Applied Soft Computing 13 (2013) 1592–1607

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

A multivariate fuzzy c-means method

Bruno A. Pimentel, Renata M.C.R. de Souza*

Centro de Informática, Av. Jornalista Anibal Fernandes, s/n – Cidade Universitária 50.740-560, Recife (PE), Brazil

ARTICLE INFO

Article history:
Received 10 July 2012
Received in revised form
27 November 2012
Accepted 30 December 2012
Available online 11 January 2013

Keywords:
Fuzzy c-means method
Unsupervised pattern recognition
Clustering
Membership degree

ABSTRACT

Fuzzy c-means (FCMs) is an important and popular unsupervised partitioning algorithm used in several application domains such as pattern recognition, machine learning and data mining. Although the FCM has shown good performance in detecting clusters, the membership values for each individual computed to each of the clusters cannot indicate how well the individuals are classified. In this paper, a new approach to handle the memberships based on the inherent information in each feature is presented. The algorithm produces a membership matrix for each individual, the membership values are between zero and one and measure the similarity of this individual to the center of each cluster according to each feature. These values can change at each iteration of the algorithm and they are different from one feature to another and from one cluster to another in order to increase the performance of the fuzzy c-means clustering algorithm. To obtain a fuzzy partition by class of the input data set, a way to compute the class membership values is also proposed in this work. Experiments with synthetic and real data sets show that the proposed approach produces good quality of clustering.

1. Introduction

A growing number of application domains such as pattern recognition, machine learning, data mining, computer vision and computational biology have used clustering algorithms [1–3]. Clustering is a method of unsupervised learning whose objective is to group a set of elements into clusters such that elements within a cluster have a high degree of similarity, while elements belonging to different clusters have a high degree of dissimilarity. Mathematically, the degree of dissimilarity can be measured using, for instance, distance, angle, curvature, symmetry, connectivity or intensity with information from the data set [4]. Hierarchical and partitioning methods are the most popular clustering techniques. Hierarchical clustering find a sequence of partitions where the algorithm starts from one group with all objects and is executed until find singletons groups, or vice versa, whereas partitioning clustering directly divides data objects into some fixed number of clusters [5] using a suitable objective function. An advantage of the partitional method is its ability to manipulate large data sets, since the construction of dendrogram by the hierarchical method may be computationally impractical in some applications.

Partitioning clustering can be divided into hard and fuzzy methods. The concept of fuzzy set was initially explored by Zadeh [6] and applied to clustering by Ruspini [7]. Works with fuzzy set applications in cluster analysis were proposed and applied in several areas [8]. The concept of fuzzy set allowed works on industrial and academic fields [9]. Termini [10] used the definition of fuzzy sets to create an interaction with human sciences. Wong and Lai [11] described the applications of the fuzzy set theory in production and operations management, for example, planning, quality control and artificial intelligence (AI) techniques. Moreover, the work of Wong and Lai is based on the information of 402 articles published on journals between 1998 and 2009 that deal with application of fuzzy set theory techniques. In the medical field, Kuo et al. [12] used fuzzy set theory and, health care failure mode and effect analysis to study patient according the decision-making factors: severity, incidence, and detection.

In the hard approach each element of the data set can be associated to only one cluster, while in the fuzzy approach each element of the data set has a possibility of belonging to all cluster but with different membership degrees. Therefore, the calculation of membership functions is an important problem in fuzzy clustering. When each pattern is associated to the cluster with the largest measure of membership, the fuzzy clustering is equivalent to hard clustering. Three examples of categories of fuzzy used in the cluster analysis are: fuzzy clustering based on fuzzy relation, fuzzy clustering based on objective functions and the fuzzy generalized k-nearest neighbor rule [8]. The most popular fuzzy clustering method based on objective functions is the Fuzzy c-means (FCMs) [1,13]. An advantage of the FCM is that it may be used in applications where the clusters are overlapping [8].

There are several papers with related works to theory and applications of the FCM algorithm such as stochastic and numerical theorems, image processing, parameter estimation and many

* Corresponding author.
E-mail address: rmcrs@cin.ufpe.br (R.M.C.R. de Souza).

1568-4946/$ – see front matter
http://dx.doi.org/10.1016/j.asoc.2012.12.024

others [8]. Hathaway et al. [14] presented fuzzy c-means methods using general L_p norm distances whose main objective is to increase robustness to outliers. Jajuga [15] presented the fuzzy clustering algorithm based on the L_1 norm. Groenen and Jajuga [16] presented a new fuzzy clustering based on squared Minkowski distance. Oh et al. [17] proposed a new fuzzy clustering algorithm for categorical multivariate data. Xu and Wunsch [18] presented a replacement of the Euclidean norm by a new robust metric in c-means clustering diminishing the weaknesses of the classical FCM. Zhang and Chen [19] proposed the substitution of a kernel-induced distance metric for the original Euclidean distance in the FCM which allows implicitly to perform a nonlinear mapping to a high dimensional feature space. Kummamuru et al. [20] showed a modified version of the algorithm proposed by Oh et al. where this version can be applied in datasets containing large number of documents or words. Pal et al. [21] showed the production of memberships simultaneously in order to avoid various problems of the FCM. De Carvalho et al. [22] presented partitional fuzzy clustering methods based on different adaptive quadratic distances defined by fuzzy covariance matrices. Liu and Xu [23] obtained kernelized fuzzy attribute c-means clustering algorithm with kernel-induced distance. Chen et al. has studied the application of fuzzy methods [24–28]. Tang et al. [29] proposed a new kind of data weighted fuzzy c-means clustering approach. Mei and Chen [30] characterized each cluster by multiple medoids with the help of prototype weight. After, Mei and Chen [31] proposed fuzzy clustering with multiple weighted medoids. An important parameter that influences the robustness of the FCM is the weighting exponent m called fuzzifier. Yu [32] proposed a theoretical approach to measuring this parameter. Wu [33] introduced a new guideline for selecting the weighting exponent m.

In a fuzzy clustering, the memberships are calculated based on distances between clusters and prototypes and is assumed that these memberships are the same for all the features, i.e., the features are considered equally important for the definition of the memberships. However, this model can be restrictive since the features can have dissimilar dispersions and the fuzzy clustering algorithm can have its performance affected. The main contribution of this work is to introduce a new FCM method where the membership values are computed based on the inherent information in each feature. The idea of this method is to find a set of prototypes and a multivariate fuzzy partition that minimizes an objective function. Here, the multivariate memberships allows to take account the intra-structure of the clustering. Section 2 describes a motivation example. In Section 3, a fuzzy c-means method based on membership values computed by feature is presented. In order to validate the proposed method, Section 4 shows experiments with synthetic and real data sets. A comparative study of this method in relation to two fuzzy c-means clustering methods with adaptive distances introduced in [22] is performed. These adaptive distances consider weights for variables according to the intra-structure of the clustering and they are capable of increasing the performance of the fuzzy clustering algorithms since they can be the same for all clusters or different from one cluster to another. Moreover, the proposed method is compared with the Gustafson–Kessel algorithm [34] whose the main feature is its capacity of identify clusters of different sizes and shapes. In Section 5, the concluding remarks are given.

2. A motivation example

Fig. 1 shows a data set with two clusters. Cluster 1 has five patterns indexed from 1 to 5 (labeled as circles) and cluster 2 has five patterns indexed by 6–10 (labeled as stars). According to the fuzzy c-means method, is expected that the item 5 has membership value in cluster 2 slightly greater than the membership value in cluster 1 since the distance between the items 3 and 5 is greater than distance between the items 5 and 8. However, analyzing the feature y_2 separately, we can verify that the membership of the item 5 in cluster 1 is close to 1 and the membership in cluster 2 is close to 0. Whereas, investigating the feature y_1, the membership of the item 5 in cluster 2 has value close to 1, but the membership in cluster 1 has not close to 0 since the dispersion of the feature y_1 is greater than that of the feature y_2. In this case, an approach that computes multivariate membership values is needed.

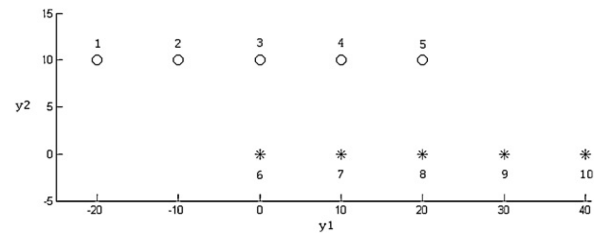

Fig. 1. Data set with two clusters.

The membership values that will be computed in this paper are such that: (a) they are able to take account the statistical information of each feature, in order to improve the performance of the algorithm, (b) they can change at each iteration and, they can be different from one feature to another and from one class to another, and (c) the algorithm is able to derive cluster prototypes optimizing an objective function based on multivariate membership values. In order to obtain a fuzzy partition by class of the input data set, a way of calculating the class membership values is also proposed in this paper.

3. Fuzzy *c*-means based on multivariate memberships

This section starts with a brief description of the classical fuzzy c-means and, subsequently, it introduces a fuzzy c-means algorithm that is able to find a multivariate fuzzy partition taking into account multiple membership matrices (denote here MFCM).

Consider Ω a set of n patterns indexed by k and formed by p features indexed by j. Each pattern k is represented by a quantitative feature vector $\mathbf{x}_k = (x_{1k}, \ldots, x_{pk})^t$. Let $L = \{\mathbf{y}_1, \ldots, \mathbf{y}_c\}$ be a set of c prototypes associated to a fuzzy partitioning into c clusters. Each prototype of a cluster $C_i (i = 1, \ldots, c)$ is also represented as a quantitative feature vector $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^t$.

The Fuzzy C-means (FCM) algorithm proposed by Bezdek [4] aims to find a prototype data set L and a fuzzy partitioning $\mathbf{U} = [u_{ik}](i = 1, \ldots, c)(k = 1, \ldots, n)$ of the data set Ω, by minimizing an objective function given by:

$$J = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m ||\mathbf{x}_k - \mathbf{y}_i||^2. \tag{1}$$

The fuzzy partitioning $\mathbf{U}$ represents a membership matrix where u_{ik} is the membership degree of a given point k which belongs to the cluster i under the following restrictions:

1. $u_{ik} \in [0, 1]$ for all i and k;
2. $0 < \sum_{k=1}^{n} u_{ik} < n$ for all i and
3. $\sum_{i=1}^{c} u_{ik} = 1$ for all k.

In the MFCM clustering algorithm, the memberships degrees are different from one feature to another and from one cluster to another. Therefore, it is necessary to consider an appropriate representation of the memberships and a way to calculate the distance between clusters and prototypes.

Let n multivariate membership matrices $\mathbf{u}_k = [u_{ijk}](k = 1, \ldots, n)$ where u_{ijk} is the membership degree of the pattern k to the cluster $i(i = 1, \ldots, c)$ on the feature $j(j = 1, \ldots, p)$. Let n distances matrices $\mathbf{D}_k = [d_{ijk}]$ where d_{ijk} is the distance between the pattern x_{kj} and the prototype y_{ij} of the cluster $i(i = 1, \ldots, c)$ on the feature $j(j = 1, \ldots, p)$.

The MFCM aims to look for a multivariate fuzzy partition $\mathbf{V} = [\mathbf{u}_k](k = 1, \ldots, n)$ of the data set Ω, by minimizing an objective function given by:

$$J^M = \sum_{i=1}^{c}\sum_{k=1}^{n}\sum_{j=1}^{p}(u_{ijk})^m d_{ijk}, \tag{2}$$

where the distance d_{ijk} is given as

$$d_{ijk} = (x_{jk} - y_{ij})^2. \tag{3}$$

The prototype y_{ij} is influenced by the degree of membership u_{ijk} and pattern x_{ikj} on the feature j. It is given as

$$y_{ij} = \frac{\sum_{k=1}^{n}(u_{ijk})^m x_{jk}}{\sum_{k=1}^{n}(u_{ijk})^m}. \tag{4}$$

Proposition 3.1. *The prototype vector* $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^t$ *of the cluster* $C_i(i = 1, \ldots, c)$*, which minimizes the clustering criterion* J^M *of the Eq. (2), is updated according to the Eq. (4).*

Proof 1. Fixed the parameter u_{ijk}, the clustering criterion J^M being additive, the problem becomes to find the prototype y_{ij} of the cluster $C_i(i = 1, \ldots, c)$ on the feature j such that minimizes

$$J_{ij} = \sum_{k=1}^{n}(u_{ijk})^m d_{ijk}^2 = \sum_{k=1}^{n}(u_{ijk})^m (x_{jk} - y_{ij})^2. \tag{5}$$

The criterion stabilizes when:

$$\frac{dJ_{ij}}{dy_{ij}} = \sum_{k=1}^{n}(u_{ijk})^m 2(x_{jk} - y_{ij})(-1) = 0, \tag{6}$$

$$\sum_{k=1}^{n}(u_{ijk})^m x_{jk} - \sum_{k=1}^{n}(u_{ijk})^m y_{ij} = 0, \tag{7}$$

$$y_{ij} = \frac{\sum_{k=1}^{n}(u_{ijk})^m x_{jk}}{\sum_{k=1}^{n}(u_{ijk})^m}. \tag{8}$$

□

The multivariate fuzzy partition $\mathbf{V}$ represents multiple membership matrices where the parameter u_{ijk} are the membership degree of the pattern k in cluster C_i on the feature j. It is based on the distance d_{ijk} such as:

$$u_{ijk} = \left[\sum_{h=1}^{c}\sum_{l=1}^{p}\left(\frac{d_{ijk}}{d_{hlk}}\right)^{(1/m-1)}\right]^{-1}. \tag{9}$$

Similarly as presented in [4], the membership u_{ijk} satisfies the following constraints:

1. $u_{ijk} \in [0, 1]$ for all i, j and k;
2. $0 < \sum_{j=1}^{p}\sum_{k=1}^{n} u_{ijk} < n$ for all i and
3. $\sum_{i=1}^{c}\sum_{j=1}^{p} u_{ijk} = 1$ for all k.

The first constraint ensures that the degree membership u_{ijk} is a probabilistic value, whatever the cluster C_i, the feature j and the pattern $\mathbf{x}_k$. The second one is a consequence of the first. The last constraint ensures that the sum of all probabilities for the pattern $\mathbf{x}_k$ is equal to 1.

Proposition 3.2. *The membership degree* u_{ijk} *of the pattern* k *belonging to the class* C_i *on the feature* j *minimizes the clustering criterion* J^M *of the Eq. (2) and it is calculated using the Eq. (9).*

Proof 2. Fixed the parameter m and the prototype vector $\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^t$ of the cluster C_i and the clustering criterion J^M being additive, the problem becomes to find for some pattern $\mathbf{x}_k(k = 1, \ldots, n)$ the membership u_{ijk} that minimizes

$$J_k = \sum_{i=1}^{c}\sum_{j=1}^{p}(u_{ijk})^m d_{ijk}. \tag{10}$$

Thus, applying the Lagrange multiplier method for finding an optimum value to J_k, we have:

$$F_k(\lambda) = \sum_{i=1}^{c}\sum_{j=1}^{p} u_{ijk}^m d_{ijk} - \lambda\left(\sum_{i=1}^{c}\sum_{j=1}^{p} u_{ijk} - 1\right). \tag{11}$$

F_k is stationary in the following situations:

$$\frac{\partial F_k}{\partial \lambda} = \sum_{i=1}^{c}\sum_{j=1}^{p} u_{ijk} - 1 = 0, \tag{12}$$

$$\frac{\partial F_k}{\partial u_{rst}} = [m(u_{rst})^{m-1} d_{rst} - \lambda] = 0. \tag{13}$$

From the Eq. (13), we can rewrite the membership of the pattern $\mathbf{x}_t$ of the cluster C_r on the feature s using the following equation:

$$u_{rst} = \left(\frac{\lambda}{md_{rst}}\right)^{(1/m-1)} = \left(\frac{\lambda}{m}\right)^{(1/m-1)}\left(\frac{1}{d_{rst}}\right)^{(1/m-1)} \tag{14}$$

Regarding the Eqs. (12) and (14) and the third constraint previously presented:

$$\sum_{a=1}^{c}\sum_{b=1}^{p} u_{abt} = \left(\frac{\lambda}{m}\right)^{(1/m-1)}\left\{\sum_{a=1}^{c}\sum_{b=1}^{p}\left(\frac{1}{d_{abt}}\right)^{(1/m-1)}\right\} = 1. \tag{15}$$

Thus, we may have the following result:

$$\left(\frac{\lambda}{m}\right)^{(1/m-1)} = \frac{1}{\sum_{a=1}^{c}\sum_{b=1}^{p}\left(\frac{1}{d_{abt}}\right)^{(1/m-1)}} \tag{16}$$

Substituting the Eq. (16) into Eq. (14) we have

$$\begin{aligned} u_{rst} &= \frac{1}{\sum_{a=1}^{c}\sum_{b=1}^{p}(1/d_{abt})^{(1/m-1)}}\left(\frac{1}{d_{rst}}\right)^{(1/m-1)}, \\ &= \frac{1}{\sum_{a=1}^{c}\sum_{b=1}^{p}(d_{rst}/d_{abt})^{(1/m-1)}}. \end{aligned} \tag{17}$$

Concluding, the membership u_{rst} of the pattern $\mathbf{x}_t$ to the cluster C_r on the feature s is as follows:

$$u_{rst} = \left[\sum_{a=1}^{c}\sum_{b=1}^{p}\left(\frac{d_{rst}}{d_{abt}}\right)^{(1/m-1)}\right]^{-1}. \tag{18}$$

□

In order to obtain class memberships, a way for calculating these memberships is proposed. Consider δ_{ik} to be the membership degree of the pattern k to cluster C_i. Here, the membership δ_{ik} represents an aggregation measure for all the p features given as:

$$\delta_{ik} = \sum_{j=1}^{p} u_{ijk}. \tag{19}$$

Regarding this class membership degree $\delta_{ik}(i=1,\ldots,c)(k=1,\ldots,n)$ a classical fuzzy partition $\mathbf{U}=[\delta_{ik}](i=1,\ldots,c)(k=1,\ldots,n)$ can be obtained where δ_{ik} satisfies the following restrictions:

1. $\delta_{ik}\in[0,1]$ for all i and k;
2. $0<\sum_{k=1}^{n}\delta_{ik}<n$ for all i and
3. $\sum_{i=1}^{c}\delta_{ik}=1$ for all k.

The multivariate fuzzy c-means clustering algorithm is given by the following steps:

(1) **Initialization**
Fix c, $2\le c\le n$; fix m, $1<m<\infty$; fix T; and fix $\varepsilon>0$; Initialize randomly $u_{ijk}(i=1,\ldots,c;\ j=1,\ldots,p$ and $k=1,\ldots,n)$ of pattern k belonging to cluster C_i on feature j such that $u_{ijk}\in[0,1]$ and $\sum_{i=1}^{c}\sum_{j=1}^{p}u_{ijk}=1$. Do $t=1$.
(2) **Representation**
(Fix the membership μ_{ijk} of the pattern k $(k=1,\ldots,n)$ belonging to class C_i on the feature j $(j=1,\ \ldots,\ p)$). Compute the prototype $\mathbf{y}_i$ of the cluster $C_i(i=1,\ldots,c)$ using the Eq. (4).
(3) **Feature membership**
(Fix the prototype $\mathbf{y}_i$ of the cluster $C_i(i=1,\ \ldots,\ c)$). Update the fuzzy membership degree u_{ijk} of the pattern k belonging to cluster C_i on the feature j using the Eq. (9).
(4) **Stopping criterion**
If $||J_{t+1}^{M}-J_{t}^{M}||\le\varepsilon$ or $t>T$, go step 6); else update $t=t+1$ and go step 3).
(5) **Class membership**
(Fix the prototype $\mathbf{y}_i$ of the cluster C_i and the membership $\mu_{ijk}(i=1,\ldots,c)$, $(j=1,\ldots,p)$, $(k=1,\ldots,n)$) Compute the fuzzy membership degree δ_{ik} of the pattern k belonging to cluster C_i using the Eq. (19).

The time complexity of this algorithm can be calculated analyzing the complexity of each step. In the initialization step, the goal is to calculate the membership degrees for all n patterns concerning c clusters and p features, thus the complexity is $O(ncp)$. In the representation step, the computation of the c prototype vectors depends on the cardinal n and p features that describe the patterns, so for this step the time complexity is $O(ncp)$. The feature membership step performs c^2p^2 operations for each pattern. Therefore, the complexity of this step is $O(nc^2p^2)$. Regarding the stopping criterion and the class membership steps, the complexity is $O(ncp)$ for each one. In conclusion, the time complexity of the algorithm is $O(nc^2p^2)$.

4. Experimental evaluation

In this paper, four synthetic data sets and five real data ones at the UCI Repository [35] were considered in order to compare the proposed method (MFCM) with the classical fuzzy c-means (FCM) method introduced in [4]. In addition, two fuzzy c-means methods based on adaptive quadratic distances defined as pooled diagonal fuzzy covariance matrices [22] are also used in this study: the first method (here denoted AFCM-S) uses a single adaptive distance and the second one (here denoted AFCM-C) considers class adaptive distances. Moreover, the Gustafson–Kessel (GK) algorithm [34] is also applied to these data sets. The GK method is able to identify clusters of different sizes and shapes [36,37] since each cluster has its own matrix that is used to calculate the Mahalanobis distance between the objects and respective prototypes. The AFCM-C and AFCM-S algorithms are only special cases of the GK one with diagonal matrices for distance measuring (here called GK-D). In this study, it is also considered the GK based on arbitrary matrices (here called GK-F) for real data sets. It is important to stress that the proposed algorithm is suitable for axes-parallel clusters only.

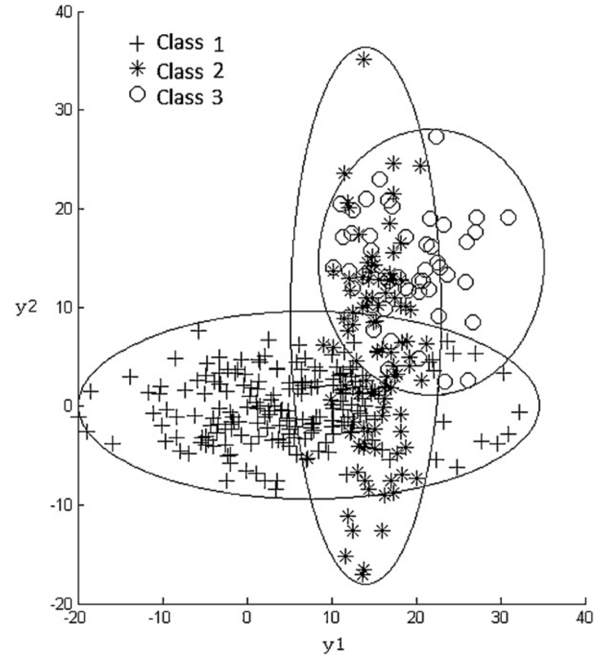

Fig. 2. Data set 1.

The evaluation of the clustering results furnished by the methods is based on the computation of an external cluster validity index. In particular, this paper uses an extension of the corrected Rand (CR) index [38] to fuzzy clustering. In the fuzzy approach, even if two fuzzy partition result in the same hard partition, the membership may represent different information. The classical Rand index does not consider this information from partitions and it is not appropriate for fuzzy clustering assessment. This extension of the corrected Rand index [39] is able to evaluate a fuzzy partition of a data set. Thus, for this reason, the Fuzzy Rand index (here called FR index) is adopted in this work. It takes its values in the interval [0,1] where the value 1 indicates a perfect agreement between the fuzzy partitions, unlike the value 0.

In our experiments, we use the parameter fuzzifier m equal to 2. The FR index is computed using the t and s norms implemented as $t=\min$ and $s=\max$.

4.1. Synthetic data sets

Synthetic data sets are described by points in the $\Re^2$ following a bi-variate normal distribution of independent components. Each data set has 350 points scattered into tree classes. Four different configuration models are considered. The goal concerns to generate overlapping among classes regarding different or equal diagonal covariance matrices for the classes.

Let $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$, be a diagonal covariance matrix and a mean vector, respectively, noted as:

$$\mathbf{\Sigma}=\begin{pmatrix}\sigma_1^2 & 0\\ 0 & \sigma_2^2\end{pmatrix}\ \boldsymbol{\mu}=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}.$$

4.1.1. Configuration model 1: $\Sigma_i\neq\Sigma_g(i,g=1,2,3)$ with $\sigma_1\neq\sigma_2$

Data set 1 (Fig. 2) consists of 350 points scattered among three classes of sizes: 200 points (class 1), 100 (class 2) and 50 (class

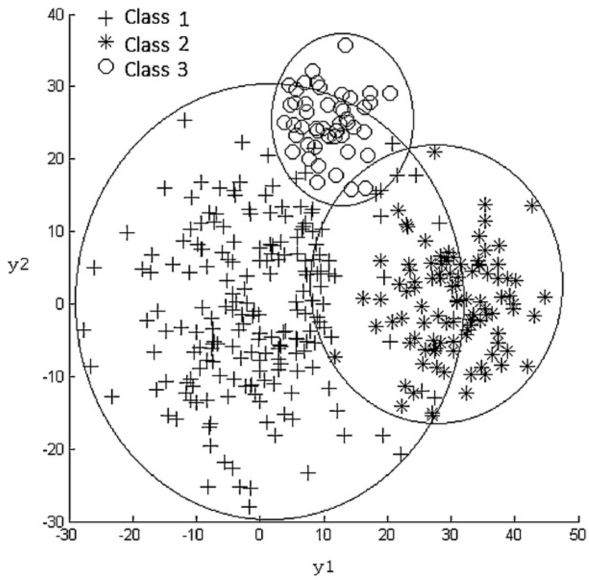

Fig. 3. Data set 2.

3). The points of each class were drawn according to the following parameters:

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 81 & 0 \\ 0 & 9 \end{pmatrix} \boldsymbol{\mu}_1 = \begin{pmatrix} 5 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 9 & 0 \\ 0 & 100 \end{pmatrix} \boldsymbol{\mu}_2 = \begin{pmatrix} 15 \\ 5 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 25 & 0 \\ 0 & 36 \end{pmatrix} \boldsymbol{\mu}_3 = \begin{pmatrix} 18 \\ 14 \end{pmatrix}.$$

4.1.2. Configuration model 2: $\boldsymbol{\Sigma}_i \neq \boldsymbol{\Sigma}_g (i, g = 1, 2, 3)$ with $\sigma_1 = \sigma_2$

Data set 2 (Fig. 3) also consists of 350 points scattered among three classes of sizes: 200 points (class 1), 100 (class 2) and 50 (class 3). The points of each class were drawn according to the following parameters:

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 100 & 0 \\ 0 & 100 \end{pmatrix} \boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 49 & 0 \\ 0 & 49 \end{pmatrix} \boldsymbol{\mu}_2 = \begin{pmatrix} 30 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 16 & 0 \\ 0 & 16 \end{pmatrix} \boldsymbol{\mu}_3 = \begin{pmatrix} 10 \\ 25 \end{pmatrix}.$$

4.1.3. Configuration model 3: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_g (i, g = 1, 2, 3)$ with $\sigma_1 \neq \sigma_2$

Data set 3 (Fig. 4) consists of 300 points scattered among three classes each one of size 100. The points of each class were drawn according to the following parameters:

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 100 & 0 \\ 0 & 4 \end{pmatrix} \boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 100 & 0 \\ 0 & 4 \end{pmatrix} \boldsymbol{\mu}_2 = \begin{pmatrix} 15 \\ 3 \end{pmatrix},$$

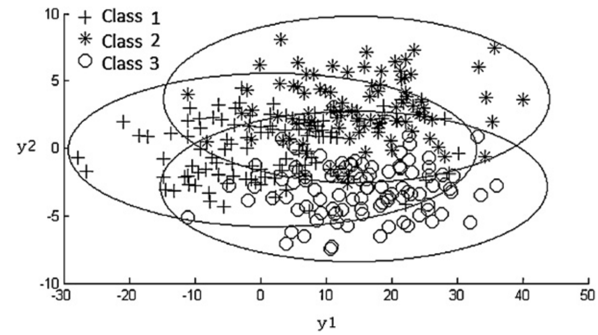

Fig. 4. Data set 3.

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 100 & 0 \\ 0 & 4 \end{pmatrix} \boldsymbol{\mu}_3 = \begin{pmatrix} 15 \\ -3 \end{pmatrix}.$$

4.1.4. Configuration model 4: $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_g (i, g = 1, 2, 3)$ with $\sigma_1 = \sigma_2$

Data set 4 (Fig. 5) consists of 300 points scattered among three classes of sizes: 150 points (class 1), 100 (class 2) and 50 (class 3). The points of each class were drawn according to the following parameters:

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 16 & 0 \\ 0 & 16 \end{pmatrix} \boldsymbol{\mu}_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 16 & 0 \\ 0 & 16 \end{pmatrix} \boldsymbol{\mu}_2 = \begin{pmatrix} 15 \\ 0 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 16 & 0 \\ 0 & 16 \end{pmatrix} \boldsymbol{\mu}_3 = \begin{pmatrix} -15 \\ 0 \end{pmatrix}.$$

4.1.5. Results

In the framework of a Monte Carlo experiment, 100 replications of each data set were carried out. For each data set, 50 random initializations of the clustering algorithm are performed. The best result from these 50 repetitions is selected according to the criterion function and the FR index is obtained. After the 100 replications, the average and standard deviation of this index are

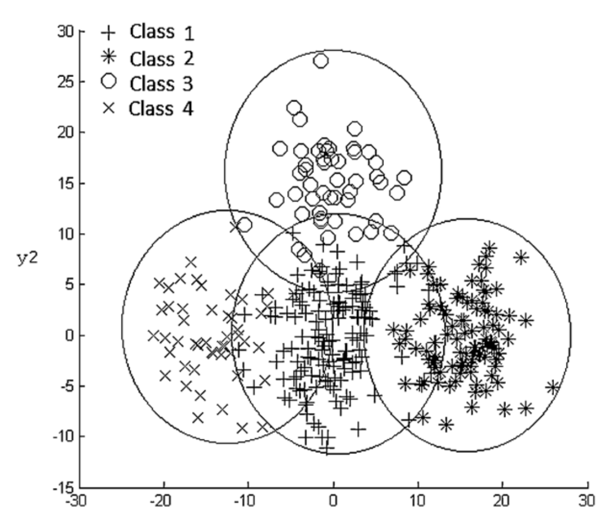

Fig. 5. Data set 4.

Table 1
Average and standard deviation (in parenthesis) of the FR index for the clustering methods and data sets 1–4.

Data set	FCM	AFCM-C	AFCM-S	MFCM
1	0.66105	0.66672	0.65552	0.72545
	(0.00918)	(0.00997)	(0.00829)	(0.01324)
2	0.69772	0.68880	0.69591	0.68921
	(0.00942)	(0.00910)	(0.01033)	(0.01035)
3	0.61239	0.59990	0.61385	0.66615
	(0.01119)	(0.00735)	(0.01381)	(0.01025)
4	0.71618	0.69825	0.69851	0.68355
	(0.00927)	(0.01404)	(0.01409)	(0.01340)

Table 2
Statistical tests comparing the clustering methods for data sets 1 and 3.

Data set	H_0: $\mu_1 \geq \mu_4$ H_1: $\mu_1 < \mu_4$	H_0: $\mu_2 \geq \mu_4$ H_1: $\mu_2 < \mu_4$	H_0: $\mu_3 \geq \mu_4$ H_1: $\mu_3 < \mu_4$
1	−39.77189	−34.53694	−44.54171
3	−33.86047	−38.89779	−39.83476

calculated. Table 1 gives the values of the average an standard deviation (in parenthesis) of the FR index obtained with the clustering methods and for data sets 1–4.

As expected, the performance of the proposed method measured by the FR index was superior to those of the FCM and AFCM-C, AFCM-S methods for the data sets 1 and 3 since $\sigma_1^2 \neq \sigma_2^2$. In order to compare these methods regarding the data sets 1 and 3, Student's t-tests for independent samples with 5% of significance are performed. Table 2 gives the values of the test statistics following a Student's t-distribution with 198° of freedom. In this table, μ_1, μ_2, μ_3 and μ_4 are, respectively, the average of the FR index for FCM and AFCM-C, AFCM-S, MFCM methods. From the results in this table, we can conclude that the proposed method is the best option. An important aspect to be highlighted is that the data sets 1 and 3 are described by features with dissimilar dispersions. Although the methods based on adaptive [22] distances are able to recognize clusters of different shapes and sizes, the partition obtained by the proposed method is closer to the original one than those provided by these methods.

With respect to the data sets 2 and 4 with $\sigma_1^2 = \sigma_2^2$, the average values of the FR index are similar for all the methods in data set 2. However, the FR index average for the proposed method is slightly inferior to those for the FCM and AFCM-C, AFCM-S methods in data set 4. The comparison is also achieved based on Student's t-tests for independent samples with 5% of significance. Table 3 gives the values of the test statistics. The results in this table show clearly that the FCM and AFCM-S, AFCM-C methods outperform the proposed MFCM one.

In conclusion, for these data configurations, the performance of the proposed method depends on the variance structure of the features. The multivariate FCM method performs well when the features have dissimilar dispersions and it is better than the adaptive FCM methods based on diagonal fuzzy covariance matrices that aim to find clusters of different shapes and sizes.

In order to make a more detailed study comparing the output of the MFCM and GK-D methods, the following tables show the centers and dispersions (standard deviations) of the clusters for each configuration and method (see Tables 4–7).

Table 3
Test statistics comparing the clustering methods for data sets 2 and 4.

Data set	H_0: $\mu_1 \leq \mu_4$ H_1: $\mu_1 > \mu_4$	H_0: $\mu_2 \leq \mu_4$ H_1: $\mu_2 > \mu_4$	H_0: $\mu_3 \leq \mu_4$ H_1: $\mu_3 > \mu_4$
2	6.05028	−0.29601	4.55886
4	19.92545	7.53612	7.65513

Table 4
Centers of the clusters for GK-D method.

Data set	Cluster 1		Cluster 2		Cluster 3	
	y_1	y_2	y_1	y_2	y_1	y_2
1	2.445	−0.075	15.871	−2.025	17.147	13.829
2	−2.260	−1.896	28.804	−0.529	10.593	22.355
3	−2.377	−0.378	15.384	3.090	19.037	−2.687
4	−0.208	0.357	15.337	0.325	−14.951	0.070

Table 5
Dispersions of the clusters for GK-D method.

Data set	Cluster 1		Cluster 2		Cluster 3	
	y_1	y_2	y_1	y_2	y_1	y_2
1	5.956	2.642	3.938	5.923	4.592	5.245
2	8.719	8.127	7.313	6.632	4.794	6.280
3	7.852	1.602	8.208	1.630	8.430	1.536
4	3.420	3.951	3.714	3.782	3.829	4.212

Table 6
Centers of the clusters for MFCM method.

Data set	Cluster 1		Cluster 2		Cluster 3	
	y_1	y_2	y_1	y_2	y_1	y_2
1	3.674	−0.112	14.472	7.110	20.569	13.518
2	2.036	−3.735	20.809	0.389	9.664	18.008
3	1.119	−0.480	13.389	3.261	13.034	−2.861
4	1.741	0.967	10.607	2.335	−0.688	−4.109

Table 7
Dispersions of the clusters for MFCM method.

Data set	Cluster 1		Cluster 2		Cluster 3	
	y_1	y_2	y_1	y_2	y_1	y_2
1	9.811	2.706	3.619	9.331	4.448	8.456
2	14.359	10.334	15.491	4.986	6.682	12.022
3	10.772	1.425	12.391	1.985	9.148	3.172
4	9.264	3.100	9.737	4.887	13.113	2.718

The comparison between the MFCM and GK-D methods is achieved by the dissimilarity measure

$$d = \sum_{i=1}^{c}\sum_{j=1}^{p}\phi_{ij}, \tag{20}$$

where ϕ_{ij} is the City-Block distance function measuring the difference between the center (or dispersion) obtained by a method and the corresponding *a priori* one for variable j.

Table 8 shows the values of d for the configurations 1–4 and GK-D and MFCM methods regarding centers and dispersions. According to these values, it can be observed that the centers and dispersions furnished by the MFCM method are more similar to the *a priori* centers and dispersions than those furnished by the GK-D one for configurations 1 and 3 (where the features have dissimilar dispersions) unlike the configurations 2 and 4 where the features have similar dispersions.

Table 8
Comparison between GK-D and MFCM methods using the d function.

Data set	d for centers		d for dispersions	
	GK-D	MFCM	GK-D	MFCM
1	11.550	7.127	9.580	5.401
2	9.119	22.679	6.909	25.902
3	7.579	5.576	6.742	5.777
4	1.346	27.857	1.516	23.183

Table 9
Average and standard deviation of the features by class for the wine data set.

Feature	Class 1 (size = 59)		Class 2 (size = 71)		Class 3 (size = 48)	
	Average	SD	Average	SD	Average	SD
Alcohol	13.7447	0.4621	12.2787	0.5379	13.1537	0.5302
Malic acid	2.0106	0.6885	1.9326	1.0155	3.3337	1.0879
Ash	2.4555	0.2271	2.2447	0.3154	2.4370	0.1846
Alcalinity of ash	17.0372	2.5463	20.2380	3.3497	21.4166	2.2581
Magnesium	106.3389	10.4989	94.5492	16.7534	99.3125	10.8904
Total phenols	2.8401	0.3389	2.2588	0.5453	1.6787	0.3569
Flavanoids	2.9823	0.3974	2.0808	0.7057	0.7814	0.2935
Nonflavanoid phenols	0.2900	0.0700	0.3636	0.1239	0.4475	0.1241
Proanthocyanins	1.8993	0.4121	1.6302	0.6020	1.1535	0.4088
Color intensity	5.5283	1.2385	3.0866	0.9249	7.3962	2.3109
Hue	1.0620	0.1164	1.0562	0.2029	0.6827	0.1144
OD280/OD315	3.1577	0.3570	2.7853	0.4965	1.6835	0.2721
Proline	1115.7118	221.5207	519.5070	157.2112	629.8958	115.0970

4.2. UCI machine learning repository data sets

The clustering methods are applied to five real data sets: Wine, Abalone, Haberman, Iris and Pima Indians Diabetes. These data sets may be found in the UCI machine learning repository [35] accessible at the link http://archive.ics.uci.edu/ml/datasets.html. The clusters obtained with these clustering methods were compared with the *a priori* cluster partition. Each method is run until the convergence and the best result of 200 repetitions is selected according to the adequacy criterion. For this best result, the FR is calculated.

In order to observe descriptive statistics of the features by class for these data sets, the mean and standard deviation were calculated. Moreover, histograms are built to show the behavior of the membership degrees computed after the convergence of each clustering method.

4.2.1. Wine data set

The Wine data set contains information about chemical analysis of wines grown in the same region in Italy. Each sample is described by 13 numerical features that are measures of components present in the sample wine: Alcohol, Malic acid, Ash, Alcalinity of ash, Magnesium, Total phenols, Flavanoids, Nonflavanoid phenols, Proanthocyanins, Color intensity, Hue, OD280/OD315 of diluted wines and Proline. 178 patterns of this data set are scattered in three qualities of wine. The average and standard deviation (SD) of the features by class are presented in Table 9. We can observe in this table that the feature dispersions represented by the corresponding standard deviations are dissimilar.

For this data set, the FR indices obtained by the FCM, AFCM-C, AFCM-S and GK-F methods were 0.713, 0.612, 0.604, 0.535, whereas for the proposed MFCM one was 0.789. Thus, the MFCM method outperforms the other fuzzy methods for this data set. Figs. 6–10 illustrate histograms of the membership degrees by class for all the clustering methods.

From these histograms, we can say that the main difference among the fuzzy methods is that, in the MFCM method, there are high frequencies of memberships close to values 0.0 and 1.0 and low frequencies between these values for all the clusters. Given a cluster, high frequencies of memberships close to the value 1 mean that the sum of these frequencies is close to the size of the cluster and high frequencies of memberships close to the value 0 mean that the sum of these frequencies is close to the size of the data set less the size of the cluster. This model contributes to improve the fuzzy clustering quality.

Table 10 displays a multivariate membership matrix $\boldsymbol{\mu} = [\mu_{ij}] (i = 1, 2, 3) (j = 1, \ldots, 13)$ obtained after the convergence of the proposed MFCM method for an object of this data set. The value δ_{ik} represents an aggregation measure for all the 13 features for class $i (i = 1, 2, 3)$ and individual $k (k = 1, \ldots, n)$. In this matrix, the highest value 0.866365 indicates that this object has a high probability of belonging to class 1. The computation of this value was strongly influenced by Nonflavanoid phenols and Hue features. An important observation is that the alcalinity of ash, magnesium and proline features for this object does not influence in the computation of the class membership degrees. In general, the information of each feature was considered and different values of feature membership degrees were obtained. This demonstrates that the features cannot be considered equally important in the computation of the class membership degree.

4.2.2. Abalone data set

This data set contains information about physical measurements of abalone whose goal is predicting the age. For this, the following continuous features are considered: length of the longest shell measurement (length), diameter measured perpendicular to length (diameter), height, whole weight, Shucked weight, viscera weight, shell weight and rings. There are 4177 patterns scattered in 3 classes. Table 11 presents the average and standard deviation of each feature for each class. The values in this table show clearly that the features have dissimilar dispersions within each class.

The FR indices calculated according to *a priori* partition of this data set for the FCM, AFCM-C, AFCM-S and GK-F methods were respectively 0.705, 0.500, 0.537 and 0.401, whereas for the proposed MFCM one was 0.822. Again, MFCM method surpass the fuzzy clustering methods for this data set. Figs. 11–15 display histograms of the degrees of membership by class for all the clustering methods. As the previous data, the histograms show that the MFCM method have high frequencies for membership degrees close to 0.0

Table 10
Degrees of membership by feature for an object of the wine data set.

Feature	Cluster 1	Cluster 2	Cluster 3
Alcohol	0.000554	0.000395	0.033675
Malic acid	0.009168	0.002222	0.000083
Ash	0.002751	0.053667	0.008406
Alcalinity of ash	0.000008	0.000003	0.000003
Magnesium	0.000008	0.000008	0.000030
Total phenols	0.012140	0.003249	0.000197
Flavanoids	0.003606	0.000684	0.000073
Nonflavanoid phenols	0.665086	0.013185	0.004094
Proanthocyanins	0.000810	0.002266	0.008738
Color intensity	0.000262	0.000213	0.000026
Hue	0.156992	0.006828	0.002218
OD280/OD315	0.007490	0.000781	0.000081
Proline	0.000000	0.000000	0.000000
δ_{ik}	0.866365	0.083501	0.057624

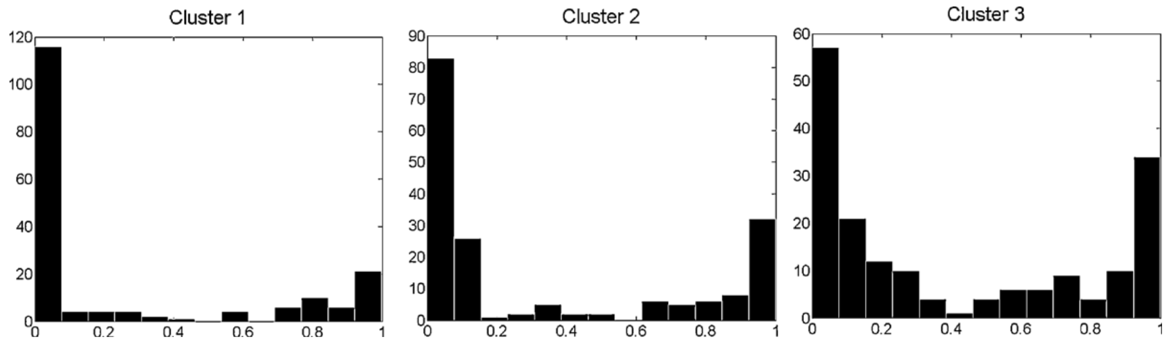

Fig. 6. Histograms of the membership degrees by class for the FCM method.

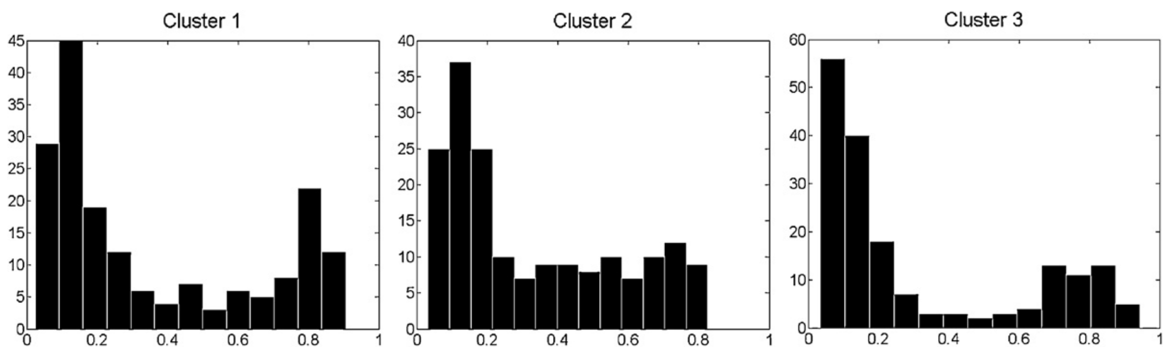

Fig. 7. Histograms of the membership degrees by class for the AFCM-C method.

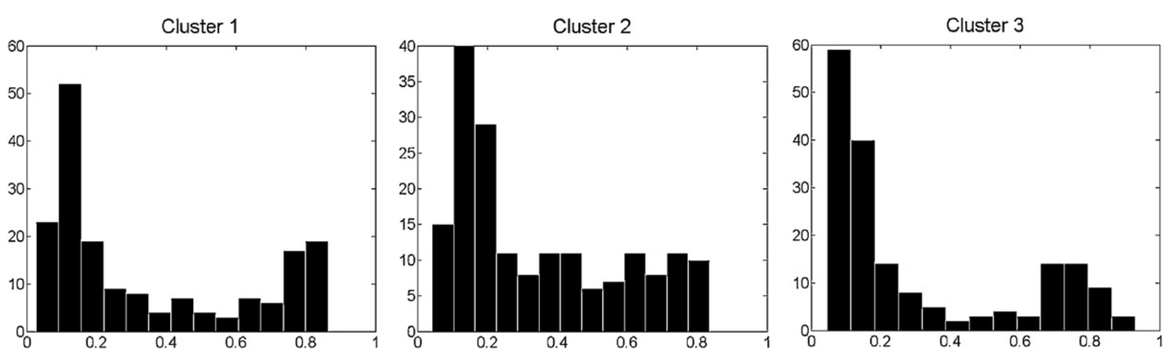

Fig. 8. Histograms of the membership degrees by class for the AFCM-S method.

and 1.0 and low frequencies for values between 0.0 and 1.0 for all the clusters. This model of memberships allowed the method to achieve a good clustering quality measure in comparison with the other methods.

4.2.3. *Haberman data set*

The data set Haberman has 306 patterns describing information of patients who had undergone surgery for breast cancer. It is part of the study conducted between 1958 and 1970 at the University

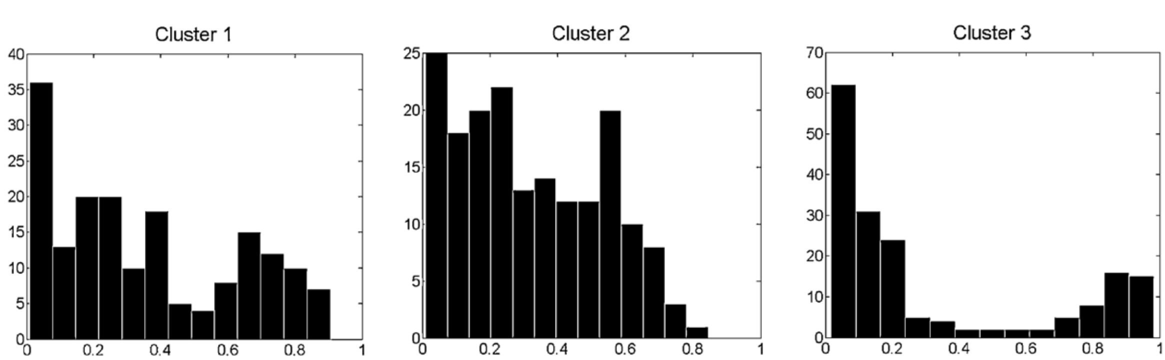

Fig. 9. Histograms of the membership degrees by class for the GK-F method.

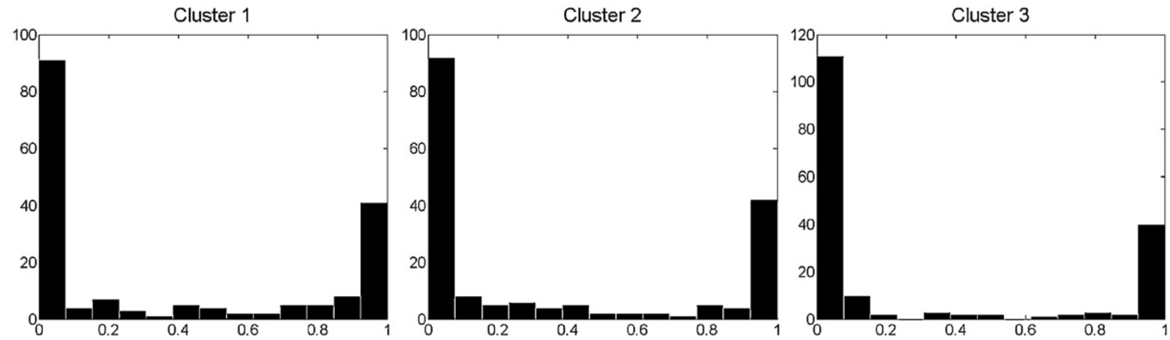

Fig. 10. Histograms of the membership degrees by class for the MFCM method.

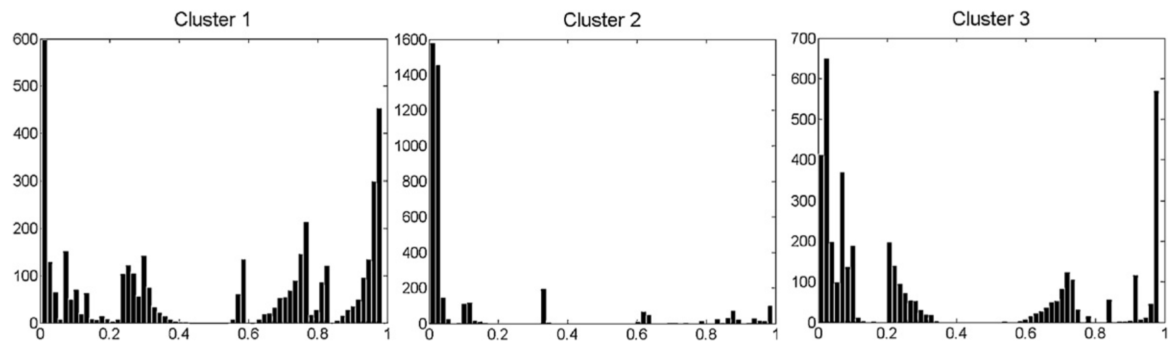

Fig. 11. Histograms of the membership degrees by class for the FCM method.

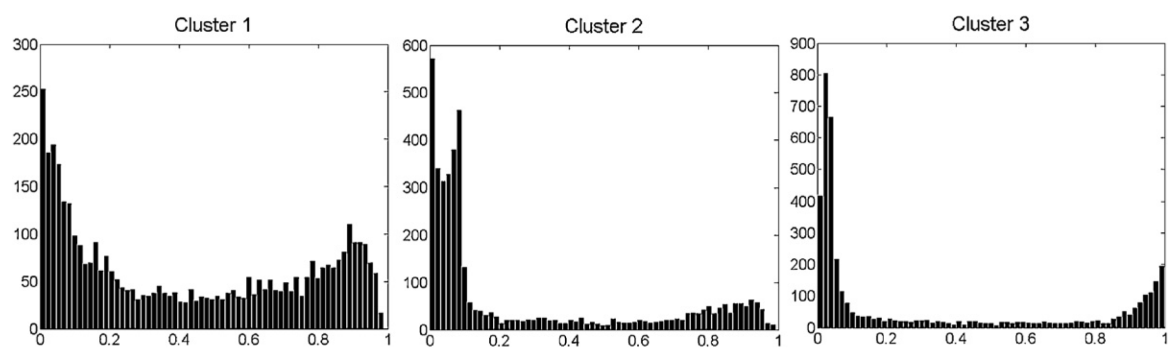

Fig. 12. Histograms of the membership degrees by class for the AFCM-C method.

of Chicago's Billings Hospital. There are three numerical features that describe the patients: age of patient at time of operation (age), patient's year of operation (year) and number of positive axillary nodes detected (nodes). The patients are scattered in two classes: those who survived five years or longer and died within five years. Table 12 shows the average and standard deviation of each feature for two classes. Here, again the features have dissimilar dispersions within each class.

In this data set, the FR indices for the FCM, AFCM-C, AFCM-S and GK-F methods and the proposed MFCM one were, respectively,

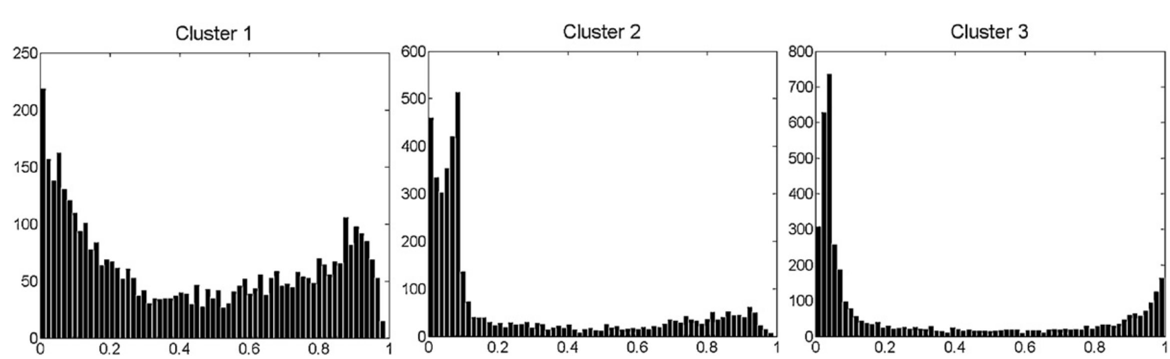

Fig. 13. Histograms of the membership degrees by class for the AFCM-S method.

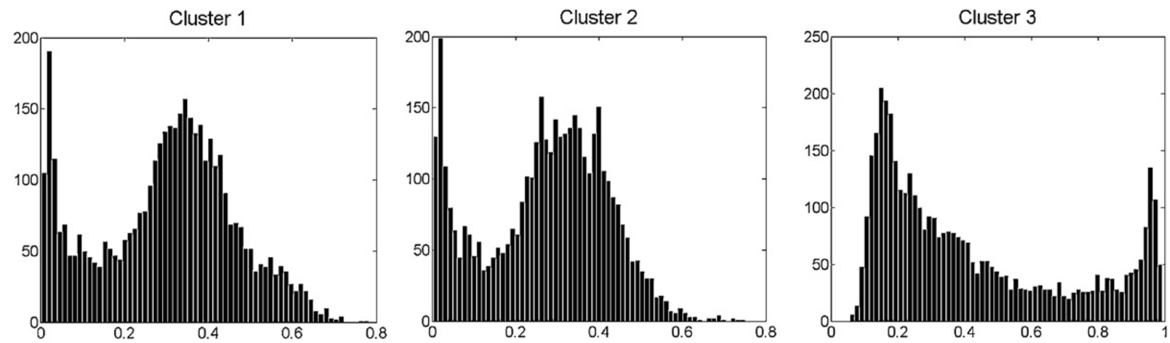

Fig. 14. Histograms of the membership degrees by class for the GK-F method.

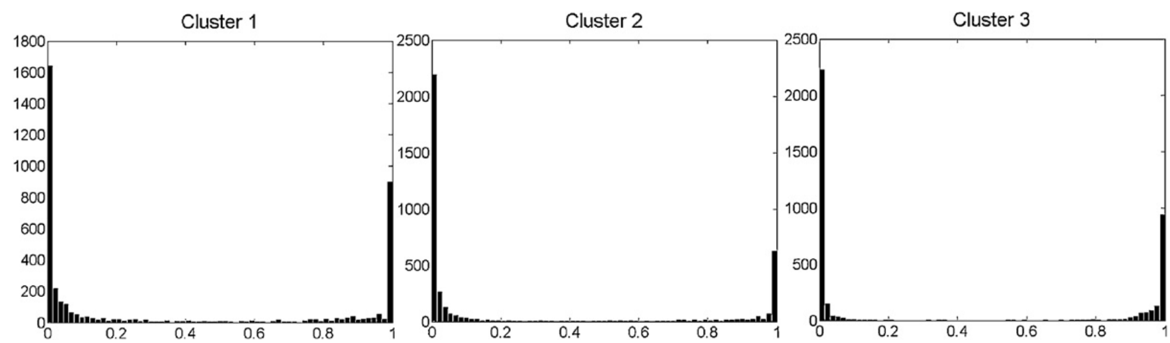

Fig. 15. Histograms of the membership degrees by class for the MFCM method.

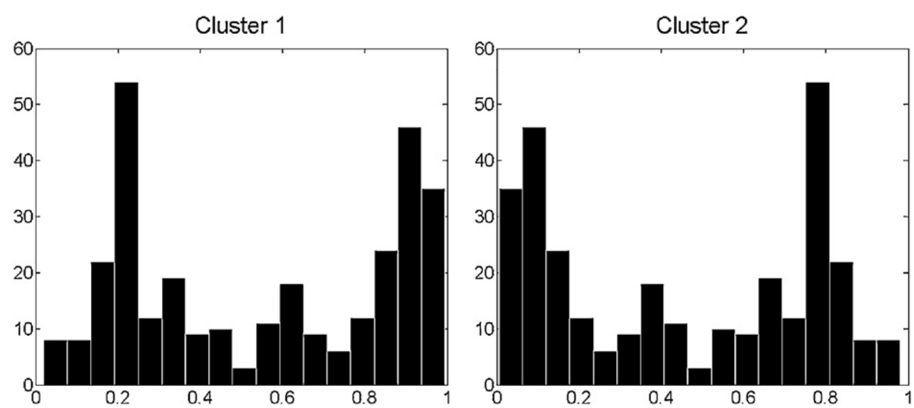

Fig. 16. Histograms of the membership degrees by class for the FCM method.

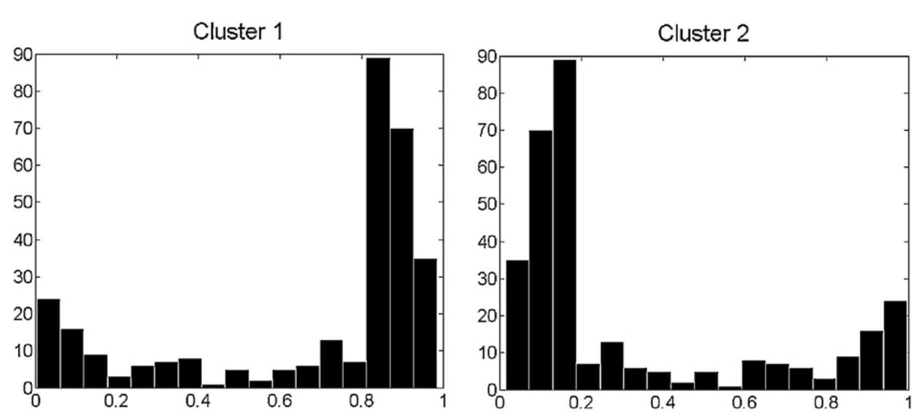

Fig. 17. Histograms of the membership degrees by class for the AFCM -C method.

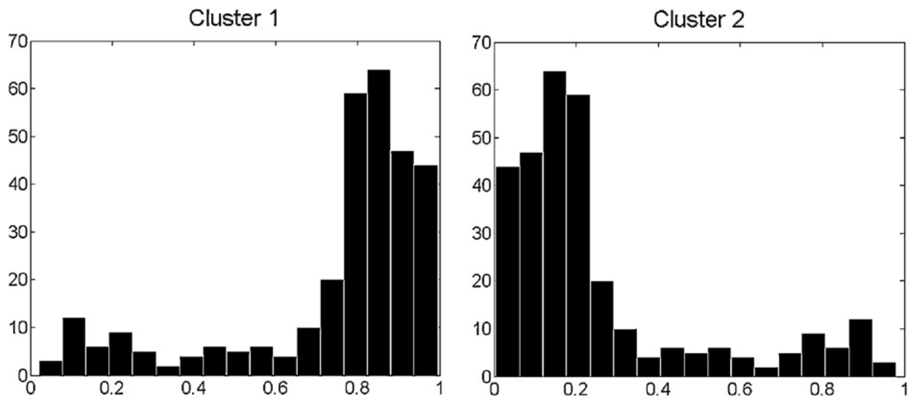

Fig. 18. Histograms of the membership degrees by class for the AFCM-S method.

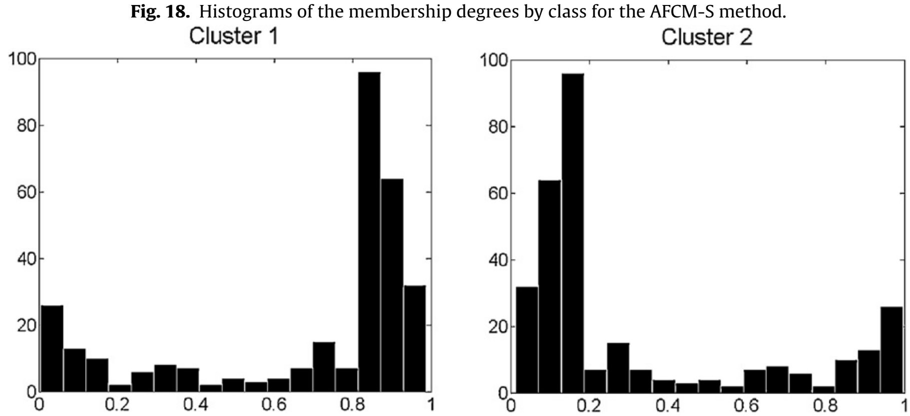

Fig. 19. Histograms of the membership degrees by class for the GK-F method.

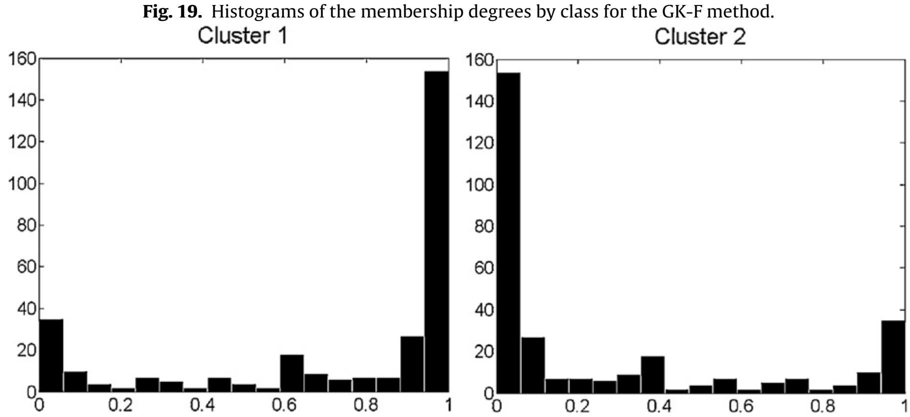

Fig. 20. Histograms of the membership degrees by class for the MFCM method.

Table 11
Average and standard deviation of the features by class for the abalone data set.

Feature	Class 1 (size = 1528)		Class 2 (size = 1307)		Class 3 (size = 1342)	
	Average	SD	Average	SD	Average	SD
Length	0.5614	0.1027	0.5791	0.0862	0.4277	0.1089
Diameter	0.4393	0.0844	0.4547	0.0710	0.3265	0.0881
Height	0.1514	0.0348	0.1580	0.0400	0.1080	0.0320
Whole weight	0.9915	0.4706	1.0465	0.4303	0.4314	0.2863
Shucked weight	0.4329	0.2230	0.4462	0.1987	0.1910	0.1284
Viscera weight	0.2155	0.1049	0.2307	0.0976	0.0920	0.0625
Shell weight	0.2820	0.1308	0.3020	0.1256	0.1282	0.0849
Rings	10.7055	3.0263	11.1293	3.1043	7.8905	2.5116

Table 12
Average and standard deviation of the features by class for the Haberman data set.

Feature	Class 1 (size = 225)		Class 2 (size = 81)	
	Average	SD	Average	SD
Age	52.0178	11.0122	53.6790	10.1671
Year	62.8622	3.2229	62.8272	3.3421
Nodes	2.7911	5.8703	7.4568	9.1857

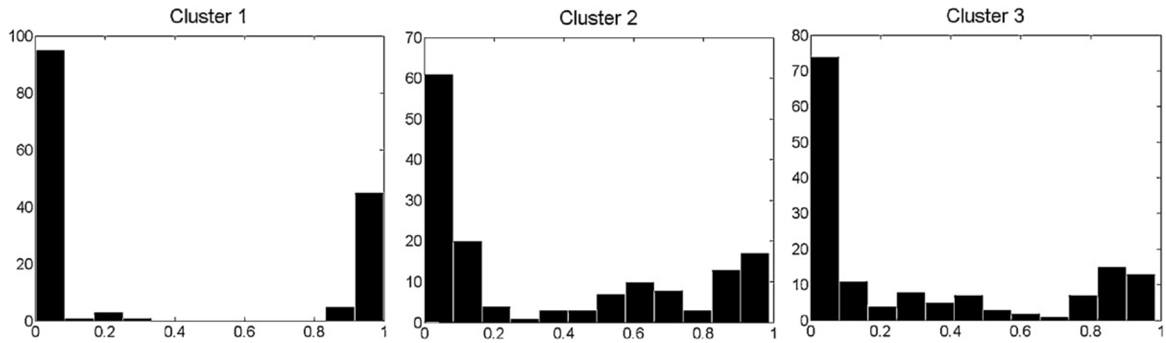

Fig. 21. Histograms of the membership degrees by class for the FCM method.

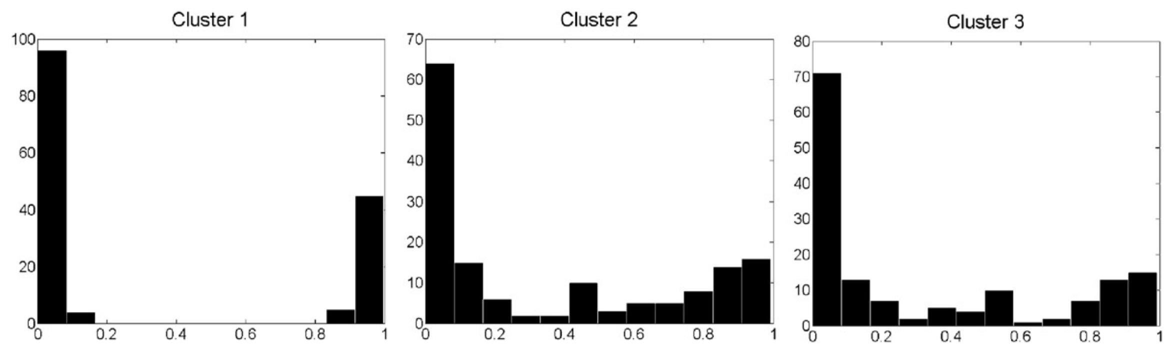

Fig. 22. Histograms of the membership degrees by class for the AFCM-C method.

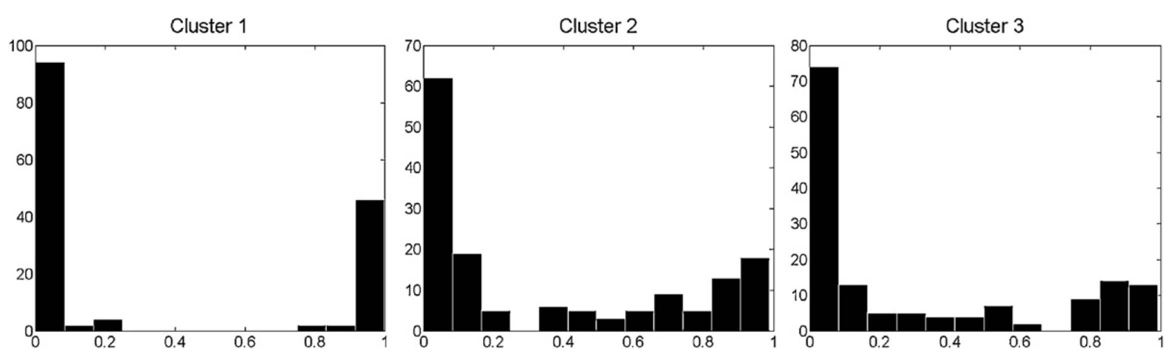

Fig. 23. Histograms of the membership degrees by class for the AFCM-S method.

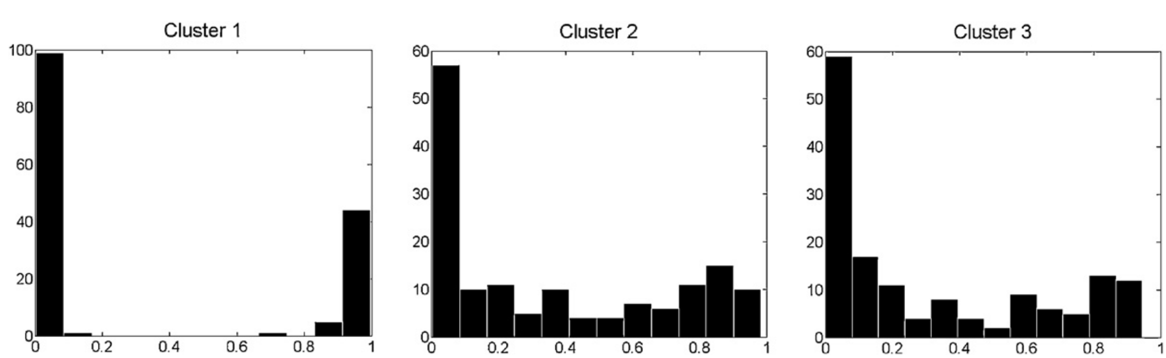

Fig. 24. Histograms of the membership degrees by class for the GK-F method.

Table 13
Average and standard deviation of the features by class for the Iris data set.

Feature	Class 1 (size = 50)		Class 2 (size = 50)		Class 3 (size = 50)	
	Average	SD	Average	SD	Average	SD
Sepal length	5.0060	0.3525	5.9360	0.5162	6.5880	0.6359
Sepal width	3.4180	0.3810	2.7700	0.3138	2.9740	0.3225
Petal length	1.4640	0.1735	4.2600	0.4699	5.5520	0.5519
Petal width	0.2440	0.1072	1.3260	0.1978	2.0260	0.2747

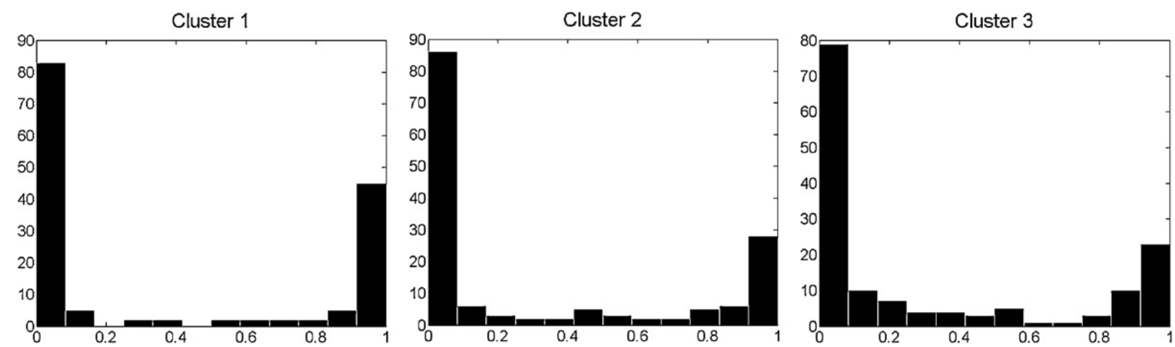

Fig. 25. Histograms of the membership degrees by class for the MFCM method.

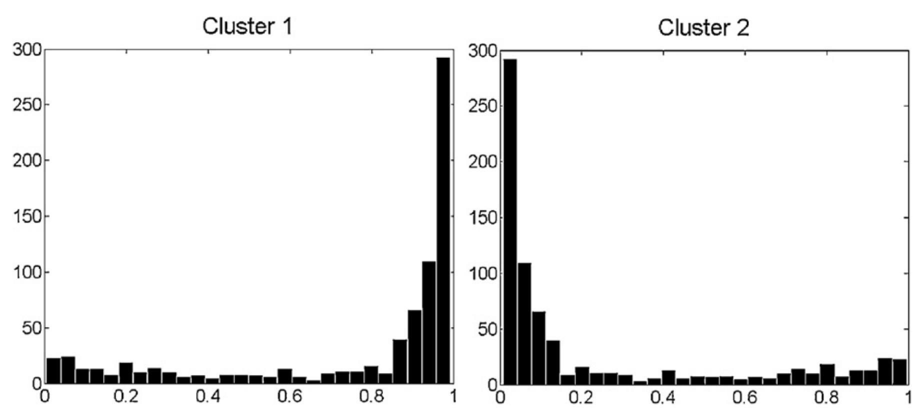

Fig. 26. Histograms of the membership degrees by class for the FCM method.

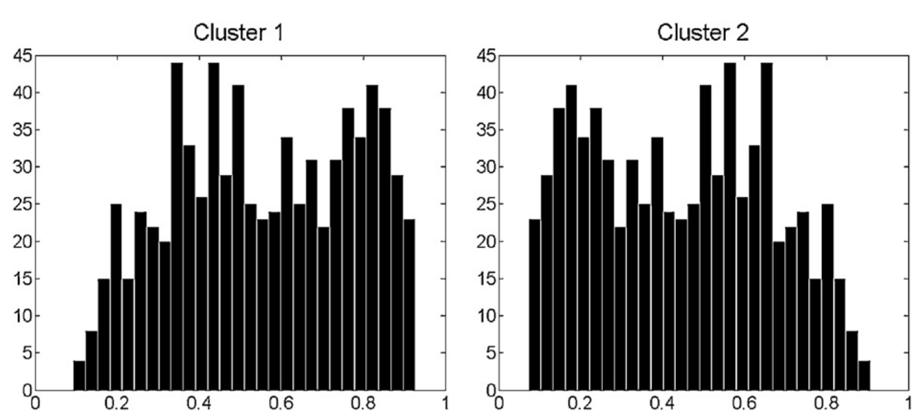

Fig. 27. Histograms of the membership degrees by class for the AFCM-C method.

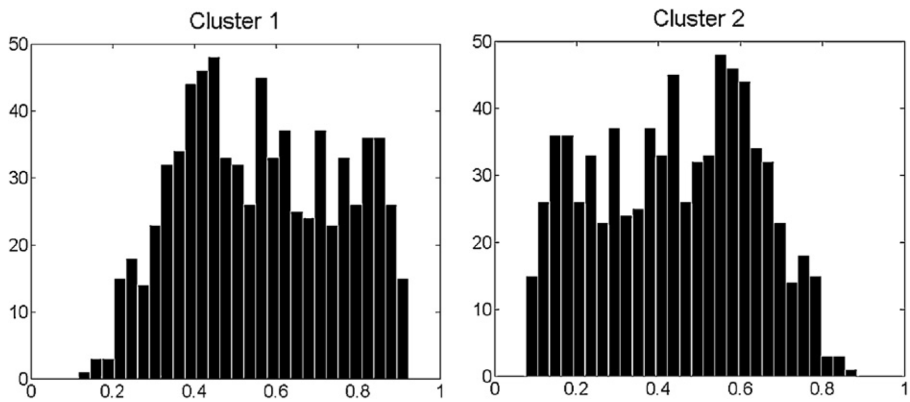

Fig. 28. Histograms of the membership degrees by class for the AFCM-S method.

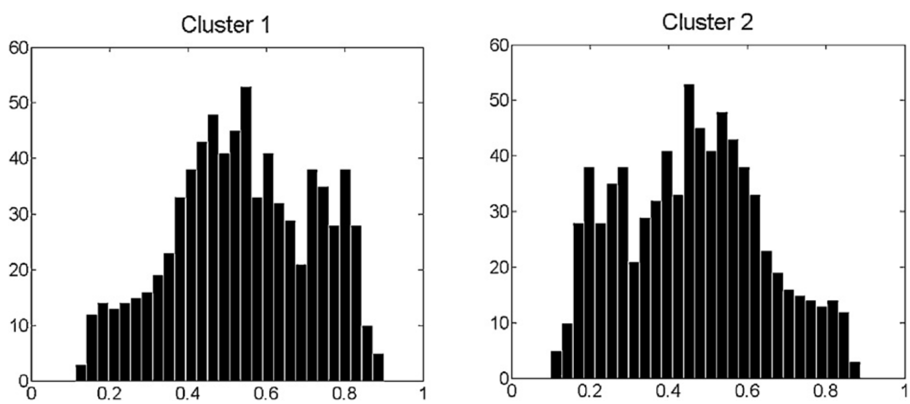

Fig. 29. Histograms of the membership degrees by class for the GK-F method.

0.645, 0.740, 0.736, 0.670 and 0.786. These indices show that the proposed method has performed better than the other ones for this data set. Figs. 16–20 present histograms of the class membership degrees for each method.

The histograms for the FCM show high frequencies close to 0.2 and 0.8 values and this method had the worst performance in terms of the FR index. The histograms for the proposed MFCM method show: for cluster 1 a high frequency of values close to 1.0 and low frequencies of values between 0.0 and 1.0, and for cluster 2 a high frequency of values close to 0.0. This is expected since the size of the class 1 (225) is much greater than the size of the class 2 (81).

4.2.4. Iris data set

This data set has 150 patterns equally scattered in three classes where each class refers to a type of iris plant: Iris Setosa, Iris Versicolour and Iris Virginica. Four features measure the dimensions of the sepal and petal of each plant: sepal length, sepal width, petal length and petal width. Table 13 displays the average and standard deviation descriptive measures for each class. From this table, we can observe that the dispersions are small and slightly dissimilar.

The clustering methods were applied to this data set and the FR indices were calculated according to the corresponding *a priori* partition. For the FCM, AFCM-C, AFCM-S and GK-F methods, the FR

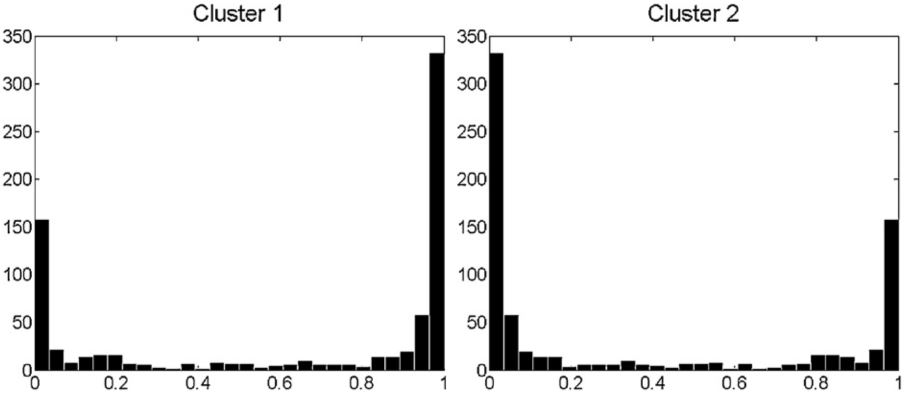

Fig. 30. Histograms of the membership degrees by class for the MFCM method.

Table 14
Average and standard deviation of the features by class for the Pima Indians diabetes data set.

Feature	Class 1 (size = 500)		Class 2 (size = 268)	
	Average	SD	Average	SD
NumTPreg	3.2980	3.0172	4.8657	3.7412
GluConc	109.9800	26.1412	141.2575	31.9396
DiasPre	68.1840	18.0631	70.8246	21.4918
TriSkinFoldThi	19.6640	14.8899	22.1642	17.6797
Insulin	68.7920	98.8653	100.3358	138.6891
Index	30.3042	7.6899	35.1425	7.2630
DiaPeFunc	0.4297	0.2991	0.5505	0.3724
Age	31.1900	11.6677	37.0672	10.9683

indices were 0.722, 0.724, 0.729 and 0.716, whereas for the proposed MFCM one was 0.773. Thus, the fuzzy clustering method by feature outperforms slightly the other fuzzy clustering methods for this data set. Figs. 21–25 display histograms of the class membership degrees for each method. For this data set, the histograms are similar since the methods have similar FR indices.

4.2.5. Pima Indians diabetes data set

This data set investigates females at least 21 years old of Pima Indian heritage patients if those show signs of diabetes according to World Health Organization criteria. The size of the data set is 768 grouped into two classes: patients who present signs of diabetes and patients who do not. To evaluate these signals, eight numerical features is analyzed: number of times pregnant (NumTPreg), plasma glucose concentration a 2 hours in an oral glucose tolerance test (GluConc), diastolic blood pressure (DiasPre), triceps skin fold thickness (TriSkinFoldThi), 2-h serum insulin (insulin), body mass index (index), diabetes pedigree function (DiaPeFunc) and age. Table 14 shows the average and standard deviation of the features for each class. In this data set, the features have very dissimilar dispersions within each class.

The clustering methods were applied to this data set and FR indices were 0.788, 0.576, 0.565 and 0.448, for AFCM, AFCM-C, FCM-S and GK-F methods, respectively, whereas for the proposed MFCM one was 0.812. Again, the fuzzy clustering method introduced in this paper is the best option. Figs. 26–30 show histograms of the class membership degrees for each method.

For this data set, the methods based on adaptive (AFCM-C and AFCM-S) and Mahalanobis (GK-F) distance had the worst performance in terms of FR index. This result is expected since we can observe in the histograms an atypical behavior for the memberships: high frequencies close to 0.5 for both clusters. The FCM and MFCM methods have similar histograms. But the MFCM method highlights high frequencies close to 0 for both clusters while the FCM method does not highlight. For this reason, the MFCM method is better than the FCM one.

5. Conclusion

In this work, we presented a new fuzzy clustering approach to handle the degree of belonging based on the information of each feature. The traditional fuzzy *c*-means clustering method calculates memberships considering that all the features have equal importance and this way of obtaining memberships may not be the most appropriate for applications whose features have dissimilar dispersions. Thus, the clustering method introduced in this paper assumes that, given a cluster, there are membership degrees for an individual that are different from one feature to another.

In order to show the usefulness of the proposed method, an experimental evaluation has been carried out. The accuracy of the results furnished by the clustering methods was assessed by the Fuzzy Rand Index. This index is able to evaluate a fuzzy partition of a data set better than the classical corrected Rand Index. Four synthetic data configurations showing different covariance matrices for the features and an application with five real data sets from the UCI Machine Learning Repository were considered in this evaluation.

For synthetic data sets, the Fuzzy Rand Index was estimated by the Monte Carlo method with 100 replications of each data set and Student's t-tests were used to compare the performance of the methods. It is important to highlight that the proposed method is more suitable for data sets with uncorrelated features. A comparative study based on the centers and dispersions of the clusters furnished by the Gustafson–Kessel and proposed methods and the centers and dispersions of the *a priori* classes was also performed. In this study, the proposed method surpassed the Gustafson–Kessel one based on diagonal matrices.

The results according to the Fuzzy Rand index showed that, regarding features with dissimilar dispersions, the proposed clustering method in many important cases was superior to the classical fuzzy *c*-means, the fuzzy methods based on adaptive distances and the Gustafson–Kessel method. The Fuzzy Rand Index indicated that the fuzzy partition obtained by the proposed method was more similar than the original partition. The superiority of the proposed method in comparison to other fuzzy clustering methods considered in this paper can be explained by the computation of the membership degrees obtained taking into account the information of the features. In this case, the proposed method supposes that the features are not considered equally important and they have dissimilar dispersions.

References

[1] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM Computing Surveys 31 (3) (1999) 264–323.
[2] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiely and Sons, New York, 2001.
[3] G. Ma Gan, C.J. Wu, Data Clustering Theory, Algorithms, and Applications, Siam, 2007.
[4] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[5] R. Xu, D.I.I. Wunsch, Survey of clustering algorithms, IEEE Transactions on Neural Networks (2005) 645–678.
[6] L.A. Zadeh, Fuzzy Sets, Information and Control 8 (3) (1965) 338–353.
[7] E.H. Ruspini, A new approach to clustering, Information and Control 15 (1969) 22–32.
[8] M.-S. Yang, A Survey of Fuzzy Clustering, Mathematical and Computer Modelling 18 (11) (1993) 1–16.
[9] C.-W. Chen, Stability conditions of fuzzy systems and its application to structural and mechanical systems, Advances in Engineering Software 37 (2006) 624–629.
[10] S. Termini, On Some "family resemblances" of fuzzy set theory and human sciences, Soft Computing in Humanities and Social Sciences 273 (2012) 39–54.
[11] B.K. Wong, V.S. Lai, A survey of the application of fuzzy set theory in production and operations management: 1998–2009, International Journal Production Economics 129 (2011) 157–168.
[12] R.-J. Kuo, Y.-H. Wu, T.-S. Hsu, Integration of fuzzy set theory and TOPSIS into HFMEA to improve outpatient service for elderly patients in Taiwan, Journal of the Chinese Medical Association 75 (2012) 341–348.
[13] H. Izakian, A. Abraham, Fuzzy, c-means and fuzzy swarm for fuzzy clustering problem, Expert Systems with Applications 38 (3) (2011) 1835–1838.

[14] R.J. Hathaway, J.C. Bezdek, Y. Hu, Generalized fuzzy *c*-means clustering strategies using Lp norm distance, IEEE Transactions on Fuzzy Systems 8 (5) (2000) 572–576.
[15] K. Jajuga, L1-norm based fuzzy clustering, Fuzzy Sets and Systems 39 (2001) 63–83.
[16] P.J.F. Groenen, K. Jajuga, Fuzzy clustering with squared Minkowsky distances, Fuzzy Sets and Systems 120 (2001) 227–237.
[17] C.-H. Oh, K. Honda, H. Ichihashi, Fuzzy clustering for categorical multivariate data, in: Joint Ninth IFSA World Congress and Twentieth, NAFIPS International Conf., vol. 4, 2001, pp. 2154–2159.
[18] K.-L. Wu, M.-S. Yang, Alternative c-means clustering algorithms, Pattern Recognition 35 (2002) 2267–2278.
[19] D.-Q. Zhang, S.-C. Chen, Kernel-Based Fuzzy and Possibilistic c-Means Clustering, in: Proc. ICANN, Istanbul, Turkey, 2003, pp. 122–125.
[20] K. Kummamuru, A. Dhawale, R. Krishnapuram, Fuzzy co-clustering of documents and keywords, in: Twelfth IEEE International Conf. on Fuzzy Systems, 2, 2003, pp. 772–777.
[21] N.R. Pal, K. Pal, J.M. Keller, J.C. Bezdek, A possibilistic fuzzy c-means clustering algorithm, IEEE Transactions on Fuzzy Systems 13 (4) (2005) 517–530.
[22] F.A.T. De Carvalho, C.P. Tenorio, N.L.C. Junior, Partitional fuzzy clustering methods based on adaptive quadratic distances, Fuzzy Sets and Systems 157 (2006) 2833–2857.
[23] J. Liu, Meizhi Xu, Kernelized fuzzy attribute c-means clustering algorithm, Fuzzy Sets and Systems 159 (2008) 2428–2445.
[24] C.-W. Chen, The stability of an oceanic structure with T–S fuzzy models, Mathematics and Computers in Simulation 80 (2) (2009) 402–426.
[25] C.Y. Chen, J.W. Lin, W.I. Lee, C.W. Chen, Fuzzy control for an oceanic structure: a case study in time-delay TLP system, Journal of Vibration and Control 16 (1) (2010) 147–160.
[26] C.W. Chen, Modeling and fuzzy PDC control and its application to an oscillatory TLP structure, Mathematical Problems in Engineering (2010), an open access.
[27] K. Yeh, C.W. Chen, Stability analysis of interconnected fuzzy systems using the fuzzy Lyapunov method, Mathematical Problems in Engineering (2010), an open access journal.
[28] C.W. Chen, Application of fuzzy-model-based control to nonlinear structural systems with time delay: an LMI method, Journal of Vibration and Control 16 (2010) 1651–1672.
[29] C. Tang, S. Wang, W. Xu, New fuzzy c-means clustering model based on the data weighted approach, Data and Knowledge Engineering 69 (9) (2010) 881–900.
[30] J.-P. Mei, L. Chen, Fuzzy clustering with weighted medoids for relational data, Pattern Recognition 43 (5) (2010) 1964–1974.
[31] J.-P. Mei, L. Chen, Fuzzy relational clustering around methods: a unified view, Fuzzy Sets and Systems 183 (2011) 44–56.
[32] J. Yu, Q. Cheng, H. Huang, Analysis of the weighting exponent in the FCM, IEEE Transactions on Systems, Man, and Cybernetics – Part B 34 (2004) 634–639.
[33] K.-L. Wu, Analysis of parameter selections for fuzzy c-means, Pattern Recognition 45 (2012) 407–415.
[34] D.E. Gustafson, W.C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, in: Proc. IEEE CDC, 1979, pp. 761–766.
[35] C.L. Blake, C.J. Merz, UCI repository of machine learning databases, University of California, Department of Information and Computer Science, Irvine, CA (accessed August, 2011).
[36] E. Egrioglu, C.H. Aladag, U. Yolcu, V.R. Uslu, N.A. Erilli, Fuzzy time series forecasting method based on Gustafson–Kessel fuzzy clustering, Expert Systems with Applications 38 (2011) 10355–10357.
[37] D. Dovan, I. Škrjanc, Recursive clustering based on a Gustafson–Kessel algorithm, Evolving Systems 2 (1) (2011) 15–24.
[38] L. Hubert, P. Arabie, Comparing partitions, Journal of Classification 2 (1985) 193–218.
[39] R.J.G.B. Campello, A fuzzy extension of the Rand index and other related indexes for clustering and classification assessment, Pattern Recognition Letters 28 (7) (2007) 833–841.