Credal clustering: relationship with rough and other clustering paradigms

Thierry Denœux and Orakanya Kanjanatarakul

Over the years, the notion of partitional clustering has been extended to several important variants, including fuzzy, possibilistic, rough [2, 4] and credal clustering [1, 3]. Contrary to classical (hard) partitional clustering, in which each object is assigned unambiguously and with full certainty to one 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 [5].

Among soft clustering paradigms, credal clustering represents the uncertainty about the membership of objects to clusters using the formalism of belief functions. In a credal partition, the membership of each object to clusters is described by a mass function, i.e., a function that assigns a mass between 0 and 1 to each set of clusters, with the constraint that the masses sum to 1. EVCLUS [1] and the Evidential c-means (ECM) [3] are two algorithms for generating a credal partition from dissimilarity or attribute data.

In this paper, we review recent work on credal clustering, and we analyze its relationships with alternative paradigms. Specifically we show that each of the alternative partitional clustering structures (i.e., hard, fuzzy, possibilistic and rough partitions) corresponds to a special form of the mass functions within a credal partition (see Figure 1). For example, a fuzzy partition is recovered when all mass functions are Bayesian (i.e., when they assign masses to singletons only), and a rough clustering structure is recovered when all mass functions are logical (i.e., when they assign a mass to only one set of clusters).

We also examine different ways in which a general credal partition can be summarized into a simpler clustering structure for easier interpretation. To this end, each mass function has to be transformed into a singleton, a probability distribution,

Thierry Denœux and Orakanya Kanjanatarakul
Sorbonne Universités, Université de Technologie de Compiègne, CNRS, France

Orakanya Kanjanatarakul
Department of Economics, Faculty of Management Sciences, Chiang Mai Rajabhat University, Thailand

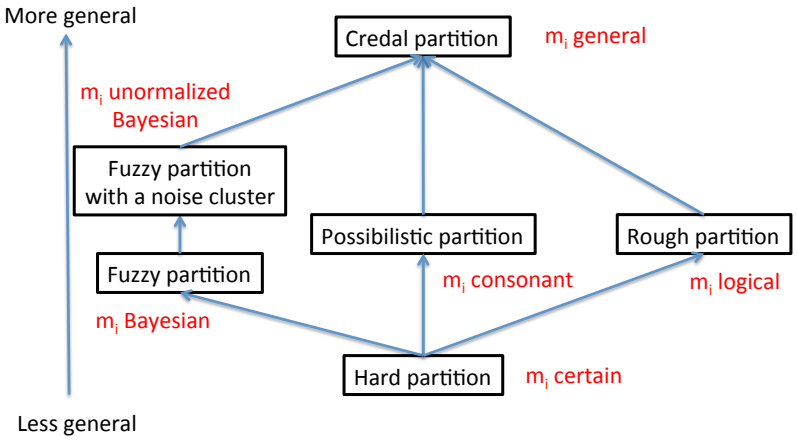

Fig. 1 Relationship between credal partitions and other clustering structures.

a possibility distribution, or a set, yielding, respectively, a hard, fuzzy, possibilistic, or rough partition.

As they build more complex clustering structures, credal clustering algorithms tend to be more computationally demanding than other algorithms. For large datasets and/or large numbers of clusters, this issue can be dealt with by restricting the form of the credal partition and by using efficient optimization algorithms.

Acknowledgement

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. T. Denœux and M.-H. Masson. EVCLUS: Evidential clustering of proximity data. *IEEE Trans. on Systems, Man and Cybernetics B*, 34(1):95–109, 2004.
2. P. Lingras and G. 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.
3. M.-H. Masson and T. Denœux. ECM: an evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41(4):1384–1397, 2008.
4. G. Peters. Is there any need for rough clustering? *Pattern Recognition Letters*, 53:31–37, 2015.
5. G. Peters, F. Crespo, P. Lingras, and R. Weber. Soft clustering fuzzy and rough approaches and their extensions and derivatives. *International Journal of Approximate Reasoning*, 54(2):307 – 322, 2013.