International Journal of Uncertainty,
Fuzziness and Knowledge-Based Systems
Vol. 26, No. 6 (2018) 893–916

DOI: 10.1142/S021848851850040X

A Generalized Multivariate Approach for Possibilistic Fuzzy C-Means Clustering

Bruno Almeida Pimentel* and Renata M. C. R. de Souza†

Universidade Federal de Pernambuco (UFPE), Centro de Informática (CIn),
Av. Jornalista Anibal Fernandes, s/n - Cidade Universitária 50.740-560,
Recife – PE, Brazil
**bap@cin.ufpe.br*
†rmcrs@cin.ufpe.br

Received 17 April 2017
Revised 24 November 2017

Fuzzy c-Means (FCM) and Possibilistic c-Means (PCM) are the most popular algorithms of the fuzzy and possibilistic clustering approaches, respectively. A hybridization of these methods, called Possibilistic Fuzzy c-Means (PFCM), solves noise sensitivity defect of FCM and overcomes the coincident clusters problem of PCM. Although PFCM have shown good performance in cluster detection, it does not consider that different variables can produce different membership and possibility degrees and this can improve the clustering quality as it has been performed with the Multivariate Fuzzy c-Means (MFCM). Here, this work presents a generalized multivariate approach for possibilistic fuzzy c-means clustering. This approach gives a general form for the clustering criterion of the possibilistic fuzzy clustering with membership and possibility degrees different by cluster and variable and a weighted squared Euclidean distance in order to take into account the shape of clusters. Six multivariate clustering models (special cases) can be derivative from this general form and their properties are presented. Experiments with real and synthetic data sets validate the usefulness of the approach introduced in this paper using the special cases.

Keywords: Fuzzy clustering; possibilistic clustering; hybridization; multivariate membership.

1. Introduction

Pattern Recognition may be considered, in a broader sense, as a problem of searching for structure in data[1] dividing the space into regions of categories or clusters.[2] There is a wide variety of methods, which include clustering methods, whose task is to find a hidden pattern as well as to discover the pattern in the cluster.[3,4] Nowadays, clustering methods are widely used in several domain of application in order to automatically understanding, processing and summarizing data.[5–7] They can be classified into two main categories: hierarchical and partitioning.

Hierarchal approach consider that groups can have sub-groups found by a divisive way (if it starts with a group containing all patterns and splits it until singleton groups) or agglomerative way (if it executes the inverse process).[8] On the other hand, partitioning methods consider that there is no sub-groups and directly divide data patterns into some fixed number of groups.[9,10]

Partitioning clustering methods may be classified into two approaches: hard and non-hard.[9] The former one considers that a pattern from data belongs to only one cluster. On the other hand, the latter one takes into account the probability or possibility of a pattern belong to a cluster. Fuzzy c-Means (FCM) proposed by Bezdek (1973)[11] is the most popular algorithm of the probabilistic approach,[12] whereas the most well-known method of possibilistic approach is the Possibilistic c-Means (PCM) proposed by Krishnapuram and Keller (1993).[13] In order to solve the noise sensitivity defect of FCM and overcome the coincident clusters problem of PCM, Pal *et al.* (2005)[14] proposed a hybridization of these methods called Possibilistic Fuzzy c-Means (PFCM). Although FCM and its versions have shown good performance in cluster detection, they do not consider that different variables may produce different membership degrees.

In order to consider multivariate membership degrees, Pimentel and Souza (2013)[17] proposed the Multivariate Fuzzy c-Means (MFCM) method where membership degrees of both clusters and variables are computed. To treat with interval data, two multivariate methods are also presented in Refs. 18 and 19, respectively. The first one generalizes the MFCM with weights for the multivariate membership degrees and the second one generalizes the PCM with multivariate possibilistic degrees. More recently, the authors introduced a version of the MFCM with weights for distance and multivariate membership degrees[20] and Himmelspach and Conrad (2016)[21] proposed a combination of MFCM and PFCM, however the possibility degrees are not multivariate.

Since the PFCM hybrid version aims to solve the noise sensitivity defect and overcome the coincident clusters problem, this paper presents a generalized multivariate approach for PFCM. The multivariate possibility can also allow the clustering method to identify which variable is more relevant to indicate if a given object is an outlier/noise and to improve the clustering quality. In this context, the proposed approach gives a general form to the clustering criterion of the possibilistic fuzzy clustering and six multivariate clustering models (special cases) can be derivative. The general model assumes that, for a given object, both membership and possibility degrees are computed for each cluster and variable. Moreover, it uses a weighted distance in order to take into account the shape of clusters.

The remainder of this paper is organized as follows. Section 2 introduces the multivariate possibilistic fuzzy c-means clustering algorithm. It is divided into two parts: the former one contains definitions to better understand the proposed method and the latter one presents the algorithm and proof of its convergence. Section 3 presents special cases derivative from proposed generalized method. In order to evaluate the performance of method presented in this paper, Section 4 shows experiments with syntectic and real datasets. Finally, in Section 5, concluding remarks are given.

2. Generalized Multivariate Possibilistic Fuzzy C-Means Clustering Method

In this section, we introduce a partitional clustering method that generalizes the possibilistic fuzzy c-means clustering method by using multivariate membership degrees, multivariate possibilistic degrees and a weighted distance in order to take into account the variability of each variable and cluster. The methodology is presented in two parts. The former one shows definitions and notation of the method and the latter one presents the algorithm and discusses time and space complexities and the proof of convergence.

2.1. *Proposed method*

Let $\Omega = \{1, \ldots, k, \ldots, n\}$ be a set of n objects indexed by k. Each object k is represented by a vector of quantitative variables $\mathbf{x}_k = (x_{k1}, \ldots, x_{kj}, \ldots, x_{kp})$ described by p variables indexed by j where $x_{kj} \in \Re$.

Consider $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_c\}$ a set of c prototypes related to c groups of partition, where each prototype of a group C_i $(i = 1, \ldots, c)$ is also represented by a vector of quantitative variables $\mathbf{y}_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{ip})$, where $y_{ij} \in \Re$. Let $\mathbb{Y}^c = \underbrace{\mathbb{Y} \times \ldots \times \mathbb{Y}}_{c}$ and $\mathbb{Y} = \underbrace{\Re \times \ldots \times \Re}_{p} = \Re^p$, where $\mathbb{Y}$ is the space of presentation of prototypes such that $\mathbf{y}_i \in \mathbb{Y}$ and $\mathbf{Y} \in \mathbb{Y}^c$.

Let $\mathbf{U} = [\mathbf{u}_k]$ be a matrix of matrices of multivariate membership degrees, where for each object k from Ω there exist a matrix $c \times p$ of multivariate membership degrees $\mathbf{u}_k = [u_{ijk}]$ $(k = 1, \ldots, n)$ where u_{ijk} is the membership degree of object k for the group C_i $(i = 1, \ldots, c)$ regarding the variable j $(j = 1, \ldots, p)$. Consider $\mathbb{U}^n = \underbrace{\mathbb{U} \times \ldots \times \mathbb{U}}_{n}$, $\mathbb{U} = \underbrace{\mathbb{V} \times \ldots \times \mathbb{V}}_{c}$ and $\mathbb{V} = \underbrace{[0,1] \times \ldots \times [0,1]}_{p}$. Thus, $\mathbb{U}$ is the space of representation of the matrix of multivariate membership degree $\mathbf{u}_k$ such that $\mathbf{u}_k \in \mathbb{U}$ and $\mathbf{U} \in \mathbb{U}^n$.

Let $\mathbf{T} = [\mathbf{t}_k]$ be a matrix of matrices of multivariate possibility degrees, where for a given object k $(k = 1, \ldots, n)$, $\mathbf{t}_k = [t_{ijk}]$ is a matrix of multivariate possibility degree $c \times p$ and $t_{ijk} \in [0,1]$ is the possibility degree of object k for group C_i $(i = 1, \ldots, c)$ and variable j $(j = 1, \ldots, p)$. Consider $\mathbb{T}^n = \underbrace{\mathbb{T} \times \ldots \times \mathbb{T}}_{n}$, $\mathbb{T} = \underbrace{\mathbb{X} \times \ldots \times \mathbb{X}}_{c}$ and $\mathbb{X} = \underbrace{[0,1] \times \ldots \times [0,1]}_{p}$, where $\mathbb{T}$ is the space of representation of the matrix of multivariate possibility degree $\mathbf{t}_k$ such that $\mathbf{t}_k \in \mathbb{T}$ and $\mathbf{T} \in \mathbb{T}^n$.

Let $\mathbf{\Lambda} = [\boldsymbol{\lambda}_i]$ be a matrix of weight vectors, where $\boldsymbol{\lambda}_i = (\lambda_{i1}, \ldots, \lambda_{ij}, \ldots, \lambda_{ip})$ and $\lambda_{ij} \in \Re^+_*$ is the weight regarding the cluster C_i $(i = 1, \ldots, c)$ and variable j

($j = 1, \ldots, p$). Consider $\mathbb{L}^c = \underbrace{\mathbb{L} \times \ldots \times \mathbb{L}}_{c}$, $\mathbb{L} = \underbrace{\Re^+_* \times \ldots \times \Re^+_*}_{p} = (\Re^+_*)^p$, where $\Re^+_*$ means the set of positive real numbers non-null, and $\mathbb{L}$ is the representation space of weight vector $\boldsymbol{\lambda}_i$ such that $\boldsymbol{\lambda}_i \in \mathbb{L}$ and $\boldsymbol{\Lambda} \in \mathbb{L}^c$. The goal of weights λ_{ij} is to take into account the variability of each variable and cluster. Let $\Delta = [\delta_{ik}]$ be a matrix $c \times n$ of standard membership degree where δ_{ik} is the membership degree of object k for group C_i.

The objective of method Multivariate Possibilistic Fuzzy c-Means (here abbreviated as MPFCM-D) is to find a matrix of prototypes $\mathbf{Y}^*$, a matrix of multivariate membership degree $\mathbf{U}^*$, a matrix of multivariate possibility degree $\mathbf{T}^*$ and a matrix of weights $\boldsymbol{\Lambda}^*$ such that:

$$J(\mathbf{Y}^*, \mathbf{U}^*, \mathbf{T}^*, \boldsymbol{\Lambda}^*) = \min\{J(\mathbf{Y}, \mathbf{U}, \mathbf{T}, \boldsymbol{\Lambda}) : \mathbf{Y} \in \mathbb{Y}^c, \mathbf{U} \in \mathbb{U}^n, \mathbf{T} \in \mathbb{T}^n, \boldsymbol{\Lambda} \in \mathbb{L}^c\}\,. \tag{1}$$

The goal of the algorithm is to minimize the objective function is given as follows:

$$J(\mathbf{Y}, \mathbf{U}, \mathbf{T}, \boldsymbol{\Lambda}) = \sum_{i=1}^{c}\sum_{j=1}^{p}\sum_{k=1}^{n} \phi_{ijk}\,, \tag{2}$$

where ϕ_{ijk} is the dissimilarity between the object k and prototype of group C_i according to variable j given as:

$$\phi_{ijk} = \lambda_{ij}\{[a(u_{ijk})^m + b(t_{ijk})^\eta]d_{ijk} + \pi_{ij}(1 - t_{ijk})^\eta\}\,, \tag{3}$$

and a and b are constants that define the relative importance of membership and possibility degree in the objective function.[14] These constants are positive and non-null, that is, $a, b \in \Re^+_*$. When b is larger than a, possibility degrees have more importance in prototypes computing. Similar reasoning can be applied to the parameters m and η. Distance d_{ijk} measures the dissimilarity between the object k and prototype of group C_i according to variable j and it is given as follows:

$$d_{ijk} = (x_{kj} - y_{ij})^2\,. \tag{4}$$

The π_{ij} value is important to balance the two terms of the equation, in such way that π_{ij} has same order of magnitude of d_{ijk}. Krishnapuram and Keller (1993)[13] proposed to compute π_i in the PCM using the average of distance object-prototype. Pal *et al.* (2005)[14] also used the same expression in the PFCM. Here, the main difference between π_{ij} of this proposed method from π_i of PFCM is that π_{ij} value is computed for each variable j, taking into account the inherent information of each one. It is given as:

$$\pi_{ij} = K\frac{\sum_{k=1}^{n}(u_{ijk})^m d_{ijk}}{\sum_{k=1}^{n}(u_{ijk})^m}\,, \tag{5}$$

where K is used as 1. In the algorithm, π_{ij} is calculated only once after the parameters initialization to avoid instabilities in the convergence.

Prototypes in MPFCM-D method takes into account the values a, b, the multivariate membership degree u_{ijk} and the multivariate possibility degree t_{ijk}. Therefore, these values exert a great influence on the representation of groups and, thereafter, in the clustering quality.

Proposition 1. *Fixing the membership degree u_{ijk}, the possibility degree t_{ijk} and the weight λ_{ij}, the prototype y_{ij} which minimizes the criterion J is updated using the following equation:*

$$y_{ij} = \frac{\sum_{k=1}^{n} \left[a(u_{ijk})^m + b(t_{ijk})^\eta\right] x_{kj}}{\sum_{k=1}^{n} \left[a(u_{ijk})^m + b(t_{ijk})^\eta\right]} . \tag{6}$$

Proof 1. The proof is given in Appendix A. □

Similarly to Multivariate Fuzzy c-Means method and its weighted versions, the membership degree is under the restriction $\sum_{i=1}^{c} \sum_{j=1}^{p} u_{ijk} = 1$, for all $k \in \Omega$.

Proposition 2. *Fixing the prototype y_{ij}, the possibility degree t_{ijk} and the weight λ_{ij}, the membership degree u_{ijk} under the restriction $\sum_{i=1}^{c} \sum_{j=1}^{p} u_{ijk} = 1$ which minimizes the criterion J is updated using the following equation:*

$$u_{ijk} = \left[\sum_{r=1}^{c} \sum_{s=1}^{p} \left(\frac{\lambda_{ij} d_{ijk}}{\lambda_{rs} d_{rsk}}\right)^{\frac{1}{m-1}}\right]^{-1} . \tag{7}$$

Proof 2. The proof is given in Appendix B. □

The possibility degree is different from membership degree by the fact that the first one is not under the sum restriction, that is, give an object k the sum for all clusters and variables is not necessarily equal to 1. This flexibility allows methods based on possibilistic approach to be more robust to noisy or aberrant data. The further away an object is the prototype, the lower the possibility degree, making this object less influential in computation of prototypes and, thus, increasing the clustering quality for datasets with noise/outliers. In this method, the multivariate possibility degree t_{ijk} take into account the weight λ_{ij}, therefore, given an object k, t_{ijk} changes according to the shape of cluster C_i.

Proposition 3. *Fixing the prototype y_{ij}, the membership degree u_{ijk} and the weight λ_{ij}, the possibility degree t_{ijk} which minimizes the criterion J is updated using the following equation:*

$$t_{ijk} = \left[1 + \left(b\frac{\lambda_{ij} d_{ijk}}{\pi_{ij}}\right)^{\frac{1}{(\eta-1)}}\right]^{-1} . \tag{8}$$

Proof 3. The proof is given in Appendix C. □

The goal of weight λ_{ij} is to take into account the variability of each variable and cluster, therefore this method is able to find clusters of different shapes and sizes. It is under restriction $\prod_{j=1}^{p} \lambda_{ij} = 1$. It is known that the use of the sum of weights as constraint is more suitable in a specific case: elongated clusters.[15,16] Here, the goal is to allow the method to be able to find both elliptical and spherical clusters. Therefore, we adopted the constraint based on the product of weights.

Proposition 4. *Fixing the prototype y_{ij}, the membership degree u_{ijk} and possibility degree t_{ijk}, the weight λ_{ij} under the restriction $\prod_{j=1}^{p} \lambda_{ij} = 1$ which minimizes the criterion J is updated using the following equation:*

$$\lambda_{ij} = \frac{\{\prod_{h=1}^{p} [\sum_{k=1}^{n} [a(u_{ihk})^m + b(t_{ihk})^\eta]\, d_{ihk}]\}^{\frac{1}{p}}}{\sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta]\, d_{ijk}} \,. \tag{9}$$

Proof 4. The proof is given in Appendix D. □

According to Eq. (9), the weight has larger values when the variability in the cluster C_i regarding variable j is lower. In other words, the more compact a cluster is, the larger weight is (according to a given variable). The objective function looks for compact and well separated clusters. In this clustering method, the goal is to use weights to measure the compactness of each cluster regarding variables. In this sense, weights contribute to improvement of the objective function through the measuring of dispersion of clusters in the partition, allowing the method recognize clusters of different shapes. In the MPFCM-D method, the matrix of memberships by cluster $\Delta = [\delta_{ik}]$ is obtained as follows:

$$\delta_{ik} = \sum_{j=1}^{p} u_{ijk} \,, \tag{10}$$

After the convergence of algorithm, a crisp cluster boundaries may be obtained by a hard partition $P = \{C_1, \ldots, C_i, \ldots, C_c\}$ where $\bigcup_{i=1}^{c} C_i = \Omega$ and $\bigcap_{i=1}^{c} C_i = \emptyset$. Each object k is allocated to cluster C_{i*} such that:

$$i* = \underbrace{\mathrm{argmax}}_{1 \le i \le c} \delta_{ik} \,. \tag{11}$$

2.2. *Algorithm, space and time complexities and convergence*

The algorithm of Multivariate Possibilistic Fuzzy c-Means method is described as follows:

Algorithm 1 MPFCM-D (Ω, c)

Require: Data set Ω and the number of clusters c;

1: Let $\varepsilon > 0$. Fix c, $2 \leq c \leq n$; fix m, $1 < m < \infty$; fix maximum number of iterations T; Initialize randomly u_{ijk} and t_{ijk} ($i = 1, \ldots, c$; $j = 1, \ldots, p$ and $k = 1, \ldots, n$) of object k belonging to cluster C_i for variable j such that $u_{ijk} \in [0, 1]$, $0 < \sum_{k=1}^{n} u_{ijk} < n$ and $\sum_{i=1}^{c} \sum_{j=1}^{p} u_{ijk} = 1$, for all $k \in \Omega$. Fix $\lambda_{ij} = 1$, ($i = 1, \ldots, c$; $j = 1, \ldots, p$). Initialize π_{ij} using Eq. (5).

2: $t \leftarrow 0$;

3: $J(t) \leftarrow 0$;

4: $J(t+1) \leftarrow \sum_{i=1}^{c} \sum_{j=1}^{p} \lambda_{ij} \sum_{k=1}^{n} \left\{ \left[a(u_{ijk})^m + b(t_{ijk})^\eta \right] d_{ijk} + \pi_{ij}(1 - t_{ijk})^\eta \right\}$;

5: **while** $|J(t) - J(t+1)| > \varepsilon$ **and** $t < T$ **do**

6: **Update matrix of prototype Y**: Fixing membership degree u_{ijk}, possibility degree t_{ijk} and weight λ_{ij} ($i = 1, \ldots, c$; $j = 1, \ldots, p$ and $k = 1, \ldots, n$), update the prototype y_{ij} using Eq. (6);

7: **Update matrix of multivariate membership degree U**: Fixing possibility degree t_{ijk}, prototype y_{ij} and weight λ_{ij} ($i = 1, \ldots, c$; $j = 1, \ldots, p$ and $k = 1, \ldots, n$), update the membership degree u_{ijk} using Eq. (7);

8: **Update matrix of multivariate possibility degree T**: Fixing membership degree u_{ijk}, prototype y_{ij} and weight λ_{ij} ($i = 1, \ldots, c$; $j = 1, \ldots, p$ and $k = 1, \ldots, n$), update the possibility degree t_{ijk} using Eq. (8);

9: **Update matrix of weight $\mathbf{\Lambda}$**: Fixing membership degree u_{ijk}, possibility degree t_{ijk} and prototype y_{ij} ($i = 1, \ldots, c$; $j = 1, \ldots, p$ and $k = 1, \ldots, n$), update the weight λ_{ij} using Eq. (9);

10: $J(t) \leftarrow J(t+1)$;

11: $J(t+1) \leftarrow \sum_{i=1}^{c} \sum_{j=1}^{p} \lambda_{ij} \sum_{k=1}^{n} \left\{ \left[a(u_{ijk})^m + b(t_{ijk})^\eta \right] d_{ijk} + \pi_{ij}(1 - t_{ijk})^\eta \right\}$;

12: $t \leftarrow t + 1$;

13: **end while**

14: **Compute matrix of memberships by cluster** Δ: Aggregate the multivariate memberships using Eq. (10).

return Matrices **Y**, **U**, **T** and $\mathbf{\Lambda}$.

The algorithm can be understood according its different steps as described below. In the first step, the algorithm sets the parameters ε, m and T. In this step, the weights and memberships are also initialized and it is necessary cpn operations for the matrices **U**, **T** and $\mathbf{\Lambda}$, thus the time complexity for initialization step is $O(cpn)$. Until the convergence, the algorithm performs iteratively three main steps. In the first one, to update the prototypes, the algorithm takes pn operations for each one of c prototypes, thus the time complexity of this step is $O(cpn)$. In the second step, memberships degrees are updated so that there are cpn membership values where each one takes cp calculations, then the time complexity of this step is $O(c^2p^2n)$. In the third step, cpn possibility values are calculated where each one takes $O(1)$, thus the complexity time of this step is $O(cpn)$. In the fourth step, weights for a given cluster and variable are calculated. It is necessary pn operations for each weight and there are cp weights, thus this step takes $O(cp^2n)$. To compute the objective function value used in the stopping criterion, the processing time is $O(cpn)$. Therefore, the final time complexity of the algorithm is $O(c^2p^2n)$.

Regarding the space complexity, the matrix of prototype $\mathbf{Y}$ requires p space for each one of c prototypes, thus it is necessary $O(cp)$ space to store this matrix. The matrix $\mathbf{U}$ is formed by n matrices of multivariate membership degrees where each one requires cp space, thus the final space complexity to store $\mathbf{U}$ is $O(cpn)$. A similar reasoning can be made for the matrix of possibility degrees $\mathbf{T}$. The matrix of weights $\mathbf{\Lambda}$ requires p space for each one of c clusters, thus it is necessary $O(cp)$ space to store this matrix. Therefore, the space complexity of this algorithm is $O(cpn)$.

Following Diday and Simon (1976),[22] properties of convergence of this kind of algorithm may be studied from two series: $\nu_t = (\mathbf{Y}^t, \mathbf{U}^t, \mathbf{T}^t, \mathbf{\Lambda}^t) \in \mathbb{Y}^c \times \mathbb{U}^n \times \mathbb{T}^n \times \mathbb{L}^c$ and $\omega_t = J(\nu_t) = J(\mathbf{Y}^t, \mathbf{U}^t, \mathbf{T}^t, \mathbf{\Lambda}^t)$, $t = 0, 1, 2, \ldots, T$. The algorithm starts from initial series $\nu_0 = (\mathbf{Y}^0, \mathbf{U}^0, \mathbf{T}^0, \mathbf{\Lambda}^0)$ and computes the different terms of the series ν_t until the convergence when the criterion J achieves a stationary value.

Proposition 5. *The series $\omega_t = J(\nu_t)$ decreases at each iteration and converges.*

Proof 5. The proof is given in Appendix E. □

Proposition 6. *The series $\nu_t = (\boldsymbol{Y}^t, \boldsymbol{U}^t, \boldsymbol{T}^t, \mathbf{\Lambda}^t)$ converges.*

Proof 6. The proof is given in Appendix F. □

3. Special Cases of the Clustering Criterion

According to the criterion function described by Eq. (2), we derive six models based on different conditions defined for the weight, membership and possibility degrees and parameters a and b.

3.1. *Case 1*

The algorithm assumes the weights $\lambda_{ij} = 1$ for $i = 1, \ldots, c;\ j = 1, \ldots, p$ and the parameter $a = 0$. This model computes only multivariate possibility degrees and it is a version of the MPCM one proposed by Pimentel and Souza (2014b)[19] for continuous data. The dissimilarity ϕ_{ijk} is rewritten by the following equation:

$$\phi_{ijk} = b(t_{ijk})^\eta d_{ijk} + \pi_{ij}(1 - t_{ijk})^\eta \,. \tag{12}$$

3.2. *Case 2*

Similar to previous case, with $a = 0$ but the weights λ_{ij} for $i = 1, \ldots, c; j = 1 \ldots, p$ is updated for each iteration of the algorithm such that $\lambda_{ij} \neq 1$ $(i = 1, \ldots, c; j = 1, \ldots, p)$. Therefore, this model represents the MPCM method with weighted distance. The dissimilarity ϕ_{ijk} is given by the following equation:

$$\phi_{ijk} = \lambda_{ij} \left\{(t_{ijk})^\eta d_{ijk} + \pi_{ij}(1 - t_{ijk})^\eta\right\}, \tag{13}$$

where the weight λ_{ij} is given as:

$$\lambda_{ij} = \frac{\{\prod_{h=1}^{p} [\sum_{k=1}^{n} (t_{ihk})^{\eta} d_{ihk}]\}^{\frac{1}{p}}}{\sum_{k=1}^{n} (t_{ijk})^{\eta} d_{ijk}} . \tag{14}$$

3.3. *Case 3*

Here, the algorithm assumes the weights $\lambda_{ij} = 1$ $(i = 1, \ldots, c; j = 1, \ldots, p)$, univariate possibility degrees and multivariate membership degrees. It is the PMFCM method proposed by Himmelspach and Conrad (2016).[21] In this case, ϕ_{ijk} is rewritten by the following equation:

$$\phi_{ijk} = [a(u_{ijk})^{m} + b(t_{ik})^{\eta}]\, d_{ijk} + \pi_i (1 - t_{ik})^{\eta} , \tag{15}$$

where π_i is calculated according to PCM method proposed by Krishnapuram and Keller (1993):[13]

$$\pi_i = K \frac{\sum_{k=1}^{n} (u_{ik})^{m} d_{ik}}{\sum_{k=1}^{n} (u_{ik})^{m}} . \tag{16}$$

u_{ik} is the univariate membership degree of the PFCM method[14] calculated as follows:

$$u_{ik} = \left[\sum_{r=1}^{c} \left(\frac{d_{ik}}{d_{rk}} \right)^{\frac{1}{m-1}} \right]^{-1} , \tag{17}$$

and t_{ik} is univariate possibility degree according to PFCM method.[14] It is updated using the following expression:

$$t_{ik} = \left[1 + \left(b \frac{d_{ik}}{\pi_i} \right)^{\frac{1}{(\eta-1)}} \right]^{-1} . \tag{18}$$

Here, the dissimilarity between object k and prototype of cluster C_i is given as:

$$d_{ik} = \sum_{j=1}^{p} d_{ijk} , \tag{19}$$

where d_{ijk} is defined as in Eq. (4).

3.4. *Case 4*

This model is a version of the Case 3 in which the weights of the distance are updated for each iteration of the algorithm, that is, there is at least one weight $\lambda_{ij} \neq 1$ $(i = 1, \ldots, c; j = 1, \ldots, p)$. The idea is to allows the algorithm be able to identify clusters of different shapes and sizes. Therefore, this case is an extension of the PMFCM method proposed by Himmelspach and Conrad (2016)[21] with weighted distance. The dissimilarity ϕ_{ijk} is rewritten by the following equation:

$$\phi_{ijk} = \lambda_{ij} \left\{ [a(u_{ijk})^{m} + b(t_{ik})^{\eta}]\, d_{ijk} + \pi_i (1 - t_{ik})^{\eta} \right\} , \tag{20}$$

where the weight λ_{ij} $(i = 1, \ldots, c; j = 1, \ldots, p)$ is given as:

$$\lambda_{ij} = \frac{\{\prod_{h=1}^{p} [\sum_{k=1}^{n} [a(u_{ihk})^m + b(t_{ik})^\eta] \, d_{ihk}]\}^{\frac{1}{p}}}{\sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ik})^\eta] \, d_{ijk}} . \tag{21}$$

3.5. *Case 5*

Here, the algorithm computes both multivariate membership and possibility degrees but the weights λ_{ij} $(i = 1, \ldots, c; j = 1, \ldots, p)$ are fixed. In addition, the constants a and b are positive and non-null $(a, b \in \Re_*^+)$, that is, the algorithm takes into account both multivariate membership and possibility degrees to compute the objective function. This method is called Multivariate Possibilistic Fuzzy C-Means method (here abbreviated by MPFCM). The dissimilarity ϕ_{ijk} is given by the following equation:

$$\phi_{ij} = [a(u_{ijk})^m + b(t_{ijk})^\eta] \, d_{ijk} + \pi_{ij}(1 - t_{ijk})^\eta . \tag{22}$$

3.6. *Case 6*

The algorithm is able to identify clusters of different shapes and sizes since it uses weighted distance such that $\lambda_{ij} \neq 1$ $(i = 1, \ldots, c; j = 1, \ldots, p)$. Weights are updated in each iteration of algorithm minimizing the objective function. Moreover, it computes both multivariate membership and possibility degrees as well as the Case 5. This method is called Multivariate Possibilistic Fuzzy C-Means with weighted distance method (here abbreviated by MPFCM-D). The dissimilarity ϕ_{ijk} is given by the Eq. (3) as described in the previous section. Therefore, the model 6 is the most general one. It generalizes all the cases 1 to 5.

4. Experiments

In this work, the goal is to evaluate the performance of multivariate possibilistic fuzzy c-means methods MPFCM (Case 5) and MPFCM-D (Case 6), that is more general cases, compared with four hybrid methods based on possibilistic fuzzy c-means approach. First one is the PFCM proposed by Pal *et al.* (2005).[14] Second one is the PFCM with weighted distance (here called PFCM-D) where weights are computed as described in Eq. (23). Third one is the PMFCM proposed by Himmelspach and Conrad (2016)[21] which is a combination of PFCM and MFCM (Case 3). And fourth one is the PMFCM-D that uses weighted distance (Case 4).

$$\lambda_{ij} = \frac{\left\{\prod_{h=1}^{p} \left[\sum_{k=1}^{n} [a(u_{ik})^m + b(t_{ik})^\eta] \, (x_{kj} - y_{ij})^2\right]\right\}^{\frac{1}{p}}}{\sum_{k=1}^{n} [a(u_{ik})^m + b(t_{ik})^\eta] \, (x_{kj} - y_{ij})^2} . \tag{23}$$

The evaluation of clustering results furnished by each method is based on the computation of an external cluster validity index of fuzzy type proposed by Hullermeier *et al.* (2012)[23] that follows the Rand index, proposed by Rand (1971)[24] (here the fuzzy index is abbreviated by FR). The Rand index takes its values in the interval $[0, 1]$ where the value 1 indicates a perfect agreement between partitions, values close to 0 correspond to insufficient recovery properties. FR according to fuzzy partitions P and Q is given as:

$$FR(P, Q) = 1 - \frac{\sum_{k=1}^{n} \sum_{k'=1}^{n} |E_P(\mathbf{x}_k, \mathbf{x}_{k'}) - E_Q(\mathbf{x}_k, \mathbf{x}_{k'})|}{n(n-1)/2}, \tag{24}$$

where $E_P(\mathbf{x}_k, \mathbf{x}_{k'}) = 1 - \|\boldsymbol{\delta}_k - \boldsymbol{\delta}_{k'}\|$ and $\boldsymbol{\delta}_k = (\delta_{1k}, \ldots, \delta_{ik}, \ldots, \delta_{ck})$ is a vector of membership degrees by cluster of object $\mathbf{x}_k$ (similar definition $\boldsymbol{\delta}_{k'}$ with respect to object $\mathbf{x}_{k'}$ and $E_Q(\mathbf{x}_k, \mathbf{x}_{k'})$ to partition Q). In this paper, the metric $\|\cdot\|$ on $[0, 1]^c$ that yields values in $[0, 1]$ is given as $d(\boldsymbol{\delta}_k, \boldsymbol{\delta}_{k'}) = \frac{1}{c} \sum_{i=1}^{c} (\delta_{ik} - \delta_{ik'})^2$.

4.1. *Synthetic data*

In this study, two synthetic datasets in $\Re^2$ are adopted. Each class is drawn according to a bi-variate normal distribution whose components are independent variables. Tables 1 and 2 show parameters of classes for datasets 1 and 2, respectively.

Table 1. Parameters for Dataset 1.

Class	μ_1	μ_2	σ_1^2	σ_2^2	# Objects
1	0	16	6	6	200
2	−8	8	13	13	100
3	−16	−5	20	20	50

Table 2. Parameters for Dataset 2.

Class	μ_1	μ_2	σ_1^2	σ_2^2	# Objects
1	7	20	2	50	100
2	0	−6	50	2	100
3	12	0	50	2	100

Figure 1 shows the scatterplots of the dataset 1 showing spherical classes and dataset 2 showing elliptical classes.

In order to build noisy datasets, an amount of noise is added to the normal data according to a percentage t of the number of objects in the partition. Here, 6 values for t were considered: 0%, 5%, 10%, 15%, 20%, 25% and 30%. The noises are drawn according bi-variate uniform distribution. The parameters a and b of uniform distributions $U(a, b)$ for both variables and datasets 1 and 2 are $a = -100$ and $b = 50$.

Figure 2 shows the scatterplots of datasets 1 and 2 with a percentage of noise equal to 30%.

In the framework of a Monte Carlo experiment, 100 replications of the previous process were carried out. For each replication, bi-variate normal and uniform distributions following the parameters were randomly drawn 50 times and clustering

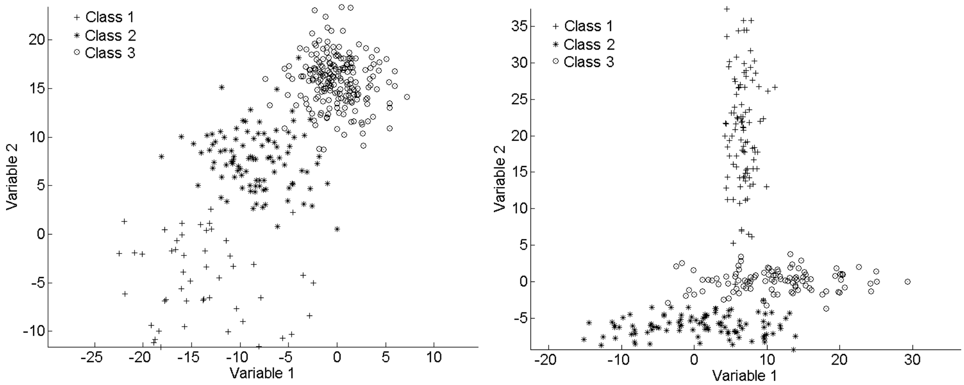

Fig. 1. Scatterplot of datasets 1 and 2.

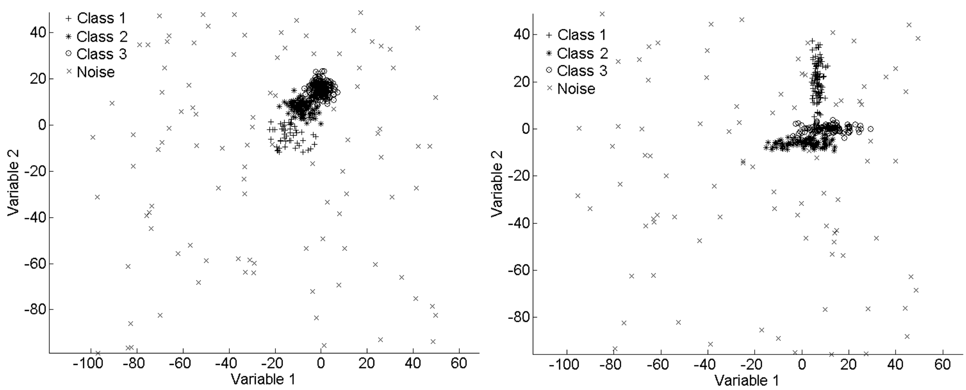

Fig. 2. Scatterplot of datasets 1 and 2 containing noise.

methods were applied to datasets. The best result according to criterion function was selected and the FR index were calculated comparing partition obtained by each clustering method and the partition known *a priori*. After 100 replications, the average and standard deviation of these indexes were calculated. This process was carried out for the 6 values of percentage of outlier. Algorithms were executed with $m = 2$, $\eta = 2$, $\varepsilon = 10^{-4}$, $a = 1$, $b = 1$ and $K = 1$.

Table 3 shows values of the average and standard deviation (in parentheses) of FR index obtained by PFCM, PFCM-D, PMFCM, PMFCM-D, MPFCM and MPFCM-D methods for synthetic dataset 1. The largest average value FR index is in bold. The ANOVA hypothesis tests was applied with a significance level of 5% and the p-value was $3.24e - 152$. So, we can say there is difference among the averages of FR index in Table 3. Thereafter, it was applied the one-sided Student's t-test for paired samples with a significance level of 5%.

Table 3. Average and standard deviation (in parenthesis) of FR index for clustering methods and synthetic dataset 1.

%	PFCM	PFCM-D	PMFCM	PMFCM-D	MPFCM	MPFCM-D
0%	0.681 (0.015)	0.687 (0.019)	0.590 (0.008)	0.593 (0.017)	**0.721** (0.019)	**0.722** (0.019)
5%	0.684 (0.016)	0.693 (0.006)	0.589 (0.007)	0.591 (0.011)	**0.728** (0.018)	**0.729** (0.019)
10%	0.684 (0.015)	0.690 (0.021)	0.592 (0.007)	0.0595 (0.009)	**0.724** (0.018)	**0.724** (0.018)
15%	0.684 (0.015)	0.683 (0.010)	0.591 (0.005)	0.590 (0.004)	**0.725** (0.016)	**0.726** (0.017)
20%	0.683 (0.015)	0.692 (0.016)	0.590 (0.008)	0.592 (0.009)	**0.731** (0.021)	**0.732** (0.021)
25%	0.678 (0.015)	0.682 (0.017)	0.594 (0.006)	0.595 (0.008)	**0.724** (0.020)	**0.725** (0.020)
30%	0.675 (0.014)	0.685 (0.012)	0.589 (0.012)	0.591 (0.013)	**0.726** (0.021)	**0.728** (0.020)

Table 4. One-sided Student's t-tests for paired samples and MPFCM and MPFCM-D methods concerning dataset 1.

%		PFCM	PFCM-D	PMFCM	PMFCM-D	MPFCM
0%	MPFCM	$6.73e-04$	$7.35e-04$	$2.42e-08$	$3.63e-08$	–
	MPFCM-D	$5.74e-04$	$3.63e-04$	$1.55e-08$	$2.45e-08$	0.479
5%	MPFCM	$3.64e-04$	$5.25e-04$	$4.63e-09$	$6.35e-09$	–
	MPFCM-D	$3.72e-04$	$5.83e-04$	$2.74e-09$	$3.63e-09$	0.481
10%	MPFCM	$6.18e-04$	$7.27e-04$	$2.53e-08$	$5.25e-08$	–
	MPFCM-D	$6.36e-04$	$7.23e-04$	$2.46e-08$	$4.24e-08$	0.483
15%	MPFCM	$5.84e-04$	$2.65e-04$	$6.35e-09$	$8.59e-09$	–
	MPFCM-D	$5.83e-04$	$1.52e-04$	$4.25e-09$	$6.35e-09$	0.477
20%	MPFCM	$3.61e-04$	$4.72e-04$	$3.78e-10$	$5.10e-10$	–
	MPFCM-D	$4.14e-04$	$4.72e-04$	$2.74e-10$	$4.18e-10$	0.482
25%	MPFCM	$3.62e-04$	$4.72e-04$	$2.67e-09$	$4.42e-09$	–
	MPFCM-D	$3.62e-04$	$4.16e-04$	$1.56e-09$	$3.52e-09$	0.487
30%	MPFCM	$2.63e-04$	$3.73e-04$	$2.31e-09$	$4.21e-09$	–
	MPFCM-D	$2.62e-04$	$3.73e-04$	$2.16e-09$	$3.52e-09$	0.472

According to results shown in Tables 3 and 4, for all percentage of noisy data, algorithms proposed MPFCM and MPFCM-D had better results compared with other methods. Concerning multivariate methods MPFCM and MPFCM-D, they obtained similar result for the data set with spherical classes. Moreover, as expected, multivariate methods showed be stable for variation of noisy data.

Table 5 shows values of the average and standard deviation (in parentheses) of FR index obtained by PFCM, PFCM-D, PMFCM, MPFCM, PMFCM-D and MPFCM-D methods for synthetic dataset 2. The largest average value FR index is in bold. Again, the ANOVA hypothesis tests was applied with a significance level of 5% and the p-value was $5.825e-273$. So, we can say there is difference among the averages of FR index in Table 5. Then, it was applied the one-sided Student's t-test for paired samples with a significance level of 5%.

Table 5. Average and standard deviation (in parenthesis) of FR index for clustering methods and synthetic dataset 2.

%	PFCM	PFCM-D	PMFCM	PMFCM-D	MPFCM	MPFCM-D
0%	0.577 (0.020)	0.598 (0.013)	0.517 (0.008)	0.535 (0.007)	0.665 (0.038)	**0.687** (0.035)
5%	0.582 (0.021)	0.602 (0.014)	0.501 (0.013)	0.517 (0.011)	0.679 (0.035)	**0.699** (0.034)
10%	0.580 (0.020)	0.606 (0.011)	0.483 (0.017)	0.502 (0.010)	0.676 (0.033)	**0.689** (0.030)
15%	0.573 (0.019)	0.594 (0.023)	0.483 (0.011)	0.495 (0.008)	0.673 (0.033)	**0.688** (0.032)
20%	0.574 (0.022)	0.603 (0.015)	0.470 (0.019)	0.492 (0.017)	0.670 (0.038)	**0.686** (0.032)
25%	0.568 (0.020)	0.593 (0.024)	0.461 (0.011)	0.479 (0.010)	0.666 (0.034)	**0.684** (0.036)
30%	0.558 (0.019)	0.597 (0.011)	0.463 (0.011)	0.477 (0.013)	0.662 (0.033)	**0.676** (0.034)

Table 6. One-sided Student's t-tests for paired samples and MPFCM and MPFCM-D methods concerning dataset 2.

%		PFCM	PFCM-D	PMFCM	PMFCM-D	MPFCM
0%	MPFCM	$3.53e-06$	$6.36e-06$	$4.63e-09$	$2.52e-07$	–
	MPFCM-D	$4.63e-08$	$7.35e-08$	$6.35e-10$	$2.62e-08$	$2.52e-03$
5%	MPFCM	$5.35e-06$	$2.52e-05$	$5.21e-11$	$1.52e-09$	–
	MPFCM-D	$2.42e-08$	$5.34e-08$	$4.24e-12$	$1.81e-10$	$4.42e-03$
10%	MPFCM	$2.13e-06$	$7.34e-06$	$4.24e-13$	$1.73e-11$	–
	MPFCM-D	$3.62e-08$	$4.24e-08$	$1.35e-14$	$2.62e-12$	$2.11e-03$
15%	MPFCM	$2.74e-06$	$5.73e-06$	$2.52e-09$	$2.83e-07$	–
	MPFCM-D	$6.35e-08$	$5.35e-08$	$2.41e-11$	$3.62e-09$	$2.04e-03$
20%	MPFCM	$2.63e-06$	$8.24e-06$	$1.41e-09$	$3.81e-07$	–
	MPFCM-D	$5.74e-08$	$6.34e-08$	$1.31e-11$	$4.63e-09$	$6.14e-03$
25%	MPFCM	$6.21e-07$	$5.52e-06$	$3.56e-10$	$3.62e-08$	–
	MPFCM-D	$4.41e-09$	$5.24e-08$	$5.83e-12$	$8.22e-10$	$5.24e-03$
30%	MPFCM	$3.51e-07$	$7.35e-06$	$6.34e-10$	$3.98e-08$	–
	MPFCM-D	$3.11e-09$	$7.23e-08$	$6.49e-12$	$2.72e-10$	$7.52e-03$

From the results in Tables 5 and 6, multivariate methods MPFCM and MPFCM-D showed be stable for the variation of the percentage of noisy data, similarly to the configuration with spherical classes. Moreover, they showed better results than the PFCM, its version with weight PFCM-D, the PMFCM method and its weighted version PMFCM-D. For this dataset with elliptical classes, MPFCM-D method achieved a better clustering quality than MPFCM for all percentages of noisy data. Regarding the results obtained by the clustering methods in both datasets, it can be highlighted that the multivariate methods based on possibilistic approach MPFCM and MPFCM-D were superior to their corresponding univariate versions (PFCM and PFCM-D) and the hybrid methods PMFCM and PMFCM-D. Therefore, MPFCM and MPFCM-D shown be more robust to the variation of the percentage noisy data.

4.2. *Real datasets*

In this study, 9 real datasets were considered: Abalone, Ecoli, Glass Identification, Haberman's Survival, Pima Indian Diabetes, Statlog Vehicle, Vertebral Column, Wine and Wine Quality. These datasets may be found in the UCI machine learning repository.[25] Table 7 describes the number of classes, number of objects, the cardinality of objects in each class and the number of variables for each UCI datasets.

Table 7. Summarized properties of the UCI datasets.

Datasets	# Classes	# Obj.	Cardinality	# Var.
Abalone	3	4177	1307, 1342, 1528	8
Breast	2	569	357, 212	30
Breast Tissue	6	106	21, 15, 18, 16, 14, 22	9
Iris	3	150	50, 50, 50	4
Parkinsons	2	197	48, 149	22
Sonar	2	208	98, 110	60
Statlog	4	946	240, 240, 240, 246	18
Wine	3	178	59, 71, 48	13
WineQlt	6	4898	10, 53, 681, 639, 200, 19	11

Each variable dataset was standardized in such a way that values are in interval $[0, 1]$. Using the Booststrap technique, we performed 1000 sampling of each dataset. For each sampling, a method is executed until its convergence 200 times with random initializations and the best result according to the objective function is selected. After the selection, the FR index is calculated for each method comparing the partition obtained by the method and the known partition *a priori*. At the end of execution of the 1000 sampling, the mean and standard deviation are calculated. The parameter m of fuzzy methods and η of possibilistic methods were fixed to 2. Table 8 shows the average and standard deviation (in parenthesis) of FR index for clustering methods and real datasets with larger values in bold.

Table 8. Average and standard deviation (in parenthesis) of FR index for real datasets.

Datasets	PFCM	PFCM-D	PMFCM	PMFCM-D	MPFCM	MPFCM-D
Abalone	0.2817 (0.0516)	0.2921 (0.0355)	0.2242 (0.0562)	0.2263 (0.0421)	0.3125 (0.0217)	**0.3252** (0.0183)
Breast	0.3696 (0.0143)	0.3924 (0.0234)	0.3325 (0.0832)	0.3521 (0.0425)	0.3852 (0.0146)	**0.4015** (0.0162)
Breast Tissue	0.6481 (0.0626)	0.6387 (0.0263)	0.6283 (0.0256)	0.6315 (0.0241)	0.6597 (0.0137)	**0.6725** (0.018)
Iris	0.5517 (0.0366)	0.5637 (0.0256)	0.5325 (0.0521)	0.5624 (0.0463)	0.5725 (0.0251)	**0.5936** (0.0121)
Parkinsons	0.2680 (0.0242)	**0.3924** (0.0287)	0.2736 (0.0252)	0.2894 (0.0226)	0.2931 (0.0351)	0.2942 (0.0221)
Sonar	0.1208 (0.0131)	0.1373 (0.0144)	0.1035 (0.0162)	0.1152 (0.0142)	0.1513 (0.0134)	**0.1669** (0.0142)
Statlog	0.5287 (0.0231)	0.5424 (0.0163)	0.5036 (0.0252)	0.5426 (0.0152)	0.5607 (0.0212)	**0.6101** (0.0231)
Wine	0.3511 (0.1286)	0.3625 (0.0592)	0.3215 (0.0325)	0.3424 (0.0286)	0.3892 (0.0173)	**0.4163** (0.0238)
Wine Q.	0.6220 (0.0312)	0.6363 (0.0256)	0.6148 (0.0324)	0.6351 (0.0271)	0.6542 (0.0173)	**0.6925** (0.0122)

Table 9. One-sided Student's t-tests for paired samples and MPFCM and MPFCM-D methods concerning real datasets.

Datasets		PFCM	PFCM-D	PMFCM	PMFCM-D	MPFCM
Abalone	MPFCM	$5.67e-08$	$8.24e-08$	$2.53e-16$	$4.25e-16$	–
	MPFCM-D	$3.63e-10$	$9.24e-10$	$4.24e-18$	$2.52e-18$	$2.82e-06$
Breast	MPFCM	$3.51e-09$	$7.35e-08$	$2.74e-12$	$7.55e-10$	–
	MPFCM-D	$4.27e-11$	$1.62e-09$	$3.93e-15$	$4.25e-13$	$2.63e-08$
Breast Tissue	MPFCM	$4.63e-06$	$5.22e-08$	$2.73e-07$	$5.35e-07$	–
	MPFCM-D	$2.52e-08$	$2.62e-11$	$4.25e-12$	$4.53e-11$	$7.21e-07$
Iris	MPFCM	$4.63e-07$	$8.34e-07$	$2.63e-09$	$7.69e-08$	–
	MPFCM-D	$4.24e-09$	$5.52e-11$	$2.51e-13$	$3.11e-11$	$4.29e-08$
Parkinsons	MPFCM	$5.34e-08$	$3.52e-18$	$2.62e-07$	$6.26e-06$	–
	MPFCM-D	$3.73e-08$	$7.26e-18$	$5.24e-07$	$2.27e-06$	0.158
Sonar	MPFCM	$3.63e-09$	$2.52e-08$	$3.94e-12$	$2.91e-10$	–
	MPFCM-D	$5.22e-10$	$7.23e-09$	$7.96e-13$	$4.27e-11$	$6.23e-07$
Statlog	MPFCM	$6.12e-10$	$2.72e-08$	$2.52e-12$	$9.21e-08$	–
	MPFCM-D	$4.72e-14$	$3.61e-13$	$5.26e-16$	$3.52e-13$	$2.67e-11$
Wine	MPFCM	$3.62e-07$	$2.53e-06$	$1.42e-13$	$3.95e-11$	–
	MPFCM-D	$7.25e-08$	$3.11e-07$	$4.82e-15$	$3.02e-13$	$5.73e-09$
Wine Q.	MPFCM	$3.66e-06$	$8.35e-05$	$7.26e-08$	$1.63e-05$	–
	MPFCM-D	$7.25e-11$	$8.61e-10$	$6.24e-14$	$1.94e-10$	$4.83e-10$

ANOVA hypothesis test was applied to results presented in Table 8. With the significance level of 5% and p-value=$1.53e-164$, we can reject the null hypothesis (which means that there is difference among the averages of FR index in Table 8). Then, it was applied the one-sided Student's t-test for paired samples with a significance level of 5% and the results for p-value are shown in Table 9.

From the results in Tables 8 and 9, it is important to highlight that MPFCM-D was superior to MPFCM method for most of real datasets. Moreover, the proposed methods MPFCM-D (Case 6) and MPFCM (Case 5) are the best options in terms of clustering quality measured by the average of fuzzy Rand index for all the datasets. The application also showed that the use of weighted distance and multivariate membership and possibilistic degrees is essential.

5. Conclusion

In this work, we presented a generalized multivariate approach based on possibilistic fuzzy clustering. Moreover, it uses weighted distance in order to take into account the variability of the clusters. This type of weight may be different for each cluster and variable and change at each iteration of algorithm, minimizing the objective function. It was highlighted that from this approach six methods can be derivative. In order to show the usefulness of the proposed method in comparison with other possibilistic fuzzy c-means clustering methods, an experimental evaluation was carried out. Synthetic datasets varying the percentage of noisy data and the shape of clusters were considered. The accuracy of results furnished by clustering methods was assessed by Fuzzy Rand Index. In addition, real datasets were also considered. Results pointed out the importance to use multivariate approach for both membership and possibility degrees and weighted distance.

Acknowledgements

The authors would like to thank CNPq and FACEPE (Brazilian Agencies) for their financial support.

Appendix A. Proof of Proposition 1

Fixing the membership degree u_{ijk} and the possibility degree t_{ijk}, the prototype y_{ij} which minimizes the criterion J is updated using the following equation:

$$y_{ij} = \frac{\sum_{k=1}^{n} \left[a(u_{ijk})^m + b(t_{ijk})^\eta\right] x_{ijk}}{\sum_{k=1}^{n} \left[a(u_{ijk})^m + b(t_{ijk})^\eta\right]}\,. \tag{A.1}$$

Proof. The criterion J^1 being additive, the problem becomes to minimize J_{ij} as follows. Let $d_{ijk} = (x_{jk} - y_{ij})^2$ be the distance between object $\mathbf{x}_k$ and prototype $\mathbf{y}_i$ for variable j:

$$J_{ij} = \lambda_{ij} \sum_{k=1}^{n} \left\{\left[a(u_{ijk})^m + b(t_{ijk})^\eta\right](x_{jk} - y_{ij})^2 + \pi_{ij}(1 - t_{ijk})^\eta\right\}. \tag{A.2}$$

J_{ij} becomes stationary when:

$$\frac{\partial J_{ij}}{\partial y_{ij}} = -2\sum_{k=1}^{n} \left[a(u_{ijk})^m + b(t_{ijk})^\eta\right](x_{jk} - y_{ij}) = 0\,. \tag{A.3}$$

That is:

$$\sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta]\, x_{jk} - y_{ij} \sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta] = 0\,. \tag{A.4}$$

From this, we can conclude:

$$y_{ij} = \frac{\sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta]\, x_{ijk}}{\sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta]}. \tag{A.5}$$

□

Appendix B. Proof of Proposition 2

Fixing the prototype y_{ij} and the possibility degree t_{ijk}, the membership degree u_{ijk} under the restriction $\sum_{i=1}^{c} \sum_{j=1}^{p} u_{ijk} = 1$ which minimizes the criterion J is updated using the following equation:

$$u_{ijk} = \left[\sum_{r=1}^{c} \sum_{s=1}^{p} \left(\frac{\lambda_{ij} d_{ijk}}{\lambda_{rs} d_{rsk}}\right)^{\frac{1}{m-1}}\right]^{-1}. \tag{B.1}$$

Proof. The criterion J being additive, the problem becomes to find for some object $\mathbf{x}_k$ the membership degree u_{ijk} that minimizes

$$J_k = \lambda_{ij} \sum_{k=1}^{n} \{[a(u_{ijk})^m + b(t_{ijk})^\eta]\, d_{ijk} + \pi_{ij}(1 - t_{ijk})^\eta\}\,. \tag{B.2}$$

The method of Lagrange multipliers can be applied to find an optimum value to J_k:

$$F_k(\lambda) = \lambda_{ij} \sum_{k=1}^{n} \{[a(u_{ijk})^m + b(t_{ijk})^\eta] d_{ijk} + \pi_{ij}(1 - t_{ijk})^\eta\} - \mu \left(\sum_{i=1}^{c} \sum_{j=1}^{p} u_{ijk} - 1\right). \tag{B.3}$$

Thus, F_k is stationary in the following situations:

(1) $\frac{\partial F_k}{\partial \mu} = \sum_{i=1}^{c} \sum_{j=1}^{p} u_{ijk} - 1 = 0$;
(2) $\frac{\partial F_k}{\partial u_{ijk}} = am(u_{ijk})^{m-1} \lambda_{ij} d_{ijk} - \mu = 0$.

From the second item, it is possible to calculate the membership degree value u_{ijk} as follows:

$$u_{ijk} = \left(\frac{\mu}{am}\right)^{\frac{1}{m-1}} \left(\frac{1}{\lambda_{ij} d_{ijk}}\right)^{\frac{1}{m-1}}. \tag{B.4}$$

According to $\sum_{i=1}^{c} \sum_{j=1}^{p} u_{ijk} = 1$ and Eq. (B.4), we have:

$$\sum_{r=1}^{c} \sum_{s=1}^{p} u_{rsk} = \left(\frac{\mu}{am}\right)^{\frac{1}{m-1}} \sum_{r=1}^{c} \sum_{s=1}^{p} \left(\frac{1}{\lambda_{rs} d_{rsk}}\right)^{\frac{1}{m-1}} = 1\,. \tag{B.5}$$

Therefore, we obtain the following expression:

$$\left(\frac{\mu}{am}\right)^{\frac{1}{m-1}} = \frac{1}{\sum_{r=1}^{c}\sum_{s=1}^{p}\left(\frac{1}{\lambda_{rs}d_{rsk}}\right)^{\frac{1}{m-1}}}\,. \tag{B.6}$$

Replacing Eq. (B.6) into Eq. (B.4), we have:

$$u_{ijk} = \frac{1}{\sum_{r=1}^{c}\sum_{s=1}^{p}\left(\frac{1}{\lambda_{rs}d_{rsk}}\right)^{\frac{1}{m-1}}}\left(\frac{1}{\lambda_{ij}d_{ijk}}\right)^{\frac{1}{m-1}}\,. \tag{B.7}$$

Finally, we obtain:

$$u_{ijk} = \left[\sum_{r=1}^{c}\sum_{s=1}^{p}\left(\frac{\lambda_{ij}d_{ijk}}{\lambda_{rs}d_{rsk}}\right)^{\frac{1}{m-1}}\right]^{-1}\,. \tag{B.8}$$

□

Appendix C. Proof of Proposition 3

Fixing the membership degree u_{ijk} and the possibility degree t_{ijk}, the prototype y_{ij} which minimizes the criterion J is updated using the following equation:

$$t_{ijk} = \left[1 + \left(b\frac{\lambda_{ij}d_{ijk}}{\pi_{ij}}\right)^{\frac{1}{(\eta-1)}}\right]^{-1}\,. \tag{C.1}$$

Proof. The criterion J^1 being additive, the problem becomes to minimize J_{ijk} as follows. Let $d_{ijk} = (x_{jk} - y_{ij})^2$ be the distance between object $\mathbf{x}_k$ and prototype $\mathbf{y}_i$ for variable j:

$$\frac{\partial J^1_{ijk}}{\partial t_{ijk}} = \eta b(t_{ijk})^{\eta-1}\lambda_{ij}d_{ijk} + (-1)\eta\pi_{ij}(1-t_{ijk})^{\eta-1} = 0\,, \tag{C.2}$$

which is equivalent to:

$$\begin{aligned} &(b\lambda_{ij}d_{ijk})^{\frac{1}{(\eta-1)}}t_{ijk} - (\pi_{ij})^{\frac{1}{(\eta-1)}} + (\pi_{ij})^{\frac{1}{(\eta-1)}}t_{ijk} = 0\,,\\ &\Rightarrow t_{ijk}\left[(b\lambda_{ij}d_{ijk})^{\frac{1}{(\eta-1)}} + (\pi_{ij})^{\frac{1}{(\eta-1)}}\right] = (\pi_{ij})^{\frac{1}{(\eta-1)}}\,, \qquad \text{(C.3)}\\ &\Rightarrow t_{ijk} = \frac{(\pi_{ij})^{\frac{1}{(\eta-1)}}}{(\pi_{ij})^{\frac{1}{(\eta-1)}} + (b\lambda_{ij}d_{ijk})^{\frac{1}{(\eta-1)}}}\,. \end{aligned}$$

From this, we can conclude:

$$t_{ijk} = \left[1 + \left(b\lambda_{ij}\frac{d_{ijk}}{\pi_{ij}}\right)^{\frac{1}{(\eta-1)}}\right]^{-1}\,. \tag{C.4}$$

□

Appendix D. Proof of Proposition 4

Fixing the prototype y_{ij}, the membership degree u_{ijk} and possibility degree t_{ijk}, the weight λ_{ij} under the restriction $\prod_{j=1}^{p} \lambda_{ij} = 1$ which minimizes the criterion J is updated using the following equation:

$$\lambda_{ij} = \frac{\{\prod_{h=1}^{p} [\sum_{k=1}^{n} [a(u_{ihk})^m + b(t_{ihk})^\eta]\, d_{ihk}]\}^{\frac{1}{p}}}{\sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta]\, d_{ijk}}\,. \tag{D.1}$$

Proof. The weight may be calculated applying the method of Lagrange multipliers:

$$F_{ij}(\lambda) = \lambda_{ij} \sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta]\, d_{ijk} - \mu \left(\prod_{h=1}^{p} \lambda_{ih} - 1 \right). \tag{D.2}$$

F_{ij} is stationary when:

$$\frac{\partial F_{ij}}{\partial \lambda_{ij}} = \sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta]\, d_{ijk} - \mu \left(\frac{\prod_{h=1}^{p} \lambda_{ih}}{\lambda_{ij}} \right) = 0\,. \tag{D.3}$$

From Eq. (D.3), we obtain:

$$\sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta]\, d_{ijk} = \mu \left(\frac{\prod_{h=1}^{p} \lambda_{ih}}{\lambda_{ij}} \right). \tag{D.4}$$

Since the restriction is $\prod_{h=1}^{p} \lambda_{ih} = 1$, then:

$$\lambda_{ij} = \frac{\mu}{\sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta]\, d_{ijk}}\,. \tag{D.5}$$

According to restriction and using Eq. (D.5), it is also true in the following sentences:

$$\begin{aligned} \prod_{h=1}^{p} \lambda_{ih} &= \prod_{h=1}^{p} \left[\frac{\mu}{\sum_{k=1}^{n} [a(u_{ihk})^m + b(t_{ihk})^\eta]\, d_{ihk}} \right] \\ &= \frac{\mu^p}{\prod_{h=1}^{p} [\sum_{k=1}^{n} [a(u_{ihk})^m + b(t_{ihk})^\eta]\, d_{ihk}]} = 1 \\ &\Rightarrow \mu = \left\{ \prod_{h=1}^{p} \left[\sum_{k=1}^{n} [a(u_{ihk})^m + b(t_{ihk})^\eta]\, d_{ihk} \right] \right\}^{\frac{1}{p}}. \end{aligned} \tag{D.6}$$

Replacing Eq. (D.6) into Eq. (D.5), we have:

$$\lambda_{ij} = \frac{\{\prod_{h=1}^{p} [\sum_{k=1}^{n} [a(u_{ihk})^m + b(t_{ihk})^\eta]\, d_{ihk}]\}^{\frac{1}{p}}}{\sum_{k=1}^{n} [a(u_{ijk})^m + b(t_{ijk})^\eta]\, d_{ijk}}\,. \tag{D.7}$$

□

Appendix E. Proof of Proposition 5

The series $\omega_t = J(\nu_t)$ decreases at each iteration and converges.

Proof. According to algorithm, the following inequalities (I), (II), (III) and (IV) hold at each iteration:

$$J(\mathbf{Y}^t, \mathbf{U}^t, \mathbf{T}^t, \mathbf{\Lambda}^t) \overbrace{\geq}^{(I)} J(\mathbf{Y}^{t+1}, \mathbf{U}^t, \mathbf{T}^t, \mathbf{\Lambda}^t) \overbrace{\geq}^{(II)}$$
$$J(\mathbf{Y}^{t+1}, \mathbf{U}^{t+1}, \mathbf{T}^t, \mathbf{\Lambda}^t) \overbrace{\geq}^{(III)} J(\mathbf{Y}^{t+1}, \mathbf{U}^{t+1}, \mathbf{T}^{t+1}, \mathbf{\Lambda}^t) \overbrace{\geq}^{(IV)} J(\mathbf{Y}^{t+1}, \mathbf{U}^{t+1}, \mathbf{T}^{t+1}, \mathbf{\Lambda}^{t+1})\,. \tag{E.1}$$

Let $d(x_{jk}, y_{ij}^{(t)}) = (x_{jk} - y_{ij}^{(t)})^2$ be the distance between object x_{jk} and prototype y_{ij} in the iteration t. The inequality (I) holds because

$$J(\mathbf{Y}^{t+1}, \mathbf{U}^t, \mathbf{T}^t, \mathbf{\Lambda}^t) = \sum_{i=1}^{c}\sum_{j=1}^{p} \lambda_{ij}(t) \sum_{k=1}^{n} \left\{ \left[a(u_{ijk}^{(t)})^m \right.\right.$$
$$\left.\left. + b(t_{ijk}^{(t)})^\eta \right] d(x_{kj}, y_{ij}^{(t+1)}) + \pi_{ij}(1 - t_{ijk}^{(t)})^\eta \right\}, \tag{E.2}$$

and according to Proposition 1,

$$y_{ij}^{(t+1)} = \underbrace{\operatorname{argmin}}_{y \in \mathfrak{R}} \lambda_{ij}^{(t)} \sum_{k=1}^{n} \left\{ \left[a(u_{ijk}^{(t)})^m + b(t_{ijk}^{(t)})^\eta \right] d(x_{kj}, y) + \pi_{ij}(1 - t_{ijk}^{(t)})^\eta \right\}. \tag{E.3}$$

Moreover, inequality (II) also holds because

$$J(\mathbf{Y}^{t+1}, \mathbf{U}^{t+1}, \mathbf{T}^t, \mathbf{\Lambda}^t) = \sum_{i=1}^{c}\sum_{j=1}^{p} \lambda_{ij}^{(t)} \sum_{k=1}^{n} \left\{ \left[a(u_{ijk}^{(t+1)})^m \right.\right.$$
$$\left.\left. + b(t_{ijk}^{(t)})^\eta \right] d(x_{kj}, y_{ij}^{(t+1)}) + \pi_{ij}(1 - t_{ijk}^{(t)})^\eta \right\}, \tag{E.4}$$

and according to Proposition 2,

$$u_{ijk}^{(t+1)} = \underbrace{\operatorname{argmin}}_{u \in [0,1]} \lambda_{ij}^{(t)} \left\{ \left[a(u)^m + b(t_{ijk}^{(t)})^\eta \right] d(x_{kj}, y_{ij}^{(t+1)}) + \pi_{ij}(1 - t_{ijk}^{(t)})^\eta \right\}. \tag{E.5}$$

inequality (III) also holds because

$$J(\mathbf{Y}^{t+1}, \mathbf{U}^{t+1}, \mathbf{T}^{t+1}, \mathbf{\Lambda}^t) = \sum_{i=1}^{c}\sum_{j=1}^{p} \lambda_{ij}^{(t)} \sum_{k=1}^{n} \left\{ \left[a(u_{ijk}^{(t+1)})^m \right.\right.$$
$$\left.\left. + b(t_{ijk}^{(t+1)})^\eta \right] d(x_{kj}, y_{ij}^{(t+1)}) + \pi_{ij}(1 - t_{ijk}^{(t+1)})^\eta \right\}, \tag{E.6}$$

and according to Proposition 3,

$$t_{ijk}^{(t+1)} = \underbrace{\text{argmin}}_{t\in[0,1]} \lambda_{ij}^{(t)} \left\{ \left[a(u_{ijk}^{(t+1)})^m + b(t)^\eta \right] d(x_{kj}, y_{ij}^{(t+1)}) + \pi_{ij}(1-t)^\eta \right\}. \quad \text{(E.7)}$$

and inequality (IV) also holds because

$$J(\mathbf{Y}^{t+1}, \mathbf{U}^{t+1}, \mathbf{T}^{t}, \mathbf{\Lambda}^{t+1}) = \sum_{i=1}^{c}\sum_{j=1}^{p} \lambda_{ij}^{(t+1)} \sum_{k=1}^{n} \left\{ \left[a(u_{ijk}^{(t+1)})^m + b(t_{ijk}^{(t)})^\eta \right] d(x_{kj}, y_{ij}^{(t+1)}) + \pi_{ij}(1-t_{ijk}^{(t)})^\eta \right\}, \quad \text{(E.8)}$$

and according to Proposition 4,

$$\lambda_{ij}^{(t+1)} = \underbrace{\text{argmin}}_{\lambda\in\Re} \lambda \sum_{k=1}^{n} \left\{ \left[a(u_{ijk}^{(t+1)})^m + b(t_{ijk}^{(t+1)})^\eta \right] \times d(x_{kj}, y_{ij}^{(t+1)}) + \pi_{ij}(1-t_{ijk}^{(t+1)})^\eta \right\}. \quad \text{(E.9)}$$

Finally, because the series ω_t decreases and it is bounded ($J(\nu_t) \geq 0$), it converges. $\square$

Appendix F. Proof of Proposition 6

The series $\nu_t = (\mathbf{Y}^t, \mathbf{U}^t, \mathbf{T}^t, \mathbf{\Lambda}^t)$ converges.

Proof. Assume that the stationarity of the series ν_t is achieved in the iteration $t = T$. Therefore, we have that $\nu_T = \nu_{T+1}$ and then $J(\nu_T) = J(\nu_{T+1})$, that is, $J(\mathbf{Y}^T, \mathbf{U}^T, \mathbf{T}^T, \mathbf{\Lambda}^T) = J(\mathbf{Y}^{T+1}, \mathbf{U}^{T+1}, \mathbf{T}^{T+1}, \mathbf{\Lambda}^{T+1})$. From this equality and Proposition 5, it is possible rewrite as equalities (I), (II), (III) and (IV):

$$J(\mathbf{Y}^T, \mathbf{U}^T, \mathbf{T}^T, \mathbf{\Lambda}^T) \overbrace{=}^{(I)} J(\mathbf{Y}^{T+1}, \mathbf{U}^T, \mathbf{T}^T, \mathbf{\Lambda}^T) \overbrace{=}^{(II)} J(\mathbf{Y}^{T+1}, \mathbf{U}^{T+1}, \mathbf{T}^T, \mathbf{\Lambda}^T) \overbrace{=}^{(III)} J(\mathbf{Y}^{T+1}, \mathbf{U}^{T+1}, \mathbf{T}^{T+1}, \mathbf{\Lambda}^T) \overbrace{=}^{(IV)} J(\mathbf{Y}^{T+1}, \mathbf{U}^{T+1}, \mathbf{T}^{T+1}, \mathbf{\Lambda}^{T+1}). \quad \text{(F.1)}$$

From the first equality (I), $\mathbf{Y}^T = \mathbf{Y}^{T+1}$ because $\mathbf{Y}$ is unique minimizing J when $\mathbf{U}^T$, $\mathbf{T}^T$ and $\mathbf{\Lambda}^T$ are fixed. From the second equality (II), $\mathbf{U}^T = \mathbf{U}^{T+1}$ because $\mathbf{U}$ is unique minimizing J when $\mathbf{Y}^{T+1}$, $\mathbf{T}^T$ and $\mathbf{\Lambda}^T$ are fixed. Moreover, from the third equality (III), $\mathbf{T}^T = \mathbf{T}^{T+1}$ because $\mathbf{T}$ is unique minimizing J when $\mathbf{Y}^{T+1}$, $\mathbf{U}^{T+1}$ and $\mathbf{\Lambda}^T$ are fixed. Finally, from the fourth equality (IV), $\mathbf{\Lambda}^T = \mathbf{\Lambda}^{T+1}$ because $\mathbf{\Lambda}$ is unique minimizing J when $\mathbf{Y}^{T+1}$, $\mathbf{U}^{T+1}$ and $\mathbf{T}^{T+1}$ are fixed.

Finally, we conclude that $\nu_T = \nu_{T+1}$. This conclusion holds for all $t \geq T$ and $\nu_t = \nu_T, \forall t \geq T$ and it follows that the series ν_t converges. $\square$

References

1. J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Springer Science & Business Media, 2013).
2. K. Fukunaga, *Introduction to Statistical Pattern Recognition* (Academic Press, 2013).
3. N. Padhy *et al.* The survey of data mining applications and feature scope, *Int. J. Computer Science, Engineering and Information Technology* (*IJCSEIT*) **2** (2012) 43–58.
4. W. Wu *et al. Clustering and Information Retrieval* (Springer Science & Business Media, 2013).
5. A. K. Jain *et al.*, Data clustering: a review, *ACM Computing Surveys* **31** (1999) 264–323.
6. R. O. Duda and P. E. Hart, *Pattern Classification* (John Wiley and Sons, 2001).
7. G. Gan *et al.*, *Data Clustering: Theory, Algorithms, and Applications*, Book 20 (SIAM, 2007).
8. F. Murtagh and P. Contreras, Algorithms for hierarchical clustering: an overview, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2**(1) (2012) 86–97.
9. R. Xu and D. I. I. Wunsch, Survey of clustering algorithms, *IEEE Transactions on Neural Networks* **16**(3) (2005) 645–678.
10. R. Xu and D. Wunsch, *Clustering* (John Wiley & Sons, 2008).
11. J. C. Bezdek *et al.*, FCM: The fuzzy c-means clustering algorithm, *Computers and Geosciences* **10**(2) (1984) 191–203.
12. M. Gong *et al.*, Fuzzy c-means clustering with local information and kernel metric for image segmentation, *IEEE Transactions on Image Processing* **22**(2) (2013) 573–584.
13. R. Krishnapuram and J. M. Keller, A possibilistic approach to clustering, *IEEE Transactions on Fuzzy Systems* **1**(2) (1993) 98–110.
14. N. R. Pal *et al.*, A possibilistic fuzzy c-means clustering algorithm, *IEEE Transactions on Fuzzy Systems* **13**(4) (2005) 517–530.
15. A. Keller and F. Klawonn, Fuzzy clustering with weighting of data variables, *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems* **8**(6) (2000) 735–746.
16. A. F. Alvim and R. M. C. R. Souza, A fuzzy weighted clustering method for symbolic interval data, in *2012 Annual Meeting of the North American Fuzzy Information Processing Society* (*NAFIPS*) (2012), pp. 1–6.
17. B. A. Pimentel and R. M. C. R de Souza, A multivariate fuzzy c-means method, *Applied Soft Computing* **13**(4) (2013) 1592–1607.
18. B. A. Pimentel and R. M. C. R de Souza, A weighted multivariate Fuzzy C-Means method in interval-valued scientific production data, *Expert Systems with Applications* **41**(7) (2014a) 3223–3236.
19. B. A. Pimentel and R. M. C. R de Souza, Possibilistic clustering methods for interval-valued data. *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems* **22**(2) (2014b) 263–291.
20. B. A. Pimentel and R. M. C. R de Souza, Multivariate fuzzy C-means algorithms with weighting, *Neurocomputing* **174** (2016) 946–965.
21. L. Himmelspach and S. Conrad, A possibilistic multivariate fuzzy c-means clustering algorithm, in *Int. Conf. Scalable Uncertainty Management* (Springer International Publishing, 2016), pp. 338–344.
22. E. Diday and J. C. Simon Clustering analysis, in K. S. Fu (Ed.), *Digital Pattern Classification* (Springer, Berlin, 1976), pp. 47–94.

23. E. Hullermeier *et al.*, Comparing fuzzy partitions: A generalization of the Rand index and related measures, *IEEE Transactions on Fuzzy Systems* **20**(3) (2012) 546–556.
24. W. M. Rand, Objective criteria for the evaluation of clustering methods, *Journal of the American Statistical Association* **66**(336) (1971) 846–850.
25. K. Bache and M. Lichman, UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science (2017). Acessible: http://archive.ics.uci.edu/ml.