



## Soft clustering - fuzzy and rough approaches and their extensions and derivatives

Georg Peters<sup>a,b,\*</sup>, Fernando Crespo<sup>c</sup>, Pawan Lingras<sup>d</sup>, Richard Weber<sup>e</sup>

<sup>a</sup> Munich University of Applied Sciences, Department of Computer Science and Mathematics, Munich, Germany

<sup>b</sup> Australian Catholic University, School of Arts and Sciences, North Sydney, Australia

<sup>c</sup> Universidad de Valparaíso, Escuela de Ingeniería Industrial, Santiago, Chile

<sup>d</sup> Saint Mary's University, Department of Mathematics and Computer Science, Halifax, Nova Scotia, Canada

<sup>e</sup> Universidad de Chile, Department of Industrial Engineering, Santiago, Chile

### ARTICLE INFO

#### Article history:

Received 8 February 2012

Received in revised form 15 October 2012

Accepted 15 October 2012

Available online 25 October 2012

#### Keywords:

k-Means

Fuzzy c-means

Rough k-means

Hybrid Soft Clustering

### ABSTRACT

Clustering is one of the most widely used approaches in data mining with real life applications in virtually any domain. The huge interest in clustering has led to a possibly three-digit number of algorithms with the k-means family probably the most widely used group of methods. Besides classic bivalent approaches, clustering algorithms belonging to the domain of soft computing have been proposed and successfully applied in the past four decades. Bezdek's fuzzy c-means is a prominent example for such soft computing cluster algorithms with many effective real life applications. More recently, Lingras and West enriched this area by introducing rough k-means. In this article we compare k-means to fuzzy c-means and rough k-means as important representatives of soft clustering. On the basis of this comparison, we then survey important extensions and derivatives of these algorithms; our particular interest here is on hybrid clustering, merging fuzzy and rough concepts. We also give some examples where k-means, rough k-means, and fuzzy c-means have been used in studies.

© 2012 Elsevier Inc. All rights reserved.

## 1. Introduction

The interest in data mining has been growing rapidly in the past decades. Mainly driven by the increasing ability of modern information technology the amount of data collected has increased substantially. Obviously, this leads to an evident need to analyze these data and derive valuable new insights out of them.

This phenomenon is not limited to a special industry or application but can be observed in virtually any industry and almost any field of application, covering such diverse areas like marketing, bioinformatics, or engineering.

Clustering is among the most widely used techniques in data mining addressing this challenge [34, 106]. The objective of clustering is to group similar objects together in the same cluster while dissimilar objects should belong to different clusters.

Classic k-means [33, 57] is probably still the most frequently used clustering algorithm. It unambiguously assigns objects to a predefined number of clusters. Therefore, the classic k-means is also called hard k-means (HKM) in differentiation to soft computing approaches. These soft clustering techniques are characterized by a relaxation of the hard borders of k-means towards soft constraints. The soft borders address some particular challenges in many typical real-life applications where overlapping clusters, outliers or uncertain cluster memberships can often be observed.

A well established soft clustering approach is Bezdek's [7, 8] fuzzy c-means (FCM). Ten years ago, Lingras and West [51] enriched this field by introducing rough k-means (RKM).

\* Corresponding author at: Munich University of Applied Sciences, Department of Computer Science and Mathematics, Munich, Germany.  
E-mail address: [georg.peters@cs.hm.edu](mailto:georg.peters@cs.hm.edu) (G. Peters).

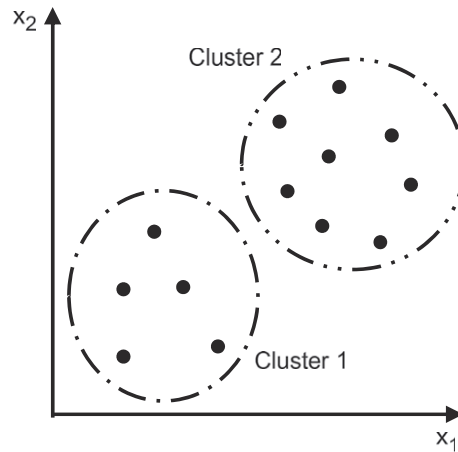


Fig. 1. Grouping objects in clusters.

In the meantime several *extensions* and *derivatives* of these clustering approaches have been suggested. In the context of our paper we define extensions and derivatives as follows:

- *Extensions* are modifications of the clustering algorithm within the same uncertainty domain.<sup>1</sup> For example, we consider a dynamic or an evolutionary version of one of these algorithms as an extension.<sup>2</sup>
- *Derivatives* are modifications that change the uncertainty dimension of the underlying clustering algorithm. E. g. Krishnapuram and Keller's [40] possibilistic *c*-means (PCM) has been directly derived from fuzzy *c*-means; the kind of uncertainty used in FCM has been generalized toward a possibilistic nature in PCM.

There have already been some comparative studies on the relationships among these clustering algorithms. The link between *k*-means and fuzzy *c*-means has been addressed in several publications, see for example Mingoti and Lima [69]. Krishnapuram and Keller [40] discussed the relationship between fuzzy and possibilistic *c*-means and Lingras et al. [53] compared conventional and rough *k*-means while Joshi et al. [36] correlated fuzzy and rough clustering. However, a current comparison and survey of soft clustering techniques, including recent hybridizations is most desirable.

The remainder of the article is organized as follows. In Section 2 we discuss the properties of hard, fuzzy, and rough clustering, including examples for extensions of these algorithms. In the subsequent section we survey derivatives of fuzzy and rough clustering approaches. Then, in Section 4, we overview application areas regarding the fundamental algorithms discussed in Section 2. The article concludes with a summary in Section 5.

## 2. Foundations of clustering

### 2.1. Introductory notes

The fundamental objective of clustering is to group similar objects in the same cluster and dissimilar objects in different clusters (see Fig. 1).

A common categorization is to distinguish between hierarchical and partitive clustering [37]. In this article, we concentrate on partitive clustering, in particular the *k*-means algorithm [57] and some of its important soft computing derivatives: Bezdek's fuzzy *c*-means [6–8] and Lingras' rough *k*-means [51]. We limit our analysis to objects that are described exclusively by numeric features where, e.g., an Euclidean distance between two objects can be defined. A discussion of more general distance measures for non-numeric features would go beyond the scope of this paper; the interested reader is referred to, e.g., Xu and Wunsch [106].

### 2.2. Algorithmic structure

The partitive clustering approaches that we survey in the paper have similar algorithmic structures as depicted in Fig. 2. In the initialization phase, the number of clusters  $K$  has to be set and further parameters where applicable (e.g., the fuzzifier parameter in fuzzy *c*-means or weights in rough *k*-means). Then the objects have to be assigned to the clusters (e.g., randomly).

<sup>1</sup> We are aware of the ongoing intense discussion of the term "uncertainty", in particular on the relationship of it with respect to probability and soft computing. It would go beyond the scope of the article discussing it here. Hence, we use "uncertainty" in an opportunistic way.

<sup>2</sup> In the meantime a huge number of extensions have been suggested in particular for *k*-means and fuzzy *c*-means. Hence, in the particular sections of our paper we can just give almost arbitrary examples for extensions without demanding completeness.

**Initialization:**

- Set initial parameters:  
the number of clusters  $K$  and  
further parameters where applicable.
- Assign the objects to the  $K$  clusters.

**Iteration:**

1. Compute new cluster centers.
2. Assign the objects to the new clusters.
3. Converged? Yes: stop / No: goto 1.

Fig. 2. Algorithmic structure of partitive clustering.

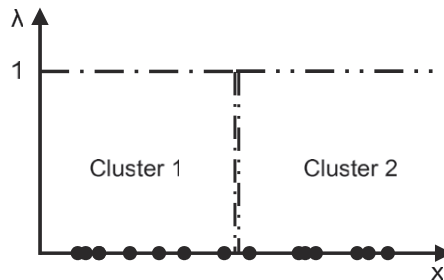


Fig. 3. Memberships for hard k-means.

After initialization a three step iteration is performed. First, the new cluster centers (means) are determined. Then, the objects are assigned to the clusters and finally the algorithms check for convergence.

However, the clustering approaches are different with respect to their details, in particular, how they address uncertainty. In the following subsections we discuss these clustering techniques in the chronological order of their introduction.

### 2.3. Clustering algorithms

#### 2.3.1. Hard k-means

*Algorithm.* The k-means algorithm [57] partitions objects  $i$  ( $i = 1, \dots, N$ ) into  $K$  disjoint subsets  $S_k$  ( $k = 1, \dots, K$ ). The objects are described by a feature vector  $x_i = (x_{i1}, \dots, x_{if}, \dots, x_{iF})$  with  $f = 1, \dots, F$  features. The objective is to minimize the sum of all squared distances  $d^2(x_i, \mu_k)$  of the clusters' objects  $x_i$  to its corresponding means  $\mu_k$  with  $d^2(x_i, \mu_k) = \|x_i - \mu_k\|^2$ :

$$\min \left\{ J = \sum_{k=1}^K \sum_{x_i \in S_k} \|x_i - \mu_k\|^2 \right\} \tag{1}$$

In k-means clustering only bivalent membership degrees of object  $i$  to cluster  $k$  are allowed:  $\lambda_{i,k} \in \{0, 1\}$ . If an object  $i$  is a member of a cluster it cannot be a member of any other cluster (see Fig. 3). Its memberships can be expressed by a membership vector  $\Lambda_i = (\lambda_{i,1}, \dots, \lambda_{i,K})$  with  $\lambda_{i,h} = 1$  for  $h \in \{1, \dots, K\}$  and  $\lambda_{i,l} = 0$  for  $l = 1, \dots, K \wedge l \neq h$ .

The classic k-means is very widely used in virtually any domain. Reasons probably include that the algorithm is easy to understand, transparent and can be implemented quickly. Furthermore, a wide variety of free as well as commercial software packages is obtainable to support analyses based on k-means.

The main strength of the k-means is in applications where the analyzed clusters are well separated and/or a hard (bivalent) decision is required.

*Extensions.* Out of the numerous extensions of k-means we only give one example which has raised great attention in the past decade. Pelleg and Moore [87] suggested the x-means that incorporates the selection of the number of clusters in the algorithm.

#### 2.3.2. Fuzzy c-means

*Algorithm.* In fuzzy clustering [7,25] the objects are assigned to  $K$  fuzzy clusters  $\tilde{S}_k$  ( $k = 1, \dots, K$ ) with a corresponding degree of membership for each cluster. Its objective is to minimize the weighted sum of Euclidean distances between the objects  $i$  and the means  $\mu_k$  of the corresponding clusters  $k$ . The weights are derived from the membership degrees of the

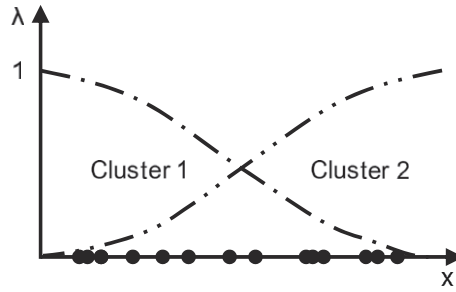


Fig. 4. Memberships for fuzzy c-means.

objects to the clusters (Fig. 4).

$$\min \left\{ J = \sum_{k=1}^K \sum_{i=1}^N \lambda_{i,k}^m \|x_i - \mu_k\|^2 \right\} \tag{2}$$

with  $\lambda_{i,k}$  the membership degree of object  $i$  to cluster  $k$ :

$$\lambda_{i,k} = \frac{1}{\sum_{j=1}^K \left( \frac{d(x_i, \mu_k)}{d(x_j, \mu_k)} \right)^{\frac{2}{m-1}}} \tag{3}$$

and  $m \in (1, \infty)$  the fuzzifier parameter which defines the clusters' fuzziness. For  $m \rightarrow \infty$  all membership degrees converge towards  $1/K$ , the most fuzziest result possible. For  $m \rightarrow 1$  each membership degree  $\lambda_{i,k}$  converges towards 0 or to 1, i.e. fuzzy c-means provides hard results as in classic k-means. Selecting an appropriate  $m$  is challenging and depends on the given data as well as on the expected results. A still well accepted compromise is to set  $m = 2.0$  [77]. However, studies also show that this selection is not appropriate for some applications (see e.g., [19,99] for studies in bioinformatics). Further methods that address the selection of  $m$  have been discussed by Yu et al. [109] and Wu [105].

The membership degrees  $\lambda_{i,k}$  show how representative an object  $i$  is for cluster  $k$  (with  $\lambda_{i,k} \in [0, 1] \forall i, k$ ). A membership degree of 0 indicates “not representative at all” while a membership degree of 1 indicates that the object represents the cluster perfectly. The sum of the membership degrees of an object  $i$  to all clusters is required to equal 1 in order to assure the algorithm's convergence [7]:

$$\sum_{k=1}^K \lambda_{i,k} = 1 \quad \forall i = 1, \dots, N \tag{4}$$

Overlapping classes constitute a main application area of fuzzy clustering. Furthermore, fuzzy membership degrees are very suitable when, e.g., in complex decision processes, information from different sources has to be aggregated to arrive at a final decision.

Optionally, after performing fuzzy c-means, the results can be defuzzified in a final step of the analysis when hard decisions are required.

*Extensions.* In our paper we can only give very few examples for extensions of fuzzy clustering. Gustafson and Kessel [27] suggested a seminal fuzzy clustering approach using fuzzy covariance matrices. Pedrycz [80] introduced conditional fuzzy clustering as an extension of fuzzy clustering. Pedrycz [82] also proposed collaborative fuzzy clustering. Coletta et al. [13] recently presented some refinements and design guidelines of collaborative fuzzy clustering. Ménard et al. [68] proposed fuzzy c+2 means addressing the ambiguity rejection in clustering. Their approach has similar objectives to Davé's noise clustering [17] and its extensions [3,18]. Crespo and Weber [15] suggested a dynamic fuzzy c-means. Rhee [98] addressed uncertainty aspects in fuzzy clustering.

### 2.3.3. Rough k-means

*Algorithm.* In rough k-means [52] a cluster is described by two hard approximations, a lower and upper approximation or respectively a lower approximation and a boundary region (Fig. 5).<sup>3</sup>

Hence, an object  $i$  has two bivalent membership degrees to a cluster  $k$ , one for its lower approximation and one for its boundary:<sup>4</sup>

$$\lambda_{i,k}^{Lower\ Approximation} = \underline{\lambda}_{i,k} \in \{0, 1\} \text{ and } \lambda_{i,k}^{Boundary} = \widehat{\lambda}_{i,k} \in \{0, 1\}.$$

<sup>3</sup> The join of the lower approximation and the boundary form the upper approximation.

<sup>4</sup> Throughout the paper we indicate lower approximations as underlined and boundaries by a  $\widehat{\text{hat}}$ .

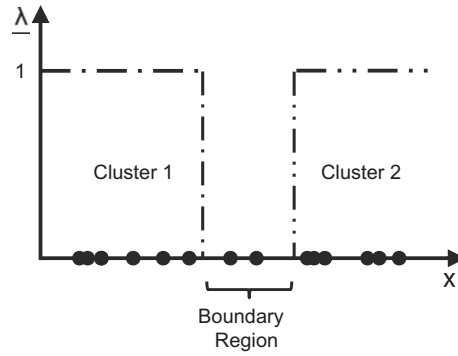


Fig. 5. Memberships for rough k-means (lower approximation).

While objects in the lower approximation surely belong to the corresponding cluster, objects in the boundary region may belong to the cluster. The decision if an object  $i$  belongs to the lower approximation of a cluster or to its boundary region is based on the object's distances to all clusters and a threshold  $\zeta$  as shown in Eq. (5):

$$T = \left\{ k : \frac{d(x_i, \mu_k)}{d(x_i, \mu_h)} \leq \zeta \wedge h \neq k \right\}. \quad (5)$$

assuming that object  $i$  is closest to cluster center  $h$  among all cluster centers, i.e.  $d(x_i, \mu_h) \leq d(x_i, \mu_k) \forall k = 1, \dots, K$ .

If set  $T$  is empty, then object  $i$  belongs to the lower approximation of cluster  $h$ . In the opposite case object  $i$  belongs to the boundary regions of cluster  $h$  and all clusters  $k$  with  $k \in T$ . The following statements hold for an object's membership relations with respect to all clusters (see also Lingras and Peters [50]):

- An object cannot be simultaneously member of a cluster's lower approximation and the same cluster's boundary region.
- If an object is member of a cluster's lower approximation it cannot belong to any other cluster, neither to its lower approximation nor to its boundary region.
- If an object is not a member of any lower approximation it must belong to the boundary regions of at least two clusters.

Regarding the membership relations of an object  $i$  to a cluster  $k$  we get three possibilities:

- $(\underline{\lambda}_{i,k}, \widehat{\lambda}_{i,k}) = (1, 0)$ : object  $i$  is a sure member of cluster  $k$ .
- $(\underline{\lambda}_{i,k}, \widehat{\lambda}_{i,k}) = (0, 1)$ : object  $i$  is a member of cluster  $k$ 's boundary region.
- $(\underline{\lambda}_{i,k}, \widehat{\lambda}_{i,k}) = (0, 0)$ : object  $i$  does not belong to cluster  $k$  at all.

Finally, the objective criterion is to minimize the weighted sum of the Euclidean distances of the objects in the lower approximations and those in the boundary regions to the respective cluster centers:<sup>5</sup>

$$\min \left\{ J = \sum_{k=1}^K \left( \frac{w}{|\underline{S}_k|} \sum_{x_i \in \underline{S}_k} \|x_i - \mu_k\|^2 + \frac{\widehat{w}}{|\widehat{S}_k|} \sum_{x_i \in \widehat{S}_k} \|x_i - \mu_k\|^2 \right) \right\} \quad (6)$$

where

- $w$ : weight for the lower approximations,
- $\widehat{w} = 1 - w$ : weight for the boundary regions,
- $|\underline{S}_k|$ : number of objects in the lower approximation of cluster  $k$ ,
- $|\widehat{S}_k|$ : number of objects in the boundary region of cluster  $k$ .

A main area of application of rough clustering is present when clear cases need to be distinguished from unclear cases. This is given, e.g., in quality control where definitely good products can pass immediately whereas products with some doubts regarding their quality need a "second look". Such a situation will be revealed by the boundary ("buffer zone") around the lower approximation.

*Extensions.* Extensions of rough k-means include Mitra's evolutionary rough k-means [70], some refinements suggested by Peters [88,89], Peters and Lampart's rough k-medoids [90] and Lingras' approach [46] based on self-organizing maps

<sup>5</sup> For simplicity reasons we assume  $|\underline{S}_k| \neq 0$  and  $|\widehat{S}_k| \neq 0$ .

**Table 1**  
Range of membership degrees.

Algorithm	Membership degrees		
	Valence	Grades $\lambda_{i,k}$	Restrictions for an object $i$
k-Means	Bivalent	$\lambda_{i,k} \in \{0, 1\}$	$\sum_{k=1}^K \lambda_{i,k} = 1$
RKM	Trivalent	$\underline{\lambda}_{i,k}, \widehat{\lambda}_{i,k} \in \{0, 1\}$	$\left( \begin{array}{l} \sum_{k=1}^K \underline{\lambda}_{i,k} = 1 \wedge \sum_{k=1}^K \widehat{\lambda}_{i,k} = 0 \\ \sum_{k=1}^K \underline{\lambda}_{i,k} = 0 \wedge \sum_{k=1}^K \widehat{\lambda}_{i,k} \geq 2 \end{array} \right) \vee$
FCM	Continuous	$\lambda_{i,k} \in [0, 1]$	$\sum_{k=1}^K \lambda_{i,k} = 1$

**Table 2**  
Objective criteria.

Algorithm	Objective criterion (minimization)
k-Means	$\sum_{k=1}^K \sum_{x_i \in S_k} \ x_i - \mu_k\ ^2$
RKM	$\sum_{k=1}^K \left( \frac{w}{ S_k } \sum_{x_i \in S_k} \ x_i - \mu_k\ ^2 + \frac{\widehat{w}}{ S_k } \sum_{x_i \in \widehat{S}_k} \ x_i - \mu_k\ ^2 \right)$
FCM	$\sum_{k=1}^K \sum_{i=1}^N \lambda_{i,k}^m \ x_i - \mu_k\ ^2$

(SOM). Mitra et al. [72] developed rough collaborative clustering and Peters et al. [92] introduced dynamic rough clustering. For an advanced review on recent developments in rough clustering the reader is referred to Lingras and Peters [50].

2.4. Characteristics of and relationships among the algorithms

In the following paragraphs, we first compare the clustering algorithms' characteristics by means of their valence, objective criteria, and initial settings. Subsequently, we briefly address under which conditions the soft clustering algorithms reduce to hard k-means.

2.4.1. Valence

An important difference between the clustering algorithms is how the membership of an object to a cluster is defined. We can identify two groups with similar characteristics: discrete vs. continuous methods:

- *Discrete Methods (Bi- and Trivalent)*. In k-means as well as in rough k-means membership values are discrete and restricted to  $\lambda \in \{0, 1\}$ . The fundamental difference between these two algorithms is that in rough k-means the core clusters (lower approximations) are surrounded by boundary regions.<sup>6</sup> Hence, the membership degrees of an object to a cluster are not bivalent like in k-means but trivalent.
- *Continuous Methods*. Fuzzy c-means provides continuous membership degrees which range from 0 to 1.

These results are summarized in Table 1.<sup>7</sup> To reflect the grouping as discussed above we rearrange the presentation of the cluster algorithms from chronological order to the two groups as defined above: clustering algorithms with discrete memberships (k-means, rough k-means) and the algorithm with continuous memberships (fuzzy c-means).

2.4.2. Objective criteria

Comparing k-means and rough k-means reveals that the rough k-means' objective function is basically the one used by k-means extended by incorporating the boundary region. In fuzzy c-means the distances between the objects and the cluster centers are weighted by their modified membership degrees  $\lambda_{i,k}^m$  with the parameter  $m$  defining the degree of fuzziness of the clusters. Table 2 summarizes the algorithms' objective criteria.

2.4.3. Initial settings

Setting the initial parameters is crucial to obtain useful clustering results. Although some guidelines are available how to set these initial parameters it generally remains a challenge (see e.g., Mitra [70]).

Table 3 shows the required parameters for the initial settings that are relevant for the clustering algorithms dealt with in Section 2.

- *Number of Clusters  $K$* . In all algorithms the number of clusters  $K$  has to be set. E.g., for hard clustering enhanced algorithms have been suggested, in particular the x-means [87], that support the determination of  $K$ . For fuzzy c-means cluster

<sup>6</sup> Note, the boundary region is generally not spatial.  
<sup>7</sup> In Table 1 we again assume  $|S_k| \neq 0$  and  $|\widehat{S}_k| \neq 0$  for simplicity reasons.

**Table 3**  
Initial settings.

Algorithm	Initial settings
k-Means	$K$
RKM	$K, \zeta, \underline{w}, \widehat{w} = 1 - \underline{w}$
FCM	$K, m, \epsilon$

**Table 4**  
Membership relationship and convergence of the clustering algorithms.

	k-Means	RKM	FCM
		$\lambda_{i,k} \in \{0, 1\}$	$\lambda_{i,k} \in \{0, 1\}$
k-Means	-	$\rightarrow$	$\rightarrow$
		$(\lambda_{i,k}, \widehat{\lambda_{i,k}}) \in \{(0, 0), (1, 0), (0, 1)\}$	$\lambda_{i,k} \in [0, 1] \wedge \sum_{k=1}^K \lambda_{i,k} = 1$
RKM	$\zeta \rightarrow 1$	-	$\not\rightarrow$
FCM	$m \rightarrow 1$	$\not\rightarrow$	-

validity measures support the identification of the optimal number of clusters (e.g., by Pal and Bezdek [77] or Windham [104]). For rough k-means cluster validity indexes have been suggested (e.g., Mitra et al. [72], Peters [91] or Lingras et al. [47]) that may support in the selection of the appropriate number of clusters. However, it should be mentioned that the “correct number of clusters” does not exist in general. The most adequate cluster number in a given application is always a compromise between the optimization of some validity measures on the one hand and practical and application driven considerations on the other hand.

- *Threshold  $\zeta$  and Weights  $\underline{w}, \widehat{w}$ .* Besides the number of clusters, in rough k-means the threshold  $\zeta$  and the weight  $\underline{w}$  has to be set. The boundary weight  $\widehat{w}$  then is  $\widehat{w} = 1 - \underline{w}$ . Up to today setting these initial parameters remains one of the great challenges of rough clustering since no well accepted guidelines have been proposed. However, Malyszko and Stepaniuk [63] proposed rough entropy to determine the weights while Wang and Zhou [103] suggested adaptive weights and a hybrid threshold. The parameters  $\underline{w}, \widehat{w}$  and  $\zeta$  also provide degrees of freedom to optimize rough k-means by tuning them adequately (see e.g., Mitra [70] or Peters et al. [91]).
- *Fuzzifier Parameter  $m$  and the Threshold  $\epsilon$ .* As already discussed the fuzzifier parameter for fuzzy c-means is commonly, due to better knowledge, set to  $m = 2.0$ . However, better substantiated methods have also been suggested (e.g., by Wu [105] or Yu [109]). The threshold  $\epsilon$  is used to decide on the convergence of fuzzy c-means; when the improvements obtained in an iteration are less than  $\epsilon$  the algorithm terminates.<sup>8</sup>

2.4.4. Convergence towards k-means

Table 4 shows the relationships among soft clustering algorithms and particularly if and how they can be converted into each other. The entries above the diagonal in the table show how the membership values  $\lambda_{i,k}$  are transformed from one cluster approach to another. The lightning symbol  $\not\rightarrow$  indicates that a direct transformation between the respective algorithms does not exist. The fields below the diagonal display the parameter settings that are required to melt down a cluster approach to another.

Note, that rough k-means as well as fuzzy c-means can melt down into k-means. However, a direct transformation of rough k-means into fuzzy c-means and vice versa is impossible.

2.5. Discussion of theoretical aspects

There is an ongoing intense debate about soft computing concepts and their relationships to classic approaches, in particular probabilistic approaches. It would go beyond the scope of this article to address the differences among these concepts in detail. However, we shall briefly discuss important insights that are relevant in the context of clustering.

In our article, hard k-means takes over the role as representative for classic approaches; sometimes it is also attributed as “statistical k-means” [102] which is in line with a common definition that statistics “may be defined as the collection, presentation, analysis, and interpretation of numerical data” [16]. However, it is not a probabilistic approach: it neither explicitly requires any probability distribution of the data to be clustered nor it does deliver any probability for an object being a membership to a certain cluster. It plainly delivers bivalent membership degrees to hard clusters on the basis of Euclidean distances.

Rough clustering enriches k-means by a boundary region that addresses uncertainty.<sup>9</sup> In classic rough set theory objects belong to a set or they do not belong to that set. With respect to this bivalent membership concept rough sets have

<sup>8</sup> In hard k-means and rough k-means the algorithm terminates when the cluster assignment of the objects has not changed in an iteration. So, a similar threshold is not required.

<sup>9</sup> Note, rough clustering belongs to the branch of rough set theory which is interval based. Hence, the given interpretation assumes the transfer of the principles of original set-based rough set theory to its interval-based interpretation.

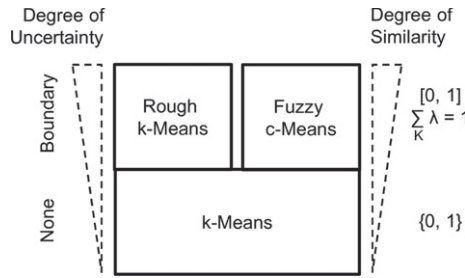


Fig. 6. Derivatives of k-means.

identical characteristics as the classic k-means. However, in rough clustering this bivalent membership concept of sure membership values is enriched by objects that still surely belong to one and only one cluster but it is not known to which one; there might be several candidate clusters for such an object. Reasons for its unclear status may include missing information or errors in the data collection process. Such a candidate object is assigned to the boundary of the corresponding clusters.

Hence, rough k-means addresses uncertainty which stems from missing information instead of simultaneous similarities to several clusters as is the case of fuzzy c-means. This enhancement of k-means constitutes one branch of soft clustering algorithms (see the left pylon in Fig. 6).

In contrast to this, in fuzzy c-means the hard membership values are generalized by allowing graded similarities of objects to clusters which range from 0 (totally dissimilar) to 1 (perfect representative of the cluster). Sometimes fuzzy c-means is denoted as “probabilistic” [4,67] in contrast to possibilistic approaches (see Section 3.1.1). A motivation behind this may be that the membership degrees formally fulfill the Kolmogorov axioms of probability theory [38]. However, we are not aware of strong supporting arguments for this probabilistic interpretation beyond these formal arguments. Nevertheless, we will use the term “probabilistic” in the above sense in the course of the paper whenever the authors of the original papers use it.

Note that – starting from k-means - fuzzy c-means addresses a specific kind of ambiguity inherent in clustering results; it cannot unambiguously be decided to which cluster an object belongs, but the respective algorithms determine a precise number (membership degree) which represents the degree of similarity an object has to a particular cluster.

This increasing flexibility to describe similarities constitutes another branch of soft clustering algorithms (see the right pylon in Fig. 6).

### 3. Derivatives of fuzzy and rough clustering

In the context of this paper derivatives of fuzzy c-means are clustering algorithms that inspired from fuzzy c-means and enhance it with respect to the kind of uncertainty considered, i.e. possibilistic and evidential approaches. We are not aware of similar derivatives of rough k-means. However, there are several hybrid derivatives merging fuzzy and rough concepts which we discuss in Section 3.3: *Hybrid Derivatives of Fuzzy and Rough Clustering*.

#### 3.1. Derivatives of fuzzy c-means

In the following paragraphs we survey possibilistic, evidential, and belief c-means as derivatives of fuzzy c-means.

##### 3.1.1. Possibilistic c-means

Krishnapuram and Keller [40] proposed possibilistic c-means as a derivative of fuzzy c-means. Both, fuzzy c-means as well as possibilistic c-means provide continuous membership degrees. The significant difference between them is that for each object the sum of its membership values to all classes must equal 1 for fuzzy c-means and has to be between 0 and 1 for possibilistic c-means:

$$\sum_{k=1}^K \lambda_{i,k} \leq 1 \quad \forall i = 1, \dots, N \tag{7}$$

This relaxation makes it possible to better describe objects that are dissimilar to all clusters, e.g., outliers.

Pal et al. [78] denote a membership degree obtained by fuzzy c-means as an indicator for the *relative typicality* of an object to a cluster while a membership degree obtained by possibilistic clustering as an indicator for the *absolute typicality*.



**Table 5**  
Example for evidential clustering [65].

$A$	$m_1(A)$	$m_2(A)$	$m_3(A)$	$m_4(A)$	$m_5(A)$
$\emptyset$	1	0	0	0	0
$\{\omega_1\}$	0	0	0	0.2	0
$\{\omega_2\}$	0	1	0	0.4	0
$\{\omega_1, \omega_2\}$	0	0	0	0	0
$\{\omega_3\}$	0	0	0.2	0.4	0
$\{\omega_1, \omega_3\}$	0	0	0.5	0	0
$\{\omega_2, \omega_3\}$	0	0	0	0	0
$\Omega$	0	0	0.3	0	1

In case of possibilistic clustering the objective function of the fuzzy c-means algorithm has to be extended by a term with a Lagrange multiplier which assures a non-trivial relaxation of the sum of the membership degrees:

$$\min \left\{ J = \sum_{k=1}^K \sum_{i=1}^N \lambda_{i,k}^m \|x_i - \mu_k\|^2 + \sum_{k=1}^K \eta_k \sum_{i=1}^N (1 - \lambda_{i,k})^m \right\} \quad (8)$$

with  $\eta_k$  a suitable positive number. Krishnapuram and Keller [40] suggested how to set  $\eta$  in possibilistic c-means. The reader is referred to their paper for further details.

As hybridizations of fuzzy and possibilistic clustering Pal et al. suggested fuzzy-possibilistic c-means [78] as well as a possibilistic-fuzzy c-means algorithm [79]. Massulli and Rovetta [67] proposed a soft transition model from probabilistic to possibilistic fuzzy clustering.

### 3.1.2. Evidential c-means

Inspired by Dempster-Shafer's theory of evidence [20,100,107] and Davé's [17] approach to detect noise in clustering Denœux and Masson [21,22] developed credal partitions. A credal partition allows an object to be assigned to more than one subset of classes.

To obtain the credal partitions Denœux and Masson suggested the EVCLUS-Evidential CLUstering algorithm [21,22]. Masson and Denœux proposed a variant of the EVCLUS for interval-valued dissimilarity data [64] and later the evidential c-means (ECM) [65].

In the context of our paper we just present the very basic idea of the ECM concentrating on the generalization of the memberships of objects to clusters. The reader is referred to the original paper of Masson and Denœux [65] for more details.

The memberships are given in the form of basic belief assignments (*bba*):  $m(A)$  with  $\sum_{A \subseteq \Omega} m(A) = 1$  and  $\Omega = \{\omega_1, \dots, \omega_K\}$  the set of  $K$  clusters.

To illustrate the ECM let us discuss an example as given by Masson and Denœux [65]: five objects ( $i = 1, \dots, N$  with  $N = 5$ ) are assigned to three clusters  $\omega_1, \dots, \omega_3$  (see Table 5). Then the degrees of memberships of an object to the clusters are given by basic belief assignments (*bba*):  $m_i(A)$ . Note, the function  $m$  is also labeled as "mass".

Table 5 can be interpreted as follows [65]:

- Object 1 ( $m_1(A)$ ): The mass  $m_1(\emptyset) = 1$  indicates that object 1 does not belong to any of the given clusters  $\Omega = \{\omega_1, \omega_2, \omega_3\}$ .
- Object 2 ( $m_2(A)$ ): The mass  $m_2(\{\omega_2\}) = 1$  indicates that object 2 certainly belongs to cluster  $\omega_2$ .
- Object 3 ( $m_3(A)$ ): The mass  $m_3$  is divided among  $\{\omega_3\}$ ,  $\{\omega_1, \omega_3\}$  and  $\Omega$ . Therefore, only partial knowledge about the membership of object 3 is available.
- Object 4 ( $m_4(A)$ ): The mass  $m_4$  is allocated to singletons only:  $\{\omega_1\}$ ,  $\{\omega_2\}$ ,  $\{\omega_3\}$ . Hence,  $m_4$  is called Bayesian mass. Obviously, as in the case before, only partial knowledge is available.
- Object 5 ( $m_5(A)$ ): The mass  $m_5(\Omega) = 1$  indicates that the corresponding cluster of object 5 is unknown. It is just known that it belongs to one of the clusters  $\Omega = \{\omega_1, \omega_2, \omega_3\}$ .

Masson and Denœux [65] showed that the ECM can be melted down to fuzzy and possibilistic clustering. Hence, it can be considered as their generalization which gives more flexibility and more degrees of freedom when complex clustering tasks have to be performed.

However, Liu et al. [55] pointed out that in ECM it can happen that cluster centers get very close and that the clusters eventually totally overlap. They addressed this issue by proposing the belief c-means approach (BCM). Further developments of ECM include a relational evidential c-means algorithm (RECM) [66] and a constrained evidential c-means algorithm (CECM) [2].

Recently Joshi and Lingras [35] discussed the relationship of evidential and rough clustering and found a great deal of agreement between the cluster assignments by ECM and RKM. In addition, ECM was able to identify outliers by assigning large *bba* values to an empty set. However, RKM, on the other hand, produced more reasonable clustering schemes for very large data sets with high dimensions.

### 3.2. Derivatives of rough k-means

As already mentioned, in the context of our paper, derivatives of rough k-means are clustering algorithms that derive from rough k-means and enhance it with respect to the kind of uncertainty. Note, we restrain from enhancements based on rough and derived concepts since we are dealing with them separately in Section 3.3: *Hybrid Derivatives of Fuzzy and Rough Clustering*. Within these limitations we are not aware of any derivatives of rough k-means.

### 3.3. Hybrid derivatives of fuzzy and rough clustering

In the past two decades increasing attention has been directed to the hybridization of fuzzy and rough concepts. Dubois and Prade [23] fundamentally addressed rough-fuzzy and fuzzy-rough hybridization already in 1990. Hybrid rough-fuzzy and fuzzy-rough approaches include e.g., Huang [32], Petrosino and Ceccarelli [93] or Petrosino and Salvi [94]. Lingras and Jensen [49] surveyed rough and fuzzy hybridization.

In the context of our paper we review hybridization of fuzzy and rough clustering approaches in the following subsections.

#### 3.3.1. Rough-fuzzy clustering

In this section we discuss two rough-fuzzy clustering algorithms that are named identically but integrate fuzzy concepts into the rough k-means slightly differently:

- Mitra's et al. Rough-Fuzzy Clustering [72]
- Maji and Pal's Rough-Fuzzy Clustering [58]

*Mitra's et al. Rough-Fuzzy Clustering.* Mitra et al. [72] introduced a hybrid rough-fuzzy clustering algorithm (RFCM) with fuzzy lower approximations and fuzzy boundaries. The RFCM mainly differs from the original RKM (i) with respect to the assignment of the objects to a lower approximation or a boundary and (ii) with respect to the calculation of the means:

- (i) *Object's assignment to a lower approximation or a boundary.* The idea is to replace the Euclidean distances used in rough k-means to distinguish between the objects in the boundary and in the lower approximations by membership degrees obtained by fuzzy clustering (see Eq. (3) in Section 2.3.2 for the original fuzzy approach).
- (ii) *Calculation of the means.* When calculating the means in rough clustering, the objects in the lower approximation as well as in the boundary are weighted by their respective fuzzy membership degrees.

As Mitra et al. found out the inclusion of fuzzy memberships into rough clustering adds robustness to the algorithm with respect to the selection of its initial parameters.

For the fuzzifier parameter  $m \rightarrow 1$  the rough-fuzzy cluster algorithm converges towards rough k-means. For a small threshold  $\zeta$  in rough k-means the boundary region becomes empty; hence, it converges towards fuzzy c-means.

In a subsequent paper Mitra and Barman [73] applied rough-fuzzy c-means successfully to medical imagery.

*Maji and Pal's Rough-Fuzzy Clustering.* Maji and Pal [58] introduced a variation of Mitra's et al. hybrid rough-fuzzy clustering algorithm. They proposed that all objects in a lower approximation should have identical influence on the determination of their means and should be independent from other clusters. Hence, they suggested a rough-fuzzy c-means with crisp lower approximations and fuzzy boundaries. Obviously, in Maji and Pal's [58] algorithm the lower approximation has a higher impact on clustering in comparison to Mitra's et al. [72] approach where the objects in lower approximation are weighted by factors between 0 and 1 (membership degrees raised to the power of  $m$ ).

Applications of Maji and Pal's RFCM are, for example, in the fields of microarray gene expression data [61,62] and image segmentation [60].

Fig. 7 summarizes Mitra's et al. [72] and Maji and Pal's [58] versions of the rough-fuzzy c-means.

#### 3.3.2. Rough-fuzzy possibilistic clustering

In 2007 Maji and Pal [59] introduced a rough set based generalized fuzzy c-means algorithm which they called RFPCM (rough fuzzy possibilistic c-means). It merges fuzzy and possibilistic approaches and rough k-means in a way that these clustering algorithms can be derived from it. Therefore, it is a unifying soft clustering approach.

Like in Maji and Pal's [58] RFCM a cluster is approximated by a crisp lower approximation and a fuzzy boundary. Additionally, in the RFPCM, the boundary can be characterized by possibilistic elements (see Fig. 7).<sup>10</sup> The reader is referred to the paper of Maji and Pal [59] for further details.

<sup>10</sup> Maji and Pal [59] use the term *probabilistic* for the memberships obtained by FCM and *possibilistic* for the memberships obtained by PCM. The term *fuzzy* subsumes both.

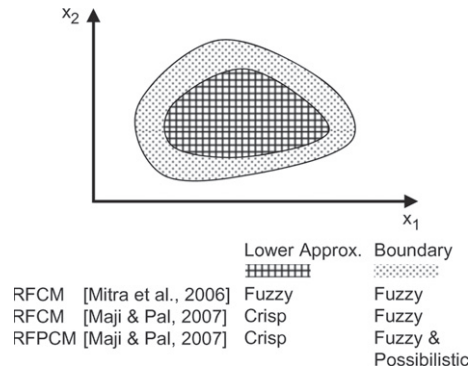


Fig. 7. Mitra's et al. [72] and Maji and Pal's [58] RFCM and Maji and Pal's RFPCM [59].

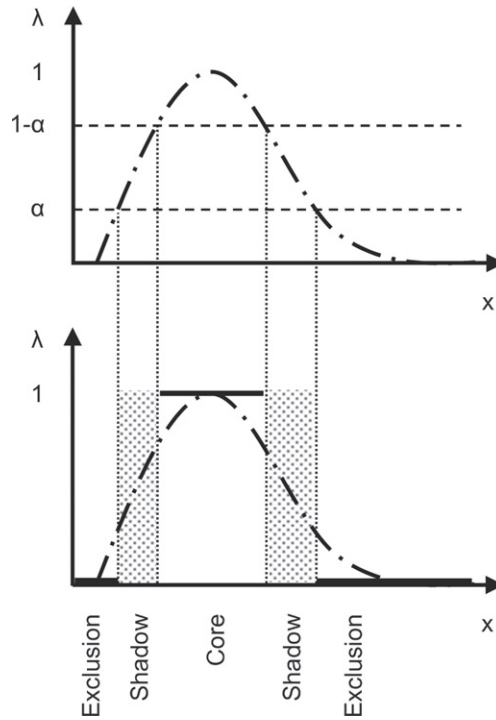


Fig. 8. Shadowed sets (adapted from [81]).

### 3.3.3. Shadowed set clustering

In the field of probability sets [31] it has been observed that membership grades around 0 or 1 can rather easily be defined while determining grades around 0.5 is particularly difficult.

The reason for this might be that objects with membership grades close to 1 can normally be regarded as members of a certain class while objects with membership grades close to 0 are normally no members of the corresponding class. In contrast to this, the objects around 0.5 form a particular challenge since they are undecided in-betweens. Any small change in their estimated grades could possibly change their anticipated memberships to a class. Hence, assigning them the right probability is much more important than in the rather clear cases close to 0 or 1.

These findings inspired Pedrycz [81] to introduce shadowed sets by transferring this fundamental idea to memberships in fuzzy set theory. He proposed to approximate a membership function by three zones: the *core*, the *shadow* and the *exclusion* (see Fig. 8):

- *Core*. Membership degrees close to 1 are simply set to 1.
- *Exclusion*. Membership degrees close to 0 are simply approximated by 0.
- *Shadow*. Membership degrees that are neither close to 0 nor close to 1 constitute a shadowed region.

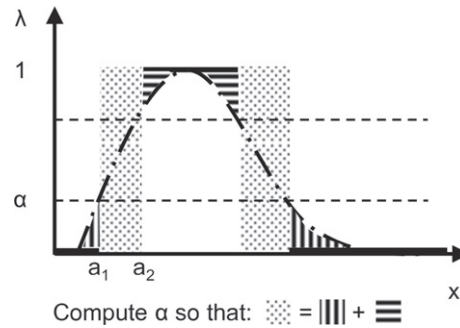


Fig. 9. Computation of the threshold  $\alpha$ .

To determine the size of the shadow Pedrycz [85] proposed to balance vagueness, i.e. the shadowed region on the one hand and the regions with the “wiped-out” detailed membership degrees on the other hand should be identically large. The borders between these regions are symmetric and defined by  $\alpha$  for the lower bound and respectively  $1 - \alpha$  for the upper bound. Then, the parameter  $\alpha$  is determined as shown in Fig. 9.

At the first sight shadowed sets seem to be identical to rough sets. But this is only true with respect to the categorization of the objects to three classes: (1) sure members, (2) sure non-members, and (3) objects in-between. A closer look reveals distinct differences between shadowed and rough sets. These include:

- In contrast to rough sets the equivalence classes are defined dynamically in shadowed sets, i.e. the respective thresholds are derived depending on the data sets, whereas in rough sets the thresholds are determined by a user.
- A similar property like in rough sets that an object must belong to a least two shadowed regions (rough sets: upper approximations) if it is not a member of any core (rough sets: lower approximation) does not exist for shadowed sets.

Obviously, the idea of shadowed sets can be applied to clustering [81,83,84]. Mitra et al. [74] suggested a shadowed c-means (SCM) as an integration of fuzzy and rough clustering. Their main idea is to weight the members in the core of a shadowed set by 1, the objects in the shadowed region like in fuzzy sets by  $\lambda_{ik}^m$  and the objects in the exclusion zones by membership degrees “double-powered” by the fuzzifier parameter  $m$ :  $\lambda_{ik}^{m^2}$ .

Hence, objects in the core count full, while objects in the shadowed regions count like in regular fuzzy clustering. Due to the “double-powering”  $m$  the objects in the exclusion regions have very little impact on the clustering results. Anticipating that outliers should normally be located in exclusion regions their influence on the clustering result is small.

Recently, Zhou et al. [86] discussed shadowed sets in the characterization of rough-fuzzy clustering.

#### 4. Applications

Since clustering is a standard approach in data mining it has been applied in virtually any domain. As mentioned before k-means is probably the most widely used clustering algorithm. But also fuzzy and possibilistic c-means have been used in a numerous number of real life applications. Since rough k-means is relatively new the number of its applications is smaller than that of the previously mentioned methods. However, it has also been applied in several real life studies already. Evidential c-means and hybrid clustering algorithms have been proposed very recently. Hence, they are mostly applied to common data sets given supporting arguments for their usefulness.

In the following paragraphs we present a small selection of applications of the fundamental clustering algorithms we discussed in Section 2 (including possible extensions).<sup>11</sup> We provide examples for the partly overlapping domains Biology and Bioinformatics, Engineering (including Pattern Recognition), Business and Economics as well as Miscellaneous (including information technology).

*Hard k-means:*

- Biology and Bioinformatics
  - Protein structure prediction [28]
  - Gene expression data analysis [56]
  - Exploring local protein sequence motifs [110]
- Engineering
  - Optical character recognition, speech recognition, and encoding/decoding [1]

<sup>11</sup> Note that considering the huge number of applications the selection must be arbitrary. The examples are only intended to give an impression of the diversity of applications of the algorithms and by no means intended to be comprehensive.

- Image retrieval system [54]
- Color image segmentation [97]
- Business and Economics
  - Exploration of shopping orientations and online purchase intention [10]
  - Applications in marketing [95]
  - Clustering companies [96]
- Miscellaneous
  - Understanding motivation effort in free/open source software projects [42]
  - Prediction of students' academic performance [76]
  - Network traffic classification [108]

#### *Rough k-means:*

- Biology and Bioinformatics
  - Identify patterns of gene expression in cancer datasets [41]
  - Analyzing microarray data [91]
  - Bioinformatics [71]
- Engineering
  - Medical imagery [73]
  - Speech recognition [70]
  - Forest cover data [70]
- Business and Economics
  - Changing retail data [14]
  - Supermarket data [48]
- Miscellaneous
  - Clustering path profiles on a website [12]
  - Traffic monitoring [45]
  - Neighborhood clustering of web users [101]

#### *Fuzzy c-means:*

- Biology and Bioinformatics
  - Survey on fuzzy pattern recognition in medicine [5]
  - Clustering analysis of microarray data [29]
  - Clustering in data analysis of metabolomics [44]
- Engineering
  - Segmentation of a thematic mapper image [11]
  - Application to remote sensing [24]
  - Color segmentation of thermal infrared breast images [26]
- Business and Economics
  - Classification of knowledge intensity in China's manufacturing industry [9]
  - Market segmentation [39]
  - Clustering in marketing [95]
- Miscellaneous
  - Detection of polluted sites [30]
  - Profiling network applications [43]
  - Soil pattern recognition [75]

Many applications of the recently proposed derivatives are in the field of bioinformatics while the “older” PCM has a longer history with many applications in various domains.

## 5. Conclusion

Soft clustering goes back to the beginning of the seventies of the last century when Dunn [25] introduced fuzzy ISODATA in 1973. In our survey we have summarized some of the important contributions to this area over the past 40 years.

Bezdek's fuzzy c-means [7] has probably become the most popular soft clustering algorithm so far with many real life applications in a very diverse range of domains. Its derivatives include possibilistic clustering [40] as well as more recent approaches based on evidence [65]. These algorithms constitute one family of soft clustering algorithms.

Lingras and West [51] suggested rough k-means ten years ago. We showed that rough k-means is much closer related to k-means than to fuzzy c-means and its derivatives. We would like to stress this point since this goes beyond a pure categorization of terminologies but might also influence the awareness for and acceptance of rough k-means, especially in domains where hard k-means is dominantly used currently.

While fuzzy c-means generalizes k-means with respect to ambiguity based on similarity between objects and clusters, rough k-means adds a dimension of uncertainty due to missing or wrong information. Hence, rough k-means constitutes a different family of soft clustering algorithms.

In the past years, intensive work has been directed towards the hybridization of clustering approaches, in particular on the hybridization of fuzzy and rough concepts (e.g., Maji et al. [58,59] and Mitra et al. [72]). The obtained cluster algorithms generally perform more robustly with respect to the initial settings and/or in the presence of outliers than their original approaches. Last but not least clustering approaches based on Pedrycz's shadowed sets [81] have been integrating fuzzy and rough clustering [74,86] recently.

In particular the recent promising developments in the hybridization of soft cluster algorithms show the need for approaches that holistically address uncertainty. Hence, soft clustering will remain in the interest of researchers and most probably attract even more practitioners in the field of data mining in support of their real life applications.

## Acknowledgments

The authors would like to thank the editor and the anonymous reviewers for their valuable comments. Support from the Chilean Instituto Sistemas Complejos de Ingenieria (ICM: P-05-004-F, CONICYT: FBO16) is greatly acknowledged ([www.sistemasdeingenieria.cl](http://www.sistemasdeingenieria.cl)).

## References

- [1] E. Alpaydin, Introduction to Machine Learning, MIT Press, Cambridge, MA, USA, 2004.
- [2] V. Antoine, B. Quost, H.M. Masson, T. Denœux, CECM: Constrained evidential c-means algorithm, *Computational Statistics & Data Analysis* 4 (1) (2012) 894–914.
- [3] A. Banerjee, R.N. Davé, Robust clustering, *WIREs Data Mining and Knowledge Discovery* 2 (2012) 29–59.
- [4] A. Baraldi, P. Blonda, A survey of fuzzy clustering algorithms for pattern recognition – part II, *IEEE Transactions on Systems Man and Cybernetics, Part B: Cybernetics* 29 (6) (1999) 785–801.
- [5] A.S. Begum, O.M. Devi, Fuzzy algorithms for pattern recognition in medical diagnosis, *Assam University Journal of Science and Technology: Physical Sciences and Technology* 7 (2) (2011) 1–12.
- [6] J.C. Bezdek, Fuzzy mathematics in pattern classification. Ph.D. thesis, Cornell University, 1973.
- [7] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Algorithms*, Plenum Press, New York, 1981.
- [8] J.C. Bezdek, J. Harris, Fuzzy partitions and relations, *Fuzzy Sets and Systems* 1 (2) (1978) 111–127.
- [9] S. Bing, Application of factor analysis and fuzzy c-means for classification of knowledge intensity in china's manufacturing industry, in: *International Conference on Management Science and Engineering (ICMSE)*, Rome, 2011, pp. 111–117.
- [10] M. Brown, N. Pope, K. Voges, Buying or browsing? An exploration of shopping orientations and online purchase intention, *European Journal of Marketing* 37 (11) (2003) 1666–1684.
- [11] R.L. Cannon, J.V. Dave, J. Bezdek, M.M. Trivedi, Segmentation of a thematic mapper image using the fuzzy c-means clustering algorithm, *IEEE Transactions on Geoscience and Remote Sensing* GE-24 (2) (1986) 400–408.
- [12] X. Cheng, R. Wu, Clustering path profiles on a website using rough k-means method, *Journal of Computational Information Systems* 8 (14) (2012) 6009–6016.
- [13] L.F.S. Coletta, L. Vendramin, E.R. Hruschka, R.J.G.B. Campello, W. Pedrycz, Collaborative fuzzy clustering algorithms: Some refinements and design guidelines, *IEEE Transactions on Fuzzy Systems* 20 (3) (2012) 444–462.
- [14] F. Crespo, G. Peters, R. Weber, Rough clustering approaches for dynamic environments, in: G. Peters, P. Lingras, D. Slezak, Y.Y. Yao (Eds.), *Rough Sets: Selected Methods and Applications in Management and Engineering*, Advanced Information and Knowledge Processing, Springer, London, 2012, pp. 39–50.
- [15] F. Crespo, R. Weber, A methodology for dynamic data mining based on fuzzy clustering, *Fuzzy Sets and Systems* 150 (2) (2005) 267–284.
- [16] F.E. Croxton, D.J. Cowden, *Applied General Statistics*, Prentice-Hall, New York, NY, 1939.
- [17] R.N. Davé, Characterization and detection of noise in clustering, *Pattern Recognition Letters* 12 (11) (1991) 657–664.
- [18] R.N. Davé, S. Sen, Generalized noise clustering as a robust fuzzy c-M-estimators model, in: *Proceedings Conference of the North American Fuzzy Information Processing Society*, Pensacola Beach, FL, 1998, pp. 256–260.
- [19] D. Dembele, P. Kastner, Fuzzy c-means method for clustering microarray data, *Bioinformatics* 19 (8) (2003) 973–980.
- [20] A.P. Dempster, A generalization of Bayesian inference, Technical Report 20, Harvard University, Department of Statistics, 1967.
- [21] T. Denœux, M.H. Masson, Clustering of proximity data using belief functions, in: B. Bouchon-Meunier, L. Foulloy, R.R. Yager (Eds.), *Intelligent Systems for Information Processing from Representation to Application*, Elsevier, Amsterdam, 2003, pp. 291–302.
- [22] T. Denœux, M.H. Masson, EVCLUS: Evidential CLUSTERing of proximity data, *IEEE Transactions System, Man, and Cybernetics – Part B: Cybernetics* 34 (1) (2004) 95–109.
- [23] D. Dubois, H. Prade, Rough fuzzy sets and fuzzy rough sets, *International Journal of General Systems* 17 (2) (1990) 191–209.
- [24] P. Dulyakarn, Y. Rangsaneri, Fuzzy c-means clustering using spatial information with application to remote sensing, in: *22nd Asian Conference on Remote Sensing*. Conference Publications, Singapore, 2001.
- [25] J. Dunn, A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, *Journal of Cybernetics* 3 (3) (1973) 32–57.
- [26] M. EtehadTavakol, S. Sadr, E.Y.K. Ng, Application of k- and fuzzy c-means for color segmentation of thermal infrared breast images, *Journal of Medical Systems* 34 (1) (2010) 35–42.
- [27] D.E. Gustafson, W.C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, in: *IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes*, vol. 17, San Diego, CA, USA, 1978, pp. 761–765.
- [28] K.F. Han, D. Baker, Recurring local sequence motifs in proteins, *Journal of Molecular Biology* 251 (1995) 176–187.
- [29] L. Han, X. Zeng, H. Yan, Fuzzy clustering analysis of microarray data, in: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 222, 2008, pp. 1143–1148.
- [30] M. Hanesch, R. Scholgera, M.J. Dekkersb, The application of fuzzy c-means cluster analysis and non-linear mapping to a soil data set for the detection of polluted sites, *Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy*, 26 (11–12) (2001) 885–891.
- [31] K. Hirota, Concepts of probabilistic sets, *Fuzzy Sets and Systems* 5 (1) (1981) 31–46.

- [32] K.Y. Huang, T.H. Chang, T.C. Chang, Determination of the threshold value  $\beta$  of variable precision rough set by fuzzy algorithms, *International Journal of Approximate Reasoning* 52 (2011) 1056–1072.
- [33] A.K. Jain, Data clustering: 50 years beyond k-means, *Pattern Recognition Letters* 31 (2010) 651–666.
- [34] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: A review, *ACM Computing Surveys* 31 (3) (1999) 264–323.
- [35] M. Joshi, P. Lingras, Evidential clustering or rough clustering: The choice is yours, in: *Rough Sets and Knowledge Technology*, vol. 7414 of LNCS, 2012, pp. 123–128.
- [36] M. Joshi, P. Lingras, C.R. Rao, Correlating fuzzy and rough clustering, *Fundamenta Informaticae XXI* (2011) 1001–1013.
- [37] J. Kaufman, P. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, John Wiley, New York, USA, 1990.
- [38] A.N. Kolmogorov, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, Berlin, 1933.
- [39] R.J. Kou, L.M. Ho, C.M. Hu, Integration of self-organizing feature map and k-means algorithm for market segmentation, *Computers & Operations Research* 29 (11) (2002) 1475–1493.
- [40] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, *IEEE Transactions on Fuzzy Systems* 1 (2) (1993) 98–110.
- [41] P. Kumar, S.K. Wasan, Comparative study of k-means, pam and rough k-means algorithms using cancer datasets, in: *Proceedings of CSIT: 2009 International Symposium on Computing, Communication, and Control (ISCCC 2009)*, vol. 1. IACSIT Press, Singapore, 2011, pp. 136–140.
- [42] K.R. Lakhani, R.G. Wolf, Why hackers do what they do: Understanding motivation effort in free/open source software projects, Working Paper 4425-03, MIT Sloan School of Management, 2003.
- [43] T. Lampinen, H. Koivisto, T. Honkanen, Profiling network applications with fuzzy c-means clustering and self-organizing map, in: *International Conference on Fuzzy Systems and Knowledge Discovery*, Singapore, 2002.
- [44] X. Li, X. Lu, J. Tian, P. Gao, H. Kong, G. Xu, Application of fuzzy c-means clustering in data analysis of metabolomics, *Artificial Intelligence in Medicine* 81 (11) (2009) 4468–4475.
- [45] P. Lingras, Unsupervised rough set classification using GAs, *Journal of Intelligent Information Systems* 16 (2001) 215–228.
- [46] P. Lingras, Applications of rough set based k-means, Kohonen SOM, GA clustering, *Transactions on Rough Sets VII* (2007) 120–139.
- [47] P. Lingras, M. Chen, D. Miao, Rough cluster quality index based on decision theory, *IEEE Transactions on Knowledge and Data Engineering* 21 (7) (2009) 1014–1026.
- [48] P. Lingras, M. Hogo, M. Snorek, C. West, Temporal analysis of clusters of supermarket customers: conventional versus interval set approach, *Information Sciences* 172 (2005) 215–240.
- [49] P. Lingras, R. Jensen, Survey of rough and fuzzy hybridization, in: *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2007)*, 2007, pp. 1–6.
- [50] P. Lingras, G. Peters, Rough clustering, *WIREs Data Mining and Knowledge Discovery* 1 (2011) 64–72.
- [51] P. Lingras, C. West, Interval set clustering of web users with rough k-means, Tech. Rep. 2002-002, Department of Mathematics and Computer Science, St. Mary's University, Halifax, Canada, 2002.
- [52] P. Lingras, C. West, Interval set clustering of web users with rough k-means, *Journal of Intelligent Information Systems* 23 (2004) 5–16.
- [53] P. Lingras, R. Yan, C. West, Comparison of conventional and rough k-means clustering, *Proceedings RSFDGrC – 9th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, LNAI*, vol. 2639, Springer-Verlag, Berlin, Germany, 2003, pp. 130–137.
- [54] H. Liu, X. Yu, Application research of k-means clustering algorithm in image retrieval system, in: *Proceedings of the Second Symposium International Computer Science and Computational Technology (ISCSCT 09)*, Huangshan, PR China, 2009, pp. 274–277.
- [55] Z.G. Liu, J. Dezert, G. Mercier, Q. Pan, Belief c-means: An extension of fuzzy c-means algorithm in belief functions framework, *Pattern Recognition Letters* 33 (2012) 291–300.
- [56] Y. Lu, S. Lu, F. Fotouhi, Y. Deng, S. Brown, Incremental genetic k-means algorithm and its application in gene expression data analysis, *BMC Bioinformatics* 5 (2004) 172.
- [57] J. MacQueen, Some methods for classification and analysis of multivariate observations, in: L.M. LeCam, J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1: Statistics. University of California Press, Berkeley and Los Angeles, CA, 1967, pp. 281–297.
- [58] P. Maji, S.K. Pal, RFCM: A hybrid clustering algorithm using rough and fuzzy sets, *Fundamenta Informaticae* 80 (4) (2007) 475–496.
- [59] P. Maji, S.K. Pal, Rough set based generalized fuzzy c-means algorithm and quantitative indices, *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics* 37 (6) (2007) 1529–1540.
- [60] P. Maji, S.K. Pal, Maximum class separability for rough-fuzzy c-means based brain mr image segmentation, *Transactions on Rough Sets IX* (2008) 114–134.
- [61] P. Maji, S. Paul, Microarray time-series data clustering using rough-fuzzy c-means algorithm, in: *IEEE International Conference on Bioinformatics and Biomedicine*, Atlanta, Georgia, 2011, pp. 269–272.
- [62] P. Maji, S. Paul, Rough-fuzzy c-means for clustering microarray gene expression data, *Perception and Machine Intelligence*, LNCS, vol. 7143, Springer, Berlin, 2012.
- [63] D. Malyszko, J. Stepaniuk, Rough entropy based k-means clustering, in: *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, vol. 5908 of LNCS, 2009, pp. 406–413.
- [64] M.H. Masson, T. Denœux, Clustering interval-valued data using belief functions, *Pattern Recognition Letters* 25 (2) (2004) 163–171.
- [65] M.H. Masson, T. Denœux, ECM: An evidential version of the fuzzy c-means algorithm, *Pattern Recognition* 41 (2008) 1384–1397.
- [66] M.H. Masson, T. Denœux, RECM: Relational evidential c-means algorithm, *Pattern Recognition Letters* 30 (11) (2009) 1015–1026.
- [67] F. Masulli, S. Rovetta, Soft transition from probabilistic to possibilistic fuzzy clustering, *IEEE Transactions on Fuzzy Systems* 14 (4) (2006) 516–527.
- [68] M. Ménard, C. Demko, P. Loonis, The fuzzy c+2-means: solving the ambiguity rejection in clustering, *Pattern Recognition* 33 (7) (2000) 1219–1237.
- [69] S.A. Mingoti, J.O. Lima, Comparing som neural network with fuzzy c-means, k-means and traditional hierarchical clustering algorithms, *European Journal of Operational Research* 174 (2006) 1742–1759.
- [70] S. Mitra, An evolutionary rough partitive clustering, *Pattern Recognition Letters* 25 (2004) 1439–1449.
- [71] S. Mitra, Computational intelligence in bioinformatics, *Transactions on Rough Sets III* (2005) 134–152.
- [72] S. Mitra, H. Banka, W. Pedrycz, Rough-fuzzy collaborative clustering, *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics* 36 (4) (2006) 795–805.
- [73] S. Mitra, B. Barmann, Rough-fuzzy clustering: An application to medical imagery, in: *RSKT 2008*, No. 5009 in LNAI, 2008, pp. 300–307.
- [74] S. Mitra, W. Pedrycz, B. Barman, Shadowed c-means: Integrating fuzzy and rough clustering, *Pattern Recognition* 43 (2010) 1282–1291.
- [75] I.O.A. Odeh, D.J. Chittleborough, A.B. McBratney, Soil pattern recognition with fuzzy c-means: Application to classification and soil-landform interrelationships, *Soil Science Society of America Journal* 56 (2) (1990) 505–516.
- [76] O.J. Oyedale, O.O. Oladipupo, I.C. Obagbuwa, Application of k-means clustering algorithm for prediction of students' academic performance, *International Journal of Computer Science and Information Security* 7 (1) (2010) 292–295.
- [77] N.R. Pal, J.C. Bezdek, On cluster validity for the fuzzy c-means model, *IEEE Transactions on Fuzzy Systems* 3 (3) (1995) 370–379.
- [78] N.R. Pal, K. Pal, J.C. Bezdek, A mixed c-means clustering model, in: *IEEE International Conference on Fuzzy Systems*, Spain, 1997, p. 1121.
- [79] N.R. Pal, K. Pal, J.M. Keller, J.C. Bezdek, A possibilistic fuzzy c-means clustering algorithm, *IEEE Transactions on Fuzzy Systems* 13 (4) (2005) 517–530.
- [80] W. Pedrycz, Conditional fuzzy c-means, *Pattern Recognition Letters* 17 (6) (1996) 625–631.
- [81] W. Pedrycz, Shadowed sets: Representing and processing fuzzy sets, *IEEE Transactions System, Man, and Cybernetics – Part B: Cybernetics* 28 (1998) 103–109.
- [82] W. Pedrycz, Collaborative fuzzy clustering, *Pattern Recognition Letters* 23 (14) (2002) 1675–1686.
- [83] W. Pedrycz, Interpretation of clusters in the framework of shadowed sets, *Pattern Recognition Letters* 26 (2005) 2439–2449.
- [84] W. Pedrycz, Knowledge-Based Clustering: From Data to Information Granules, John Wiley, Hoboken, NJ, 2005.
- [85] W. Pedrycz, From fuzzy sets to shadowed sets: Interpretation and computing, *International Journal of Intelligent Systems* 24 (1) (2009) 48–61.
- [86] W. Pedrycz, Shadowed sets in the characterization of rough-fuzzy clustering, *Pattern Recognition* 44 (8) (2011) 1738–1749.

- [87] D. Pelleg, A. Moore, X-means: Extending k-means with efficient estimation of the number of clusters, in: *Proceedings ICML 2000 – 17th International Conference on Machine Learning*, Stanford, CA, USA, 2000, pp. 727–734.
- [88] G. Peters, Outliers in rough k-means clustering, *Proceedings PreMI 2005 – 1st International Conference on Pattern Recognition and Machine Intelligence*, LNCS, vol. 3776, Springer-Verlag, Berlin, Germany, 2005, pp. 702–707.
- [89] G. Peters, Some refinements of rough k-means, *Pattern Recognition* 39 (2006) 1481–1491.
- [90] G. Peters, M. Lampart, A partitive rough clustering algorithm, *Proceedings RSCTC 2006 – 5th International Conference on Rough Sets and Current Trends in Computing*, LNAI, vol. 4259, Springer-Verlag, Berlin, Germany, 2006, pp. 657–666.
- [91] G. Peters, M. Lampart, R. Weber, Evolutionary rough k-medoids clustering, *Transactions on Rough Sets VIII* (2008) 289–306.
- [92] G. Peters, R. Weber, R. Nowatzke, Dynamic rough clustering and its applications, *Applied Soft Computing* 12 (2012) 3193–3207.
- [93] A. Petrosino, M. Ceccarelli, Unsupervised texture discrimination based on rough fuzzy sets and parallel hierarchical clustering, in: *ICPR 2000 – 15th International Conference on Pattern Recognition*, 2000, p. 7100.
- [94] A. Petrosino, G. Salvi, Rough fuzzy set based scale space transforms and their use in image analysis, *International Journal of Approximate Reasoning* 41 (2006) 212–228.
- [95] G. Punj, D.W. Stewart, Cluster analysis in marketing research: Review and suggestions for application, *Journal of Marketing Research* 20 (2) (1983) 134–148.
- [96] Y. Qian, K-means algorithm and its application for clustering companies listed in Zhejiang province, in: A. Zanasi, C. Brebbia, N. Ebecken (Eds.), *Data Mining VII: Data, Text and Web Mining and their Business Applications*, WIT Transactions on Information and Communication Technologies, Southampton, UK, 2006, pp. 35–44.
- [97] S. Ray, R.H. Turi, Determination of number of clusters in k-means clustering and application in colour image segmentation, in: N.R. Pal, A.K. De, J. Das (Eds.), *Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99)*, Narosa Publishing House, New Delhi, India, 1999, pp. 137–143.
- [98] F.C.H. Rhee, Uncertain fuzzy clustering: Insights and recommendation, *IEEE Computational Intelligence Magazine* 2 (1) (2007) 44–56.
- [99] V. Schwämmle, N.O. Jensen, A simple and fast method to determine the parameters for fuzzy c-means cluster analysis, *Bioinformatics* 26 (2010) 2841–2848.
- [100] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, 1976.
- [101] R. Soni, R. Nanda, Neighborhood clustering of web users with rough k-means, in: *Proceedings of 6th WSEAS International Conference on Circuits, Systems, Electronics, Control & Signal Processing*, Cairo, Egypt, 2007, pp. 570–574.
- [102] A. Ultsch, C. Vetter, Selforganizing feature maps versus statistical clustering: A benchmark, *Tech. Rep. 9*, Dept. of Mathematics and Computer Science, University of Marburg, Germany, 1994.
- [103] H. Wang, M. Zhou, A refined rough k-means clustering with hybrid threshold, in: *Rough Sets and Current Trends in Computing*, vol. 7413 of LNCS, 2012, pp. 26–35.
- [104] M. Windham, Cluster validity for fuzzy clustering algorithms, *Fuzzy Sets and Systems* 5 (1981) 177–185.
- [105] K.L. Wu, Analysis of parameter selections for fuzzy c-means, *Pattern Recognition* 45 (2012) 407–415.
- [106] R. Xu, D. Wunsch, Survey of clustering algorithms, *IEEE Transactions on Neural Networks* 16 (3) (2005) 645–678.
- [107] R.R. Yager, L. Liu (Eds.), *Classic Works of the Dempster–Shafer Theory of Belief Functions*, *Studies in Fuzziness and Soft Computing*, vol. 219, Springer, Berlin, 2008.
- [108] L. Yingqiu, L. Wei, L. Yungchun, Network traffic classification using k-means clustering, in: *2nd International Multi-Symposiums on Computer and Computational Sciences (IMSCCS 2007)*, Beijing, 2007, pp. 360–365.
- [109] J. Yu, Q. Cheng, H. Huang, Analysis of the weighting exponent in the FCM, *IEEE Transactions on Systems, Man, and Cybernetics, Part B* 34 (2004) 634–639.
- [110] W. Zhong, G. Altun, R. Harrison, P.C. Tai, Y. Pan, Improved k-means clustering algorithm for exploring local protein sequence motifs representing common structural property, *IEEE Transactions on NanoBioscience* 4 (3) (2005) 255–265.