Bilateral Weighted Fuzzy C-Means Clustering

A. H. Hadjahmadi*, M. M. Homayounpour** and S. M. Ahadi***

Abstract: Nowadays, the Fuzzy C-Means method has become one of the most popular clustering methods based on minimization of a criterion function. However, the performance of this clustering algorithm may be significantly degraded in the presence of noise. This paper presents a robust clustering algorithm called Bilateral Weighted Fuzzy C-Means (BWFCM). We used a new objective function that uses some kinds of weights for reducing the effect of noises in clustering. Experimental results using, two artificial datasets, five real datasets, viz., Iris, Cancer, Wine, Glass and a speech corpus used in a GMM-based speaker identification task show that compared to three well-known clustering algorithms, namely, the Fuzzy Possibilistic C-Means, Credibilistic Fuzzy C-Means and Density Weighted Fuzzy C-Means, our approach is less sensitive to outliers and noises and has an acceptable computational complexity.

Keywords: Fuzzy Clustering, Fuzzy Possibilistic C-Means, Credibilistic Fuzzy C-Means, Density Weighted Fuzzy C-Means.

1 Introduction

Clustering can be considered as the most important unsupervised learning problem. Clustering algorithms try to partition a set of unlabeled input data into a number of clusters such that data in the same cluster are more similar to each other than to data in the other clusters [1]. Clustering has been applied in a wide variety of fields ranging from engineering (machine learning, artificial intelligence, pattern recognition, mechanical engineering, electrical engineering) [2-4], computer sciences (web mining, spatial database, analysis, textual document collection, image segmentation) [5,6], life and medical sciences (genetics, biology, microbiology, paleontology, psychiatry, clinic, pathology) [7-9], to earth sciences (geography, geology, remote sensing) [7], social sciences (sociology, psychology, archeology, education) and economics (marketing, business) [2, 10]. A large number of clustering algorithms have been proposed for various applications. These may be roughly categorized into two classes: hard and fuzzy (soft) clustering [1]. In fuzzy clustering, a given pattern does not necessarily belong to only one cluster but can have varying degrees of memberships to several clusters [11].

Among soft clustering algorithms, Fuzzy C-Means (FCM) is the most famous clustering algorithm. However, one of the greatest disadvantages of this method is its sensitivity to noises and outliers in the data [12-14]. Since the membership values of FCM for an outlier data is the same as real data, outliers have a great effect on the centers of the clusters [14].

There exist different methods to overcome this problem. Among them, three well-known robust clustering algorithms, namely, the Fuzzy Possibilistic C-Means (FPCM) [15, 16], Credibilistic Fuzzy C-Means (CFCM) [12, 13] and Density Weighted Fuzzy C-Means (DWFCM) [13] have attracted more attention.

In this paper we attempt to decrease the noise sensitivity in fuzzy clustering by using different kinds of weights in objective function, in order to decrease the effect of noisy samples and outliers on centroids. For the purpose of comparing different methods for computation of clustering weights and in order to compare our proposed and conventional clustering methods, two artificial datasets, their noisy versions and five real datasets were produced. Experimental results confirm the robustness of the proposed method.

This paper is organized as follows: Section 2 presents the mentioned Fuzzy C-Means motivated algorithms and their disadvantages. Our proposed Bilateral Weighted Clustering algorithm is described in Section 3. Section 4 describes experimental datasets and

Iranian Journal of Electrical & Electronic Engineering, 2012.
Paper first received 24 July 2011 and in revised form 24 April 2012.
* The Author is with the Department of Computer Engineering, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran.
E-mail: hadjahmadi@vru.ac.ir
** The Author is with the Department of Computer Engineering and Information Technology, Amirkabir University of Technology, Tehran, Iran.
E-mail: homayoun@aut.ac.ir
*** The Author is with the Department of Electrical Engineering and Information Technology, Amirkabir University of Technology, Tehran, Iran.
E-mail: sma@aut.ac.ir

Section 5 presents experimental results for clustering of datasets including outliers. Finally, conclusions are drawn in Section 6.

2 Some Fuzzy C-Means Motivated Algorithms

In this section, classical Fuzzy C-Means and three robust Fuzzy C-Means motivated algorithms will be studied and their performance and possible advantages and disadvantages discussed.

Given a set of input patterns $X = \{x_1, x_2, \ldots, x_N\}$, where $x_i = (x_{i1}, x_{i2}, \ldots, x_{in})^T \in \Re^n$. C-Means motivated clustering algorithms attempt to seek C cluster centroid vectors $\{v_1, v_2, \ldots, v_C \mid v_i \in \Re^n\}$, such that the similarity between patterns within a cluster is larger than the similarity between patterns belonging to different clusters.

2.1 The Fuzzy C-Means Clustering (FCM)

Fuzzy C-Means clustering (FCM) is the most popular fuzzy clustering algorithm. It assumes that the number of clusters, is known a priori, and minimizes

$$J_{fcm} = \sum_{i=1}^{C}\sum_{k=1}^{N} u_{ik}^m d_{ik}^2, \quad d_{ik} = \|x_k - v_i\| \tag{1}$$

with the constraint of:

$$\sum_{i=1}^{C} u_{ik} = 1 \quad ; k = 1, \ldots, N \tag{2}$$

Here, $m > 1$ is known as the fuzzifier parameter and any norm, $\| \ \|$, can be used (we use the Euclidean norm) [2,16]. The algorithm provides the fuzzy membership matrix U and the fuzzy cluster center matrix V.

Using the Lagrange multiplier method, the problem is equivalent to minimizing the following equation with constraints [2, 16]:

$$L(U, \lambda) = \sum_{i=1}^{C}\sum_{k=1}^{N} u_{ik}^m d_{ik}^2 + \sum_{k=1}^{N} \lambda_k \left(1 - \sum_{i=1}^{C} u_{ik}\right) \tag{3}$$

From Eq. (3), we readily obtain the following update equation:

$$u_{ik} = \left(\sum_{j=1}^{C} \left(d_{ik} / d_{jk} \right)^{2/(m-1)} \right)^{-1} \tag{4}$$

Then, we can assume that u_{ik} is a fixed number and plug it into Eq. (1) to obtain:

$$v_i = \sum_{k=1}^{N} \left(u_{ik}^m x_k \right) \Big/ \sum_{k=1}^{N} \left(u_{ik}^m \right) \tag{5}$$

Thus, we observe two main deficiencies associated with FCM [12]:

1. Inability to distinguish outliers from non-outliers by weighting the memberships.
2. Attraction of the centroids towards the outliers.

A noise-robust clustering technique should have the following properties [12]:

1. Should assign the outliers with low memberships to all the C clusters.
2. Centroids on a noisy set should not deviate significantly from those generated for the corresponding noiseless set, obtained by removing the outliers.

2.2 The Fuzzy Possibilistic C-Means Clustering (FPCM)

FPCM is a mixed C-Means technique which generates both probabilistic membership and typicality for each vector in the data set [12, 15, 16, 17]. FPCM [12, 15, 16, 17] minimizes the objective function

$$J_{fcm} = \sum_{i=1}^{C}\sum_{k=1}^{N} (u_{ik}^m + t_{ik}^\eta) d_{ik}^2 \quad , d_{ik} = \|x_k - v_i\| \tag{6}$$

where η is a parameter for controlling the effect of typicality on clustering and with constraints similar to Eq. (2), and also

$$\sum_{k=1}^{N} t_{ik} = 1 \tag{7}$$

must be satisfied [12, 15, 17].

Using the Lagrange multiplier method, the algorithm provides the fuzzy membership matrix U, the fuzzy typicality matrix T and the fuzzy cluster center matrix V respectively by equations Eq. (4) [12,15]

$$t_{ik} = \left[\sum_{j=1}^{N} \left(d_{ik} / d_{ij} \right)^{2/(\eta-1)} \right]^{-1} \tag{8}$$

$$v_i = \sum_{k=1}^{N} \left(u_{ik}^m + t_{ik}^\eta \right) x_t \Big/ \sum_{k=1}^{N} \left(u_{ik}^m + t_{ik}^\eta \right) \tag{9}$$

Due to the constraint Eq. (7), if the number of input samples (N) in a dataset is large, the typicality of samples will degrade and then the FPCM will not be insensitive to outliers. Therefore, a modified version of FPCM, called MFPCM is used [18]. In MFPCM the sum of typicality values of a cluster i, for all the input samples, is equal to the number of data that belongs to this cluster [19].

2.3 The Credibilistic Fuzzy C-Means Clustering (CFCM)

The idea of CFCM is to decrease the noise sensitivity in fuzzy clustering by modifying the probabilistic constraint Eq. (2) so that the algorithm generates low memberships for outliers [15]. To distinguish an outlier from a non-outlier, in [15], Chintalapudi and Kam introduced a new variable, credibility. Credibility of a vector represents its typicality to the data set, not to any particular cluster. If a vector has a low value of credibility, it is atypical to the data set and is considered as an outlier [15]. Thus, the credibility ψ_k, of a vector x_k is defined as:

$$\Psi_k = 1-(1-\theta)\alpha_k \Big/ \max_{j=1\ldots N}(\alpha_j) \quad ,0 \le \theta \le 1 \tag{10}$$

where $\alpha_k = \max(d_{ik})$ for $i = 1,\ldots,C$. Here, α_k is the distance of vector x_k from its nearest centroid. The parameter θ controls the minimum value of v_k so that the noisiest vector gets credibility equal to θ [15]. Hence, CFCM partitions X by minimizing Eq. (1) (the FCM objective function) subject to the constraint

$$\sum_{i=1}^{C} u_{ik} = \Psi_k \tag{11}$$

Likewise, the Lagrange multiplier method is used to derive the update Eq. (12) and Eq. (5) for CFCM.

$$u_{ik} = \Psi_k \Big/ \sum_{j=1}^{C} \left(d_{ik}/d_{jk}\right)^{2/(m-1)} \tag{12}$$

We note that since the original update equation for prototype (see [15]) in CFCM is host identical to that of FCM, we simply use Eq. (5) here. Also, it has been mentioned in [15] that using Eq. (5) may result in oscillations for noise-free data and for overlapped clusters, but the original update equation will not.

Although CFCM usually converges rapidly, it has some drawbacks. The major drawback of CFCM is that this method uses Eq. (10) for computing the credibility of an input vector. When there is at least one far outlier, the credibility values of weak outliers will be near to clean data by Eq. (10). Fig. 1 shows this phenomenon by a simple example. There are 8 clean samples and 2 outliers in this Figure, where v_1 is the cluster centroid vector, x_1 is a clean sample, x_2 is a weak outlier, x_3 is a far outlier and $d_i, i = 1,2,3$ are the distances between i 'th sample and v_1. Assume that d_1, d_2 and d_3 are respectively equal to 1, 3 and 20. Then according to Eq. (10), the credibility values of them equals to 0.95, 0.85 and 0 respectively. Therefore, equation Eq. (11) is faint to distinguish between weak outliers and clean data in the presence of at least one far outlier.

2.4 The Density Weighted Fuzzy C-Means Clustering (DWFCM)

In the DWFCM algorithm [10], Chen and Wang aim to identify the less important data point by using the potential measurement before clustering process, not during the convergence process of clustering. To achieve this goal, they modify Eq. (1) into

$$J = \sum_{k=1}^{N}\sum_{i=1}^{C} w_k (u_{ik})^m \left\| x_k - v_i \right\|^2 \tag{13}$$

where w_k is a density measurement,

$$w_k = \sum_{y=1}^{N} \exp(-h \times \left\| x_k - x_y \right\| / \sigma) \tag{14}$$

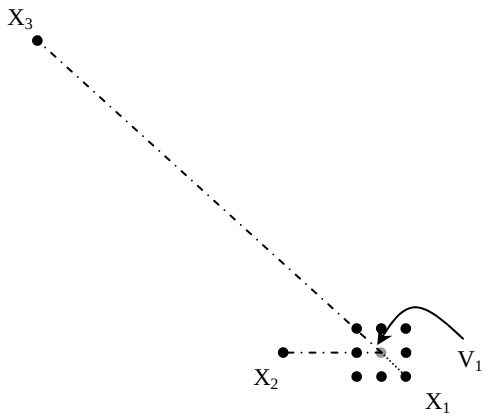


Fig. 1 A simple dataset with one cluster and 2 outliers. X_1 : A clean data sample. X_2 : A weak outlier. X_3 : A far outlier.

for which h is a resolution parameter and σ is the standard deviation of input data. Likewise, in [10] the Lagrange multiplier method is used to derive the following update equations for U and V for DWFCM.

$$v_i = \left(\sum_{k=1}^{N} w_k \left(u_{ik}^m\right) x_t\right) \Big/ \left(\sum_{k=1}^{N} w_k \left(u_{ik}^m\right)\right) \tag{15}$$

Although DWFCM also usually converges rapidly, it has its own drawbacks. The major drawback of DWFCM is that this method uses density motivated weights as clustering weights. According to Eq. (14), density weights can reduce the effect of the whole dataset outliers, not the outliers from a particular cluster. In other words, when the ratio of inter-cluster deviations to the whole data deviation is small, DWFCM works deplorably. Fig. 2 shows this drawback by a simple example.

There are two main clusters named C_1 and C_2 in Fig. 2. Cluster C_1 contains 1000 and cluster C_2 contains 100 samples. x_1 is a sample data from C_1, x_2 is a sample data from C_2 and x_3 is an outlier. The normalized distance between k 'th and y 'th data sample is defined as follows:

$$h_{ky} = \frac{h \times \left\| x_k - x_y \right\|}{STD} \tag{16}$$

If we assume that the ratio of inter-cluster deviations to the whole data deviation is small, then we can use h_{12} as the normalized distance between all data samples of C_1 and C_2, h_{13} as the normalized distance between all data samples of C_1 and x_3, and h_{23} as the normalized distance between all data samples of C_2 and x_3. In Fig. 2, three suppositional values for h_{ky}, where $ky \in \{12,13,23\}$, are shown. The density weights

according to Eq. (14) for these h_{ky} are depicted in Table 1.

Table 1 The density weights according to (14) for suppositional values of h_{ky} depicted in Fig. 1

k	1	2	3
w_k	1.0915e+003	1.0057e+003	1.0708e+003

From Table 1, we can observe that density weight of x_3 is greater than density weight of x_2. In other words, density weights increase the effect of outliers on the centroid of C_2, leading to poor performance of DWFCM, when the ratio of inter-cluster deviations to the whole data deviation is small.

3 Robust Weighted Fuzzy C-Means Clustering (RWFCM)

We attempt to decrease the noise sensitivity in fuzzy clustering by using different kinds of weights in objective function, so that the noisy samples and outliers have less effect on centroids. The basic idea of this approach is similar to DWFCM and RWFCM [20]. Two general kinds of weights can be used for achieving this aim; Cluster-independent weights (weights that are independent of a particular cluster) and cluster-dependent weights (those that depend on a particular cluster). Here we combine both kinds of weights and proposed a new robust weighted clustering.

3.1 The Cluster-Independent Weights

This clustering method minimizes the objective function Eq. (13) with the constraint of Eq. (2). Similar to DWFCM, the weights (w_k) are independent of a particular cluster, but contrary to DWFCM, whose weights were constant during the clustering, in this approach, the weights can change and be updated during the clustering process.

Using the Lagrange multiplier method, the problem is equivalent to minimizing the following equation satisfying the constraint Eq. (2):

$$L(U,\lambda)=\sum_{i=1}^{C}\sum_{t=1}^{N}w_k u_{ik}^m d_{ik}^2+\sum_{k=1}^{N}\lambda_k\left(1-\sum_{i=1}^{C}u_{ik}\right) \tag{17}$$

For the sake of simplicity in computations, we use an assumption that $\partial w_k / \partial v_k \approx 0$. Therefore by setting $\partial L / \partial u_{ik} = 0$, the following equation will be obtained.

$$\begin{aligned}\partial L/\partial u_{ik}=0 &\Rightarrow m(w_k)u_{ik}^{m-1}\left\|x_k-v_i\right\|^2-\lambda=0\\ &\Rightarrow u_{ik}=\left(\lambda\Big/ mw_k\left(\left\|x_k-v_i\right\|^2\right)\right)^{1/(m-1)}\end{aligned} \tag{18}$$

And replacing u_{ik}, found in Eq. (18), in Eq. (2), would lead to

$$\begin{aligned}&\sum_{j=1}^{C}\left(\lambda\Big/ mw_k\left(\left\|x_k-v_j\right\|^2\right)\right)^{1/(m-1)}=1\\ &\Rightarrow\ \left(\lambda/mw_k\right)^{\frac{1}{m-1}}=\left(1\Big/\sum_{j=1}^{C}\left(1\Big/\left\|x_k-v_j\right\|^2\right)^{1/(m-1)}\right)\end{aligned} \tag{19}$$

Combining Eq. (18) and Eq. (19), u_{ik} can be rewritten as

$$u_{ik}=\left(\sum_{j=1}^{C}\left(\left\|x_k-v_i\right\|^2\Big/\left\|x_k-v_j\right\|^2\right)^{1/(m-1)}\right)^{-1} \tag{20}$$

Also, by letting $\partial L/\partial v_i = 0$, updating of the equation for centroids can be carried out as

$$\begin{aligned}\partial L/\partial v_i=0 &\Rightarrow -2\sum_{k=1}^{N}w_k u_{ik}^m\left(x_k-v_i\right)=0\\ &\Rightarrow v_i=\sum_{k=1}^{N}w_k u_{ik}^m x_k\Big/\sum_{k=1}^{N}w_k u_{ik}^m\end{aligned} \tag{21}$$

Calculation of weights were carried out in our experiments using density weights presented in Eq. (14) and credibility weights presented in Eq. (10). Using weights presented in Eq. (14), the proposed method is the same as DWFCM.

3.2 The Cluster-Dependent Weights

This type of clustering minimizes the objective function

$$J=\sum_{i=1}^{C}\sum_{t=1}^{N}w_{ik}u_{ik}^m d_{ik}^2 \tag{22}$$

with the constraint of Eq. (2). Contrary to the weights w_k in Section 3.1, the weights w_{ik} in Eq. (22) depend on a particular cluster. These weights can change and be updated during the clustering process.

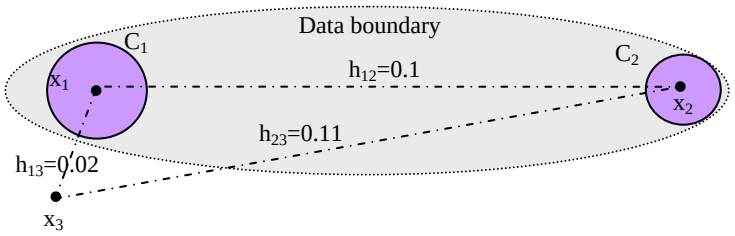


Fig. 2 A simple dataset with 2 clusters where the ratio of inter-cluster deviation to the whole data deviation is small. X_1: A clean data sample from C_1. X_2: A clean data sample from C_2. X_3: An outlier.

Using the Lagrange multiplier method, the problem is equivalent to minimizing the following equation with constraints

$$L(U,\lambda)=\sum_{i=1}^{C}\sum_{t=1}^{N}w_{ik}u_{ik}^{m}d_{ik}^{2}+\sum_{k=1}^{N}\lambda_k(1-\sum_{i=1}^{C}u_{ik}) \quad (23)$$

For simplicity in computations and related equations, once again, we use the assumption that $\partial w_{ik}/\partial v_k \approx 0$. Therefore, setting $\partial L/\partial u_{ik}=0$, we'll obtain

$$\begin{aligned}\partial L/\partial u_{ik}=0 &\Rightarrow m(w_{ik})u_{ik}^{m-1}\|x_k-v_i\|^2-\lambda=0\\ &\Rightarrow u_{ik}=\left(\lambda/mw_{ik}(\|x_k-v_i\|^2)\right)^{1/(m-1)}\end{aligned} \quad (24)$$

Replacing u_{ik} in Eq. (2) with that in Eq. (24), we get:

$$\begin{aligned}&\sum_{j=1}^{C}\left(\lambda/mw_{jk}(\|x_k-v_j\|^2)\right)^{1/(m-1)}=1\\ &\Rightarrow\ (\lambda/m)^{1/(m-1)}=\left(1/\sum_{j=1}^{C}\left(1/w_{jk}\|x_k-v_j\|^2\right)^{1/(m-1)}\right)\end{aligned} \quad (25)$$

Further replacing Eq. (25) in Eq. (24), u_{ik} can be rewritten as follows:

$$u_{ik}=\left(\sum_{j=1}^{C}\left(\frac{w_{ik}\|x_k-v_i\|^2}{w_{jk}\|x_k-v_j\|^2}\right)^{1/(m-1)}\right)^{-1} \quad (26)$$

Setting $\partial L/\partial v_i=0$, the updating equation for the centroids will be:

$$\begin{aligned}\partial L/\partial v_i=0 &\Rightarrow -2\sum_{k=1}^{N}w_{ik}u_{ik}^{m}(x_k-v_i)=0\\ &\Rightarrow v_i=\sum_{k=1}^{N}w_{ik}u_{ik}^{m}x_k\Big/\sum_{k=1}^{N}w_{ik}u_{ik}^{m}\end{aligned} \quad (27)$$

For computing weights, we use the typicality given by Eq. (8). However, as mentioned before, due to its computational complexity, which is $O(CN^2)$, we propose a simplified type of typicality weights, computed as follows:

$$w_{ik}=(1/d_{ik})^{2/(\eta-1)} \quad (28)$$

where $\eta>1$ is a parameter depending on the variation of outliers.

The order of computational complexity of this kind of weights is $O(CN)$ and seems to be acceptable for large datasets such as speech signals or images.

3.3 Robust Bilateral Weighted Fuzzy C-Means

This clustering method minimizes the following objective function

$$J=\sum_{k=1}^{N}\sum_{i=1}^{C}(r_k w_{ik})u_{ik}^{m}\|x_k-v_i\|^2 \quad (29)$$

with the constraint of Eq. (2). The weights r_k are independent of a particular cluster but the weights w_{ik} depend on the i'th cluster. Both kind of weights can change and be updated during the clustering process. Using the Lagrange multiplier method, the problem is equivalent to minimizing the following equation with constraints

$$\begin{aligned}&L(U,\lambda)=\\ &\sum_{k=1}^{N}\sum_{i=1}^{C}r_k w_{ik}u_{ik}^{m}\|x_k-v_i\|^2-\sum_{k=1}^{N}\lambda\left(\sum_{i=1}^{C}u_{ik}-1\right)\end{aligned} \quad (30)$$

For the sake of simplicity in computations and related equations, once again we consider an assumption that $\partial r_k/\partial v_k\approx 0$, $\partial w_{ik}/\partial v_k\approx 0$ and $\partial(r_k w_{ik})/\partial v_k\approx 0$. Therefore, setting $\partial L/\partial u_{ik}=0$ would lead to

$$\begin{aligned}\partial L/\partial u_{ik}=0 &\Rightarrow m(r_k w_{ik})u_{ik}^{m-1}\|x_k-v_i\|^2-\lambda=0\\ &\Rightarrow u_{ik}=\left(\lambda/\left(m(r_k w_{ik})(\|x_k-v_i\|^2)\right)\right)^{1/(m-1)}\end{aligned} \quad (31)$$

Replacing u_{ik} in Eq. (2) by Eq. (31), one would get

$$\begin{aligned}&\sum_{j=1}^{C}\left(\lambda/\left(mr_k w_{ik}(\|x_k-v_j\|^2)\right)\right)^{1/(m-1)}=1\ \Rightarrow\\ &(\lambda/m)^{1/(m-1)}=\left(1/\sum_{j=1}^{c}\left(\frac{1}{r_k w_{ik}\|x_k-v_j\|^2}\right)^{1/(m-1)}\right)\end{aligned} \quad (32)$$

Therefore, Eq. (31) could be rewritten as

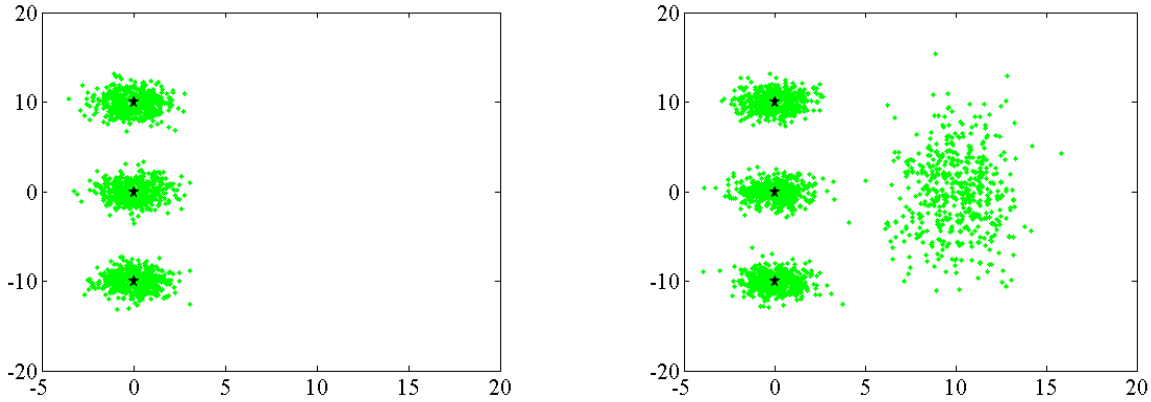
$$u_{ik}=\left(\sum_{j=1}^{C}\left(\frac{\left(w_{ik}\|x_k-v_i\|^2\right)}{\left(w_{jk}\|x_k-v_j\|^2\right)}\right)^{1/(m-1)}\right)^{-1} \quad (33)$$

Furthermore, by setting $\partial L/\partial v_i=0$, the updating equation for centroids would be:

$$\begin{aligned}\partial L/\partial v_i=0 &\Rightarrow -2\sum_{k=1}^{N}(r_k w_{ik})u_{ik}^{m}(x_k-v_i)=0\\ &\Rightarrow v_i=\left(\sum_{k=1}^{N}(r_k w_{ik})u_{ik}^{m}x_k\right)\Big/\left(\sum_{k=1}^{N}(r_k w_{ik})u_{ik}^{m}\right)\end{aligned} \quad (34)$$

4 Robust Weighted Fuzzy C-Means Clustering (RWFCM)

In order to be able to compare our proposed method with other clustering methods, two artificial datasets X1500, and X1000, and five real datasets, viz., Iris, Cancer, Wine, Glass and an speech corpus named TFARSDAT (Persian) were used in this research.



(a) (b)

Fig. 3 The X1500 datasets: (a) without any noise samples, (b) with 400 Gaussian noise samples.

Table 2 X1000 clean dataset and its noisy versions with uniform noise distribution.

Dataset	Kind of uniform outliers	No. of outliers with uniform distribution
1	-	0
2	concentrated	50
3	concentrated	100
4	Dispersed	50
5	Dispersed	100

X1500: This artificial dataset contained three Gaussian clusters. Each cluster had 500 data samples. The central vectors of the clusters were $[0,-10]^T$, $[0,0]^T$ and $[0,10]^T$ and all three clusters had an identity covariance matrix. In order to generate noisy versions of X1500, different noise samples, considered as outliers, were added to this dataset. The noisy versions had 50, 100, 200, 300, 400 and 500 noise samples. Noise samples (outliers) had a Gaussian distribution with a mean vector $[10,0]^T$ and a diagonal covariance matrix with diagonal values of $[2,21]^T$. Fig. 3 shows the clean X1500 dataset and its noisy version with 400 outliers.

X1000: The second artificial dataset was named X1000 and contained two clusters with Gaussian distributions. One of these clusters contained 600 and the other one 400 data samples. The central vectors of these clusters were $[0,6]^T$ and $[0,6]^T$. All clusters had an identity covariance matrix. In order to make a noisy version of X1000, two kinds of noises were added to this dataset. Noise samples had uniform (as background noise) and Gaussian distributions. Table 2 shows the different versions of X1000 datasets. According to this table, outliers with uniform distribution may be concentrated or dispersed and their number may be 50 or 100. Gaussian noise outliers (0, 300 and 600) were also added to the 5 datasets mentioned in Table 2. Two Gaussian noise distributions with central vectors of $[-1, 10]^T$ and $[1, -10]^T$ and diagonal covariance matrices of $2 \times I$ were used. Fig. 4(a) shows the clean version of X1000 and Fig. 4(b) shows its noisy version with 100 concentrated uniform and 300 Gaussian noises. Fig. 4(c) depicts the noisy version of X1000 with 100 dispersed uniform and 300 Gaussian noises.

Iris: The iris dataset is one of the most popular datasets to examine the performance of novel methods in pattern recognition and machine learning [21]. Iris represents different categories of Iris plants having four feature values. The four feature values represent the sepal length, sepal width, petal length and the petal width in centimeters. It has three classes Setosa, Versicolor and Virginica, with 50 samples per class. It is known that two classes Versicolor and Virginica have some amount of overlap while the class Setosa is linearly separable from the other two [22].

Cancer: This breast cancer database was obtained from the University of Wisconsin Hospitals, Madison from Dr. William and H. Wolberg. It consists of 699 samples of which 458 are benign and 241 are malignant, all in a 9-dimensional real space. These 9 features are: Clump Thickness, Size Uniformity, shape Uniformity, Marginal Adhesion, cell size, Bare Nuclei, Bland Chromatin, Normal Nucleoli and Mitoses [23].

Wine: The wine data includes three classes, 13 features and 178 samples of which 59 are first class, 71 are second class and 48 are third class [22].

Glass: The Glass dataset has two main clusters, 9 features and 214 samples. Its 9 features are refractive index, Sodium, Magnesium, Aluminum, Silicon, Potassium, Calcium, Barium and Iron [22].

TFARSDAT: This is a speech corpus that is collected from 64 male and female adult speakers from 10 different Iranian accents uttering some Persian

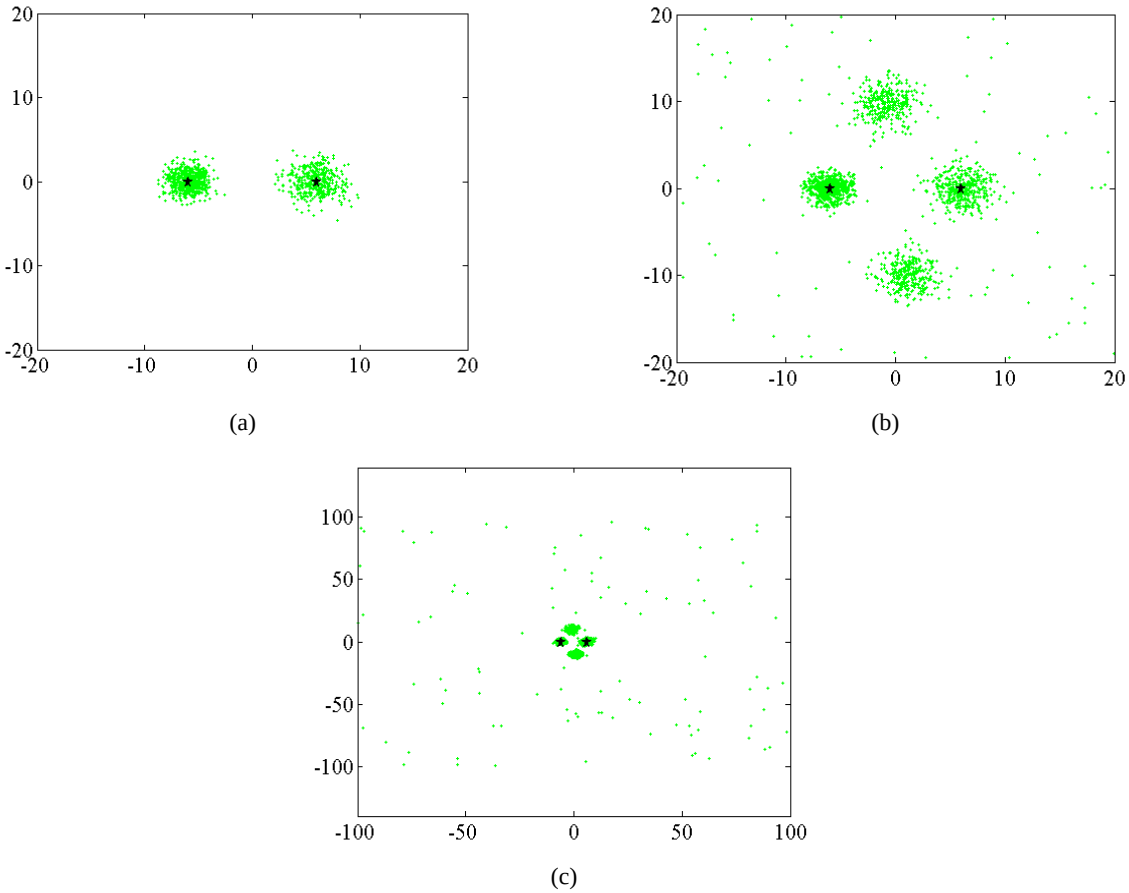


Fig. 4 The X1000 datasets: (a) without any noise samples, (b) with 100 concentrated uniform noise and 300 Gaussian noise samples, (c) with 100 dispersed uniform and 300 Gaussian noise samples.

cardinal numbers, days of week, names of months of the year, Persian alphabet letters, 50 frequent Persian words and 6 Persian sentences [24]. The data was collected in normal office conditions with SNRs of 25 dB.

5 Experimental Results

In this section, initially, we present experimental results comparing the performance of different clustering weight computation methods. Two artificial datasets X1000, X1500 and their noisy versions were used for this purpose. In the second part, conventional and proposed robust clustering methods are compared. To compare the performance of different robust clustering methods two artificial and four real datasets (Iris, Cancer, Glass and Wine) were used. For the evaluation of clustering methods on artificial datasets, cluster centers obtained using clustering methods and the real cluster centers were compared and their Mean Square Error (MSE) was used as the comparison criterion. Also, for the evaluation of clustering methods on real datasets, different clustering methods were run on the mentioned real datasets and the number of misclassified samples and the average cluster purities were used for evaluation [25]. Cluster Purity is a unary classification criterion. Here, purity denotes the fraction of the cluster taken up by its predominant class label. The purity of a cluster c can be defined as follows:

$$\rho(c) = \sum_{p=1}^{P} \left(n_{cp} / n_c \right)^2 \tag{35}$$

where $c \in \{1,\ldots,C\}$ is a cluster, C is the total number of clusters, $p \in \{1,\ldots,P\}$ is a class, P is the number of classes, n_{cp} is the number of samples in cluster c belonging to class p and n_c is the number of samples in cluster c. Then average purity can be computed as follows:

$$\bar{p} = \frac{1}{N} \sum_{c=1}^{C} n_{cp} . \rho_c \tag{36}$$

Therefore, bad clustering has an average purity value close to 0 and perfect clustering has a purity of 1.

At the third experimental part, conventional and proposed robust clustering methods are used in a real Gaussian mixture model (GMM)-based speaker identification application. Since the performance of Gaussian mixture models are very sensitive to the initial mixture mean vectors, using a C-Means motivated clustering method for computing the initial mean

vectors is highly recommended. Therefore, we use different mentioned clustering methods for computing the initial Gaussian mixtures mean vectors. At the end of this section, in the fourth experimental part, the conventional and proposed clustering methods are compared regarding their computation time and convergence rate.

5.1 Comparison of Clustering Weight Computation Methods

Three main categories of clustering weight computation methods including cluster-dependent weights, cluster-independent weights and bilateral weights are compared using BWFCM clustering method. Cluster-dependent weights are credibility weights Eq. (10) and density weights Eq. (14). Cluster-independent weights are typicality weights Eq. (8) and simplified typicality weights Eq. (28). Bilateral weights are joint credibility and simplified typicality weights, and joint density and simplified typicality weights. Both of the artificial datasets, X1000 and X1500 were used in this comparison. The results are depicted in Tables 3 and 4.

The experimental results show that after clustering, the MSE degrades when the number of outliers increases. Also it can be seen that the cluster-dependent

Table 3 Comparison of different kinds of clustering weight computation methods on X1500 dataset with different number of outliers, in terms of MSE.

		Unilateral weights				**Bilateral weights**	
		Independent of a particular cluster		Dependent on a particular cluster		Credibility and Simplified typicality weights	Density and Simplified typicality weights
		Credibility weights	Density weights	Typicality weights	Simplified typicality weights		
Number of uniform outliers	0	0.06	0.07	0.07	0.10	0.07	0.08
	50	0.14	0.09	0.15	0.12	0.12	0.11
	100	0.27	0.13	0.26	0.20	0.12	0.11
	200	0.44	0.22	0.38	0.23	0.13	0.11
	300	0.72	0.40	0.59	0.35	0.14	0.12
	400	0.83	0.56	0.66	0.32	0.19	0.15
	500	1.25	0.91	0.97	0.59	0.33	0.28

Table 4 Comparison of different kinds of clustering weight computation on X1000 dataset and its uniform noisy versions with different number of Gaussian outliers in terms of MSE.

Dataset version	**No. of Gaussian Outliers**	**Unilateral weights**				**Bilateral weights**	
		Independent of a particular cluster		Dependent on a particular cluster		Joint Credibility and Simplified typicality weights	Joint Density and Simplified typicality weights
		Credibility weights	Density weights (DWFCM)	Typicality weights	Simplified typicality weights		
1	0	0.07	0.10	0.09	0.06	0.083	0.10
	300	0.51	0.81	0.51	0.55	0.12	0.14
	600	1.67	3.62	1.09	1.07	0.53	0.55
2	0	0.24	0.20	0.25	0.19	0.08	0.081
	300	1.01	0.85	0.51	0.64	0.25	0.26
	600	2.49	3.37	0.90	1.16	0.40	0.31
3	0	0.14	0.15	0.09	0.14	0.12	0.09
	300	0.89	0.76	0.44	0.59	0.27	0.12
	600	3.10	3.46	0.98	1.24	0.51	0.38
4	0	0.21	0.09	0.16	0.11	0.12	0.09
	300	1.49	1.02	0.50	0.64	0.40	0.24
	600	6.96	3.45	1.09	1.34	0.91	0.64
5	0	0.37	0.12	0.16	0.11	0.12	0.09
	300	1.47	1.31	0.61	0.78	0.66	0.59
	600	7.01	3.46	1.64	1.28	0.73	0.65

weights are more robust than the cluster-independent weights and the bilateral weights are the most robust clustering weights. It also appears that the joint density and simplified typicality weights are more robust than the other kinds of weights.

5.2 Comparison of Clustering Methods

In this section the proposed Bilateral Weighted Fuzzy C-Means clustering method (BWFCM) is compared to the conventional methods including classical FCM, MFPCM, CFCM and DWFCM. This is carried out by running all of these clustering methods on both artificial and real data sets. In BWFCM both joint Credibility and simplified typicality weights named BWFCM1, and joint Density and simplified typicality weights named BWFCM2 are used.

5.2.1 Artificial Datasets Results

Results of this comparison using X1500, X1000 and their noisy versions are presented respectively in Tables 5 and 6, and also in Fig. 5. It is evident from these results that the proposed bilateral weighted clustering methods, in general, provide better values of the validity indices. Also note that except for the first uniform noisy version of X1000, in most of the other cases, the MSE values obtained using joint credibility and simplified typicality weights are greater than joint Density and simplified typicality weights. From Fig. 5, it can be seen that among the conventional clustering methods, CFCM outperforms FCM, MFPCM and DWFCM except for the two cases of uniform noisy versions of the X1000 where far outliers exist (see results for versions 4 and 5 of X1000 dataset in Fig. 5).

5.2.2 Real Dataset Results

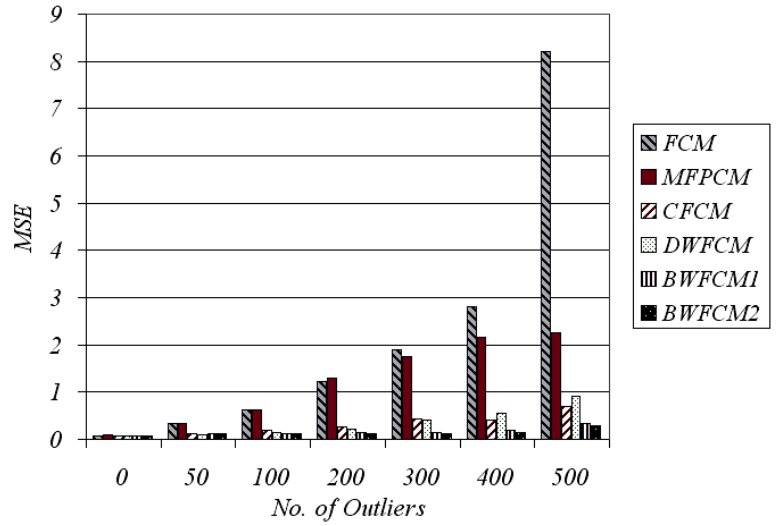
Similar to the experiment in the previous section, we compared FCM, MFPCM, CFCM, DWFCM and both bilateral-weighted FCM methods (BWFCM1 and BWFCM2) using four real datasets, namely Iris, Cancer, Wine and Glass. Results are presented in Table 7 in terms of purity index and in Fig. 6 in terms of misclassified data. The number of misclassifications is

Table 5 Comparison of conventional and proposed clustering methods on X1500 dataset with different number of outliers in terms of MSE.

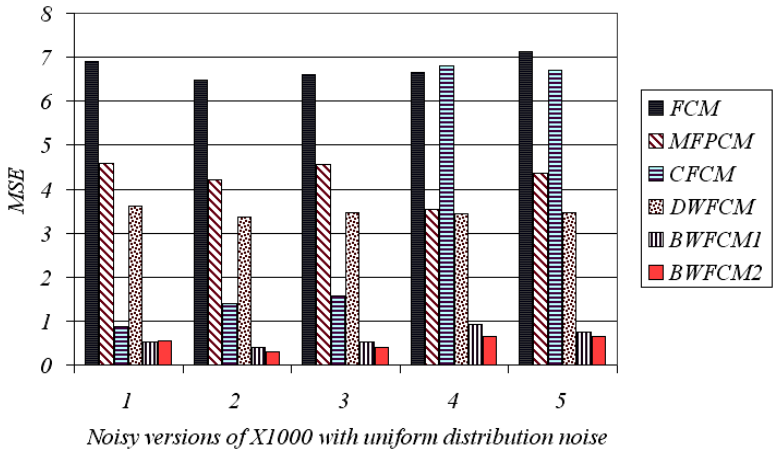
		Conventional robust clustering methods				**Bilateral Weighted FCM**	
		FCM	MFPCM	CFCM	DWFCM	Joint Credibility and Simplified typicality weights	Joint Density and Simplified typicality weights
Number of Outliers	0	0.07	0.08	0.07	0.07	0.07	0.08
	50	0.32	0.34	0.11	0.09	0.12	0.11
	100	0.63	0.63	0.19	0.13	0.14	0.11
	200	1.21	1.30	0.25	0.22	0.13	0.11
	300	1.89	1.76	0.42	0.40	0.12	0.12
	400	2.81	2.16	0.40	0.56	0.19	0.15
	500	8.19	2.25	0.69	0.91	0.33	0.28

Table 6 Comparison of conventional and proposed clustering methods on X1000 dataset and its noisy versions (uniform noise) with different number of Gaussian outliers in terms of MSE.

Dataset version	**No. of Gaussian Outliers**	**Conventional robust clustering methods**				**Bilateral Weighted FCM**	
		FCM	MFPCM	CFCM	DWFCM	Joint Credibility and Simplified typicality weights	Joint Density and Simplified typicality weights
1	0	0.05	0.10	0.08	0.10	0.08	0.10
	300	1.70	1.70	0.25	0.81	0.12	0.14
	600	6.89	4.59	0.86	3.62	0.53	0.55
2	0	0.28	0.26	0.22	0.20	0.08	0.08
	300	1.58	1.76	0.71	0.85	0.25	0.26
	600	6.48	4.20	1.40	3.37	0.40	0.31
3	0	0.18	0.34	0.12	0.15	0.12	0.12
	300	1.40	1.89	0.61	0.76	0.27	0.12
	600	6.59	4.55	1.57	3.46	0.51	0.38
4	0	0.38	0.62	0.17	0.09	0.12	0.09
	300	1.54	1.74	1.45	1.02	0.40	0.24
	600	6.65	3.53	6.79	3.45	0.91	0.64
5	0	0.94	1.03	0.22	0.12	0.12	0.09
	300	1.25	1.18	1.45	1.31	0.66	0.59
	600	7.13	4.35	6.70	3.46	0.73	0.65



(a)



(b)

Fig. 5 Comparison of different kinds of computing clustering weights on (a) X1500 datasets with different number of outliers and (b) X1000 with different uniform noisy versions and 600 Gaussian outliers.

Table 7 Comparison of conventional and proposed Bilateral weighted clustering methods on real datasets in terms of average cluster purity index.

dataset	FCM	MFPCM	CFCM	DWFCM	*STWFCM	†BWFCM1	‡BWFCM2
Iris	0.8272	0.8272	0.8592	0.8695	0.8510	0.8592	0.8715
Cancer	0.9161	0.9377	0.9268	0.9322	0.9464	0.9464	0.9552
Wine	0.6022	0.5808	0.6110	0.6226	0.6226	0.6114	0.6225
Glass	0.8391	0.8159	0.8159	0.8391	0.8575	0.8573	0.8684

*STWFCM denotes simplified typicality weighted FCM

†BWFCM1 denotes joint credibility and simplified typicality weighted FCM.

called the re-substitution error rate [26]. It is evident from Table 7 that the proposed BWFCM approach with joint Density and simplified typicality weights (BWFCM2), in general, presents better values for the purity index.

5.2.3 Real Application

According to the theoretical considerations above, we also present, in this experimental part, the results of a GMM-based speaker identification experiment [27]. The available TFARSDAT speech data corpus is used to compare these algorithms. We use 35 speakers, including 12 female and 23 male speakers. The data

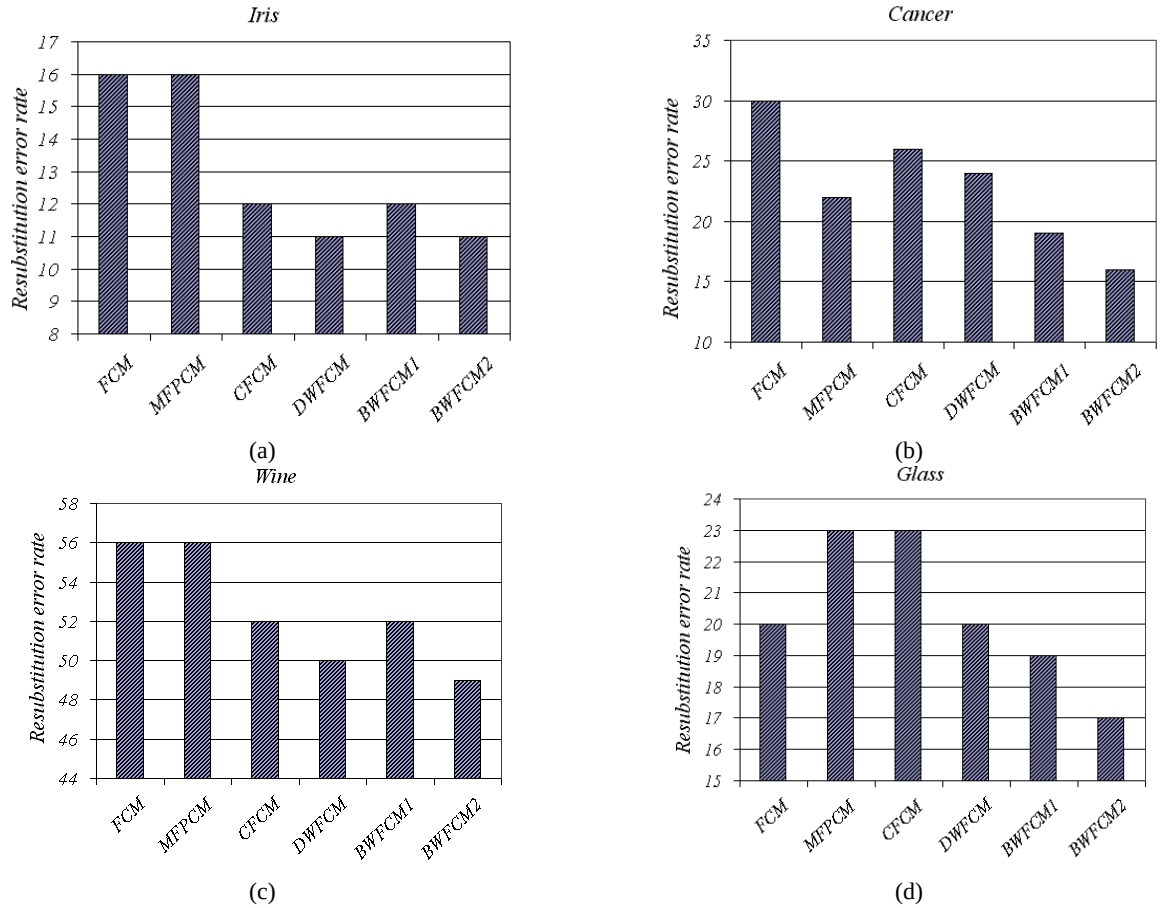


Fig. 6 Comparison of conventional and proposed Bilateral weighted clustering methods on real datasets in terms of number of misclassified samples.

were processed in 25 ms frames at a frame rate of 200 frames per second. Frames were Hamming windowed and pre-emphasized with $\mu = 0.975$. For each frame, 24 Mel-spectral bands were used and 13 Mel-frequency cepstral coefficients (MFCC) were extracted. Since the performance of Gaussian mixture models are very sensitive to the selection of initial mixture mean vectors, using a C-Means motivated clustering method for computing the initial mean vectors is highly recommended. Therefore, we use different mentioned clustering methods for computing the initial mean vectors of the Gaussian mixtures.

In the training phase, 40 seconds utterances from each speaker were used to train GMMs with 4, 8, 16 and 32 mixture components.

Speaker identification was carried out by testing 980 test tokens (35 speakers with 28 utterances each) against the GMMs of all 35 speakers in the database. The experimental results are shown in Table 8. These results also show the superiority of BWFCM against different clustering methods for initialization of centroids of Gaussian mixtures in tasks such as speaker identification.

Table 8 GMM-based speaker identification error rates, while centroids of Gaussian mixture are initialized using different clustering methods.

Number of Gaussian mixtures	C-Means (HCM)	FCM	CFCM	DWFCM	*BWFCM
4	5.61	5.41	5.41	5.20	5.00
8	4.79	4.69	4.69	4.29	3.67
16	3.67	3.37	3.37	3.269	3.16
32	2.75	2.75	2.75	2.759	2.35
Average	4.21	4.06	4.06	3.88	3.55

*BWFCM denotes joint density and proposed weights, Weighted FCM.

5.3 Computational Time and Convergence Rate

In this section BWFCM and the conventional clustering methods are compared regarding their computational times and convergence rates. The necessary time of one Iteration step of clustering and the number of iterations for convergence of the algorithms are computed and compared. Our experimental platform was a Microsoft Windows-based personal computer with a 2 GHz AMD processor and 1 GB of RAM memory. Experiments were performed on X1500 dataset with different number of additional Gaussian outliers. The computational complexity of each iteration step in FCM, CFCM, DWFCM and both types of BWFCM is $O(C^2N^2)$, and in MFPCM is $O(C^2N^3)$.

The computational time of one Iteration step for FCM, MFPCM, CFCM, DWFCM, BWFCM1, BWFCM2 are presented in Table 9. The numbers of convergence iteration steps for these clustering methods are also depicted in Fig.7. According to Table 9, the necessary time for one iteration step in FCM, DWFCM and BWFCM2 is almost the same. This time is nearly twice for CFCM and BWFCM1. But this time is considerably higher for MFPCM, since this method has a computational complexity of $O(C^2N^3)$.

Based on Fig. 7, the convergence rates of FCMT, DWFCM, BWFM1 and BWFCM2 for different amounts of added outlier samples are nearly equal. More number of iterations is needed for convergence of CFCM compared to FCMT, DWFCM, BWFCM1 and BWFCM2. Convergence rate of FCM degrades when the number of outliers increases. MFPCM has the worth convergence rate among all clustering methods.

6 Conclusion

In this paper, a new algorithm for robust fuzzy clustering named Bilateral Weighted Fuzzy C-Means (BWFCM) was proposed. Our main concern in presenting this algorithm is to reduce the influence of outliers on clustering. In order to achieve this target, BWFCM attempts to decrease the noise sensitivity in

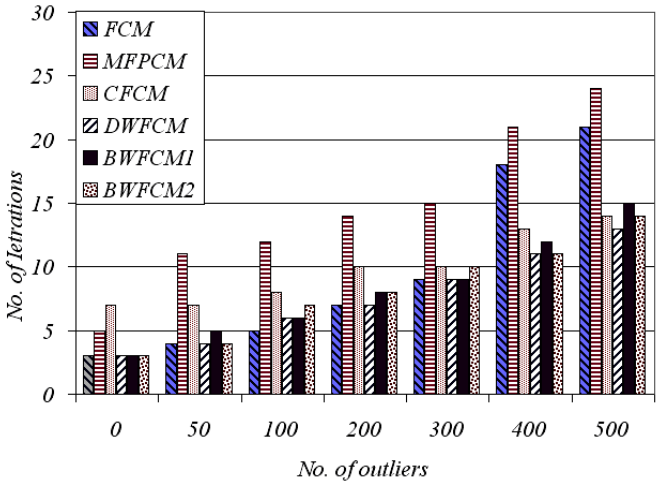


Fig. 7 Comparison of convergence rates for FCM, MFPCM, CFCM, DWFCM, BWFCM1 and BWFCM2 on X1500 dataset.

Table 9 Computational time of one iteration step in FCM, MFPCM, CFCM, DWFCM, BWFCM1 and BWFCM2 using X1500 (in milliseconds).

dataset	FCM	MFPCM	CFCM	DWFCM	*BWFCM1	†BWFCM2
0	5.00	1327.00	7.80	5.16	13.08	5.33
50	6.20	1368.80	7.80	5.33	13.35	7.00
100	6.94	2046.90	8.38	7.00	13.42	7.50
200	7.00	2086.00	9.87	7.09	13.88	7.66
300	7.42	2116.66	10.07	7.50	14.10	7.83
400	8.00	2324.25	11.14	7.83	14.33	8.54
500	8.85	2401.76	17.85	8.45	14.50	8.66
Average	7.05	1953.10	10.41	6.90	13.80	7.50

*BWFCM1 denotes joint credibility and simplified typicality weighted FCM.

†BWFCM2 denotes joint density and simplified typicality weighted FCM.

fuzzy clustering by using different kinds of weights in its objective function, so that the noisy samples and outliers have less effect on centroids. Three main categories of weights including cluster-dependent, cluster-independent and bilateral weights are also investigated theoretically and empirically in this research. The proposed method is compared to other well-known robust clustering methods such as Possibilistic Fuzzy C-Means, Credibilitistic Fuzzy C-Means and Density Weighted Fuzzy C-Means. Experimental results on two artificial and five real datasets demonstrate the high performance of the proposed method, while its order of computational complexity is comparable to many conventional clustering methods. Subject of our future work is to use the proposed method in applications such as robust speaker clustering and robust image segmentation.

References

[1] Leski J., "Towards a robust fuzzy clustering", *Fuzzy Sets and Systems*, Vol. 137, pp. 215–233, July 2003.

[2] Duda R. O., Hart P. E. and Stork G. D., "*Pattern Classification*", chapter 10, Wiley-Interscience, 1997.

[3] Michie D., Spiegelhalter D. J. and Taylor C. C., *Machine Learning, Neural and Statistical Classification*, Ellis Horwood, 1994.

[4] Nilsson N. J., *Introduction to Machine Learning*, chapter 9, Department of Computer Science, Stanford University, 1996.

[5] Cai W., Chen S. and Zhang D., "Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation", *Pattern Recognition* Vol. 40, pp. 825–838, March 2007.

[6] Cinque L., Foresti G. and Lombardi L., "A clustering fuzzy approach for image segmentation", *Pattern Recognition*, Vol. 37, pp. 1797–1807, Sep. 2004.

[7] Liew A. W. C. and Yan H., "An Adaptive Spatial Fuzzy Clustering Algorithm for 3-D MR Image Segmentation", *IEEE Trans Med Imaging*, Vol. 22, pp. 1063-1075, Sep. 2003.

[8] Pham D. L. and Prince J. L., "Adaptive Fuzzy Segmentation of Magnetic Resonance Images", *IEEE Trans Med Imaging*, Vol. 18, pp. 737–752, Sep. 1999.

[9] Turcan A., Ocelikova E. and Madarasz L., "Fuzzy Clustering Applied on Mobile Agent Behaviour Selection", *3rd Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence*, Hungary any, Slovakia, pp. 21-22, 2005.

[10] Chen J. L. and Wang J. H., "A new robust clustering algorithm-density-weighted fuzzy c-means", *IEEE International Conference on Systems*, Man, and Cybernetics, Vol. 3, pp. 90–94, 1999.

[11] Wang S., Chung K. F. L., Zhaohong D. and Dewen H., "Robust fuzzy clustering neural network based on ε-insensitive loss function", *Applied Soft Computing*, Vol. 7, pp. 577-584, 2007.

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

[13] Pal N. R., Pal K. and Bezdek J. C., "A Mixed c-Means Clustering Model", *IEEE International Conference on Fuzzy Systems*, Vol. 1, pp. 11-21, 1997.

[14] Yang T. N., Wang S. D. and Yen S. J., "Fuzzy algorithms for robust clustering", *International Computer Symposium*, pp. 18-21, 2002.

[15] Chintalapudi K. K. and Kam M., "A noise-resistant fuzzy c-means algorithm for clustering", *IEEE Conference on Fuzzy Systems Proceedings*, Vol. 2, pp. 1458–1463, 1998.

[16] Yang M. Sh. and Wu K. L., "Unsupervised possibilistic clustering", *Pattern Recognition*, Vol. 39, pp. 5–21, 2006.

[17] Boudouda H., Seridi H. and Akdag H., "The Fuzzy Possibilistic C-Means Classifier", *Asian Journal of Information Technology*, Vol. 11, pp. 981-985, 2005.

[18] Hadjahmadi, A. H., Homayounpour M. M. and Ahadi S. M., "A modified fuzzy possibilistic c-means clustering for large and noisy datasets (MFPCM)", *First Joint Congress on Fuzzy and Intelligent Systems*, 2007.

[19] Wang X. Y. and Garibaldi J. M., "Simulated Annealing Fuzzy Clustering in Cancer Diagnosis", *Informatica*, Vol. 29, pp. 61-70, 2005.

[20] Hadjahmadi, A. H., Homayounpour M. M. and Ahadi S. M., "Robust weighted fuzzy c-means clustering", *IEEE International Conference on Fuzzy Systems* , 2008.

[21] Xu R., and Wunsch D., "Survey of Clustering Algorithms", *IEEE Trans. on Neural Networks*, Vol. 16, pp. 645-678, 2005.

[22] Bezdek J. C., Keller J. M., Krishnapuram R., Kuncheva L. I. and Pal N. R., "Will the real Iris data please stand up?", *IEEE Trans. on Fuzzy Systems*, Vol. 7, pp. 368–369, 1999.

[23] Asuncion A. and Newman D. J., UCI Machine Learning Repository [http://www.ics.uci.edu/ ~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, 2007.

[24] Bijankhan M., Sheikhzadegan J., Rohani M. R., Zarintareh R., Ghasemi S. Z., and Ghasedi M. A., "Persian Monologue Telephone Speech Database: TFARSDAT", *The First Workshop on Persian Language and Computer*, 2004.

[25] Hand J. K., "Multi objective approaches to the data-driven, analysis of biological systems", Thesis, University of Manchester for the degree of Doctor of Philosophy, 2006.

[26] Zhang J. Sh. and Leung Y. W., "Robust Clustering by Pruning Outliers", *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 33, pp. 983–998, Dec. 2003.

[27] Kinnunen T., Sidoroff I., Tuononen M. and Fränti P., "Comparison of clustering methods: A case study of text-independent speaker modeling", *Pattern Recognition Letters*, Vol. 32, pp. 1604-1617, 2011.

Amirhossein Hadjahmadi was born in Rafsanjan, Iran, on March 27, 1983. He received the B.Sc degree in Software Engineering from Yazd University, Yazd, Iran in 2005 and the M.Sc degree in Artificial Intelligence from Amirkabir University of Technology, Tehran, Iran in 2008. He is currently a lecturer with the department of Computer Engineering, Vali-e-Asr University of Rafsanjan, Rafsanjan, Iran. His research interests are in the fields of data mining, machine learning, speech processing, image processing and artificial intelligence.

Mohammad Mehdi Homayounpour received his B.Sc. and M.Sc. in electrical engineering from Amirkabir University of Technology and Khajeh Nasir Toosi University of Technology, Tehran, Iran, respectively, and his PhD degree in electrical engineering from Université de Paris Sud, Paris, France. He is currently an associate professor of the Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran. His research interests include digital signal processing, speech processing, speech recognition, text-to-speech, and speaker recognition.

Seyed Mohammad Ahadi received the B.Sc. and M.Sc. degrees in Electronics from the Electrical Engineering Department, Amirkabir University of Technology, Tehran, Iran, in 1984 and 1987 respectively, and his Ph.D. in Engineering from the University of Cambridge, Cambridge, England in 1996. He was involved in several electronic projects in private sector as well as part-time teaching at the Electrical Engineering Department, Amirkabir University of Technology during the period 1985-1988. He was appointed as a faculty member at the Electrical Engineering Department of Amirkabir University of Technology in 1988, where he started his teaching profession as well as involvement in projects. During the period 1992-1996, he pursued his studies toward the Ph.D. degree, working in the field of speech recognition. Since 1996 he has been with the Electrical Engineering Department of Amirkabir University of Technology, teaching several courses and doing research in the fields of Electronics and Communications. He is now the head of Electronics group at that department.