Computational Statistics and Data Analysis 56 (2012) 894–914

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

CECM: Constrained evidential C-means algorithm

V. Antoine [a,*], B. Quost [a], M.-H. Masson [b], T. Denœux [a]

[a] *Université de Technologie de Compiègne, Heudiasyc, UMR CNRS 6599, BP 20529, 60205 Compiègne, France*

[b] *Université de Picardie Jules Verne, IUT de l'Oise, Heudiasyc, UMR CNRS 6599, BP 20529, 60205 Compiègne, France*

ARTICLE INFO

Article history:
Received 15 January 2010
Received in revised form 20 September 2010
Accepted 20 September 2010
Available online 29 September 2010

Keywords:
Clustering
Semi-supervised learning
Pairwise constraints
Adaptive metric
Active learning
Belief functions
Dempster–Shafer theory
Evidence theory

ABSTRACT

In clustering applications, prior knowledge about cluster membership is sometimes available. To integrate such auxiliary information, constraint-based (or semi-supervised) methods have been proposed in the hard or fuzzy clustering frameworks. This approach is extended to evidential clustering, in which the membership of objects to clusters is described by belief functions. A variant of the Evidential C-means (ECM) algorithm taking into account pairwise constraints is proposed. These constraints are translated into the belief function framework and integrated in the cost function. Experiments with synthetic and real data sets demonstrate the interest of the method. In particular, an application to medical image segmentation is presented.

1. Introduction

Clustering methods aim at grouping objects into clusters based on similarity between their descriptors. However, there are some situations in which some background knowledge about the problem is available. Making use of this extra information in a clustering algorithm can help us guide the method towards a desired solution and to improve the classification accuracy (Berget et al., 2008; Gordon, 1996). Prior information can be exploited at different levels such as: the *cluster* level with, for instance, a minimum distance neighborhood (Davidson and Ravi, 2005), the *model* level with the requirement of balanced clusters (Zhong and Ghosh, 2003) or the specification of non desired solutions (Gondek and Hofmann, 2007), or at the *instance* level.

Wagstaff et al. (2001) proposed to introduce two types of instance-level constraints: the first one specifies that two objects have to be in the same cluster (*must-link* constraint) while the second one specifies that two objects should not be put in the same cluster (*cannot-link* constraint). Such pairwise constraints have been considered and integrated in many unsupervised algorithms such as the hard or the fuzzy c-means (FCM), and have recently become a topic of great interest (Xing et al., 2002; Basu et al., 2006; Wagstaff, 2007; Davidson and Ravi, 2005). They have been incorporated in many different ways, generally by including a penalty term in the objective function (Basu et al., 2004; Grira et al., 2008) or by altering the distances between objects with respect to the constraints (Xing et al., 2002).

In the FCM algorithm, each object may belong to one or more clusters with different degrees of membership. These degrees of membership are stored into a fuzzy partition matrix $U = (u_{ik})$ and are calculated by minimizing a suitable

* Corresponding address: Université de Technologie de Compiègne, Laboratoire Heudiasyc, UMR CNRS 6599, BP 20529, 60205 Compiègne, France.
E-mail address: violaine.antoine@hds.utc.fr (V. Antoine).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2010.09.021

objective function with respect to the constraints

$$u_{ik} \geq 0 \quad \forall i, k, \tag{1}$$

and

$$\sum_{k=1}^{c} u_{ik} = 1, \tag{2}$$

where $u_{ik} \in [0, 1]$ denotes the degree of membership of object i to cluster k, and c is the number of clusters. Nevertheless the method sometimes produces counterintuitive results and has poor robustness against noise and outliers. This is the reason why possibilistic methods (Dring et al., 2006; Krishnapuram and Keller, 1993; Davé, 1991) and, more recently, evidential clustering methods grounded in the theory of belief functions (Denœux and Masson, 2004; Masson and Denœux, 2004, 2008, 2009) have been proposed.

Evidential clustering is based on a new concept of partition, referred to as a *credal* partition, which extends the existing concepts of hard, fuzzy and possibilistic partitions. This is done by allocating, for each object, a *mass of belief*, not only to single clusters, but also to any subset of the set of clusters $\Omega = \{\omega_1, \dots, \omega_c\}$. As shown in the experiments reported in Denœux and Masson (2004) and Masson and Denœux (2008), this additional flexibility can be exploited to construct meaningful and robust summaries of the data. For instance, it is possible to compute, for each cluster, a set of objects that *surely* belong to it, and a larger set of objects that *possibly* belong to it. Such qualitative summaries may be argued to be more intuitive and easier to interpret than purely numerical results such as fuzzy partitions, while being much richer than classical hard partitions. Robustness is achieved by assigning outliers to the empty set.

One of the algorithms designed to derive a credal partition from data, called Evidential *C*-Means (ECM), can be considered as a direct extension of FCM (Masson and Denœux, 2008). In this paper, we propose to introduce pairwise constraints in the ECM algorithm, in order to create a new algorithm, called CECM, which combines the advantages of adding background knowledge and using belief functions. Furthermore, we present a formulation of ECM that adapts the metric using a Mahalanobis distance so that the constraints may be more easily satisfied. Finally, we propose an active learning scheme, based on the credal partition, which makes it possible to select efficient pairwise constraints.

The remainder of this paper is organized as follows. Section 2 first recalls the necessary background on belief functions, fuzzy clustering and the ECM algorithm. The basic version of the constrained ECM (CECM) algorithm with Euclidean distance and a more sophisticated version with an adaptive Mahalanobis distance are then introduced in Sections 3 and 4, respectively. Section 5 describes the experimental settings and the results. Finally, we conclude and present some perspectives in Section 6.

2. Background

In this section, the necessary background on the theory of belief functions (Section 2.1), fuzzy clustering (Section 2.2) and the ECM algorithm (Section 2.3) will first be recalled.

2.1. Belief functions

The Dempster–Shafer theory of evidence (Baudrit and Dubois, 2006; Shafer, 1976; Smets and Kennes, 1994) (or the belief function theory) is a theoretical framework for representing partial and unreliable information.

Let us consider a variable ω taking values in a finite set $\Omega = \{\omega_1, \dots, \omega_c\}$ called the frame of discernment. Partial knowledge regarding the actual value taken by ω can be represented by a *mass function* m, which is an application from the power set of Ω in the interval $[0, 1]$ such that

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{3}$$

The subsets A of Ω such that $m(A) > 0$ are called the *focal sets* of m. The value of the focal set $m(A)$ can be interpreted as a fraction of a unit mass of belief that is allocated to A and that cannot be allocated to any subset of A. Complete ignorance is obtained when Ω is the only focal set, and full certainty when the whole mass of belief is assigned to a unique singleton of Ω (m is then said to be a *certain* mass function). If all the focal sets of m are singletons, m is similar to a probability distribution: it is then called a *Bayesian* mass function. A mass function m such that $m(\emptyset) = 0$ is said to be normalized. Under the *open-world* assumption, a mass function $m(\emptyset) > 0$ is interpreted as a quantity of belief given to the hypothesis that the actual value of ω might not belong to Ω (Smets, 1998).

Given a mass function m, it is possible to define a plausibility function $\mathrm{pl} : 2^{\Omega} \to [0, 1]$ and a belief function $\mathrm{bel} : 2^{\Omega} \to [0, 1]$ by:

$$\mathrm{pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad \forall A \subseteq \Omega, \tag{4}$$

and

$$\mathrm{bel}(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B) \quad \forall A \subseteq \Omega. \tag{5}$$