Pattern Recognition Letters 33 (2012) 291–300

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

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

Belief C-Means: An extension of Fuzzy C-Means algorithm in belief functions framework

Zhun-ga Liu [a,c,*], Jean Dezert [b], Grégoire Mercier [c], Quan Pan [a]

[a] School of Automation, Northwestern Polytechnical University, Xi'an, China
[b] Onera – The French Aerospace Lab, F-91761 Palaiseau, France
[c] Télécom Bretagne, Technopôle Brest-Iroise 29238, France

A R T I C L E I N F O

Article history:
Received 8 June 2011
Available online 31 October 2011
Communicated by D. Dembele

Keywords:
Unsupervised clustering
Data clustering
BCM
ECM
FCM
Belief functions

A B S T R A C T

The well-known Fuzzy C-Means (FCM) algorithm for data clustering has been extended to Evidential C-Means (ECM) algorithm in order to work in the belief functions framework with credal partitions of the data. Depending on data clustering problems, some barycenters of clusters given by ECM can become very close to each other in some cases, and this can cause serious troubles in the performance of ECM for the data clustering. To circumvent this problem, we introduce the notion of imprecise cluster in this paper. The principle of our approach is to consider that objects lying in the middle of specific classes (clusters) barycenters must be committed with equal belief to each specific cluster instead of belonging to an imprecise meta-cluster as done classically in ECM algorithm. Outliers object far away of the centers of two (or more) specific clusters that are hard to be distinguished, will be committed to the imprecise cluster (a disjunctive meta-cluster) composed by these specific clusters. The new Belief C-Means (BCM) algorithm proposed in this paper follows this very simple principle. In BCM, the mass of belief of specific cluster for each object is computed according to distance between object and the center of the cluster it may belong to. The distances between object and centers of the specific clusters and the distances among these centers will be both taken into account in the determination of the mass of belief of the meta-cluster. We do not use the barycenter of the meta-cluster in BCM algorithm contrariwise to what is done with ECM. In this paper we also present several examples to illustrate the interest of BCM, and to show its main differences with respect to clustering techniques based on FCM and ECM.

1. Introduction

In the data clustering analysis, the credal partition based on the belief functions theory has been introduced recently in (Denœux and Masson, 2003, 2004; Masson and Denœux, 2004, 2008). The credal partition is a general extension of the fuzzy (probabilistic) (Bezdek, 1981, 2000), possibilistic partition (Krishnapuram and Keller, 1996) and hard partition (Lloyd, 1982), and it allows the object not only to belong to single clusters, but also to belong to any subsets of the frame of discernment $\Omega = \{w_1, \ldots, w_c\}$ by allocating a mass of belief of each object to all elements of the power-set of Ω denoted 2^{Ω}. So the credal partitioning provides more refined partitioning results than the other partitioning techniques. This makes it very appealing for solving data clustering problems in practice.

The evidential clustering (EVCLUS) algorithm (Denœux and Masson, 2004) for relational data and the Evidential C-Means (ECM) (Masson and Denœux, 2008) for object data have been proposed originally by Denœux and Masson for the credal partitioning of data. In this paper, we focus on the problem of computing a credal partition from object data as in ECM context but using a different approach. ECM (Masson and Denœux, 2008) has been inspired from the Fuzzy C-Means (FCM) (Bezdek, 1981) and Dave's Noise-Clustering algorithm (Dave, 1991), and it can been seen as a direct extension of FCM in the belief functions framework. The mass of belief for each object is computed based on the distance between the object and the barycenters of focal elements that are subsets of Ω. The focal element composed by more than one singleton element of Ω is called an imprecise element and its corresponding cluster is called a meta-cluster. The cluster associated with a singleton element (a single class) is called a specific cluster (or a precise cluster). In ECM algorithm, the barycenter of a meta-cluster is obtained in averaging the centers of the specific clusters involved in the meta-cluster it is related with. It implies that the objects lying in the middle of the several specific clusters will be considered to belong to the meta-cluster represented by the union (disjunction) of these specific clusters. This way of processing is questionable because it can happen that the centers of different

* Corresponding author at: Télécom Bretagne, Technopôle Brest-Iroise 29238, France.

E-mail addresses: liuzhunga@gmail.com (Z.-g. Liu), jean.dezert@onera.fr (J. Dezert), Gregoire.Mercier@telecom-bretagne.eu (G. Mercier).

0167-8655/$ - see front matter
doi:10.1016/j.patrec.2011.10.011

clusters are very close, and eventually that the centers overlap with each other, which is not efficient of course for data clustering with ECM. Thus, there is a serious difficulty for clustering the objects close to these similar/overlapped centers of meta-cluster and specific clusters.

For example, let's consider a set of data to be classified in three distinct classes $\Omega = \{w_1, w_2, w_3\}$ with the prototypes[1] $\mathbf{v_1}, \mathbf{v_2}$ and $\mathbf{v_3}$. In ECM, the center of the cluster $w_1 \cup w_3$ is given by $\mathbf{v_{1,3}} = \frac{\mathbf{v_1}+\mathbf{v_3}}{2}$, and the "ignorance center" is $\mathbf{v}_\Omega = \frac{\mathbf{v_1}+\mathbf{v_2}+\mathbf{v_3}}{3}$. However, if the centers of $\mathbf{v_2}$, $\mathbf{v_{1,3}}$ and $\mathbf{v}_\Omega$ are very close to each other, mathematically represented by $\mathbf{v_2} \approx \frac{\mathbf{v_1}+\mathbf{v_3}}{2}$, then $\mathbf{v_2} \approx \frac{\mathbf{v_1}+\mathbf{v_3}}{2} \approx \frac{\mathbf{v_1}+\mathbf{v_2}+\mathbf{v_3}}{3}$, the classification results about w_2, $w_1 \cup w_3$ and Ω will be difficult to be distinguished. Particularly, the data close to these centers can possibly be associated with the distinct cluster w_2, or with $w_1 \cup w_3$, or with Ω by ECM, and this seems not very reasonable.

In the new Belief C-Means (BCM) algorithm that we propose in this paper, the mass of belief of the specific cluster for each object is computed from the distance between the object and the center of the cluster, and the mass of belief of a meta-cluster is computed both from the distances between object and prototypes of the involved specific clusters, and the distances among these prototypes. In BCM, there is no need to compute the barycenter of the meta-clusters. At the end of this paper, we give some simple examples to show the interest of BCM with respect to FCM and ECM approaches.

2. Basics of Evidential C-Means (ECM)

ECM is a direct extension of FCM and it is based on a general model of partitioning called credal partitioning that refers to the framework of belief functions. The class membership of an object $\mathbf{x}_i = (x_{i_1}, \ldots, x_{i_p})$ is represented by a bba $m_i(.)$ over a given frame of discernment $\Omega = \{w_1, \ldots, w_c\}$, where $|\Omega| = c$ is known. $p \geqslant 1$ is the dimension of the attribute vector $\mathbf{x}_i$ associated with the ith object. 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 prototype vector $\bar{\mathbf{v}}_j$ of the element A_j. Note that A_j can either be a single class, an union of single classes, or the whole frame Ω. 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

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

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} is defined by:

$$d_{ij} = \|\mathbf{x}_i - \bar{\mathbf{v}}_j\|^2, \tag{2}$$

where $\|\mathbf{z}\| = \sqrt{z_1^2 + z_2^2 + \cdots + z_n^2}$ denotes the Euclidean norm of a n-dimensional vector.

In ECM, the determination of $m_{ij} \triangleq = m_{\mathbf{x}_i}(A_j)$ from d_{ij} is done in such a way that m_{ij} is low (resp. high) when d_{ij} is high (resp. low). Actually, m_{ij} is obtained by the minimization of the following objective function under a constraint to obtain the best credal partitioning problem (see Masson and Denœux, 2008 for justifications and details):

$$J_{ECM} = \sum_{i=1}^{n}\sum_{A_j \subseteq \Omega, A_j \neq \emptyset} c_j^{\alpha} m_{ij}^{\beta} d_{ij}^2 + \sum_{i=1}^{n}\delta^2 m_{i\emptyset}^{\beta}. \tag{3}$$

[1] A prototype is a typical attribute vector characterizing a class. Usually the prototype is chosen as the center of the given class under consideration.

Because m_{ij} must be a basic belief assignment, the following constraint must be satisfied for any object $\mathbf{x}_i$

$$\sum_{A_j \subseteq \Omega, A_j \neq \emptyset} m_{ij} + m_{i\emptyset} = 1 \tag{4}$$

The solution of the minimization of (3) under the constraint (4) has been established by Masson and Denœux (2008) and it is given for each object $\mathbf{x}_i$, $(i = 1, 2, \ldots, n)$ by:

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

$$m_{ij} = \frac{c_j^{-\alpha/(\beta-1)} d_{ij}^{-2/(\beta-1)}}{\sum_{A_k \neq \emptyset} c_k^{-\alpha/(\beta-1)} d_{ik}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}}, \tag{5}$$

where α is a tuning parameter allowing to control the degree of penalization; β is a weighting exponent (its suggested default value in (Masson and Denœux, 2008) is $\beta = 2$); δ is a given threshold tuning parameter for the filtering of the outliers; $c_j = |A_j|$ is a weighting coefficient for penalizing the subsets with high cardinality.

- For $A_j = \emptyset$,

$$m_{i\emptyset} \triangleq m_{\mathbf{x}_i}(\emptyset) = 1 - \sum_{A_j \neq \emptyset} m_{ij}. \tag{6}$$

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

$$V_{c\times p} = H_{c\times c}^{-1}.B_{c\times p}, \tag{7}$$

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

$$B_{lq} = \sum_{i=1}^{n} x_{i_q} \sum_{w_l \in A_j} c_j^{\alpha-1} m_{ij}^{\beta}, \tag{8}$$

$$H_{lk} = \sum_{i=1}^{n} \sum_{\{w_k, w_l\} \subseteq A_j} c_j^{\alpha-2} m_{ij}^{\beta}. \tag{9}$$

3. Belief C-Means (BCM) approach

3.1. Basic principle of BCM

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, as shown in (1). ECM method is of course relatively easy to apply, but it yields to serious problems in some difficult cases of data clustering where the prototype vectors of the specific clusters overlap with the meta-clusters. This problem will cause troubles in the association of an object with a particular specific cluster or the meta-cluster the object may also belong to. That is why a better approach must be developed to circumvent this problem. This is the purpose of our BCM algorithm.

In BCM approach, we consider that when a data belongs to a meta-cluster (i.e. to an imprecise class corresponding to the disjunction of several single classes), this means that the prototypes of the single classes in the meta-cluster are quite difficult to be distinguished (discerned) from the object under analysis. More clearly, if the prototype vectors of the classes included in a given meta-cluster are close to each other and in the meantime they are far from the object attribute vector, then it seems more reasonable and natural to commit this object rather to the meta-cluster, than to each of these specific classes as if they were considered separately.

To illustrate this very reasonable BCM principle, let's consider only two objects $\mathbf{x}_1$ and $\mathbf{x}_2$ and three possible centers of clusters (prototypes) $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$ corresponding to the classes w_1, w_2 and w_3 as shown in Fig. 1.