

Beyond Fuzzy, Possibilistic and Rough: An Investigation of Belief Functions in Clustering

Thierry Denœux¹ and Orakanya Kanjanatarakul²

Abstract In evidential clustering, uncertainty about the assignment of objects to clusters is represented by Dempster-Shafer mass functions. The resulting clustering structure, called a credal partition, is shown to be more general than hard, fuzzy, possibility and rough partitions, which are recovered as special cases. Different algorithms to generate a credal partition are reviewed. We also describe different ways in which a credal partition, such as produced by the EVCLUS or ECM algorithms, can be summarized into any of the simpler clustering structures.

1 Introduction

Clustering is one of the most important tasks in data analysis and machine learning. It aims at revealing some structure in a dataset, so as to highlight groups (clusters) of objects that are similar among themselves, and dissimilar to objects of other groups. Traditionally, we distinguish between *partitional* clustering, which aims at finding a partition of the objects, and *hierarchical* clustering, which finds a sequence of nested partitions.

Over the years, the notion of partitional clustering has been extended to several important variants, including fuzzy [3], possibilistic [11], rough [13,17] and evidential clustering [7, 8, 15]. Contrary to classical (hard) partitional clustering, in which each object is assigned unambiguously and with full certainty to a cluster, these variants allow ambiguity, uncertainty or doubt in the assignment of objects to clusters. For this reason, they are referred to as *soft* clustering methods, in contrast with classical, *hard* clustering [18].

Sorbonne Universités, Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, France, email: Thierry.Denoeux@utc.fr · Faculty of Management Sciences, Chiang Mai Rajabhat University, Thailand, email: orakanyaa@gmail.com

Among soft clustering paradigms, evidential clustering describes the uncertainty in the membership of objects to clusters using the formalism of *belief functions* [20]. The theory of belief functions is a very general formal framework for representing and reasoning with uncertainty. Roughly speaking, a belief function can be seen as a collection of sets with corresponding masses, or as a non additive measure generalizing a probability measure. Recently, evidential clustering has been successfully applied in various domains such as machine prognosis [19], medical image processing [12, 14] and analysis of social networks [22].

Because of its generality, the theory of belief functions occupies a central position among theories of uncertainty. The purpose of this paper is to show that, similarly, the evidential paradigm occupies a central position among soft clustering approaches. More specifically, we will show that hard, fuzzy, possibilistic and rough clustering can be all seen as special cases of evidential clustering. We will also study different ways in which a credal partition can be summarized into any of the other hard or soft clustering structures to provide the user with more synthetic views of the data.

The rest of this paper is structured as follows. In Section 2, the notion of credal partition will first be recalled, and algorithms to construct a credal partition will be reviewed. The relationships with other clustering paradigms will then be discussed in Section 3. Finally, Section 4 will conclude the paper.

2 Credal partition

We first recall the notion of credal partition in Section 2.1. In Section 2.2, we briefly review the main algorithms for constructing credal partitions.

2.1 Credal partition

Assume that we have a set $\mathcal{O} = \{o_1, \dots, o_n\}$ of n objects, each one belonging to one and only one of c groups or clusters. Let $\Omega = \{\omega_1, \dots, \omega_c\}$ denote the set of clusters. If we know for sure which cluster each object belongs to, we can give a (hard) partition of the n objects. Such a partition may be represented by binary variables u_{ik} such that $u_{ik} = 1$ if object o_i belongs to cluster ω_k , and $u_{ik} = 0$ otherwise.

If objects cannot be assigned to clusters with certainty, then we can quantify cluster-membership uncertainty by mass functions m_1, \dots, m_n , where each mass function m_i is a mapping from 2^Ω to $[0, 1]$, such that $\sum_{A \subseteq \Omega} m_i(A) = 1$. Each mass $m_i(A)$ is interpreted as a degree of support attached to the proposition “the true cluster of object o_i is in A ”, and to no more specific

proposition. A subset A of Ω such that $m_i(A) > 0$ is called a *focal set* of m_i . The n -tuple m_1, \dots, m_n is called a *credal partition* [8].

Example 1 Consider, for instance, the “Butterfly” dataset shown in Figure 1(a). This dataset is adapted from the classical example by Windham [21], with an added outlier (point 12). Figure 1(b) shows the credal partition with $c = 2$ clusters produced by the Evidential c -means (ECM) algorithm [15]. In this figure, the masses $m_i(\emptyset)$, $m_i(\{\omega_1\})$, $m_i(\{\omega_2\})$ and $m_i(\Omega)$ are plotted as a function of i , for $i = 1, \dots, 12$. We can see that $m_3(\{\omega_1\}) \approx 1$, which means that object o_3 almost certainly belongs to cluster ω_1 . Similarly, $m_9(\{\omega_2\}) \approx 1$, indicating almost certain assignment of object o_9 to cluster ω_2 . In contrast, objects o_6 and o_{12} correspond to two different situations of maximum uncertainty: for object o_6 , we have $m_6(\Omega) \approx 1$, which means that this object might as well belong to clusters ω_1 and ω_2 . The situation is completely different for object o_{12} , for which the largest mass is assigned to the empty set, indicating that this object does not seem to belong to any of the two clusters.

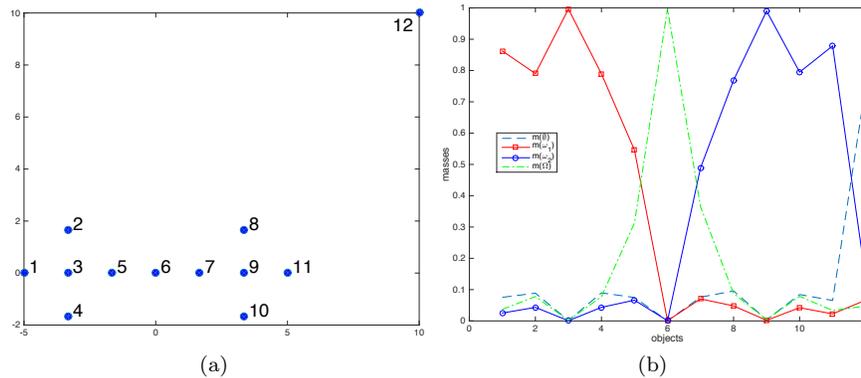


Fig. 1 Butterfly dataset (a) and a credal partition (b).

2.2 Evidential clustering algorithms

Three main algorithms have been proposed to generate credal partitions:

1. The EVCLUS algorithm, introduced in [8], applies ideas from Multidimensional Scaling (MDS) [4] to clustering: given a dissimilarity matrix, it finds a credal partition such that the degrees of conflict between mass functions match the dissimilarities, dissimilar objects being represented by highly conflicting mass functions; this is achieved by iteratively minimizing a stress function. A variant of EVCLUS allowing one to use prior

- knowledge in the form of pairwise constraints was later introduced in [1], and several improvements to the original algorithm making it capable of handling large dissimilarity datasets have been reported in [9].
2. The Evidential c -means (ECM) algorithm [15] is a c -means-like algorithm that minimizes a cost function by searching alternatively the space of prototypes and the space of credal partitions. Unlike the hard and fuzzy c -means algorithms, ECM associates a prototype not only to clusters, but also to sets of clusters. The prototype associated to a set of clusters is defined as the barycenter of the prototypes of each single cluster in the set. The cost function to be minimized insures that objects close to a prototype have a high mass assigned to the corresponding set of clusters. A variant with adaptive metrics and pairwise constraints was introduced in [2], and a relational version for dissimilarity data (called RECM) has been proposed in [16].
 3. The E_k -NNclus algorithm [7] is a decision-directed clustering procedure based on the evidential k -nearest neighbor (EK -NN) rule [6]. Starting from an initial partition, the algorithm iteratively reassigns objects to clusters using the EK -NN rule, until a stable partition is obtained. After convergence, the cluster membership of each object is described by a Dempster-Shafer mass function assigning a mass to each cluster and to the whole set of clusters. The mass assigned to the set of clusters can be used to identify outliers. The procedure can be seen as searching for the most plausible partition of the data.

Each of these three algorithms have their strengths and limitations, and the choice of an algorithm depends on the problem at hand. Both ECM and EK -NN are very efficient for handling attribute data. EK -NN has the additional advantage that it can determine the number of clusters automatically, while EVCLUS and ECM produce more informative outputs (with masses assigned to any subsets of clusters). EVCLUS was shown to be very effective for dealing with non metric dissimilarity data, and the recent improvements reported in [9] make it suitable to handle very large datasets.

3 Relationships with other clustering paradigms

In this section, we discuss the relationships between the notion of credal partition and other clustering structures. In Section 3.1, we show that hard, fuzzy, possibilistic and rough partitions are all special kinds of credal partitions. In Section 3.2, we describe how a general credal partition can be summarized in the form of any of the simpler structures mentioned previously.

3.1 Generality of the notion of credal partition

The notion of credal partition is very general, in the sense that it boils down to several alternative clustering structures when the mass functions composing the credal partition have some special forms (see Figure 2).

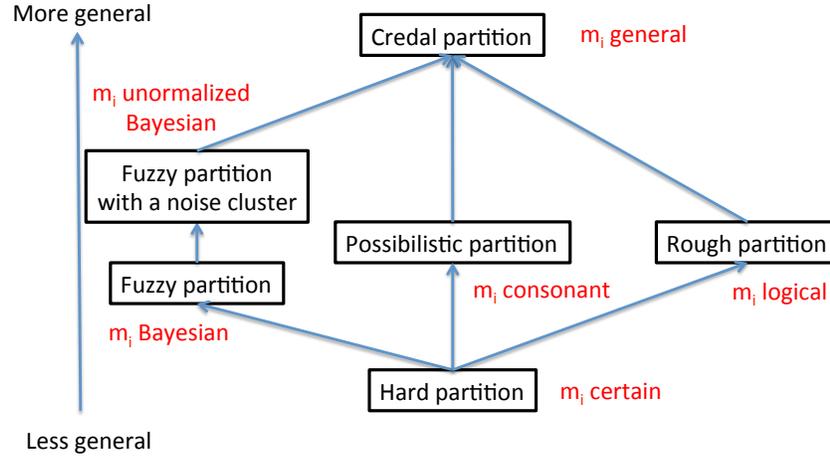


Fig. 2 Relationship between credal partitions and other clustering structures.

Hard partition: If all mass functions m_i are *certain* (i.e., have a single focal set, which is a singleton), then we have a hard partition, with $u_{ik} = 1$ if $m_i(\{\omega_k\}) = 1$, and $u_{ik} = 0$ otherwise.

Fuzzy partition: If the m_i are *Bayesian* (i.e., they assign masses only to singletons, in which case the corresponding belief function becomes additive), then the credal partition is equivalent to a fuzzy partition; the degree of membership of object i to cluster k is $u_{ik} = m_i(\{\omega_k\})$.

Fuzzy partition with a noise cluster: A mass function m such that each focal set is either a singleton, or the empty set may be called an *unnormalized Bayesian mass function*. If each mass function m_i is unnormalized Bayesian, then we can define, as before, the membership degree of object i to cluster k a $u_{ik} = m_i(\{\omega_k\})$, but we now have $\sum_{k=1}^c u_{ik} \leq 1$, for $i = 1, \dots, n$. We then have $m_i(\emptyset) = u_{i*} = 1 - \sum_{k=1}^c u_{ik}$, which can be interpreted as the degree of membership to a “noise cluster” [5].

Possibilistic partition: If the mass functions m_i are *consonant* (i.e., if their focal sets are nested), then they are uniquely described by their contour functions

$$pl_i(\omega_k) = \sum_{A \subseteq \Omega, \omega_k \in A} m_i(A), \quad (1)$$

which are possibility distributions. We then have a possibilistic partition, with $u_{ik} = pl_i(\omega_k)$ for all i and k . We note that $\max_k pl_i(\omega_k) = 1 - m_i(\emptyset)$.
 Rough partition: Assume that each m_i is *logical*, i.e., we have $m_i(A_i) = 1$ for some $A_i \subseteq \Omega$, $A_i \neq \emptyset$. We can then define the *lower approximation* of cluster ω_k as the set of objects that *surely* belong to ω_k ,

$$\omega_k^L = \{o_i \in \mathcal{O} | A_i = \{\omega_k\}\}, \quad (2)$$

and the *upper approximation* of cluster ω_k as the set of objects that *possibility* belong to ω_k ,

$$\omega_k^U = \{o_i \in \mathcal{O} | \omega_k \in A_i\}. \quad (3)$$

The membership values to the lower and upper approximations of cluster ω_k are then, respectively, $\underline{u}_{ik} = Bel_i(\{\omega_k\})$ and $\bar{u}_{ik} = Pl_i(\{\omega_k\})$. If we allow $A_i = \emptyset$ for some i , then we have $\bar{u}_{ik} = 0$ for all k , which means that object o_i does not belong to the upper approximation of any cluster.

3.2 Summarization of a credal partition

A credal partition is a quite complex clustering structure, which often needs to be summarized in some way to become interpretable by the user. This can be achieved by transforming each of the mass functions in the credal partition into a simpler representation. Depending on the representation used, each of clustering structures mentioned in Section 3.1 can be recovered as different partial views of a credal partition. Some of the relevant transformations are discussed below.

Fuzzy and hard partitions: A fuzzy partition can be obtained by transforming each mass function m_i into a probability distribution p_i using the plausibility-probability transformation defined as

$$p_i(\omega_k) = \frac{pl_i(\omega_k)}{\sum_{\ell=1}^c pl_i(\omega_\ell)}, \quad k = 1, \dots, c, \quad (4)$$

where pl_i is the contour function associated to m_i , given by (1). By selecting, for each object, the cluster with maximum probability, we then get a hard partition.

Fuzzy partition with noise cluster: In the plausibility-probability transformation (4), the information contained in the masses $m_i(\emptyset)$ assigned to the empty set is lost. However, this information may be important if the dataset contains outliers. To keep track of it, we can define an unnormalized plausibility transformation as $\pi_i(\omega_k) = (1 - m_i(\emptyset))p_i(\omega_k)$, for $k = 1, \dots, c$. The degree of membership of each object i to cluster k can

then be defined as $u_{ik} = \pi_i(\omega_k)$ and the degree of membership to the noise cluster as $u_{i*} = m_i(\emptyset)$.

Possibilistic partition: A possibilistic partition can be obtained from a credal partition by computing a consonant approximation of each of the mass functions m_i [10]. The simplest approach is to approximate m_i by the consonant mass function with the same contour function, in which case the degree of possibility of object o_i belonging to cluster ω_k is $u_{ik} = pl_i(\omega_k)$.

Rough partition: As explained in Section 3.1, a credal partition becomes equivalent to a rough partition when all mass functions m_i are logical. A general credal partition can thus be transformed into a rough partition by deriving a set A_i of clusters from each mass function m_i . This can be done either by selecting the focal set $A_{max} = \arg \max_{A \subseteq \Omega} m(A)$ with maximum mass as suggested in [15], or by the interval dominance decision rule

$$A^*(m_i) = \{\omega \in \Omega \mid \forall \omega' \in \Omega, pl_i^*(\omega) \geq m_i^*(\{\omega'\})\}, \quad (5)$$

where pl_i^* and m_i^* are defined, respectively, by $pl_i^* = pl_i / (1 - m_i(\emptyset))$ and $m_i^* = m_i / (1 - m_i(\emptyset))$. If the interval dominance rule is used, we may account for the mass assigned to the empty set by defining A_i as follows,

$$A_i = \begin{cases} \emptyset & \text{if } m_i(\emptyset) = \max_{A \subseteq \Omega} m_i(A) \\ A^*(m_i) & \text{otherwise.} \end{cases} \quad (6)$$

4 Conclusions

The notion of credal partition, as well as its relationships with alternative clustering paradigms have been reviewed. Basically, each of the alternative partitional clustering structures (i.e., hard, fuzzy, possibilistic and rough partitions) correspond to a special form of the mass functions within a credal partition. We have also examined different ways in which a credal partition can be transformed into a simpler clustering structure for easier interpretation. As they build more complex clustering structures, credal clustering algorithms such as EVCLUS and ECM tend to be more computationally demanding than alternative algorithms. This issue can be dealt with by using efficient optimization algorithms and by restricting the form of the credal partition, making it possible to apply evidential clustering to large datasets with large numbers of clusters. First results along these lines have been reported in [9].

Acknowledgements This research was supported by the Labex MS2T, which was funded by the French Government, through the program “Investments for the future” by the National Agency for Research (reference ANR-11-IDEX-0004-02).

References

1. V. Antoine, B. Quost, M.-H. Masson, and T. Denceux. CEVCLUS: evidential clustering with instance-level constraints for relational data. *Soft Computing*, 18(7):1321–1335, 2014.
2. V. Antoine, B. Quost, M.-H. Masson, and T. Denceux. CECM: Constrained evidential c-means algorithm. *Computational Statistics & Data Analysis*, 56(4):894–914, 2012.
3. J.C Bezdek. *Pattern Recognition with fuzzy objective function algorithm*. Plenum Press, New-York, 1981.
4. I. Borg and P. Groenen. *Modern multidimensional scaling*. Springer, New-York, 1997.
5. R.N. Davé. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12:657–664, 1991.
6. T. Denceux. A k -nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. on Systems, Man and Cybernetics*, 25(05):804–813, 1995.
7. T. Denceux, O. Kanjanatarakul, and S. Sriboonchitta. EK-NNclus: a clustering procedure based on the evidential k -nearest neighbor rule. *Knowledge-based Systems*, 88:57–69, 2015.
8. T. Denceux and M.-H. Masson. EVCLUS: Evidential clustering of proximity data. *IEEE Trans. on Systems, Man and Cybernetics B*, 34(1):95–109, 2004.
9. T. Denceux, S. Sriboonchitta, and O. Kanjanatarakul. Evidential clustering of large dissimilarity data. *Knowledge-Based Systems*, vol. 106, pages 179-195, 2016.
10. D. Dubois and H. Prade. Consonant approximations of belief measures. *International Journal of Approximate Reasoning*, 4:419–449, 1990.
11. R. Krishnapuram and J.M. Keller. A possibilistic approach to clustering. *IEEE Trans. on Fuzzy Systems*, 1:98–111, May 1993.
12. Benoît Lelandais, Su Ruan, Thierry Denceux, Pierre Vera, and Isabelle Gardin. Fusion of multi-tracer PET images for dose painting. *Medical Image Analysis*, 18(7):1247–1259, 2014.
13. Pawan Lingras and Georg Peters. Applying rough set concepts to clustering. In G. Peters, P. Lingras, D. Ślezak, and Y. Yao, editors, *Rough Sets: Selected Methods and Applications in Management and Engineering*, pages 23–37. Springer-Verlag, London, UK, 2012.
14. Nasr Makni, Nacim Betrouni, and Olivier Colot. Introducing spatial neighbourhood in evidential c-means for segmentation of multi-source images: Application to prostate multi-parametric MRI. *Information Fusion*, 19:61–72, 2014.
15. M.-H. Masson and T. Denceux. ECM: an evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41(4):1384–1397, 2008.
16. M.-H. Masson and T. Denceux. RECM: relational evidential c-means algorithm. *Pattern Recognition Letters*, 30:1015–1026, 2009.
17. Georg Peters. Is there any need for rough clustering? *Pattern Recognition Letters*, 53:31–37, 2015.
18. Georg Peters, Fernando Crespo, Pawan Lingras, and Richard Weber. Soft clustering: fuzzy and rough approaches and their extensions and derivatives. *International Journal of Approximate Reasoning*, 54(2):307–322, 2013.
19. Lisa Serir, Emmanuel Ramasso, and Nouredine Zerhouni. Evidential evolving Gustafson-Kessel algorithm for online data streams partitioning using belief function theory. *International Journal of Approximate Reasoning*, 53(5):747–768, 2012.
20. G. Shafer. *A mathematical theory of evidence*. Princeton University Press, Princeton, N.J., 1976.
21. M. P. Windham. Numerical classification of proximity data with assignment measures. *Journal of classification*, 2:157–172, 1985.
22. Kuang Zhou, Arnaud Martin, Quan Pan, and Zhun-Ga Liu. Median evidential c-means algorithm and its application to community detection. *Knowledge-Based Systems*, 74(0):69–88, 2015.