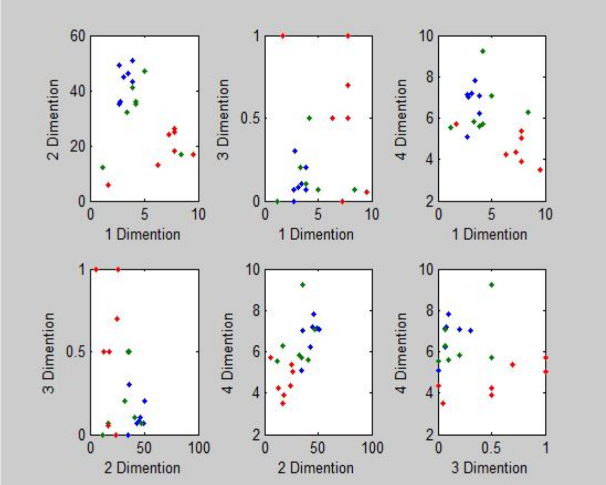


Fig2. The clustering results for Crude_oil data set

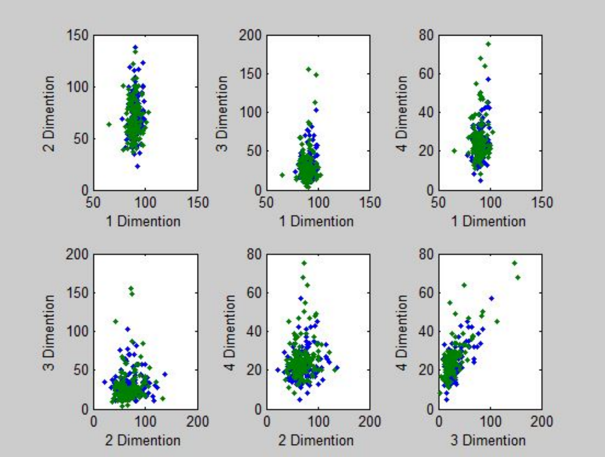


Fig3. The clustering results for Bupa data set

5. Conclusions

The fuzzy c-means algorithm is a widely applied clustering technique, but the implicit assumption that each attribute of the object data has equal importance, affects the clustering performance. At present, attributed weighted fuzzy clustering has become a very active area of research, and numerous approaches that developed numerical weights have been combined into fuzzy clustering. The proposed method has been studied over synthetic and real data sets with different characteristics. Experimental results show that the introduced method reveals that the proposed method is superior to the existing feature-weighted fuzzy c-means methods.

References

[1] Mitchell, T.M., "Machine Learning. McGraw-Hill",NewYork., 1997.

[2] Dervis Karaboga and C. Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm". Applied Soft Computing, 2011. 11: p. pp. 652-657.

[3] L.A. Zadeh, "Fuzzy sets". Information and Control 8, 1965: p. pp. 338-353.

[4] Bezdek, J.C., "Fuzzy Mathematics in Pattern Classification". Ph. D. Thesis, Applied Math. Center, Cornell University, Ithaca, 1973.

[5] K.R. Zalik, "Cluster validity index for estimation of fuzzy clusters of different sizes and densities", Pattern Recognition 43 (10) (2010) 3374–3390.

[6] R.K. Brouwer, A. "Groenwold, Modified fuzzy C-means for ordinal valued attributes with particle swarm for optimization", Fuzzy Sets and Systems 161 (13) (2010) 1774–1789.

[7] Yao, H., et al., "An improved k-means clustering algorithm for fish image segmentation". Mathematical and Computer Modelling, 2013. 58(3–4): p. 790-798.

[8] Karami, A. and M. Guerrero-Zapata, "A fuzzy anomaly detection system based on hybrid PSO-Kmeans algorithm in

content-centric networks". Neurocomputing, 2015. 149, Part C(0): p. 1253-1269.

[9] Dunn, J.C., "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well Separated Clusters". Journ. Cybern, 1974: p. 95-104.

[10] Bezdek, J.C., 1975. "Mathematical models for systematics and taxonomy". In: Estabrook, G.

[11] Ed.., Proc. 8th Internat. Conf. "Numerical Taxonomy". Freeman, San Francisco, CA, pp. 143–166.

[12] E.A. Maharaj, P. D'Urso, "Fuzzy clustering of time series in the frequency domain", Inform. Sci. 181 (2011) 1187–1211.

[13] Benaichouche, A.N., H. Oulhadj, and P. Siarry, "Improved spatial fuzzy c-means clustering for image segmentation using PSO initialization", Mahalanobis distance and post-segmentation correction. Digital Signal Processing, 2013. 23(5): p. 1390-1400.

[14] D.W. Kim, K.H.L., D. Lee,, "A novel initialization scheme for the fuzzy c-means algorithm for color clustering". pattern Recognition Letters, 2004. 25: p. pp. 227–237.

[15] Chenglong Tang, S.W., Wei Xu, "New fuzzy c-means clustering model based on the data weighted approach". Data & Knowledge Engineering, 2010. 69: p. pp. 881–900.

[16] Zhao, F., H. Liu, and J. Fan, "A multiobjective spatial fuzzy clustering algorithm for image segmentation". Applied Soft Computing, 2015. 30(0): p. 48-57.

[17] Jung-Yi Jiang, Ren-Jia Liou, and S.-J. Lee, "A Fuzzy Self-Constructing Feature Clustering Algorithm for Text Classification". IEEE Transactions Knowledge and Data Engineering, 2011. 23(3): p. 335 - 349

[18] Stetco, A., X.-J. Zeng, and J. Keane, "Fuzzy C-means++: Fuzzy C-means with effective seeding initialization". Expert Systems with Applications, 2015. 42(21): p. 7541-7548.

[19] D. Graves, W. Pedrycz, "Kernel-based fuzzy clustering and fuzzy clustering: a comparative experimental study", Fuzzy Sets Syst. 161 (2010) 522–543.

[20] Y. Yue, D. Zeng, L. Hong, "Improving fuzzy c-means clustering by a novel feature-weight learning", in: Proceedings of 2008 IEEE Pacific–Asia Workshop on Computational Intelligence and Industrial Application, 2008, pp. 173–177.

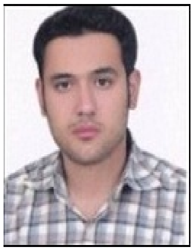
[21] Xing, H.-J. and M.-H. Ha, "Further improvements in Feature-Weighted Fuzzy C-Means". Information Sciences, 2014. 267(0): p. 1-15.

[22] W.L. Hung, M.S.Y., D.H. Chen, "Bootstrapping approach to feature-weight selection in fuzzy c-means algorithms with an application in color image segmentation". pattern Recognition Letters, 2004. 29: p. pp. 1317-1325.

[23] H. Frigui, O. Nasraoui, "Unsupervised learning of prototypes and attribute weights", Patt. Recog. 37 (2004) 567–581.

[24] H. Shen, J. Yang, S. Wang, X. Liu, "Attribute weighted mercer kernel based fuzzy clustering algorithm for general non-spherical datasets", Soft Comput. 10 (2006) 1061–1073.

[25] X.Z. Wang, Y.D.W., L.J. Wang, "Improving fuzzy c-means clustering based on feature-weight learning". pattern Recognition Letters, 2004. 25: p. pp. 1123–1132.



Amirhadi Jahanbakhsh received the B.S. degrees in Computer Engineering from Islamic Azad university of Tuyserkan ,Hamedan ,Iran in 2008-2011 M.S. degrees in Engineering Artificial Intelligence from Islamic Azad university of Buin zahra ,Qazvin ,Iran in 2013-2015, respectively.

Mojtaba Seyedzadegan received the M.S. degrees in Computer Networks from Universiti Putra Malaysia in 2005-2007 P.h.D. degrees in Computer Networks from Universiti Putra Malaysia in 2007-2011, respectively.