

A new evidential c-means clustering method

Zhun-ga Liu¹, Jean Dezert², Quan Pan¹, Yong-mei Cheng¹

1. School of Automation, Northwestern Polytechnical University, Xi'an, China.

Email: liuzhunga@gmail.com

2. ONERA - The French Aerospace Lab, F-91761 Palaiseau, France.

Email: jean.dezert@onera.fr

Abstract—Data clustering methods integrating information fusion techniques have been recently developed in the framework of belief functions. More precisely, the evidential version of fuzzy c-means (ECM) method has been proposed to deal with the clustering of proximity data based on an extension of the popular fuzzy c-means (FCM) clustering method. In fact ECM doesn't perform very well for proximity data because it is based only on the distance between the object and the clusters' center to determine the mass of belief of the object commitment. As a result, different clusters can overlap with close centers which is not very efficient for data clustering. To overcome this problem, we propose a new clustering method called belief functions c-means (BFCM) in this work. In BFCM, both the distance between the object and the imprecise cluster's center, and the distances between the object and the centers of the involved specific clusters for the mass determination are taken into account. The object will be considered belonging to a specific cluster if it is very close to this cluster's center, or belonging to an imprecise cluster if it lies in the middle (overlapped zone) of some specific clusters, or belonging to the outlier cluster if it is too far from the data set. Pignistic probability can be applied for the hard decision making support in BFCM. Several examples are given to illustrate how BFCM works, and to show how it outperforms ECM and FCM for the proximity data.

Keywords: data clustering, information fusion, belief functions, BFCM, ECM, FCM.

I. INTRODUCTION

Belief functions theory [1–3] also called evidence theory has been widely applied in the information fusion field [4–7] for dealing with the uncertain and imprecise information. The classical data clustering (classification) techniques like fuzzy¹ c-means (FCM) [8] were developed in the probability framework. However, in the clustering of the close data sets, some data points (objects) are close to each other, but really originate from different classes. In such cases, it would be hard to correctly classify such objects into a particular cluster, which indicates that the probabilistic framework can not well model the imprecise information. Now several clustering and classification methods [5, 9, 10] have been extended to work in the framework of belief functions for dealing with the close data sets.

The c-means clustering method and its variants remain so far the most popular data clustering methods. Recently a new version of c-means method has been developed by Kurihara and Welling [11] in the Bayesian framework. In this paper,

we get out of the Bayesian framework and we propose a new clustering method inspired from FCM and ECM, and based on the framework of belief functions. FCM generalizes the fuzzy partition for the data set. In order to well model the imprecise information in the data clustering, an evidential version of fuzzy c-Means (ECM) clustering method has been proposed recently in [9] based on the belief functions framework. ECM is able to produce the credal partition [10] including outlier cluster, specific (singleton) clusters² and imprecise clusters (meta-clusters) as the disjunction of the associated singleton clusters. Let us consider a frame of discernment $\Omega = \{w_1, \dots, w_k\}$. The credal partition as an extension of fuzzy partition allows the object not only to belong to the specific clusters, but also to belong to any subsets (with respect to meta-cluster) of frame of discernment Ω with a mass of belief defined on the power-set of Ω (denoted by 2^Ω). The power-set 2^Ω is constructed by all the subsets of Ω . Therefore, the credal partition provides more enlarged partition results than the fuzzy partition techniques, and provides a deeper insight to the data. This made ECM appealing for solving imprecise (uncertain) data clustering problems in applications.

In ECM, the mass of belief for each object is based on the distance between the object and the centers of focal elements that are subsets of Ω . The singleton focal element corresponds to the specific cluster, and the focal element composed by the union of more than one singleton element of Ω is called an imprecise element and it corresponds to (imprecise) meta-cluster. The meta-clusters' centers are obtained by the simple average of the involved specific clusters' centers. The smaller distance between the object and the clustering center, the bigger mass of belief of the object committed to the corresponding cluster. The data point in (or close to) the middle of several specific clusters will be likely in the meta-cluster as the disjunction (union) of these associated specific clusters, since it is close to the average center of these specific clusters. However, the clustering centers of the singleton clusters may overlap with some centers of meta-clusters, and the centers of the different meta-clusters can still be overlapped in some cases. This problem will cause troubles in the association of an object with a particular specific (singleton) cluster, or the (different) meta-cluster the object may also belong to.

For example, let's consider a 4-classes data set and the frame

¹The "fuzzy" epithet is misleading and should better be replaced by "probabilistic" in fact.

²Singleton cluster means there is only one class (element) in this cluster, and it is the same as the cluster in FCM.

$\Omega = \{w_1, w_2, w_3, w_4\}$ with the corresponding centers $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3,$ and \mathbf{v}_4 . It is possible that³ in some cases to have $\mathbf{v}_2 \approx (\mathbf{v}_1 + \mathbf{v}_3)/2 = \mathbf{v}_{1,3}$, and to have also $\mathbf{v}_{1,4} = (\mathbf{v}_1 + \mathbf{v}_4)/2 \approx (\mathbf{v}_2 + \mathbf{v}_3)/2 = \mathbf{v}_{2,3}$. Then with ECM, a trouble occurs in the association of an object with w_2 or $w_1 \cup w_3$ if the object is close to \mathbf{v}_2 (or $\mathbf{v}_{1,3}$), and also with $w_1 \cup w_4$ and $w_2 \cup w_3$ if the object is close to $\mathbf{v}_{1,4}$ (or $\mathbf{v}_{2,3}$), since in ECM the determination of the mass only depends on the distance between the object and the centers of clusters. So we probably get overlapped clusters (i.e. w_2 and $w_1 \cup w_3$, $w_1 \cup w_4$ and $w_2 \cup w_3$) with ECM, which seems not a very efficient and reasonable clustering result. We also have noticed that the penalizing weighting factor $|A|^\alpha$ ($|A|$ being the cardinality of the set A) to control the number of data point in the cluster A in ECM is usually very big and especially for the meta-clusters. Big $|A|^\alpha$ factors lead that the clusters with different cardinality have different weights in the determination of masses of belief. This principle seems ill-adapted (at least theoretically unjustified) for the meta-clusters with big cardinality.

In order to overcome the limitation of ECM, a method called belief c-means (BCM) was recently developed in [12]. In BCM, the meta-clusters are given a distinct interpretation as in ECM, and the meta-clusters refer to the objects that are neither close to the associated clustering centers, nor far from them. The computation of mass of belief on the meta-clusters is a bit complex in BCM. In this work, we propose a simpler evidential c-means clustering method called belief functions c-means (BFCM). The meta-clusters are used to represent the objects that are in the middle of different clusters and difficult to be correctly committed into a particular cluster, and the mass of belief on the meta-clusters is determined in a new and simple way. In the determination of the mass of belief focused on the meta-cluster, we take into account not only the distance between the object and the center of the meta-cluster, but also the distances between the object and the centers of these specific clusters involved in the meta-cluster. Then, even if the centers of the different clusters may be overlapped, the distances between the object and the centers of the associated specific clusters may be different. Thus, the data will be committed to the cluster whose associated specific clusters are close to the data. Some few data points far from the data set with respect to the outlier threshold will be considered as outlier. We will represent the outlier class by \emptyset . The objective function of BFCM is designed according to this basic principle, and the clustering centers and the mass of belief for the object are acquired by the optimization of the objective function. The credal partition can be reduced to fuzzy partition using the Pignistic probability transformation $BetP(\cdot)$ [3] for hard decision-making support if necessary. After a brief recall of ECM approach in section II, the new BFCM approach is introduced in section III. Then some simple examples are given in section IV to illustrate the effectiveness of BFCM with respect to FCM and ECM approaches, before concluding this paper in section V.

³We denote by $\mathbf{v}_{i,j}$ the center of the meta-cluster $w_i \cup w_j$.

II. BASICS OF EVIDENTIAL C-MEANS (ECM)

ECM [9] is an extension of FCM based on credal partition using the theoretical framework of belief functions. The class membership of an object $\mathbf{x}_i = (x_{i_1}, \dots, x_{i_p})$ is represented by a basic belief assignment (bba) $m_i(\cdot)$ over a given frame of discernment $\Omega = \{w_1, \dots, w_k\}$. This representation is able to model all situations ranging from complete ignorance to full certainty concerning the class of \mathbf{x}_i . In ECM, the mass of belief for associating the object \mathbf{x}_i with an element A_j of 2^Ω denoted by $m_{ij} \triangleq m_{\mathbf{x}_i}(A_j)$, is determined from the distance d_{ij} between \mathbf{x}_i and the (center) prototype vector $\bar{\mathbf{v}}_j$ of the element A_j . Note that A_j can either be a single class, or an union of single classes. The prototype vector $\bar{\mathbf{v}}_j$ of A_j , is defined as the mean vector of the prototype attribute vectors of the singletons of Ω included in A_j . $\bar{\mathbf{v}}_j$ is defined mathematically by

$$\bar{\mathbf{v}}_j = \frac{1}{c_j} \sum_{i=1}^c s_{kj} \mathbf{v}_k \quad \text{with} \quad s_{kj} = \begin{cases} 1, & \text{if } w_k \in A_j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where \mathbf{v}_k is the prototype attribute vector of (i.e. the center of the single cluster associated with) the single class w_k , and $c_j = |A_j|$ denotes the cardinality of A_j , and d_{ij} denotes the Euclidean distance between \mathbf{x}_i and $\bar{\mathbf{v}}_j$.

In ECM, the determination of $m_{ij} \triangleq m_{\mathbf{x}_i}(A_j)$ is based on d_{ij} as in FCM. Actually, m_{ij} is obtained by the minimization of the following objective function under a constraint:

$$J_{ECM} = \sum_{i=1}^n \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} c_j^\alpha m_{ij}^\beta d_{ij}^2 + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta \quad (2)$$

Subject to

$$\sum_{A_j \subseteq \Omega, A_j \neq \emptyset} m_{ij} + m_{i\emptyset} = 1 \quad (3)$$

The solution of the minimization of (2) under the constraint (3) has been established by Masson and Denœux in [9] and it is given for each object \mathbf{x}_i , ($i = 1, 2, \dots, n$) by:

- For all $A_j \subseteq \Omega$ and $A_j \neq \emptyset$,

$$m_{ij} = \frac{c_j^{-\alpha/(\beta-1)} d_{ij}^{-2/(\beta-1)}}{\sum_{A_k \neq \emptyset} c_k^{-\alpha/(\beta-1)} d_{ik}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}} \quad (4)$$

where α is a tuning parameter allowing to control the degree of penalization; β is a weighting exponent (its suggested default value in [9] is $\beta = 2$); δ is a given threshold tuning parameter for the filtering of the outliers; $c_j = |A_j|$ is the cardinality of the set A_j .

- For $A_j = \emptyset$,

$$m_{i\emptyset} \triangleq m_{\mathbf{x}_i}(\emptyset) = 1 - \sum_{A_j \neq \emptyset} m_{ij}. \quad (5)$$

The centers of the class are given by the rows of the matrix $V_{C \times p}$

$$V_{C \times p} = H_{C \times C}^{-1} \cdot B_{C \times p} \quad (6)$$

where the elements B_{lq} of $B_{c \times p}$ matrix for $l = 1, 2, \dots, c$, $q = 1, 2, \dots, p$, and the elements H_{lk} of $H_{c \times c}$ matrix for $l, k = 1, 2, \dots, c$ are given by:

$$B_{lq} = \sum_{i=1}^n x_{i_q} \sum_{w_l \in A_j} c_j^{\alpha-1} m_{ij}^\beta \quad (7)$$

$$H_{lk} = \sum_{i=1}^n \sum_{\{w_k, w_l\} \subseteq A_j} c_j^{\alpha-2} m_{ij}^\beta \quad (8)$$

III. NEW BELIEF FUNCTIONS C-MEANS (BFCM)

A. Basic principle of BFCM

In ECM, the prototype vector (i.e. the center) of an imprecise (i.e. a meta) cluster is obtained by averaging the prototype vectors of the specific clusters included in it. The smaller distance between the object and the center of the (specific or imprecise) cluster will lead to the bigger mass of belief committed to this corresponding cluster. Whereas, a specific cluster's center may overlap with (or very close to) the center of a meta-cluster in some cases, and the clustering centers of the distinct meta-clusters can still overlap sometimes. If so, it will cause a problem in the association of an object with a particular specific cluster or the meta-cluster the object may also belong to, and these different clusters will also overlap. Moreover, the penalized weighting factor α of ECM is usually very big in the applications to avoid the problem that a number of data points are committed to the meta-clusters. The big α leads that the clusters with different cardinality take quite different weights in the determination of mass of belief, which seems unfair (unjustified) especially for the meta-clusters with big cardinality. To circumvent these problems, a new extension of fuzzy c-means clustering method under belief functions framework is proposed, and we call it *belief functions c-means* method (and denoted by BFCM acronym).

In BFCM, the data committed to a specific (singleton) cluster must be close to the center of this cluster. The data belonging to the imprecise/meta-cluster cluster (i.e. the disjunction of these singleton clusters) should be not only close to the average prototype of these involved singleton clusters (as done in ECM), but also close to prototypes of these singleton clusters. The outlier cluster represents the few data points very far from the data set with respect to the outlier threshold. The penalized weighting factor α in BFCM is always very small and even zero, and the cardinality of the clusters is not necessarily taken into account in the determination of belief. So the meta-clusters and the specific clusters can be considered with the (almost) same weight in BFCM. Thus, in BFCM, the mass of belief committed to a singleton cluster for an object will depend on the distance between the object and the center of the singleton cluster. The belief committed to a meta-cluster will depend on the distances between the object and the prototypes of the singleton clusters involved in the meta-cluster, as well as on the distances between the object and the average center of these singleton cluster centers. The belief of the outlier cluster is determined according to a chosen outlier threshold.

B. The objective function of BFCM

Let us consider a set of $n \geq 1$ objects. Each object (data point) is indexed by a number i and is represented by a given attribute vector \mathbf{x}_i of dimension $1 \times p$ with $p \geq 1$. These objects will be clustered in a given frame of discernment (a finite set of specific classes) $\Omega = \{w_1, w_2, \dots, w_k\}$ with the corresponding centers $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$. The meta-clusters are generalized by the disjunction of these specific classes in the frame of discernment Ω .

In BFCM, the mass of belief $m_{\mathbf{x}_i}(w_j)$ of \mathbf{x}_i committed to the "singleton" cluster w_j (i.e. a specific class), is assumed to increase with the decrease of the distance $d_{\mathbf{x}_i \mathbf{v}_j}$ between \mathbf{x}_i and the center \mathbf{v}_j . The smaller $d_{\mathbf{x}_i \mathbf{v}_j}$ leads to the bigger $m_{\mathbf{x}_i}(w_j)$. The determination of the belief committed to the imprecise clusters takes into account not only the distance between the object x_i and the prototypes of the meta-cluster, but also the distances between x_i and the prototypes of the singleton clusters involved in the imprecise clusters. This is done as follows. If the object \mathbf{x}_i is closer to the centers $\mathbf{v}_j, \mathbf{v}_{j+1}, \dots, \mathbf{v}_t$ for some $j, t < k$, which indicates that \mathbf{x}_i has potentially more chance to belong to the classes w_j, w_{j+1}, \dots, w_t than to other clusters, and in the meanwhile the distances between the object and the average center of singleton clusters $\bar{\mathbf{v}}_s = (\mathbf{v}_j + \dots + \mathbf{v}_t)/(t - j)$ is smaller, then the value committed to $m_{\mathbf{x}_i}(w_j \cup w_{j+1} \dots \cup w_t)$ by BFCM will increase. $m_{\mathbf{x}_i}(w_j \cup w_{j+1} \dots \cup w_t)$ will be obtained according to the average of the distances $d_{\mathbf{x}_i \mathbf{v}_j}, \dots, d_{\mathbf{x}_i \mathbf{v}_t}$ and $d_{\mathbf{x}_i \bar{\mathbf{v}}_s}$. Moreover, the particular distance between the object and the average center of singleton clusters can be weighted by a chosen weighting factor which can be tuned according to the application. The objects too far from all the centers of clusters (according to a chosen outlier threshold δ) will be considered as outliers and committed to the outlier-class represented by \emptyset .

In BFCM, we minimize the objective function J_{BFCM} to compute the masses of belief of the matrix $M = (m_1, \dots, m_n) \in \mathbb{R}^{n \times 2^{|\Omega|}}$ associated with the credal partition, and to compute the matrix $V_{c \times p}$ of cluster centers. The minimization is done under the constraint (10). The objective function J_{BFCM} of BFCM is different from J_{ECM} (2) and it is defined by

$$\begin{aligned} J_{BFCM}(M, V) = & \sum_{i=1}^n \sum_{\substack{j|A_j \subseteq \Omega, \\ |A_j|=1}} m_{ij}^\beta d_{ij}^2 \\ & + \sum_{i=1}^n \sum_{\substack{j|A_j \subseteq \Omega, \\ |A_j|>1}} |A_j|^\alpha m_{ij}^\beta \frac{\sum_{A_k \in A_j, |A_k|=1} d_{ik}^2 + \gamma d_{i\bar{\mathbf{v}}_s}^2}{|A_j| + \gamma} \\ & + \sum_{i=1}^n m_{i\emptyset}^\beta \delta^2 \quad (9) \end{aligned}$$

Subject to

$$\sum_{j|A_j \subseteq \Omega} m_{ij} + m_{i\emptyset} = 1 \quad (10)$$

where $d_{ij} \triangleq d_{\mathbf{x}_i \mathbf{v}_j}$ is the distance between the data point \mathbf{x}_i and the class center \mathbf{v}_j corresponding to the cluster A_j . If A_j is a meta-cluster, its center is the mean value of the centers of the involved specific clusters, and it is given by

$$\bar{\mathbf{v}}_j = \frac{1}{|A_j|} \cdot \sum_{A_k \in A_j, |A_k|=1} \mathbf{v}_k. \quad (11)$$

The tuning parameters α , β , and δ have the same meaning as in ECM. γ is the weighting factor of the distance between the object and the center of the meta-clusters. The objective function J_{BFCM} can be justified by the following expected properties:

- If the object (a data point) is quite far from all the centers of the singleton clusters (beyond the given threshold δ), the most mass of belief will be committed to outlier cluster represented by \emptyset .
- The belief of a data point associated to a singleton cluster is proportional to the distance between the data point and the center of the singleton cluster.
- The belief of a data point associated to a meta-cluster is proportional to the average distance between the data point and the involved singleton clusters' centers, and also to the distance between the object and the center of the meta-cluster with the tuned weighting factor γ .

We also notice that if A_j is a specific (singleton) cluster, then

$$\frac{\sum_{A_k \in A_j, |A_k|=1} d_{ik}^2 + \gamma d_{ij}^2}{|A_j| + \gamma} = \frac{d_{ij}^2 + \gamma d_{ij}^2}{1 + \gamma} = d_{ij}^2 \quad (12)$$

Therefore, the objective function given in (9) can be rewritten as in (13) for the convenience of optimization in the sequel.

$$J_{BFCM}(M, V) = \sum_{i=1}^n \sum_{j|A_j \subseteq \Omega} |A_j|^\alpha m_{ij}^\beta D_{ij}^2 \quad (13)$$

with

$$D_{ij}^2 = \begin{cases} \delta^2, & A_j = \emptyset \\ \frac{\sum_{A_k \in A_j, |A_k|=1} d_{ik}^2 + \gamma d_{ij}^2}{|A_j| + \gamma}, & A_j \neq \emptyset \end{cases} \quad (14)$$

As with FCM [8] or ECM [9] methods, $\beta = 2$ can be used as default value, and we have used this value in our simulations presented in the sequel. α allows to control the number of points assigned to the meta-clusters. The higher α leads to the less imprecision of the clustering result. However, since we consider that the meta-clusters must not be greatly penalized with respect to the singleton clusters, the α parameter should be very small and even zero (i.e $\alpha = 0$ is default in this work). δ is the threshold of the outliers and it is strongly data-dependent. The bigger dispersion of the data set leads to the bigger δ . The bigger γ generally produces a bigger meta-cluster. We suggest to take $1 \leq \gamma$ in practice. Meanwhile, γ must not be too big. If γ is too big, the belief of the meta-cluster will mainly be determined according to the distance between the object and the average center of the intersecting cluster. Then, the belief committed to the meta-cluster and to the singleton cluster may be very similar when the center

for the meta-cluster is close to the center of the singleton cluster. To implement BFCM algorithm, we need to minimize J_{BFCM} . This is explained in the next subsection.

C. Minimization of J_{BFCM} objective function

To minimize J_{BFCM} , we use Lagrange multipliers method. In the first step, the centers of the clusters V are considered fixed. Lagrange multipliers λ_i are used to solve the constrained minimization problem with respect to M as follows:

$$\mathcal{L}(M, \lambda_1, \dots, \lambda_n) = J_{BFCM}(M, V) - \sum_{i=1}^n \lambda_i \left(\sum_{j|A_j \subseteq \Omega} m_{ij} - 1 \right) \quad (15)$$

By differentiating the Lagrangian with respect to the m_{ij} , $m_{i\emptyset}$ and λ_i and setting the derivatives to zero, we obtain:

$$\frac{\partial \mathcal{L}}{\partial m_{ij}} = |A_j|^\alpha \beta m_{ij}^{\beta-1} D_{ij}^2 - \lambda_i = 0 \quad (16)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = \sum_{j|A_j \subseteq \Omega} m_{ij} - 1 = 0 \quad (17)$$

From (16), one gets

$$m_{ij} = \left(\frac{\lambda_i}{\beta} \right)^{\frac{1}{\beta-1}} \left(\frac{1}{|A_j|^\alpha D_{ij}^2} \right)^{\frac{1}{\beta-1}} \quad (18)$$

and using (17) and (18), one has

$$\left(\frac{\lambda_i}{\beta} \right)^{1/(\beta-1)} = \frac{1}{\sum_{j|A_j \subseteq \Omega} \frac{1}{|A_j|^{\alpha/(\beta-1)} D_{ij}^{2/(\beta-1)}}} \quad (19)$$

Returning in (18), one obtains the necessary condition of optimality for M :

$$m_{ij} = \frac{|A_j|^{-\alpha/(\beta-1)} D_{ij}^{-2/(\beta-1)}}{\sum_{A_k \subseteq \Omega} |A_k|^{-\alpha/(\beta-1)} D_{ik}^{-2/(\beta-1)}} \quad (20)$$

Using (14), we get the expressions of the mass of belief respectively committed to different focal elements including empty set (corresponding to outlier), singleton cluster, imprecise cluster.

$$m_{ij} = \frac{\delta^{-2/(\beta-1)}}{\delta^{\frac{-2}{\beta-1}} + \sum_{A_j \subseteq \Omega} |A_j|^{\frac{-\alpha}{\beta-1}} \left(\frac{\sum_{A_k \in A_j, |A_k|=1} d_{ik}^2 + \gamma d_{ij}^2}{|A_j| + \gamma} \right)^{\frac{-2}{\beta-1}}}; \quad A_j = \emptyset \quad (21)$$

$$m_{ij} = \frac{|A_j|^{\frac{-\alpha}{\beta-1}} \left(\frac{\sum_{A_k \in A_j, |A_k|=1} d_{ik}^2 + \gamma d_{ij}^2}{|A_j| + \gamma} \right)^{\frac{-2}{\beta-1}}}{\delta^{\frac{-2}{\beta-1}} + \sum_{A_j \subseteq \Omega} |A_j|^{\frac{-\alpha}{\beta-1}} \left(\frac{\sum_{A_k \in A_j, |A_k|=1} d_{ik}^2 + \gamma d_{ij}^2}{|A_j| + \gamma} \right)^{\frac{-2}{\beta-1}}}; \quad A_j \neq \emptyset \quad (22)$$

More precisely, if A_j is a specific cluster, one gets:

$$m_{ij} = \frac{d_{ij}^{-2/(\beta-1)}}{\delta^{\frac{-2}{\beta-1}} + \sum_{A_j \subseteq \Omega} |A_j|^{\frac{-\alpha}{\beta-1}} \left(\frac{\sum_{A_k \in A_j, |A_k|=1} d_{ik}^2 + \gamma d_{ij}^2}{|A_j| + \gamma} \right)^{\frac{-2}{\beta-1}}};$$

$$|A_j| = 1, A_j \neq \emptyset \quad (23)$$

Now let us consider that M is fixed. The minimization of J_{BFCM} with respect to V is an unconstrained optimization problem. The partial derivatives of J_{BFCM} with respect to the centers are given by:

$$\frac{\partial J_{BFCM}}{\partial \mathbf{v}_l} = \sum_{i=1}^n \sum_{A_l \cap A_j \neq \emptyset} |A_j|^\alpha m_{ij}^\beta \frac{\partial D_{ij}^2}{\partial \mathbf{v}_l} \quad (24)$$

with

$$\frac{\partial D_{il}^2}{\partial \mathbf{v}_l} = 2(\mathbf{v}_l - \mathbf{x}_i), |A_l| = 1 \quad (25)$$

$$\frac{\partial D_{ij}^2}{\partial \mathbf{v}_l} = \frac{2(\mathbf{v}_l - \mathbf{x}_i) + \frac{2\gamma}{|A_j|} \left(\frac{\sum_{A_g \subseteq A_j} \mathbf{v}_g}{|A_j|} - \mathbf{x}_i \right)}{|A_j| + \gamma_2}, A_l \in A_j \quad (26)$$

Therefore,

$$\frac{\partial J_{BCM}}{\partial \mathbf{v}_l} = \sum_{i=1}^n 2m_{il}^\beta (\mathbf{v}_l - \mathbf{x}_i) + \sum_{i=1}^n \sum_{A_l \in A_j} |A_j|^\alpha m_{ij}^\beta \frac{2(\mathbf{v}_l - \mathbf{x}_i) + \frac{2\gamma}{|A_j|} \left(\frac{\sum_{A_g \subseteq A_j} \mathbf{v}_g}{|A_j|} - \mathbf{x}_i \right)}{|A_j| + \gamma_2} \quad (27)$$

Setting these derivatives to zero gives c linear equations that can be written as:

$$\left(\sum_{i=1}^n m_{il}^\beta + \sum_{i=1}^n \sum_{A_l \in A_j} |A_j|^\alpha m_{ij}^\beta \frac{1 + \frac{\gamma}{|A_j|}}{|A_j| + \gamma_2} \right) \mathbf{x}_i = \sum_{i=1}^n m_{il}^\beta \mathbf{v}_l + \sum_{i=1}^n \sum_{A_l \in A_j} |A_j|^\alpha m_{ij}^\beta \frac{\mathbf{v}_l + \frac{\gamma \sum_{A_g \subseteq A_j} \mathbf{v}_g}{|A_j|^2}}{|A_j| + \gamma_2} \quad (28)$$

This system of linear equations can be concisely rewritten as

$$B_{c \times n} X_{n \times p} = H_{c \times c} V_{c \times p} \quad (29)$$

where $X_{n \times p} \triangleq [\mathbf{x}_1, \dots, \mathbf{x}_n]'$ (i.e. the set of all attribute vectors of data). The elements of $B_{c \times n}$ and $H_{c \times c}$ are defined by

$$B_{li} \triangleq m_{il}^\beta + \sum_{A_l \in A_j} |A_j|^\alpha m_{ij}^\beta \frac{1 + \frac{\gamma}{|A_j|}}{|A_j| + \gamma} \quad (30)$$

$$H_{ll} \triangleq \sum_{i=1}^n m_{il}^\beta + \sum_{i=1}^n \sum_{A_l \in A_j} |A_j|^\alpha m_{ij}^\beta \frac{1 + \frac{\gamma}{|A_j|^2}}{|A_j| + \gamma} \quad (31)$$

$$H_{lq} \triangleq \sum_{i=1}^n \sum_{A_l \in A_k, A_q \in A_k} |A_k|^\alpha m_{ik}^\beta \frac{\gamma}{|A_k|^2 (|A_k| + \gamma)}, l \neq q \quad (32)$$

V is the solution of the linear system (29) which can be calculated by a standard linear system solver. The pseudo-code of the BFCM algorithm is given in Table 1 for convenience.

Table 1. Belief functions c-means algorithm

Input:	Data to cluster: $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ in \mathbb{R}^p
Parameters:	c : number clusters, $2 \leq c < n$ $\delta > 0$: outlier threshold $\alpha \geq 0$: penalization factor ($\alpha_d = 0$) $\epsilon > 0$: termination threshold ($\epsilon_d = 10^{-3}$) $\gamma > 0$: weight of the distance ($\gamma_d = 1$)
Initialization:	Choose randomly initial mass M_0
	$t \leftarrow 0$
	Repeat
	$t \leftarrow t + 1$
	Compute B_t and H_t by (30)-(32);
	Compute V_t by solving (29);
	Compute M_t using (21)-(22);
	until $\ V_t - V_{t-1}\ < \epsilon$

$\alpha_d, \beta_d, \epsilon_d$ and γ_d are the default values of the parameters applied in BFCM. The initial bba M_0 can be randomly generated, and the final clustering results are not very sensitive to the initialization of M_0 after the process of optimization.

D. Hard decision-making support for BFCM

When necessary in applications, the credal partition in BFCM can be reduced to the fuzzy partition for hard decision making. The meta-clusters need to be eliminated from credal partition in order to get a fuzzy partition. The mass of belief committed to the removed clusters should be redistributed to the other focal elements. In this work, the Pignistic probability transformation [3], denoted $BetP(\cdot)$, is applied for the redistribution of the belief of meta-clusters to the specific clusters. $BetP(\cdot)$ is defined by

$$BetP(w) = \sum_{w \in A} \frac{m(A)}{|A|}, w \in \Omega \quad (33)$$

where $|A|$ is the cardinality of the set A , i.e. the number of singleton elements A contains.

The lower and upper bounds of imprecise probability associated with bba's correspond to the belief function $Bel(\cdot)$ and the plausibility function $Pl(\cdot)$ [1]. $[Bel(\cdot), Pl(\cdot)]$ is interpreted as the imprecise interval of the unknown probability $P(\cdot)$. These bounds are given for all X in 2^Ω by

$$Bel(X) = \sum_{A \subseteq X, A \neq \emptyset} m(A) \quad (34)$$

$$Pl(X) = \sum_{A \cap X \neq \emptyset} m(A) \quad (35)$$

$Bel(\cdot)$ and $Pl(\cdot)$ can also be used for decision-making support when adopting pessimistic or optimistic attitudes if necessary.

IV. NUMERICAL EXAMPLES

Example 1: Here we consider a 3-class data set in a line to show the limitation of ECM and FCM with respect to BFCM. There are $3 * 50 = 150$ data points to cluster as shown in Fig. 1a. The data points have been generated from three 2D Gaussian distributions characterized by the following mean vectors and covariance matrices (I being the identity matrix):

$$\begin{aligned}\mu_1 &= (0, 0), \Sigma_1 = I \\ \mu_2 &= (4, 0), \Sigma_2 = 1.5I \\ \mu_3 &= (9, 0), \Sigma_3 = 1.5I\end{aligned}$$

The number of clusters has been set to $C = 3$. The outlier threshold used in ECM and BFCM was $\delta^2 = 36$, and termination threshold $\epsilon = 0.001$. One has tested BFCM with $\alpha = 0$ and $\gamma = 1$ or $\gamma = 2$ to show the effect of the parameter γ on the clustering result. In ECM, we have tested $\alpha = 0$ and $\alpha = 2$ to show its effect on the result. In the legends of figures, we denoted $w_{\{i, \dots, k\}} \triangleq w_i \cup \dots \cup w_k$ for notation convenience. The clustering results obtained with FCM, ECM and BFCM are shown in Fig. 1b–1f.

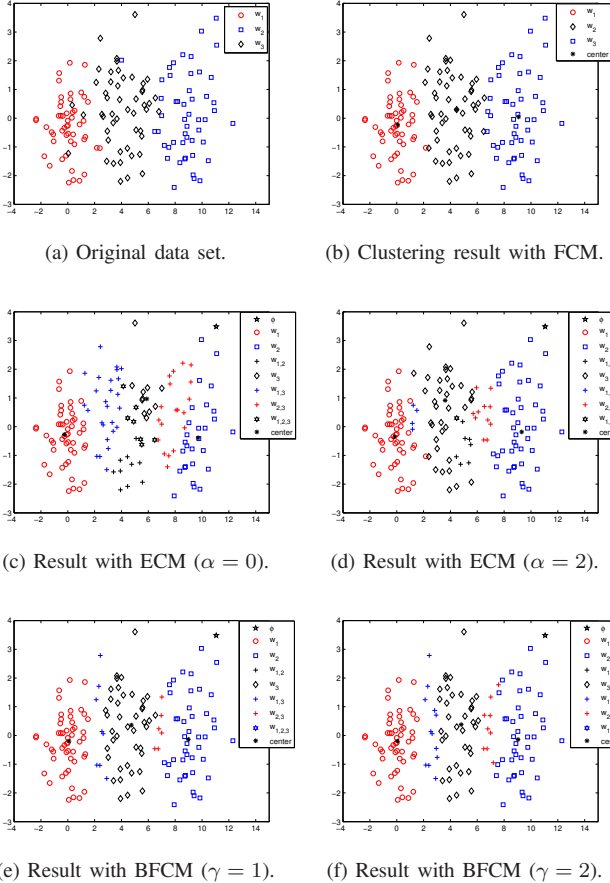


Figure 1: Clustering results by different methods.

The three original data sets are close to each other, and they are partly overlapped at the margins. FCM produces 3

singleton clusters w_1 , w_2 and w_3 based on the probability framework, and some points in the overlapped zone are wrongly clustered. ECM provides the credal partitions in belief function framework, but there are too many data points committed to meta-clusters when $\alpha = 0$ as shown in Fig. 1c. If one takes $\alpha = 2$, the meta-clusters will become much smaller as shown in Fig. 1d. Nevertheless, we can find that some points originated from w_3 are considered belonging to $w_1 \cup w_2$, whereas $w_1 \cup w_2$ and w_3 are incompatible. So these results are not very well justified. This problem arises in ECM approach because the center $(\mathbf{v}_1 + \mathbf{v}_2)/2$ of the meta-cluster $w_1 \cup w_2$ is close to the center \mathbf{v}_3 of the specific cluster w_3 , and because the distances between the objects and the centers of the involved specific clusters (i.e. w_1 and w_2) are not considered in the determination of the belief of the meta-clusters (i.e. $w_1 \cup w_2$). As shown in Fig. 1e, when $\gamma = 1$ is selected in BFCM, several points in the middle of the different classes are clustered into the meta-clusters $w_1 \cup w_3$ and $w_2 \cup w_3$, and there is no point belonging to $w_1 \cup w_2$ since w_1 and w_2 are not overlapped. Several points which are wrongly classified by FCM are committed to the imprecise class. For example, in FCM, one point from w_2 is classified into w_3 and another one point from w_3 is classified into w_2 , whereas these two points are committed to $w_2 \cup w_3$ by BFCM. This indicates that BFCM can reduce the risk of the false clustering but of course it will increase the imprecision of the clustering results, which is the price we have to pay to get more robust results. If BFCM runs with $\gamma = 2$ as shown in Fig. 1f, we see that the meta-clusters become bigger than when taking $\gamma = 1$, and there are still no point committed to $w_1 \cup w_2$. This example illustrates the effectiveness of BFCM in the clustering of the proximity data sets.

Example 2: In this example, the 4-classes of artificial data generated from two 2D Gaussian distributions are used to test BFCM, ECM and FCM. The Gaussian distributions were characterized by the following mean vectors μ_i and covariance matrices Σ_i , $i = 1, 2, 3, 4$:

$$\begin{aligned}\mu_1 &= (0, 0), \Sigma_1 = I \\ \mu_2 &= (5, 0), \Sigma_2 = 2I \\ \mu_3 &= (10, 0), \Sigma_3 = I \\ \mu_4 &= (16, 0), \Sigma_4 = 1.5I\end{aligned}$$

Fifty samples were generated for each class, and there were in total $50 * 4 = 200$ data points to cluster as shown in Fig. 2a. The number of the clusters has been chosen to $C = 4$. BFCM and ECM used $\delta^2 = 25$ and $\epsilon = 0.001$. In BFCM, one has taken $\alpha = 0$, and $\gamma = 1$ or $\gamma = 2$. One has tested ECM with $\alpha = 0$ and with $\alpha = 2$. The clustering results obtained with FCM, ECM and BFCM are shown in Fig. 2b–2f.

As we can see on Fig. 2a, the four original data sets partly overlap on the margins. Nevertheless, FCM just gives four

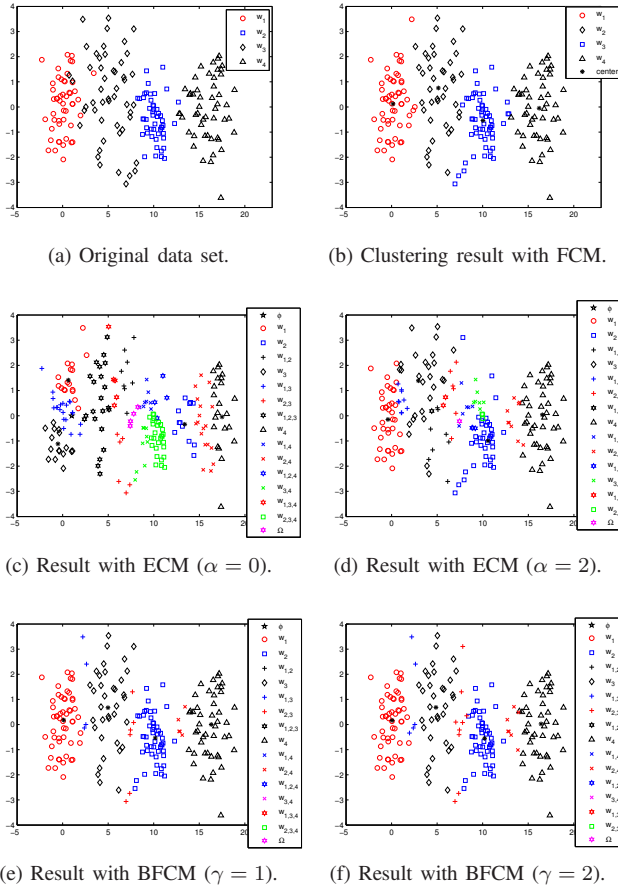


Figure 2: Clustering results by different methods.

specific clusters since it is based on the probability framework. In ECM, the clustering results as in Fig. 2c are very imprecise when one applies $\alpha = 0$. The data points originated from w_2 and w_3 are almost all committed to the distinct meta-clusters, and these meta-clusters are almost overlapped, which is bad for decision-making. If we use ECM with $\alpha = 2$, the results in Fig. 2d slightly improve. However, there remain many points from w_2 and w_3 belonging to the distinct meta-clusters. Particularly, some points are respectively committed to $w_1 \cup w_4$ and $w_2 \cup w_3$. Whereas, $w_1 \cup w_4$ and $w_2 \cup w_3$ are totally incompatible and distinct, but the points in these two clusters are close and even overlap to each other. So these ECM results do not appear reasonable in our opinion. This problem arises mainly because the clustering centers of $w_1 \cup w_4$ and $w_2 \cup w_3$ are almost overlapped, and the distances between the object and the centers of the involved specific clusters are not taken into account in the derivation of bba's by ECM. It indicates that the tuning of the penalization parameter α cannot solve the problem that some distinct clusters may overlap. In such example, BFCM produces only three meta-clusters and these meta-clusters respectively lie in the middle of the associated specific clusters, which intuitively makes perfect sense. The three meta-clusters are not overlapped,

which shows that BFCM has better performances than ECM in the clustering of the proximity data. When BFCM works with $\gamma = 2$, it provides bigger meta-clusters than with $\gamma = 1$, but these meta-clusters still do not overlap at all. Fig. 2f illustrates that bigger γ values increase the size of meta-clusters.

Example 3 : The iris flower real data set shown in Fig. 3a is a typical test case for data clustering [13]. This data set consists of 50 samples from each of three species of Iris flowers (Iris setosa, Iris virginica and Iris versicolor), and the number of the total samples is $3 \times 50 = 150$. Four features were measured from each sample. They are the length and the width of sepal and petal (in centimeters). For the notation convenience, we denote $w_1 \triangleq$ Iris setosa, $w_2 \triangleq$ Iris virginica, and $w_3 \triangleq$ Iris versicolor. The number of singleton clusters is set to $C = 3$. In ECM and BFCM, we have taken $\delta^2 = 100$, and $\epsilon = 0.001$. The best clustering results can be obtained if we select $\alpha = 2$ in ECM as Fig. 3c, and $\gamma = 1.3$ in BFCM as Fig. 3d. The clustering results of FCM are shown in Fig. 3b.

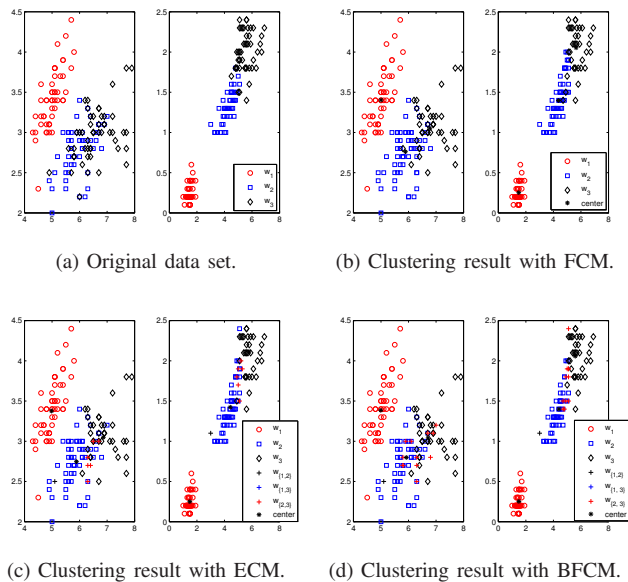


Figure 3: Clustering results of IRIS data set.

As we can see in the original data sets shown in Fig. 3a, w_1 is obviously distinct with w_2 and w_3 , but w_2 and w_3 partly overlap. So w_1 can be easily separated from w_2 and w_3 by the three methods. Whereas, w_2 and w_3 are difficult to separate from each other in the overlapped zone.

There is no meta-cluster in FCM, and all the samples are committed to the specific clusters. It causes that many samples in the overlapped zone of w_2 and w_3 are wrongly classified. More precisely, there are 13 samples of w_3 considered as w_2 , and 3 samples of w_2 considered as w_3 by FCM. In ECM, several samples in the overlapped zone are classified into the meta-clusters. There are 6 samples in w_2 or w_3 belonging to $w_2 \cup w_3$, and 1 sample in w_2 belonging to $w_1 \cup w_2$. 12 samples in the overlapped zone are miss-classified by ECM. The number of samples with false classification reduces to 7

when using BFCM, but of course the number of the samples in the meta-clusters increases to 13. 12 samples are committed to $w_2 \cup w_3$ and 1 sample is committed to $w_1 \cup w_2$ by BFCM. Our analysis indicates that the samples lying in the overlapped zones which were wrongly classified by ECM and FCM are considered as imprecision and committed to the meta-cluster $w_2 \cup w_3$ by BFCM. Hence BFCM reduces the rate of false classification at the necessary price of imprecision increase. It shows that length and the width of sepal and petal are not sufficient for the specific classification of the Iris data set especially for samples in the overlapped zone of w_2 and w_3 . If we have to take a hard decision in applications, some other available information sources should be applied to fuse with the clustering results of BFCM for the final decision-making support. We should be more cautious to the objects in the meta-clusters, since they are quite similar to each other and easily classified by mistake.

V. CONCLUSION

A new c -means clustering method based on the framework of belief functions (denoted by BFCM) has been proposed in this work to overcome the limitations of the evidential c -means (ECM) clustering method. In ECM, the belief of each cluster is only determined (keeping tuning parameters aside) by the distances between the object and the corresponding cluster's centers. If the centers of different clusters overlap or are very close, this will cause trouble in the association of an object with a particular specific cluster or a meta-cluster. Therefore as a final result with ECM we get possibly overlapped clusters, which is not a very efficient solution of the clustering problem. To overcome this problem, in BFCM we take into account both the distance between the object and the meta-cluster's center and also the distances between the object and the centers of the involved specific clusters in order to compute the belief of the (imprecise) meta-clusters. With BFCM, the different clusters cannot overlap even if their centers are very close or overlapped. When one data point is very close to a singleton (specific) cluster's center, it is committed to this specific cluster as done with FCM and ECM. When a data point is close to the middle of some singleton clusters and also close to these clusters themselves, then it is clustered into the meta-cluster defined by the disjunction of these singleton clusters. This implies that this point likely lies in the overlapped zone of these specific clusters and it is not easy to be correctly classified into a specific cluster. Any data point too far from the other data set with respect to the outlier threshold is considered as an outlier. If the hard decision of the clustering is necessary, the Pignistic probability transformation can be used to reduce the credal partition into a fuzzy partition as when working with FCM. The effectiveness of BFCM have been shown by some simple examples using both artificial and real data set with respect to FCM and ECM.

Acknowledgements

The authors want to thank anonymous reviewers for their remarks which helped us to improve the quality of this paper.

This work has been partially supported by National Natural Science Foundation of China (No. 61075029) and PhD Thesis Innovation Fund from Northwestern Polytechnical University (No.cx201015).

REFERENCES

- [1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton Univ. Press, 1976.
- [2] F. Smarandache, J. Dezert (Editors), *Advances and Applications of DSMT for Information Fusion*, American Research Press, Rehoboth, Vol. 1-3, 2004-2009, <http://fs.gallup.unm.edu/DSMT.htm>.
- [3] P. Smets, R. Kennes. *The Transferable Belief Model*, Artificial Intelligence, Vol. 66, 191-243, 1994.
- [4] J. Dezert, Z. Liu, G. Mercier, *Edge Detection in Color Images Based on DSMT*, in Proceedings of International Conference on Information Fusion 2011, Chicago, USA, July 2011.
- [5] P.-E. Dore, A. Fiche, A. Martin, *Models of belief functions - Impacts for patterns recognitions*, in Proceedings of International Conference on Information Fusion 2010, Edinburgh, July 2010.
- [6] Z. Liu, J. Dezert, G. Mercier, Q. Pan, Y. Cheng, *Change detection from remote sensing images based on evidential reasoning*, in Proceedings of International Conference on Information Fusion 2011, Chicago, USA, July 2011.
- [7] Z. Liu, J. Dezert, Q. Pan, and G. Mercier, *Combination of sources of evidence with different discounting factors based on a new dissimilarity measure*, Decision Support Systems, Vol. 52, pp. 133-141, 2011.
- [8] J. C. Bezdek, *Pattern Recognition with fuzzy objective function algorithms*, Plenum Press, New-York, 1981.
- [9] M.-H. Masson, T. Denœux, *ECM: An evidential version of the fuzzy c -means algorithm*, Pattern Recognition, Vol. 41(4), 1384-1397, 2008.
- [10] T. Denœux, M.-H. Masson, *EVCLUS: Evidential CLUSTERING of proximity data*, IEEE Trans. on Systems, Man and Cybernetics Part B, Vol. 34(1), 95- 109, 2004.
- [11] K. Kurihara, M. Welling, *Bayesian k -Means as a 'Maximization-Expectation' Algorithm*, Neural computation, Vol. 21(4), 1145-1172, 2009.
- [12] Z. Liu, J. Dezert, G. Mercier, and Q. Pan, *Belief c -means: An extension of fuzzy c -means algorithm in belief functions framework*, Pattern Recognition Letters, Vol. 33, pp. 291-300, 2012.
- [13] R.A Fisher, *The use of multiple measurements in taxonomic problems*, Annals of Eugenics 7, 179-188, 1936. http://en.wikipedia.org/wiki/Iris_flower_data_set