Journal Pre-proof

Fuzzy k-Means: history and applications

Maria Brigida Ferraro

PII: S2452-3062(21)00139-8
DOI: https://doi.org/10.1016/j.ecosta.2021.11.008
Reference: ECOSTA 302

To appear in: *Econometrics and Statistics*

Received date: 18 February 2021
Revised date: 21 November 2021
Accepted date: 21 November 2021

Please cite this article as: Maria Brigida Ferraro, Fuzzy k-Means: history and applications, *Econometrics and Statistics* (2021), doi: https://doi.org/10.1016/j.ecosta.2021.11.008

Fuzzy k-Means: history and applications

Maria Brigida Ferraro

Dipartimento di Scienze Statistiche, Sapienza Università di Roma, P.le Aldo Moro, 5 - 00185, Rome, Italy

Abstract

The fuzzy approach to clustering arises to cope with situations where objects have not a clear assignment. Unlike the hard/standard approach where each object can only belong to exactly one cluster, in a fuzzy setting, the assignment is soft; that is, each object is assigned to all clusters with certain membership degrees varying in the unit interval. The best known fuzzy clustering algorithm is the fuzzy k-means (FkM), or fuzzy c-means. It is a generalization of the classical k-means method. Starting from the FkM algorithm, and in more than 40 years, several variants have been proposed. The peculiarity of such different proposals depends on the type of data to deal with, and on the cluster shape. The aim is to show fuzzy clustering alternatives to manage different kinds of data, ranging from numeric, categorical or mixed data to more complex data structures, such as interval-valued, fuzzy-valued or functional data, together with some robust methods. Furthermore, the case of two-mode clustering is illustrated in a fuzzy setting.

Keywords: Fuzzy clustering, Fuzzy k-Means, mixed data, fuzzy data, functional data, double clustering

1. Introduction

The hard/standard approach to clustering consists in assigning each object to one and only one cluster. Thus, some objects are forced to be assigned to a given cluster despite being far from the prototype. The soft approach to clustering allows us to assign units to all the clusters with a degree ranging in $[0, 1]$. In this case, the assignment is referred to as a membership and not as a simply allocation, as in the hard approach. There exist several types of soft clustering: fuzzy, possibilistic, and rough clustering, among others (see, for a review, [35]). In a fuzzy setting, the membership degrees are not based on probabilistic assumptions but just on the distances between objects and prototypes. The possibilistic approach differs from the fuzzy one only because some constraints on membership degrees are relaxed, while in the rough clustering, there are not degrees taking values in the unit interval but objects with no clear assignment (belonging to the boundary region between two clusters) are associated to more than one cluster. Even if it is not always recognized as a soft approach, model-based clustering also provides a soft partition. Unlike the previous ones, it is based on probabilistic assumptions. In such a clustering approach, each cluster is viewed as a component of a mixture model, and the the posterior probability of a component membership may play the same role as the membership degree in the fuzzy clustering.

In this work, the focus is on the fuzzy approach. The first and most known fuzzy clustering algorithm is the generalization of the standard k-Means [66, 6], the fuzzy k-Means (FkM). It was introduced in [19] but deeply analyzed and improved in [4, 5]. Fuzzy clustering has also been shown to be a valuable tool from a computational point of view, making the clustering algorithm to become more efficient [59]. There exist about 1114000 documents containing "Fuzzy k-means" (or "Fuzzy c-means") on Google, and 178740 on Google Scholar. The time series of the number of documents on Google Scholar and Scopus is reported in Figure 1, showing an exponential growth. Focusing on Scopus, there are near 16000 documents (about 2500 Open Access) containing "Fuzzy k-Means" (or "Fuzzy c-Means") in the Article title, in the Abstract or in the Keywords (standard default search in Scopus). Regarding the areas in which this method is more popular, Figure 2 reports the standard search on Scopus containing "Fuzzy k-Means" (or "Fuzzy c-Means") by subjects. About 33% of the documents are in computer science, 23.7% in engineering and 12.3% in mathematics plus some application areas that underline its multidisciplinarity.

A review of the Fuzzy c-Means algorithm from 2000 to 2014 is provided

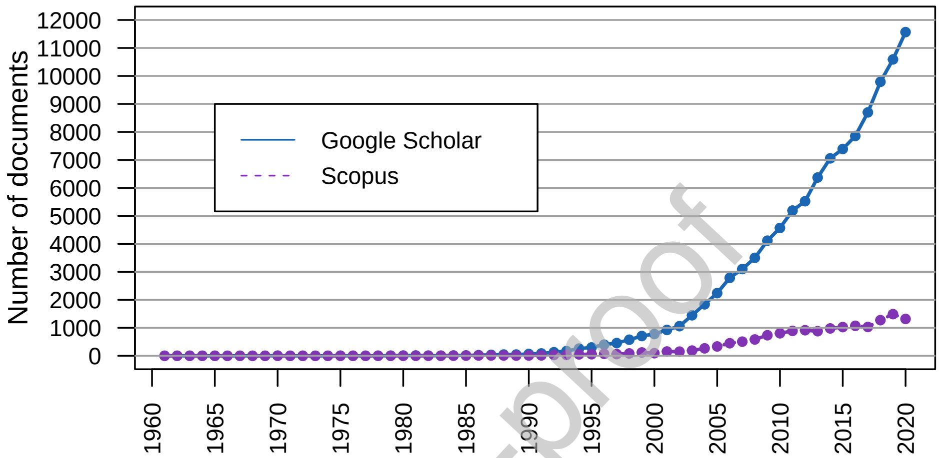

Figure 1: Number of documents on Google Scholar and Scopus containing "Fuzzy k-Means" (or "Fuzzy c-Means").

in [67]. Starting from the FkM, several variants have been proposed based on different distance measures, different prototype definitions and different kinds of data. Most proposals are devoted to numerical data: the Gustafson-Kessel Fuzzy k-Means (GK-FkM) [44], the Entropic Fuzzy k-Means (EFkM) [64, 65], the Fuzzy k-Means with Polynomial Fuzzifier (FkMPF) [58, 94], the Fuzzy k-Medoids (FkMed) [60], the Fuzzy k-Means with Noise cluster (FkMN) [13], among others.

While fuzzy cluster analysis techniques for object data (units by variables) matrices have received significant interest, fuzzy clustering of relational data has received less attention, because in most cases, an object matrix, and not pure relational data, is available. Relational data are represented by a measure of similarity (or dissimilarity) between the elements that is sometimes obtained from the objects themselves. An example of this

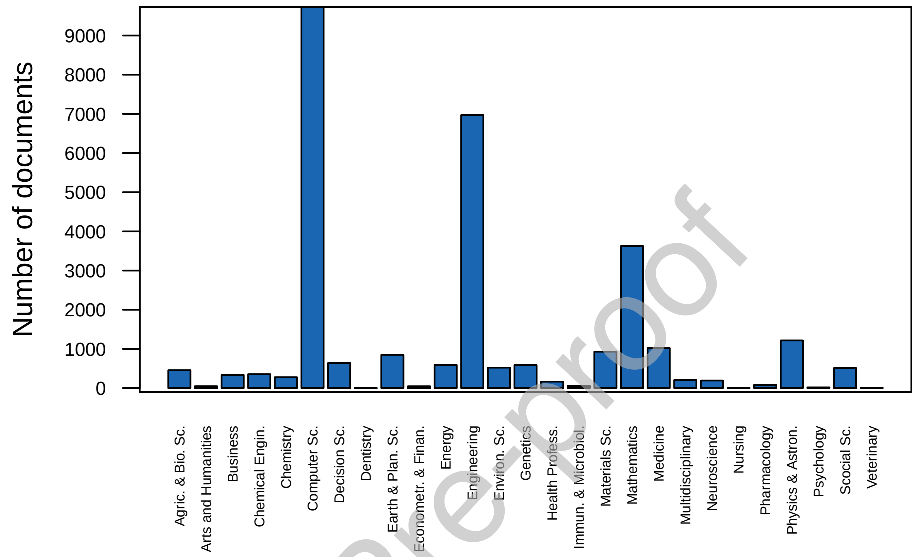

Figure 2: Documents on Scopus containing "Fuzzy k-Means" (or "Fuzzy c-Means") in the Article title, in the Abstract or in the Keywords, by Subject Area.

case is represented by the Euclidean distance between objects. Proposals of fuzzy clustering techniques for relational data are provided, for example, in [80, 93, 48, 56, 46, 47, 14].

The above fuzzy clustering methods for relational data can also be used for other kinds of data, for example, categorical or mixed data, as long as an appropriate dissimilarity measure is defined, but there also exist specific proposals for categorical or mixed data. This is the case of the fuzzy k-modes [51] for categorical and the fuzzy k-prototypes [54] for mixed data, among others.

In more than 40 years, several variants of FkM have been extended to more complex data structures such as fuzzy-valued, interval-valued and functional data.

In many real-life situations, some measurements may be imprecise and some observations may be vaguely defined. In such contexts, it is appropriate to represent the information by means of either interval data or fuzzy data instead of considering crisp/hard values [99]. Several statistical techniques have been introduced for fuzzy/interval data (see[11] for a review in statistics and econometrics). In a clustering framework, extensions of the FkM algorithm for fuzzy data are provided in [82, 47, 96, 74, 83, 97, 84, 1, 52, 23, 75, 43, 12, 38, 20, 42, 31, 25, 78]. If we consider intervals as a particular type of fuzzy data, we can find proposals of FkM-type algorithms in [39, 24, 21]. On the other hand, if intervals are managed as symbolic data [7], some fuzzy clustering methods are illustrated in [15, 17, 76, 16].

Among complex data, functional data also deserve attention. Advances in data collection and storage have led to an increasing amount of this kind of data. Hence, appropriate statistical methods are needed [79, 37]. Nowadays, it is usual to encounter functional data in many fields such as engineering, economics, finance biology, medicine, or meteorology. Some proposals of fuzzy clustering of functional data are based on transforming the data before applying a fuzzy k-means type algorithm (see, for example, [18] and [41]). A different proposal is provided in [89]. The authors suggest using a dissimilarity matrix that is itself a function. Consequently, the cluster prototypes and the membership degrees are also defined as functions.

The above proposals of clustering techniques are also known as one-mode (or one-way) clustering to distinguish them from two-mode clustering. Two-mode clustering is helpful to synthesize more complex data structures. In particular, we refer to data matrices whose two modes have an interchangeable role. The approach is symmetrical, and the aim is to look for blocks (or sub-matrices) characterized by internal cohesion and external separation. In some situations, for example, units can be similar only on a subgroup of features, and some variables can be associated only within a subgroup of units. This is the case of products and customers in market basket analysis or genes and samples in DNA microarray analysis. In this setting, an extension of the FkM is the Fuzzy double k-Means introduced in [36]. Further variants and robust versions are provided in [34].

The paper is structured as follows. Section 2 is devoted to the FkM method and its variants for numerical data. Furthermore, the proposals for relational data are briefly described in Subsection 2.1. Section 3 contains a review of the modifications of FkM for categorical or mixed data. Fuzzy clustering of fuzzy and interval data are reported in Section 4 and Subsection 4.1, respectively. The case of functional data is addressed in Section 5, whilst Section 6 contains some extensions of the FkM algorithm for two-mode clustering of a data matrix. Some applications are reported in Subsections 2.2, 4.2 and 5.1. Finally, some concluding remarks and open problems are addressed in Section 7.

2. Fuzzy *k*-Means and its variants

The fuzzy approach to clustering is based on a soft assignment of units to clusters. Most of the proposals consider object data, a typical unit-variable data matrix, with numerical variables. As already stated, the first and best known fuzzy clustering algorithm is the Fuzzy k-Means (FkM) [4, 5]. Given an $(n \times p)$ matrix, $\mathbf{X}$, where n and p are the number of units and variables, respectively, the aim is to partition the units into k groups, where each group is characterized by a prototype (centroid). Each row vector $\mathbf{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ip}]$ represents the i-th observation. The optimization problem of the FkM can be formalized as:

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{H}} J_{\mathrm{F}k\mathrm{M}} = \sum_{i=1}^{n} \sum_{g=1}^{k} u_{ig}^{m} d^2\left(\mathbf{x}_i, \mathbf{h}_g\right), \\ &\text{s.t.} \quad u_{ig} \in [0,1]\,, \quad i = 1, \ldots, n, \quad g = 1, \ldots, k, \\ &\qquad \sum_{g=1}^{k} u_{ig} = 1, \quad i = 1, \ldots, n, \end{aligned} \tag{1}$$

where $d^2\left(\mathbf{x}_i, \mathbf{h}_g\right) = \|\mathbf{x}_i - \mathbf{h}_g\|^2$ is the squared Euclidean distance between unit i and prototype g.

In (1), the $(n \times k)$ matrix $\mathbf{U}$ denotes the membership degree matrix, where each element $u_{ig} \in [0, 1]$ represents the membership degree of unit i to cluster g. The row-wise sums of $\mathbf{U}$ are equal to 1. The $(k \times p)$ matrix $\mathbf{H}$ denotes the prototype (centroid) matrix, where each row $\mathbf{h}_g = [h_{g1}, h_{g2}, \ldots, h_{gp}]$ $(g = 1, \ldots, k)$ is the prototype of cluster g. Finally, the parameter $m > 1$ is used to tune the fuzziness of the obtained partition. The greater the value of m, the further away from 1 and 0 the membership degrees are. For $m \to 1$ FkM reduces to the k-means (kM) algorithm. Empirical results depend on the selection of this parameter, which is commonly chosen between 1.5 and 2 in practice, as shown in [70].

The optimal solution can be found by means of an iterative algorithm, where the updates are obtained through the Lagrangian multiplier method. The centroid and the membership degree updates are

$$\mathbf{h}_g = \frac{\sum_{i=1}^{n} u_{ig}^m \mathbf{x}_i}{\sum_{i=1}^{n} u_{ig}^m}, \quad g = 1, \ldots, k, \tag{2}$$

and

$$u_{ig} = \frac{1}{\sum_{g'=1}^{k} \left(\frac{d^2(\mathbf{x}_i, \mathbf{h}_g)}{d^2\left(\mathbf{x}_i, \mathbf{h}_{g'}\right)} \right)^{\frac{1}{m-1}}}, \quad i = 1, \ldots, n, \quad g = 1, \ldots, k. \tag{3}$$

Starting from the FkM, several variants have been proposed in the last decades. Since the initial FkM usually identifies clusters of spherical shape, some of the variants are based on a non-Euclidean distance measure in order to overcome such a drawback. This is the case of the Gustafson and Kessel proposal [44]: the FkM with covariance matrices (FkM-GK). It consists in replacing the Euclidean distance with the Mahalanobis one:

$$d_M^2 (\mathbf{x}_i, \mathbf{h}_g) = (\mathbf{x}_i - \mathbf{h}_g)' \mathbf{F}_g (\mathbf{x}_i - \mathbf{h}_g),$$

where $\mathbf{F}_g$ is a symmetric and definite positive matrix. The optimal solution of $\mathbf{F}_g$ depends on the inverse of the fuzzy covariance matrix of the g-th cluster. To avoid numerical problems that may arise when updating $\mathbf{F}_g$, an improved version of the FkM-GK is suggested in [2].The improvement consists in constraining the condition number of $\mathbf{F}_g$ to be higher than a pre-specified threshold.

A further proposal to deal with different cluster shapes is provided in [40]. In this case, the Euclidean distance is replaced with a distance norm based on fuzzy maximum likelihood estimates: the Gauss distance.

Since the FkM algorithm involves the parameter m lacking a physical meaning, an alternative, based on an entropic regularization to replace the fuzziness parameter m, has been proposed in [64, 65]. The entropic FkM has been proven to be connected to the EM algorithm used for mixture models [45]. An entropic version of the FkM-GK is proposed in [28]. It should be used in the case of non-spherical shape clusters.

A limitation of the above FkM type algorithms is the assignment of the units to all the clusters with non-zero membership degrees. In order to overcome it, a generalization of the FkM algorithm is introduced in [58]. In particular, a polynomial fuzzifier is used in the minimization problem. In general, a fuzzifier function is a continuous, strictly increasing function $f : [0,1] \longrightarrow [0,1]$ with $f(0) = 0$ and $f(1) = 1$. In the FkM case, $f(u_{ig}) = u_{ig}^m$. The polynomial fuzzifier function is defined as $f(u_{ig}) = \left(\frac{1-\beta}{1+\beta}u_{ig}^2 + \frac{2\beta}{1+\beta}u_{ig}\right)$, with $\beta \in [0,1]$. For $\beta = 0$ we obtain the FkM with parameter m equal to 2 and for $\beta = 1$ the hard/classical kM.

Most of the above algorithms are implemented in the R package **fclust** [30, 33].

As it happens for the k-means type methods, their fuzzy versions are not robust to outliers. This is due to the unit-sum constraints of the membership degrees that force outliers to be assigned to clusters. Furthermore, since the centroids are weighted means, they are affected by anomalous points.

There exist several proposals of robust fuzzy clustering methods. A timid kind of robustification consists in replacing the centroids (means) with the medoids. The Fuzzy k-Medoids (FkMed) algorithm [60] is a generalization of the classical k-Medoids [56]. In [95], however, an alternative FkM is proposed, by replacing the Euclidean distance in (1) with a "robust" distance measure, the exponential distance: $d_{exp}\left(\mathbf{x}_i, \mathbf{h}_g\right) = 1 - \exp\left(-\gamma d\left(\mathbf{x}_i, \mathbf{h}_g\right)\right)$, where γ is a positive constant. Another robust approach consists in adding a noise cluster containing all the units considered as outliers [13]. It is important to highlight that the noise cluster is not a proper cluster characterized by homogeneity (compactness). On the other hand, a trimmed approach of fuzzy clustering is suggested in [38]. The idea is to trim a fixed proportion of observations. In this case, the trimming proportion is to be fixed, whilst in the previous one (noise cluster) we have to fix the distance of the observations from the noise prototype.

The possibilistic approach to clustering relaxes the unit-sum constraint, which boosts robustness. The membership degrees do not longer depend on the distance to all prototypes but only on the distance to the prototype of the belonging cluster. This explains why they are also called degrees of typicality. The best known possibilistic clustering algorithm is the possibilistic extension of the k-Means, the Possibilistic k-Means [61, 62]. Further possibilistic clustering methods are provided in [98, 86, 92]. The first one [98] is based on the exponential distance, robust to noise and outliers. The second one [86] is based on the introduction of a repulsion term, in order to overcome the risk of obtaining a trivial solution with coincident clusters [3]. The third proposal [92] is an extension of the method introduced in [98].

Finally, in order to take into account the benefits of the fuzzy and possibilistic approaches and to overcome their drawbacks, some hybridizations are proposed: the Fuzzy Possibilistic k-Means [71], the Possibilistic Fuzzy k-Means [72] and the Modified Fuzzy and Possibilistic k-Means [81], among others. The possibilistic and hybrid proposals are implemented in the R package **ppclust** [9].

2.1. Fuzzy clustering algorithms for relational data

In several practical applications, the information is not available in terms of object data but only as relational data. Relational data are pair-wise relations (similarity or dissimilarity/distance) between units. There exist different proposals of fuzzy clustering algorithms for such a kind of data. A Relational duals of the FkM algorithm (RFkM) is proposed in [48]. It is based on the following minimization problem:

$$\begin{aligned}
&\min_{\mathbf{U}} J_{\mathrm{RF}k\mathrm{M}} = \sum_{g=1}^{k} \frac{\sum_{i=1}^{n}\sum_{i'=1}^{n} u_{ig}^{m} u_{i'g}^{m} r(\mathbf{x}_i,\mathbf{x}_{i'})}{2\sum_{i=1}^{n} u_{ig}^{m}}, \\
&\text{s.t.} \quad u_{ig} \in [0,1], \quad i=1,\ldots,n, \quad g=1,\ldots,k, \\
&\qquad \sum_{g=1}^{k} u_{ig} = 1, \quad i=1,\ldots,n.
\end{aligned} \tag{4}$$

Unfortunately, the above method requires that $r(\mathbf{x}_i, \mathbf{x}_{i'})$ is the squared Euclidean distance. This is too restrictive in several situations. A generalization of the RFkM algorithm is introduced in [46]. It allows the use of arbitrary dissimilarity data and consists in a modification of the algorithm by means of transformation of the dissimilarity data. Further proposals are provided in [56] and in [47]. The first one is known as FANNY. It is characterized by a parameter of fuzziness $m = 2$ and the relational data usually are computed by using an L_1 norm. The second one illustrates how the FkM method can be applied to relational data. A modification of FANNY, called FRC, is proposed in [14]. In particular, a general fuzzifier is used and the relational data may be obtained from any dissimilarity measure. As for FkM, the optimal solution is obtained by means of a Lagrangian function where only the unit-sum constraints of the membership degrees are involved. When the relational data are Euclidean, the non-negativity condition of the membership degrees is automatically satisfied, as in FkM. In the case of non-Euclidean distances, neither RFkM nor FRC automatically satisfy that constraint. A solution is proposed in [14]. A robust version of fuzzy relational clustering is also provided by means of the introduction of a noise cluster [13]. Obviously, the FRC algorithm and its robust version can be applied to all kinds of data. Depending on the data, different dissimilarity matrices may be used as input argument of the function. This includes categorical, mixed or more complex data. In addition, even when data are numerical, several non-Euclidean distances can ben used in order to take into account, for example, different cluster shapes. In this respect, a proposal of fuzzy clustering for nonlinearly separable data based on the geodesic distance is introduced in [32].

2.2. FkM: a real-case study

This section is devoted to a real-case study. We consider a food balance sheet provided by the Food and Agriculture Organization (FAO). In particular, we analyze the “wheat and products” of European countries in 2018. The food balance sheet shows for each food item the sources of supply and its utilization. The aim is to find homogeneous groups of the 39 European countries characterized by similar behaviour related to production, imports and exports of wheat, fat supply quantity (g/capita/day), food supply (kcal/capita/day), food supply quantity (kg/capita/yr) and protein supply quantity (g/capita/day). The first three variables are normalized (divided) by population. We use the function `FKM` of the **fclust** package [33]. By inspecting the values of Fuzzy Silhouette [10], a cluster validity index used to evaluate the partition quality, the optimal number of clusters, corresponding to the maximum value of the index, is $k = 4$. The first cluster is composed by 7 countries: Belarus, Czechia, Denmark, Estonia, Latvia, Netherlands and Republic of Moldova. Austria, Bosnia and Herzegovina, Croatia, Finland, Germany, North Macedonia, Norway, Poland, Portugal, Slovakia, Spain, Sweden, Switzerland, United Kingdom of Great Britain and Northen Ireland are all contained in the second cluster. The third cluster contains Albania, Belgium, Greece, Iceland, Ireland, Italy, Luxembourg and Malta. The remaining 8 countries belong to the forth cluster: Bulgaria, France, Hungary, Lithuania, Romania, Russian Federation, Serbia and Ukraine.

The prototypes/centroids of the clusters are reported in Table 1. As we can notice, countries in Cluster 1 are characterized by average values of imports, exports and production of wheat and related products, and the lowest values of supplies. Cluster 4 contains countries with the highest production and exports and the lowest import quantity. Countries in Cluster 2 and Cluster 3 present a similar export quantity, but the first ones have a lower value of imports and higher production. Furthermore, Cluster 3 is also characterized by the highest values of fat, protein and food supply quantities.

By inspecting the membership degree matrix (not reported here for the sake of brevity), there are three countries not clear assigned (highest membership degree lower than 0.5): Belgium, Czechia and the Netherlands. In particular, Belgium is assigned to Cluster 3 with a membership degree equal to 0.37 and its membership degree to Cluster 2 is 0.36. Czechia has intermediate characteristics between Cluster 1 ($u_{ig} = 0.47$) and Cluster 2 ($u_{ig} = 0.46$). Finally, the Netherlands is assigned to Cluster 1 ($u_{ig} = 0.46$),

Table 1: FAO dataset: prototypes of the four clusters obtained with FkM

	Exports	Imports	Product.	PCFat	SupplyDay	SupplyYear	PCProtein
Clus 1	0.22	0.13	0.35	3.69	504.02	63.33	15.58
Clus 2	0.08	0.11	0.15	4.21	707.69	91.13	21.79
Clus 3	0.07	0.18	0.09	7.59	924.79	117.35	27.57
Clus 4	0.37	0.04	0.59	3.90	888.57	115.30	27.46

but the membership degree to Cluster 2 is 0.39. This can also be seen by looking at the feature values assumed by these three countries, reported in Table 2.

Table 2: Feature values of Belgium, Czechia and Netherlands.

	Exports	Imports	Product.	PCFat	SupplyDay	SupplyYear	PCProtein
Belgium	0.23	0.47	0.14	2.56	781	110.51	23.37
Czechia	0.21	0.05	0.41	4.80	647	84.36	18.44
Netherlands	0.09	0.43	0.06	2.46	584	68.42	18.30

3. Extensions of Fuzzy *k*-Means for categorical/mixed data

This section is devoted to fuzzy clustering proposals for categorical or mixed data. As already observed in Subsection 2.1, the fuzzy relational clustering methods can be applied by using specific dissimilarity/similarity measures for categorical or mixed data. However, specific algorithms exist. An extension of the k-Means algorithm to clustering large data sets with categorical variables is proposed in [50]. It is known as k-Modes. A fuzzy version is introduced in [51], the Fuzzy k-Modes (FkMo). It consists in using a simple matching measure for categorical data and replacing the means with the modes. The optimization problem is

$$\min_{\mathbf{U},\mathbf{H}} J_{\mathrm{F}k\mathrm{Mo}} = \sum_{i=1}^{n}\sum_{g=1}^{k} u_{ig}^{m} d_c\left(\mathbf{x}_i, \mathbf{h}_g\right),$$
$$\text{s.t.} \quad u_{ig} \in [0,1], \quad i = 1, \ldots, n, \quad g = 1, \ldots, k, \tag{5}$$
$$\sum_{g=1}^{k} u_{ig} = 1, \quad i = 1, \ldots, n,$$

where $d_c\left(\mathbf{x}_i, \mathbf{h}_g\right)$ is the simple matching dissimilarity measure. It is defined as

$$d_c\left(\mathbf{x}_i, \mathbf{h}_g\right) = \sum_{j=1}^{p} \delta(x_{ij}, h_{gj}), \tag{6}$$

where

$$\delta(x_{ij}, h_{gj}) = \begin{cases} 0, & \text{if } x_{ij} = h_{gj}, \\ 1, & \text{if } x_{ij} \neq h_{gj}. \end{cases} \tag{7}$$

The matrix $\mathbf{H}$ in (5) contains the modes. The main difference with the FkM algorithm is the update of $\mathbf{H}$. In detail, denoting with $DOM_j = \{a_j^{(1)}, a_j^{(2)}, \cdots, a_j^{(n_j)}\}$ the set of n_j categories of the j-th categorical variable, $j = 1, \cdots, p$, the loss function $J_{\mathrm{F}k\mathrm{Mo}}$ is minimized iff $h_{gj} = a_j^{(r)} \in DOM_j$ where

$$\sum_{i, x_{ij} = a_j^{(r)}} u_{ig}^{m} \geq \sum_{i, x_{ij} = a_j^{(t)}} u_{ig}^{m} \qquad 1 \leq t \leq n_j \tag{8}$$

for $1 \leq j \leq p$.

An extension of the FkMo is proposed in [57]. In particular, the clusters of categorical data are represented by means of fuzzy centroids instead of the hard ones.

In the case of mixed data (units described by both numerical and categorical variables) neither the FkM nor the FkMo can be used. For this reason, a combination of both methods is introduced [49]: the (hard) k-Prototype algorithm. A fuzzy extension, the Fuzzy k-Prototype (FkP), is proposed in

[54]. In this case, each unit vector is $\mathbf{x}_i = [x^r_{i1}, x^r_{i2}, \cdots, x^r_{ip_1}, x^c_{i,p_1+1}, \cdots x^c_{ip}]$, where the first p_1 are numerical variables, denoted by superscript r, and the remaining $p - p_1$ are the categorical ones, denoted by superscript c.

The minimization problem is

$$\begin{aligned} &\min_{\mathbf{U},\mathbf{H}} J_{\mathrm{F}k\mathrm{P}} = \sum_{i=1}^{n}\sum_{g=1}^{k} u_{ig}^m d_{rc}\left(\mathbf{x}_i, \mathbf{h}_g\right), \\ &\text{s.t.} \quad u_{ig} \in [0,1]\,, \quad i = 1,\ldots,n, \quad g = 1,\ldots,k, \\ &\qquad \sum_{g=1}^{k} u_{ig} = 1, \quad i = 1,\ldots,n, \end{aligned} \tag{9}$$

where $d_{rc}\left(\mathbf{x}_i, \mathbf{h}_g\right)$ is defined as

$$d_{rc}\left(\mathbf{x}_i, \mathbf{h}_g\right) = \sum_{j=1}^{p_1} (w_j(x^r_{ij} - h^r_{gj}))^2 + \sum_{j=p_1+1}^{p} \delta(x^c_{ij}, h^c_{gj}). \tag{10}$$

The membership degree update is the same as the FkM algorithm with dissimilarity $d_{rc}\left(\mathbf{x}_i, \mathbf{h}_g\right)$. The prototype consists of two parts. The first p_1 elements are computed as weighted means of the numerical variables (as in FkM) and the remaining $p - p_1$ ones are the fuzzy modes introduced in [57].

4. Fuzzy *k*-Means for fuzzy data and its variants

In various practical applications in economics, social science, biology, ecology or medical science, many useful variables are vague or imprecise, and it is easier to capture the vagueness/imprecision by means of more complex data than to discard it and obtain precise data. Imprecise data may be formalized by means of fuzzy numbers [99]. The space of fuzzy numbers, denoted by $\mathcal{F}_c(\mathbb{R})$, is composed by the mappings $\widetilde{U} : \mathbb{R} \to [0,1]$ such that for each $\alpha \in (0,1]$ the so-called α-level set (or α-cut) $\widetilde{U}_\alpha = \{x \in \mathbb{R} | U(x) \geq \alpha\}$ belongs to the class of nonempty compact intervals in $\mathbb{R}$ (denoted by $\mathcal{K}_c(\mathbb{R})$). The 0-level, U_0, is the closure of the support of $\widetilde{U}$.

The most used family of fuzzy numbers is the so-called class of *LR-fuzzy numbers*. An LR-fuzzy number $\widetilde{U}$ is determined by four values, $\widetilde{U} \equiv (c_1, c_2, r, l)_{LR}$. In detail, c_1 and c_2 are the left and the right centers of the

1-level of $\widetilde{U}$ and represent the location of the fuzzy number, while the right and left spreads of $\widetilde{U}$, r and l are associated with the imprecision of $\widetilde{U}$.

The membership degree of x to $\widetilde{U}$ is defined as

$$\mu_{\widetilde{U}}(x) = \begin{cases} L\left(\frac{c_1 - x}{l}\right), & \text{if } x < c_1, \\ 1, & \text{if } c_1 \leq x \leq c_2, \\ R\left(\frac{x - c_2}{r}\right), & \text{if } x > c_2, \end{cases} \tag{11}$$

where $L : \mathbb{R} \to [0,1]$ (and R) is a convex upper semi-continuous function so that $L(0) = 1$ and $L(x) = 0$, for all $x \in \mathbb{R} \setminus [0,1]$ (see [101]). If $L(z) = 1 - z$ and $R(z) = 1 - z$ for $0 \leq z \leq 1$, then $\widetilde{U}$ is a trapezoidal fuzzy number when $c_1 \neq c_2$ and a triangular fuzzy number when $c_1 = c_2 = c$. Furthermore, LR fuzzy numbers are a generalization of intervals. An interval is got when $c_1 \neq c_2$ and $l = r = 0$ (see Figure 3).

The usual arithmetic between fuzzy numbers is a level-wise extension of the standard non-linear arithmetic for intervals ([68, 99]). Given $\widetilde{U}, \widetilde{V} \in \mathcal{F}_c(\mathbb{R})$ and $\lambda \in \mathbb{R}$, the sum and the product by a scalar can be defined so that for each $\alpha \in [0,1]$ it is fulfilled that

$$\left(\widetilde{U} + \lambda\widetilde{V}\right)_\alpha = \widetilde{U}_\alpha + \lambda\widetilde{V}_\alpha = \left\{u + \lambda v : u \in \widetilde{U}_\alpha, v \in \widetilde{V}_\alpha\right\}. \tag{12}$$

Given n units described by p LR fuzzy variables, a fuzzy data matrix can be defined as $\widetilde{\mathbf{X}} = \left\{\widetilde{x}_{ij} \equiv (c_{1ij}, c_{2ij}, l_{ij}, r_{ij})_{LR}, i = 1, \ldots, n, j = 1, \ldots, p\right\}$, where $\widetilde{x}_{ij}$ is the value of the LR fuzzy variable j observed on the i-th unit with left center c_{1ij}, right center c_{2ij}, and left and right spreads l_{ij} and r_{ij}, respectively. The matrix $\widetilde{\mathbf{X}}$ can be characterized by four matrices: $\widetilde{\mathbf{X}} \equiv (\mathbf{C}_1, \mathbf{C}_2, \mathbf{L}, \mathbf{R})_{LR}$. Each observation i is expressed as $\widetilde{\mathbf{x}}_i \equiv (\mathbf{c}_{1i}, \mathbf{c}_{2i}, \mathbf{l}_i, \mathbf{r}_i)_{LR}$.

Let $d_F(\cdot, \cdot)$ be a distance measure between fuzzy numbers, the Fuzzy k-Means for Fuzzy data (FkM-F) is formalized by means of the following optimization problem:

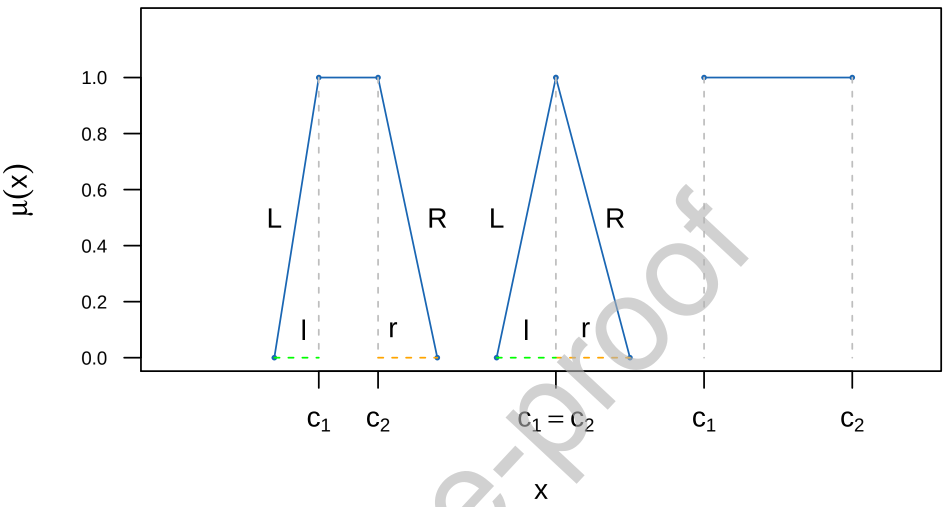

Figure 3: Three examples of LR fuzzy data: a trapezoidal fuzzy number (left), a triangular fuzzy number (center), an interval (right).

$$\begin{aligned} \min_{\mathbf{U},\widetilde{\mathbf{H}}} J_{FkM-F} &= \sum_{i=1}^{n}\sum_{g=1}^{k} u_{ig}^{m} d_F^2\left(\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{h}}_g\right), \\ \text{s.t.} \quad & u_{ig} \geq 0, \quad i = 1,\ldots,n, \quad g = 1,\ldots,k, \\ & \sum_{g=1}^{k} u_{ig} = 1, \quad i = 1,\ldots,n, \end{aligned} \tag{13}$$

where u_{ig} is the membership degree of observation i to cluster g, stored in the $(n\times k)$ matrix $\mathbf{U}$, and $\widetilde{\mathbf{H}} = \left\{\widetilde{h}_{gj} \equiv \left(h_{gj}^{C_1}, h_{gj}^{C_2}, h_{gj}^{L}, h_{gj}^{R}\right)_{LR}, g = 1, ..., k, j = 1, ..., p\right\}$ is the prototype matrix. In particular, $\widetilde{h}_{gj} \equiv \left(h_{gj}^{C_1}, h_{gj}^{C_2}, h_{gj}^{L}, h_{gj}^{R}\right)_{LR}$ represents the value of the j-th LR fuzzy variable of the g-th centroid with left center $h_{gj}^{C_1}$, right center $h_{gj}^{C_2}$, left spread h_{gj}^{L} and right spread h_{gj}^{R}. Hence, $\widetilde{\mathbf{h}}_g \equiv (\mathbf{h}_g^{C_1}, \mathbf{h}_g^{C_2}, \mathbf{h}_g^{L}, \mathbf{h}_g^{R})_{LR}$ is the fuzzy vector of length p for centroid g.

The proposal of FkM-F in [12] is based on a weighted dissimilarity measure for fuzzy data: $d_w^2(\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{x}}_{i'})$. Given two LR fuzzy observations, $\widetilde{\mathbf{x}}_i$ and $\widetilde{\mathbf{x}}_{i'}$, it is defined as a weighted sum of the squared Euclidean distances between centers and spreads:

$$\begin{aligned} d_w^2(\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{x}}_{i'}) = &\ w_C^2[d^2(\mathbf{c}_{1i}, \mathbf{c}_{1i'}) + d^2(\mathbf{c}_{2i}, \mathbf{c}_{2i'})] \\ + &\ w_S^2[d^2(\mathbf{l}_i, \mathbf{l}_{i'}) + d^2(\mathbf{r}_i, \mathbf{r}_{i'})], \end{aligned} \tag{14}$$

where $d(\cdot,\cdot)$ is the standard Euclidean distance (for non-fuzzy data), and w_C and w_S are weights for the center component and the spread component.

As for FkM, the iterative solution of the constrained quadratic minimization problem (13) is obtained through the Lagrangian multiplier method [12].

A particular case of the FkM-F method is introduced in [23] for LR_1 symmetric fuzzy data. A symmetric LR_1 fuzzy number is an LR fuzzy number with $c_1 = c_2$ and $l = r$. A weighted dissimilarity measure taking into account two components, the center and the spread, is also adopted in this case.

As for the numerical data case, since the dissimilarity (14) is based on the Euclidean distance, it does not allow us to recognize non-spherical shape clusters. To overcome this limitation, a generalization of $d_w^2(\cdot,\cdot)$ is introduced in [78]. It entails the cluster covariance matrices and is defined as

$$\begin{aligned} d_{M,w}^2(\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{x}}_{i'}) &= w_C^2[d_M^2(\mathbf{c}_{1i}, \mathbf{c}_{1i'}) + d_M^2(\mathbf{c}_{2i}, \mathbf{c}_{2i'})] \\ &+ w_S^2[d_M^2(\mathbf{l}_i, \mathbf{l}_{i'}) + d_M^2(\mathbf{r}_i, \mathbf{r}_{i'})], \end{aligned} \tag{15}$$

where $d_M(\cdot,\cdot)$ is the usual Mahalanobis distance.

In the case of outliers, as stated in Section 2, the fuzzy k-means type methods fail due to the unit-sum constraints of the membership degrees. When dealing with fuzzy data, we have to face with three kinds of outliers: outliers with respect to centers (location), outliers with respect to spreads (imprecision/size) and outliers with respect to both centers and spreads (location and imprecision/size).

Robust clustering methods are provided in [20]. The authors introduce a generalization of the FkMed [60] to the case of fuzzy data (FkMed-F), by using the dissimilarity $d_w^2(\cdot,\cdot)$ [12]. This is the starting point of the other three robust proposals. The first one is based on the following 'robust' (squared) distance measure:

$$d^2_{w-exp}(\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{x}}_{i'}) = 1 - \exp\left(-\gamma d^2_w(\widetilde{\mathbf{x}}_i, \widetilde{\mathbf{x}}_{i'})\right), \tag{16}$$

where γ is a positive constant determined according to the variability of the data. It is an extension of the distance introduced in [95]. The method is called the Smoothed Fuzzy k-Medoids for Fuzzy Data. The second proposal deals with outliers by means of the noise cluster [13] whilst the third one is a Trimmed Fuzzy k-Medoids for Fuzzy Data. In addition, a Fuzzy k-Medoids for fuzzy data based on a combination of Huber's M-estimator and Yager's ordered weighted averaging operators is provided in [25].

Further robust proposals for fuzzy data are addressed in [53] and in [100].

As for the numerical data, also in the case of fuzzy data, by relaxing the unit-sum constraints of the membership degrees we obtain the Possibilistic k-Means clustering method for Fuzzy data (PkM-F). In [31] the PkM-F has been defined by using the distance $d_w(\cdot, \cdot)$. There exists another possibilistic clustering method for LR fuzzy data proposed in [12] taking inspiration from [98]. The two proposals can be formulated in the same way except for the second term of the cost function.

In order to avoid the coincident cluster problem [3], a possibilistic clustering method with repulsion constraints for symmetric triangular fuzzy data is developed in [29]. A different strategy for preventing coincident clusters is the hybridization of the fuzzy and possibilistic approaches [31]. The last proposal exploits the benefits of both approaches, fuzzy and possibilistic. On the one hand, the fuzzy approach is helpful to find the best fuzzy partition. On the other hand, the possibilistic one helps us to identify outliers.

Finally, a mention should be made of the proposals of fuzzy clustering of fuzzy data based on hypothesis tests [43, 42].

4.1. Fuzzy k-Means for interval data and its variants

In several contexts, we face with interval data: fluctuations, ranges, grouped data, among others. An interval I is formalized by $[\inf_I, \sup_I]$. Intervals can be seen as a particular case of LR fuzzy sets. In particular, when $l = r = 0$, c_1 and c_2 represent the infimum and the supremum of the interval, respectively. An interval can also be formalized by means of the mid, $(c_1 + c_2)/2$, and the spread $(c_2 - c_1)/2$. The clustering methods reviewed in Section 4 can be obviously used for interval data. Some of them are previously introduced for this kind of data and then generalized to the case of fuzzy data. In a multidimensional setting, in the case of interval data, each observation is represented by a hyperrectangle in $\mathbb{R}^p$.

In [39] an FkM clustering algorithm for interval-valued (and fuzzy-valued) data is proposed. In particular, the authors propose to preprocess the data by a feature mapping technique.

A robust FkM clustering method for interval data is provided in [24]. The optimization problem of the FkM of interval data is analogous to that in (13) except for the dissimilarity measure. Furthermore, the authors adopt the noise cluster approach [13] to make robust the clustering algorithm and to take into account the presence of outliers. The same dissimilarity measure is also used in [21]. The authors propose two fuzzy clustering methods. The first one is the FkMed for interval-valued data, a timid kind of robustification. The second one is a more robust proposal: a trimmed FkMed algorithm.

For the sake of completeness, intervals can also be seen as a special case of symbolic data [7]. There exist several proposals of fuzzy k-means for symbolic interval data (see, for example, [15, 17, 76, 16]).

4.2. FkM-F: a real-case study

This section is devoted to an application of the FkM algorithm for fuzzy data. In particular, the data refers to a survey about the personal evaluation of a set of $n = 27$ students regarding different aspects of 9 specialized courses on Soft Computing which has been carried out in the European Centre for Soft Computing (Mieres, Spain) in 2008. The courses are: “Soft-Computing: A History of an Interdisciplinary Field”, “Fuzzy Set Theory-Fuzzy Systems”, “Neural Networks and Neuro-Fuzzy Systems”, “Evolutionary Computation and Genetic Fuzzy Systems”, “Probability and Statistics for Soft-Computing”, “Fuzzy Classification and Ensembles”, “Regression and System Modeling”, “Time Series”, “Frequent Item-Set Mining”. The students are asked to answer the questions of the survey by using trapezoidal fuzzy sets. Here we focus our attention on the overall rating of each course.

We partition the data into two clusters. By inspecting the membership degree matrix (not reported here for the sake of brevity), we can note that a cluster is composed by 9 students and the remaining 18 students belong to the other one. There are 3 students, 1 in Cluster 1 and 2 in Cluster 2, with a membership degree around 0.6, thus they have intermediate characteristics between the two clusters. Analyzing the following values of the prototypes, we can characterize the obtained partition:

$$\mathbf{H}^{C_1} = \begin{bmatrix} 53.73 & 63.96 & 49.24 & 72.68 & 57.43 & 63.37 & 67.50 & 74.41 & 55.03 \\ 67.70 & 72.12 & 72.43 & 81.81 & 76.31 & 70.39 & 75.15 & 84.24 & 74.74 \end{bmatrix}$$

$$\mathbf{H}^{C_2} = \begin{bmatrix} 57.63 & 69.72 & 55.82 & 82.16 & 62.26 & 70.50 & 72.23 & 81.24 & 61.19 \\ 75.50 & 78.45 & 77.53 & 88.35 & 80.67 & 77.91 & 82.75 & 90.09 & 81.66 \end{bmatrix}$$

$$\mathbf{H}^{L} = \begin{bmatrix} 10.82 & 8.28 & 7.92 & 8.35 & 8.67 & 10.82 & 8.57 & 9.08 & 9.64 \\ 7.37 & 7.51 & 8.44 & 7.17 & 7.84 & 7.54 & 7.26 & 7.26 & 6.66 \end{bmatrix}$$

$$\mathbf{H}^{R} = \begin{bmatrix} 13.01 & 9.79 & 8.15 & 8.31 & 9.01 & 11.66 & 9.56 & 8.52 & 11.12 \\ 6.99 & 6.49 & 7.44 & 6.28 & 7.56 & 7.64 & 7.03 & 6.05 & 5.89 \end{bmatrix}$$

In particular, the overall evaluation of the courses expressed by students in Cluster 1 are, in general, lower than those of students in Cluster 2. Looking at the values of the spreads (stored in $\mathbf{H}^L$ and $\mathbf{H}^R$) of the prototypes, we can note that the evaluations of students in Cluster 1 are more imprecise.

5. Fuzzy *k*-Means for functional data

In the last decades, functional data analysis has received a great deal of attention. Functional data encounter a complexity that is not easy to manage. The observations are supposed to be functions (on a continuous domain such as time or space), but in practice, the sampled curves are observed on a finite set of points. The usual methods for multivariate data are not suitable for functional data but can be applied to discrete measurements. In [22], for example, some fuzzy clustering algorithms are proposed for multivariate time-varying data, corresponding to discrete time data instead of functional data. On the other hand, there also exist proposals of fuzzy clustering methods for times series (see, for example, [26]). Although conceptually different, a time series can be seen as a realization of a functional random element: namely, when the time series generation process, involving concepts such as autocorrelation, is not considered, they are just (sampled) functions defined over a time domain, and clustering methods developed for functional data can be applied and conversely. In contrast, when the clustering method regards the time series generation process, it cannot be applied for general functional data.

Most of the clustering proposals for functional data are based on dissimilarities that do not depend on time t. So, the time dependency is neglected and an L_2 distance measure is used. Some dissimilarities for functional data are introduced in [88] and also in [37]. In this case, the FkM method and its variants can be easily adapted. On the other hand, functional data can be transformed and then the usual fuzzy clustering algorithms can be employed. This is, for example, the case of the proposals in [18] and [41]. In the first one, the FkM algorithm is used for clustering functional principal components of selected relevant process variables. In [41] the authors introduce an FkMed method for functional data based on two steps. In the first one, the functions are fitted to the observed data by means of B-splines. Secondly, the usual FkMed algorithm is applied to the B-spline coefficients obtained in the previous step.

A different approach is provided in [89]. In detail, the dissimilarity between functional data is considered as a function. The authors address crisp and fuzzy k-means clustering algorithms for multivariate functional data. The aim is to partition n objects, which are represented as a p-vector of continuous functions of $t \in [a, b]$,

$$\mathbf{x}_i(t) = (x_{i1}(t), x_{i2}(t), \cdots, x_{ip}(t)), \quad (i = 1, \cdots, n), \tag{17}$$

into k clusters. The optimization problem is

$$\begin{aligned}
\min_{\mathbf{U},\mathbf{H}} J_{\mathrm{F}k\mathrm{M-Func}} &= \int_a^b \sum_{i=1}^{n}\sum_{g=1}^{k} u_{ig}^m d^2(\mathbf{x}_i(t), \mathbf{h}_g(t))dt,\\
\text{s.t.} \quad u_{ig}(t) &\in [0,1], \quad i = 1,\dots,n, \quad g = 1,\dots,k,\\
\frac{1}{b-a}\int_a^b \sum_{g=1}^{k} u_{ig}(t)dt &= 1, \quad i = 1,\dots,n,
\end{aligned}$$

where $\mathbf{h}_g(t)$ is the centroid of the g-th cluster. By minimize the loss function, the obtained cluster prototypes

$$\mathbf{h}_g(t) = (h_{g1}(t), h_{g2}(t), \cdots, h_{gp}(t)) \quad (g = 1,\cdots,k,\ t \in [a,b])$$

and the membership degrees

$$u_{ig}(t) \quad (i = 1,\cdots,n,\ g = 1,\cdots,k,\ t \in [a,b])$$

are defined as functions of $t \in [a,b]$.

5.1. FkM-Func: a real-case study

This subsection includes the application of the FkM-Func algorithm to the Velib dataset. It contains data from the bike sharing system of Paris. The data are loading profiles of the bike stations over one week. The data were collected every hour during the period from Sunday, September 1, 2014, to Sunday, September 7, 2014. The data consists of the loading profiles (number of available bikes / number of bike docks) of the 1189 stations at 181 time points. We use the dataset `velib` contained in the R package **funFEM** [8]

The functions are transformed by means of B-splines and then the FkM algorithm is applied to B-spline coefficients. The optimal number of clusters according to the fuzzy silhouette index is $k = 4$. The obtained partition is composed of clusters of size 446, 266, 221 and 256, respectively. The

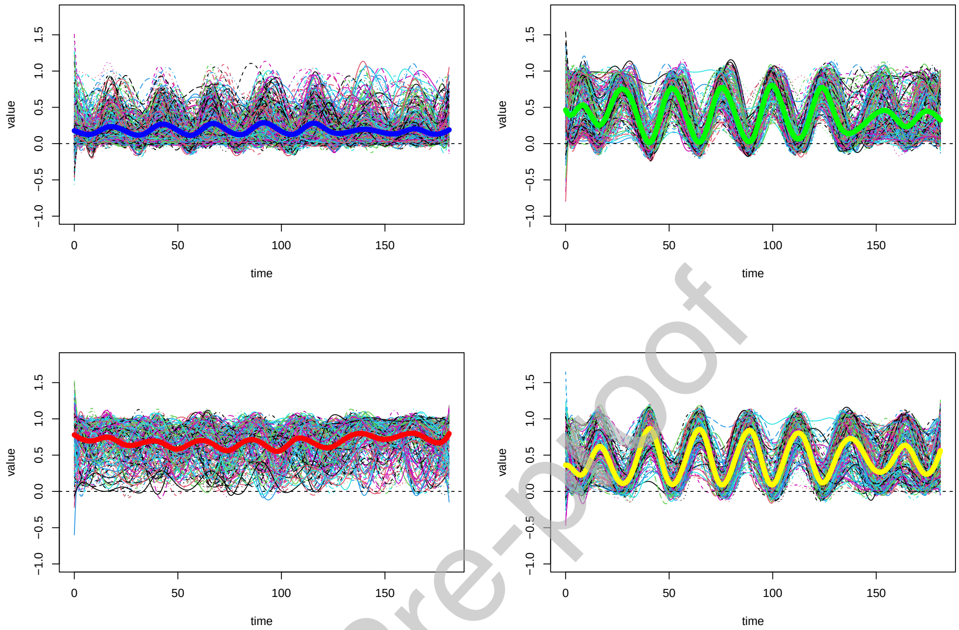

Figure 4: Velib dataset: Cluster 1 (top left), Cluster 2 (top right), Cluster 3 (bottom left) and Cluster 4 (bottom right), obtained with FkM-Func. The bold curves refer to the corresponding centroids.

4 clusters and the corresponding centroids (bold line) are represented in Figure 4.

The percentage of unclear assignments (object assigned with a membership degree lower than 0.5) is 2.19%. The membership degree plays an important role in recognizing objects (curves) with intermediate characteristics between two clusters. For example, in Figure 5 are reported two objects (black solid line and dashed red line) whose membership degrees to Cluster 1 are 0.497 and 0.497, and to Cluster 3 0.413 and 0.428, respectively.

6. Fuzzy Double *k*-Means

Double clustering, also known as biclustering, two-mode clustering or co-clustering, consists in simultaneously clustering modes (e.g., units, variables)

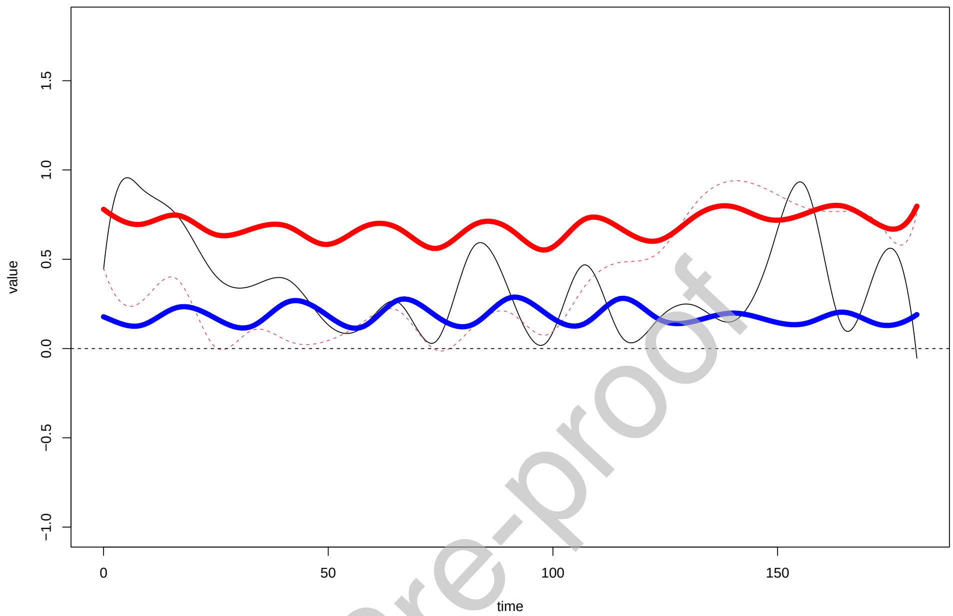

Figure 5: Velib dataset: two curves with unclear assignment (black solid line and dashed red line) and centroids of Cluster 1 and Cluster 3 (bold solid lines).

of a two-mode data matrix. Let $\mathbf{X}$ be a data matrix of order $(n \times p)$, the objective of a double clustering algorithm is to simultaneously partition n rows (e.g., units) into k clusters and p columns (e.g., variables) into c clusters.

Starting from the Double k-Means [90], the fuzzy version has been briefly introduced in [36] and deeply analyzed in [34]. The Fuzzy Double k-Means (FDkM) algorithm consists in solving the following constrained minimization problem:

$$\begin{aligned} J_{FDkM} &= \sum_{i=1}^{n} \sum_{j=1}^{p} \sum_{g=1}^{k} \sum_{f=1}^{c} \left(x_{ij} - h_{gf}\right)^2 u_{ig}^m v_{jf}^l, \\ \text{s.t.} \quad & u_{ig}, v_{jf} \in [0,1], \sum_{g=1}^{k} u_{ig} = 1, \sum_{f=1}^{c} v_{jf} = 1, \end{aligned} \tag{18}$$

where $\mathbf{U} = [u_{ig}]$ is the $(n \times k)$ membership degree matrix for the rows,

$\mathbf{V} = [v_{jf}]$ is the $(p \times c)$ membership degree matrix for the columns, and $\mathbf{H} = [h_{gf}]$ is the prototype matrix of order $(k \times c)$. In a two-mode setting, the prototypes have size depending on the numbers of clusters for rows and columns. In this way, the roles of rows and columns are interchangeable. The parameters $m > 1$ and $l > 1$ tune the fuzziness of the two partitions. The solution of (18) leads to a decomposition of the data matrix into kc blocks.

The FDkM algorithm includes some special cases. When m and l tend to 1, the FDkM solution approaches the DkM one. When $c = p$, the columns are not partitioned and FDkM reduces to the FkM algorithm. Furthermore, if $c = p$ and m tends to 1, it corresponds to the standard k-Means.

In [34], a more general Fuzzy Double k-means algorithm with polynomial fuzzifiers is also addressed, and robust versions are illustrated. The robustness is covered through the noise cluster approach, but we have to stress that, in this case, the structure is more complex, so three noise clusters are considered.

There also exist in literature proposals of fuzzy double clustering for categorical datasets. In this case, the data are stored in tables, whose rows and columns are the units and the categories of all the variables, respectively. These tables are known as cross-classification tables, contingency tables or in general co-occurrence matrices. In [69] the fuzziness is represented by an entropy regularization. The constraints on the unit membership degrees and on the category ones are different. For each unit, the constraint is the usual one: the sum of its membership degrees to all the clusters is equal to 1. On the other hand, for each cluster, the sum of the membership degrees of all the categories to this cluster have to be equal to 1. This leads to optimize the loss function when only one variable in each cluster is completely relevant and the remaining ones are irrelevant. Hence, the proposal can be seen as a variable selection procedure rather than a clustering of variables. In order to overcome numerical instabilities of the above method in the presence of large numbers of units and categories, a further proposal is introduced in [63]. Finally, in [87], a single term fuzzifier is used in the optimization problem.

7. Concluding remarks and future directions

Starting from the FkM method, first several variants together with robust proposals for standard object numerical data have been reviewed. Extensions for different kinds of data are described. In particular, the case of categorical and mixed data are considered. Then, more complex data structures, such as fuzzy, interval or functional data, are encountered. The last part is devoted to double clustering of heterogeneous datasets characterized by blocks of rows and columns. It consists in simultaneously clustering rows and columns. Even if there are huge amount of papers on this topic, there are still several open problems to deal with.

Data grow so large and complex that it becomes crucial to adapt the existing method to them. On the one hand, the algorithms have to take into account memory limitations. In a classical clustering setting, a memory-efficient k-means algorithm is implemented in the R package **biganalytics** [27] (related to R package **bigmemory** [55]). It would be interesting to address a fuzzy version of it. On the other hand, the clustering methodology have to be adapted to the intrinsic characteristics of the data. In terms of data complexity, there is increasing interest, for example, in network data. Networks represent a powerful model to describe problems and applications in various fields, such as economics, biology and technology, among others. Networks can be formalized in several ways and, depending on the formalization, extensions of FkM type algorithms can be introduced.

Furthermore, the complexity of a structure also refers to heterogeneous two-mode or multi-mode datasets, whose features are both categorical and numerical ones. In this connection, some proposals could be found in [91, 73]. It would be very useful to provide two-mode or three-mode FkM type algorithms.

References

[1] Auephanwiriyakul, S., Keller, J.M., 2002. Analysis and efficient implementation of a linguistic fuzzy c-means. IEEE Transactions on Fuzzy Systems 10, 563–582.

[2] Babuska, R., Van der Veen, P.J., Kaymak, U., 2002. Improved covariance estimation for Gustafson-Kessel clustering. In: Proceedings of the 2002 IEEE International Conference on Fuzzy Systems, pp. 1081–1085.

[3] Barni, M., Cappellini, V., Mecocci, A., 1996. Comments on 'A possibilistic approach to clustering'. IEEE Transactions on Fuzzy Systems 4, 393–396 .

[4] Bezdek, J.C., 1974. Cluster validity with fuzzy sets. Journal of Cybernetics 3, 58–73.

[5] Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithm. Plenum Press, New York.

[6] Bock, H.H., 2007. Clustering methods: a history of k-means algorithms. Selected contributions in data analysis and classification, pp. 161–172.

[7] Bock, H.H., Diday, E., 2000. Analysis of symbolic data. Exploratory methods for extracting statistical information from complex data, Springer-Verlag, Heidelberg.

[8] Bouveyron, C., Come, E., Jacques, J., 2014. The discriminative functional mixture model for the analysis of bike sharing systems, Preprint HAL n.01024186, University Paris Descartes.

[9] Cebeci, Z., 2019. Comparison of internal validity indices for fuzzy clustering. Journal of Agricultural Informatics 10, 1–14.

[10] Campello, R.J.G.B., Hruschka, E.R., 2006. A fuzzy extension of the silhouette width criterion for cluster analysis. Fuzzy Sets and Systems 157, 2858–2875.

[11] Colubi, A., 2021. Fuzzy sets and (fuzzy) random sets in Econometrics and Statistics. Econometrics and Statistics (in press.)

[12] Coppi, R., D'Urso, P., Giordani, P., 2012. Fuzzy and possibilistic clustering for fuzzy data. Computational Statistics & Data Analysis 56, 915–927.

[13] Davé, R. N., 1991.Characterization and detection of noise in clustering. Pattern Recognition Letters 12, 657–664.

[14] Davé, R. N., Sen, S., 2002. Robust fuzzy clustering of relational data. IEEE Transactions on Fuzzy Systems 10, 713–727.

[15] de Carvalho, F.A.T., 2007. Fuzzy c-means clustering methods for symbolic interval data. Pattern Recognition Letters 28, 423–437.

[16] de Carvalho, F.A.T., Simoes, E.C., 2017. Fuzzy clustering of interval-valued data with City-Block and Hausdorff distances. Neurocomputing 266, 659–673.

[17] de Carvalho, F.A.T., Tenorio, C.P., 2010. Fuzzy k-means clustering algorithms for interval-valued data based on adaptive quadratic distances. Fuzzy Sets and Systems 161, 2978–2999.

[18] Di Maio, F., Secchi, P., Vantini, S., Zio, E., 2011. Fuzzy C-Means Clustering of Signal Functional Principal Components for Post-Processing Dynamic Scenarios of a Nuclear Power Plant Digital Instrumentation and Control System. IEEE Transactions on Reliability 60, NO. 2.

[19] Dunn, J., 1973. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters, Cybernetics and Systems 3, 32–57.

[20] D'Urso, P., De Giovanni, L., 2014. Robust clustering of imprecise data. Chemometrics and Intelligent Laboratory Systems 136, 58–80.

[21] D'Urso, P., De Giovanni, L., Massari, R., 2015. Trimmed fuzzy clustering for interval-valued data. Advances in Data Analysis and Classification 9, 21–40.

[22] D'Urso, P., De Giovanni, L., Massari, R., 2018. Robust fuzzy clustering of multivariate time trajectories. International Journal of Approximate Reasoning 99, 12–38.

[23] D'Urso, P., Giordani, P., 2006. A weighted fuzzy c-means clustering model for fuzzy data. Computational Statistics & Data Analysis 50, 1496–1523.

[24] D'Urso, P., Giordani, P., 2006. A robust fuzzy k-means clustering model for interval valued data. Computational Statistics 21, 255–269.

[25] D'Urso, P., Leski, J.M., 2020. Fuzzy clustering of fuzzy data based on robust loss functions and ordered weighted averaging. Fuzzy Sets and Systems 389, 1–28.

[26] D'Urso, P., Maharaj, E.A., Alonso, A.M., 2017. Fuzzy clustering of time series using extremes. Fuzzy Sets and Systems 318, 56–79.

[27] Emerson, J.W., Kane, M. J., Chandra, S., 2020. biganalytics: Utilities for 'big.matrix' Objects from Package 'bigmemory'. Version: 1.1.21 `https://cran.r-project.org/web/packages/biganalytics/index.html`

[28] Ferraro, M.B., Giordani, P., 2013. A new fuzzy clustering algorithm with entropy regularization In: Proceedings of the 9th meeting of the Classification and Data Analysis Group, pp. 195–198.

[29] Ferraro, M.B., Giordani, P., 2013. On possibilistic clustering with repulsion constraints for imprecise data. Information Sciences 245, 63–75.

[30] Ferraro, M.B., Giordani, P., 2015. A toolbox for fuzzy clustering using the R programming language. Fuzzy Sets and Systems 279, 1–16.

[31] Ferraro, M. B., Giordani, P., 2017. Possibilistic and fuzzy clustering methods for robust analysis of non-precise data. International Journal of Approximate Reasoning 88, 23–38.

[32] Ferraro, M.B., Giordani, P., 2019. A review and proposal of (fuzzy) clustering for nonlinearly separable data. International Journal of Approximate Reasoning 115, 13–31.

[33] Ferraro, M.B., Giordani, P., Serafini, A., 2019. fclust: an R package for fuzzy clustering. The R Journal 11, 205–233.

[34] Ferraro, M. B., Giordani, P., Vichi, V., 2021. A class of two-mode clustering algorithms in a fuzzy setting. Econometrics and Statistics 18, 63–78.

[35] Ferraro, M. B., Giordani, P., 2020. Soft Clustering. Wiley Interdisciplinary Reviews: Computational Statistics 12, e1480.

[36] Ferraro, M.B., Vichi, M., 2015. Fuzzy double clustering: a robust proposal. In: Grzegorzewski P., Gagolewski M., Hryniewicz O., Gil M. (Eds.), Advances in Intelligent Systems and Computing 315, Springer, Cham., pp 225–232.

[37] Ferraty, F. and Vieu, P., 2006. Non Parametric Functional Data Analysis. Theory and Practice. Springer, Berlin.

[38] Fritz, H., Garcia-Escudero, L. A., Mayo-Iscar, A., 2013. Robust constrained fuzzy clustering. Information Sciences 245, 38–52.

[39] Gao, X., Ji, H., Xie, W., 2000. A novel FCM clustering algorithm for interval-valued data and fuzzy-valued data. In: 5th International Conference on Signal Processing Proceedings. 16th World Computer Congress 2000, Beijing, China, pp. 1551-1555.

[40] Gath, I., Geva, A.B., 1989. Fuzzy clustering for the estimation of the parameters of the components of mixtures of normal distributions. Pattern Recognition Letters 9, 77–86.

[41] Giordani, P., Perna, S., Bianchi, A., Pizzulli, A., Tripodi, S., Matricardi, P. M., 2020. A study of longitudinal mobile health data through fuzzy clustering methods for functional data: The case of allergic rhinoconjunctivitis in childhood. PLoS One 15, e0242197.

[42] Giordani, P., Ramos-Guajardo, A.B., 2016. A fuzzy clustering procedure for random fuzzy sets. Fuzzy Sets and Systems 305, 54–69.

[43] Gonzalez-Rodriguez, G., Colubi, A., D'Urso, P., Montenegro, M., 2009. Multi-sample test-based clustering for fuzzy random variables. International Journal of Approximate Reasoning 50, 721–731.

[44] Gustafson, D.E., Kessel, W.C., 1979. Fuzzy clustering with a fuzzy covariance matrix. In: Proceedings of the 1978 IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes, pp. 761-766.

[45] Hathaway, R.J., 1986. Another interpretation of the EM algorithm for mixture distributions. Statistics & Probability Letters 4, 53–56.

[46] Hathaway, R.J., Bezdek, J.C., 1994. NERF c-means: Non-Euclidean relational fuzzy clustering. Pattern Recognition 27, 429–437.

[47] Hathaway, R.J., Bezdek, J.C., Devenport, J.W., 1996. On relational data versions of c-means algorithms. Pattern Recognition Letters 17, 607–612.

[48] Hathaway, R.J., Devenport, J.W., Bezdek, J.C., 1989. Relation duals of the c-means clustering algorithms. Pattern Recognition 22, 205–212.

[49] Huang, Z., 1997. Clustering large data sets with mixed numeric and categorical values. In: Proceedings of the First Pacific Asia Knowledge Discovery and Data Mining Conference, World Scientific, Singapore, pp. 21–34.

[50] Huang, Z., 1998. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. Data Mining and Knowledge Discovery 2, 283–304.

[51] Huang, Z., Ng, M.K., 1999. A Fuzzy k-Modes Algorithm for Clustering Categorical Data. IEEE Transactions on Fuzzy Systems 7, n. 4.

[52] Hung, W.-L., Yang, M.-S., 2005. Fuzzy clustering on LR-type fuzzy numbers with an application in Taiwanese tea evaluation. Fuzzy Sets and Systems 150, 561–577.

[53] Hung, W., Yang, M., Lee, E., 2010. A robust clustering procedure for fuzzy data. Computers & Mathematics with Applications 60, 151–165.

[54] Ji, J., Pang, W., Zhou, C., Han, X., Wang, Z., 2012. A fuzzy k-prototype clustering algorithm for mixed numeric and categorical data. Knowledge-Based Systems 30, 129–135.

[55] Kane, M. J., Emerson, J., Weston, S., 2013. Scalable Strategies for Computing with Massive Data. Journal of Statistical Software 55, 1–19.

[56] Kaufman, L., Rousseuw, P. J., 1990. Finding Groups in Data: An Introduction to Cluster Analysis. Wiley, New York.

[57] Kim, D.-W., Lee, K. H., Lee, F., 2004. Fuzzy clustering of categorical data using fuzzy centroids. Pattern Recognition Letters 25, 1263–1271.

[58] Klawonn, F., Höppner, F., 2003. An alternative approach to the fuzzifier in fuzzy clustering to obtain better clustering. In: Proceedings of Eusflat Conference, pp. 730–734.

[59] Klawonn, F., Kruse, R., Winkler, R., 2015. Fuzzy clustering: more than just fuzzification. Fuzzy Sets and Systems 281, 272–279.

[60] Krishnapuram, R., Joshi, A., Nasraoui, O., Yi, L., 2001. Low-complexity fuzzy relational clustering algorithms for web mining. IEEE Transactions on Fuzzy Systems 9, 595–607.

[61] Krishnapuram, R., Keller, J. M., 1993. A possibilistic approach to clustering. IIEEE Transactions on Fuzzy Systems 1, 98–110.

[62] Krishnapuram, R., Keller, J. M., 1996. The possibilistic c-means algorithm: insights and recommendations. IEEE Transactions on Fuzzy Systems 4, 385–393.

[63] Kummamuru K., Dhawale A., Krishnapuram R., 2003. Fuzzy Co-clustering of Documents and Keywords. In: Proceedings of the 12th IEEE International Conference on Fuzzy Systems, 2003, pp. 772–777.

[64] Li, R.-P., Mukaidono, M., 1995. A maximum-entropy approach to fuzzy clustering. In: Proceedings of 1995 IEEE International Conference on Fuzzy Systems, pp. 2227-2232.

[65] Li, R.-P., Mukaidono, M., 1999. Gaussian clustering method based on maximum-fuzzy-entropy interpretation. Fuzzy Sets and Systems 102, 253–258.

[66] MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics and Probability 1, pp. 281–298.

[67] Nayak, J., Naik, B., Behera, H., 2015. Fuzzy C-means (FCM) clustering algorithm: a decade review from 2000 to 2014. Computational intelligence in data mining 2, 133–149.

[68] Nguyen, H.T., 1978. A note on the extension principle for fuzzy sets. Journal of Mathematical Analysis and Applications 64, 369–380.

[69] Oh C.H., Honda K., Ichihashi H., 2001. Fuzzy Clustering for Categorical Multivariate Data. In: Proceedings of Joint 9th IFSA World Congress and 20th NAFIPS International Conference, pp. 2154–2159.

[70] Pal, N. R., Bezdek, J. C., 1995. Cluster validity for the fuzzy c-means model. IEEE Transactions on Fuzzy Systems 3, 370–379.

[71] Pal, N.R., Pal, K., Bezdek, J.C., 1997. A mixed c-means clustering model. In: Proceedings of FUZZ-IEEE'97, pp. 11-21.

[72] Pal, N.R., Pal, K., Keller, J.M., Bezdek, J.C., 2005. A possibilistic fuzzy c-means clustering algorithm. IEEE Transactions on Fuzzy Systems 13, 517–530.

[73] Papalexakis, E.E., Sidiropoulos, N. D., Bro, R., 2013. From K-Means to Higher-Way Co-Clustering: Multilinear Decomposition With Sparse Latent Factors. IEEE Transactions on Signal Processing 61, 493–506.

[74] Pedrycz, W., Bezdek, J.C., Hathaway, R.J., Rogers, G.W., 1998. Two nonparametric models for fusing heterogeneous fuzzy data. IEEE ransactions on Fuzzy Systems 6, 411–425.

[75] Pelekis, N., Iakovidis, D. K., Kotsifakos, E. E., Kopanakis, I., 2008. Fuzzy clustering of intuitionistic fuzzy data. International Journal of Business Intelligence and Data Mining 3, No. 1.

[76] Pimentel, B. A., de Souza, R.M.C.R., 2014. A weighted multivariate Fuzzy C-Means method in interval-valued scientific production data. Expert Systems with Applications 41, 3223–3236.

[77] Puri, M.L., Ralescu, D.A., 1986. Fuzzy random variables. Journal of Mathematical Analysis and Applications 114, 409–422.

[78] Ramos-Guajardo, A.B., Ferraro, M.B., 2020. A fuzzy clustering approach for fuzzy data based on a generalized distance. Fuzzy Sets and Systems 389, 29–50.

[79] Ramsay, J., Silverman, B.W., 2005. Functional Data Analysis (2nd edn.). Springer, New York. .

[80] Roubens, M., 1978. Pattern classification problems and fuzzy sets. Fuzzy Sets and Systems 1, 239–253.

[81] Saad, M.F., Alimi, A.M., 2009. Modified fuzzy possibilistic c-means. In: Proceedings of the International Multiconference of Engineers and Computer Scientists, pp. 18–20.

[82] Sato, M., Sato,Y., 1995. Fuzzy clustering model for fuzzy data. In: Proceedings of 1995 IEEE International Conference on Fuzzy Systems, pp. 2123–2128.

[83] Takata, O., Miyamoto, S., Umayahara, K., 1998. Clustering of data with uncertainties using Hausdorff distance. In: Proceedings of the 2nd IEEE International Conference on Intelligence Processing Systems, pp. 67–71.

[84] Takata, O., Miyamoto, S., Umayahara, K., 2001. Fuzzy clustering of data with uncertainties using minimum and maximum distances based on L1 metric. In: Proceedings of Joint 9th IFSA World Congress and 20th NAFIPS International Conference, pp. 2511–2516.

[85] Tang, Y. G., Sun, F. C., Sun, Z. Q., 2005. Improved validation index for fuzzy clustering. In: Proceedings of the 2005 American Control Conference, pp. 1120–1125.

[86] Timm, H., Borgelt, C., Döring, C., Kruse, R., 2004. An extension to possibilistic fuzzy cluster analysis. Fuzzy Sets and Systems 147, 3–16.

[87] Tjhi, W.-C., Chen, L., 2005. Fuzzy Co-clustering of Web Documents. In: Proceedings of the 2005 International Conference on Cyberworlds.

[88] Tokushige, S., Inada, K., Yadohisa, H., 2002. Dissimilarity and related methods for functional data. Journal of Japanese Society of Computational Statistics 15, 319–326.

[89] Tokushige, S., Yadohisa, H., Inada, K., 2007. Crisp and fuzzy k-means clustering algorithms for multivariate functional data. Computational Statistics 22, 1–16.

[90] Vichi, M., 2001. Double k-means clustering for simultaneous classification of objects and variables. In: Borra, S., Rocci, R., Vichi, M., Schader, M. (eds.) Advances in Classification and Data Analysis. Studies in Classification, Data Analysis, and Knowledge Organization, pp. 43–52. Springer, Berlin, Heidelberg.

[91] Vichi, M., Rocci, R., Kiers, H.A.L., 2007. Simultaneous Component and Clustering models for three-way data: Within and Between Approaches. Journal of Classification 24, 71–98.

[92] Wachs, J., Shapira, O., Stern, H., 2006. A method to enhance the 'possibilistic C-means with repulsion' algorithm based on cluster validity index. In: Abraham, A., de Baets ,B., Köppen, M., Nickolay, B. (eds) Applied

Soft Computing Technologies: The Challenge of Complexity. Advances in Soft Computing, vol 34. Springer, Berlin, Heidelberg.

[93] Windham, M.P., 1985. Numerical classification of proximity data with assignment measures. Journal of Classification 2,157–172.

[94] Winkler, R., Klawonn, F., Kruse, R., 2011. Fuzzy clustering with polynomial fuzzifier function in connection with m-estimators. Applied and Computational Mathematics 10, 146–163.

[95] Wu, K.-L., Yang, M.-S., 2002. Alternative c-means clustering algorithms. Pattern Recognition 35, 2267–2278.

[96] Yang, M.S., Ko, C.H., 1996. On a class of fuzzy c-numbers clustering procedures for fuzzy data. Fuzzy Sets and Systems 84, 49–60.

[97] Yang, M.S., Liu, H.H., 1999. Fuzzy clustering procedures for conical fuzzy vector data. Fuzzy Sets and Systems 106, 189–200.

[98] Yang, M.-S., Wu, K.-L., 2006. Unsupervised possibilistic clustering. Pattern Recognition 39, 5–21.

[99] Zadeh, L.A., 1965. Fuzzy sets. Information and Control 8, 338–353.

[100] Zarandi, M. ., Razaee, Z.S., 2011. A fuzzy clustering model for fuzzy data with outliers. International Journal of Fuzzy Systems 1, 29–42.

[101] Zimmermann, H.J., 1996. Fuzzy Set Theory and its Applications (3rd edition). Kluwer Academic Publishers.