Evaluating and Comparing Soft Partitions: an Approach Based on Dempster-Shafer Theory

Thierry Denœux, Shoumei Li and Songsak Sriboonchitta

***Abstract*—In evidential clustering, cluster-membership uncertainty is represented by Dempster-Shafer mass functions. The notion of evidential partition generalizes other soft clustering structures such as fuzzy, possibilistic or rough partitions. In this paper, we propose two extensions of the Rand index for evaluating and comparing evidential partitions, called similarity and consistency indices. The similarity index is suitable for measuring the closeness of two soft partitions, while the consistency index allows one to assess the agreement, or lack of conflict, between a soft partition and the true hard partition. Simulation experiments illustrate some applications of these indices.**

***Index Terms*—Clustering, evidence theory, cluster validity measure, belief functions.**

I. Introduction

Clustering plays an important role in data mining and machine learning. Basically, it consists in finding groups in data. One of the main issues in clustering, as in other tasks such as classification or regression, is the quantification of uncertainty. After some groups have been discovered, how sure are we about the cluster-membership of objects? For instance, in Figure 1, object o_i might as well belong to any of the three clusters, while object o_j does not seem to belong to any of the clusters: these are two different situations of high uncertainty. In the former case, there is *ambiguity* because object o_i is close to several clusters. In the latter, there is a *lack of evidence* to assign object o_i to any of the clusters, because it is far from all of them.

Over the years, a growing number of clustering concepts and algorithms have been proposed. Early methods, such as the k-means algorithm, unambiguously assign each object to a single cluster, resulting in a *hard partition*; this approach amounts to neglecting cluster-membership uncertainty. Clustering approaches that explicitly describe cluster-membership uncertainty are collectively referred to as *soft clustering* [1]. In *probabilistic* [2] and *fuzzy* [3] clustering, information about the membership of object o_i to clusters is represented by c numbers $(u_{i1}, \ldots, u_{ic}) \in [0,1]^c$, where c is the total number of clusters, with the following constraint:

$$\sum_{k=1}^{c} u_{ik} = 1. \quad (1)$$

The interpretation of the u_{ik}'s (as probabilities or as degrees of membership) is different under the probabilistic and fuzzy frameworks, but the mathematical models are formally equivalent. In *possibilistic* clustering [4], [5], condition (1) is relaxed; each number u_{ik} can then be interpreted as a degree of possibility that object i belongs to cluster k. Yet another formalism is *rough clustering* [6], [7], in which each cluster is represented by two nested sets of objects: a lower approximation $\underline{\omega}_k$, defined as the set of objects that *surely* belong to cluster k, and an upper approximation $\overline{\omega}_k$, defined as the set of objects that *possibly* belong to that cluster. This is clearly a special case of possibilistic clustering, where the u_{ik}'s are constrained to be either 0 or 1; then,

$$\overline{\omega}_k = \{o_i \in \mathcal{O}, u_{ik} = 1\} \quad (2)$$

and

$$\underline{\omega}_k = \{o_i \in \mathcal{O}, u_{ik} = 1 \text{ and } u_{i\ell} = 0, \forall \ell \neq k\}. \quad (3)$$

The most general framework to date is *evidential clustering* [8]–[10], which represents cluster-membership information using Dempster-Shafer mass functions [11], forming an evidential partition (see Section II-A below). Evidential clustering encompasses probabilistic, fuzzy, possibilistic and rough clustering as special cases. Recently, it has been successfully applied in various domains such as machine prognosis [12], medical image processing [13], [14] and analysis of social networks [15]. Evidential clustering algorithms include the evidential c-means (ECM) [9] and EVCLUS [8], [10] algorithms.

When studying and evaluating clustering algorithms, we need objective methods for comparing different clustering structures. As noted by Rand [16], such a comparison may try to achieve different goals:

1) We may wish to compare a given hard or soft (fuzzy, possibilistic, rough,...) partition to ground truth (usually, a hard partition); this problem is often referred to as *external evaluation* in the clustering literature;

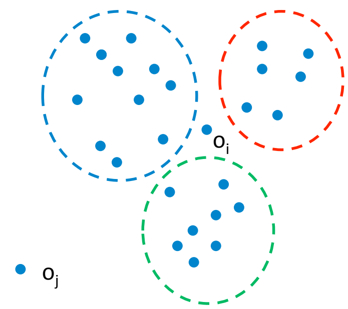

Fig. 1. A dataset with three clusters. Objects o_i and o_j cannot be assigned to a cluster with certainty.

T. Denœux is with Sorbonne Universités, Université de Technologie de Compiègne, CNRS, Heudiasyc (UMR 7253), France.

Shoumei Li is with Beijing University of Technology, College of Applied Sciences, Beijing, China.

Songsak Sriboonchitta is with Chiang Mai University, Faculty of Economics, Chiang Mai, Thailand.

2) We may be interested in measuring the sensitivity of the clustering structure to small changes of the data: in this case, we need to compare the clustering structures obtained before and after perturbing the data by some random noise;
3) A variant of the previous situation occurs when we want to evaluate the sensitivity of the clustering to missing data. In that case, we need to compare the clustering of n objects with the cluster assignment of the same objects after n' new objects have been added, and the clustering is performed on the $n + n'$ objects.
4) Finally, we often need to determine to what extent two different algorithms agree or disagree when applied to the same data.

Yet another context in which we need to compare partitions is when we wish to compare different representations of the same objects [17]: for instance, we may have initially p attributes, and search for a subset of $q < p$ attributes that allows us to obtain a similar clustering structure.

As noted by Anderson et al. [18], the number of comparison indices proposed so far is so large that it is impossible to give an exhaustive account. Even a non-exhaustive, but extensive survey would require a full paper. Here, we will focus on the Rand index [16], which is probably the most widely used comparison index, and which has recently gained revived interest in the bioinformatics community [19], [20], among others. The Rand index measures the similarity between two hard partitions: it is defined as the ratio of the number of objects pairs on which the two partitions agree, to the total number of object pairs. It is thus very simple and easy to interpret, which explains its popularity. In recent years, fuzzy extensions of the Rand index have been proposed (see, e.g., [17], [18], [21], [22]). As discussed in [17], most of these extensions are more or less arbitrary (as there are several ways to fuzzify a formula) and are not suitable for all the applications mentioned above. The index proposed in [17] has some advantages over previous proposals (in particular, it exhibits desirable metric properties), but it is essentially limited to the comparison of fuzzy partitions. In this paper, we propose new indices, grounded in Dempster-Shafer theory, making it possible to comparing evidential partitions. As probabilistic, fuzzy, possibilistic and rough partitions are all special evidential partitions, our approach can be used to evaluate and compare the results of any soft clustering methods.

The rest of this paper is organized as follows. Section II reviews previous work on evidential clustering and fuzzy extensions of the Rand index. Our new indices will then be introduced in Section III. Finally, numerical experiments will be presented in Section IV, and Section V will conclude the paper.

II. Background

This section introduces the main background material need in the rest of the paper. In Section II-A, we first recall the notion of evidential partition, and we show that it encompasses all other soft partition types as special cases. In Section II-B, we then review the Rand index and some recent proposals to extend it to fuzzy partitions, and we show that these extended Rand indices are not suitable for comparing arbitrary evidential partitions.

Throughout this paper, it is assumed that the reader has some familiarity with Dempster-Shafer theory. To make the paper self-contained, a reminder of the main concepts of this theory is included as an appendix.

A. Evidential clustering

Let $\mathcal{O} = \{o_1, \ldots, o_n\}$ be a set of n objets. We assume that each object belongs to at most one of c clusters. The set of clusters is denoted by $\Omega = \{\omega_1, \ldots, \omega_c\}$. In evidential clustering, the uncertainty about the cluster membership of each object o_i is represented by a mass function m_i in Ω. The n-tuple $M = (m_1, \ldots, m_n)$ is called an *evidential* (or *credal*) *partition*.

Example 1: Consider, for instance, the "Butterfly" dataset shown in Figure 2(a), composed of $n = 12$ objects. Figure 2(b) shows the credal partition with $c = 2$ clusters produced by the Evidential c-means (ECM) algorithm [9]. 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 $i = 1, \ldots, 12$. We can see that $m_3(\{\omega_1\}) \approx 0.8$, which means that object o_3 is strongly believed to belong to cluster ω_1. Similarly, $m_9(\{\omega_2\}) \approx 0.8$, indicating strong evidence of object o_9 belonging to cluster ω_2. In contrast, objects o_6 and o_{12} correspond to two different situations of maximum uncertainty. Object o_6 has a large mass assigned to Ω: this reflects ambiguity in the class membership of this object, which means that it might belong to ω_1 as well as to 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. □

Because of the generality of the notion of mass function, the concept of evidential partition encompasses all other notions of hard and soft partition:

- If all mass functions m_i are certain, then we have a hard partition, with $u_{ik} = 1$ if $m_i(\{\omega_k\}) = 1$, and $u_{ik} = 0$ otherwise.
- If all mass functions m_i are Bayesian, then the evidential partition is equivalent to a fuzzy partition, with $u_{ik} = m_i(\{\omega_k\})$, for $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, c\}$.
- If all mass functions m_i are logical with a single focal set $A_i \subseteq \Omega$, then we get a rough partition. The lower and upper approximations of cluster k can be defined, respectively, as

$$\underline{\omega}_k = \{o_i \in \mathcal{O} | A_i = \{\omega_k\}\}, \tag{4a}$$

and

$$\overline{\omega}_k = \{o_i \in \mathcal{O} | \omega_k \in A_i\}. \tag{4b}$$

- If each m_i is consonant, then it is equivalent to a possibility distribution, and it can be uniquely represented by the plausibility of the singletons $pl_{ik} = Pl_i(\{\omega_k\})$ for $i \in \{1, \ldots, n\}$ and $k \in \{1, \ldots, c\}$. Each number pl_{ik} is the plausibility that object i belongs to cluster k; these numbers form a possibilistic partition of the n objects.

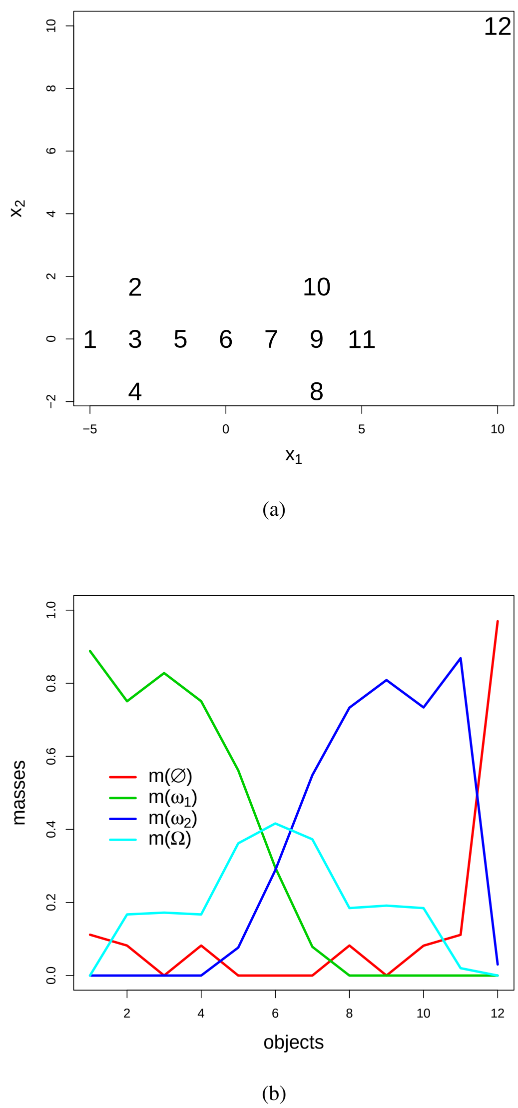

Fig. 2. Butterfly dataset (a) and a credal partition (b).

Several algorithms have been proposed for constructing evidential partitions of attribute or proximity data. The Evidential c-means (ECM) algorithm [9] is a variant of the hard and fuzzy c-means algorithms, in which not only clusters, but also sets of clusters are represented by prototypes. A relational version of ECM, the Relation Evidential c-Means (RECM) algorithm, can handle proximity data [23]. This is also the case for another procedure, EVCLUS, which constructs an evidential partition is such a way that the degrees of conflict between any two mass functions m_i and m_j match the dissimilarity d_{ij} between objects o_i and o_j [8]. Whereas the initial version of EVCLUS was limited to small datasets and small numbers of clusters, a recent version, k-EVCLUS, overcomes these limitations and can handle large proximity datasets [10]. A third method, EK-NNclus [24], is an unsupervised version of the evidential K nearest neighbor rule [25]. All these algorithms are available in the R package `evclust` [26].

An evidential 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 m_i in the evidential partition into a simpler representation. For instance, a possibilistic partition can be obtained by equating the membership degrees to the plausibilities: $u_{ik} = pl_i(\omega_k)$ for all i and all k. After normalization, we get a fuzzy partition, and a hard partition is obtained by selecting, for each object, the cluster with maximum plausibility. To get a rough partition, we have to select a *set* of possible clusters for each object. This can be achieved either by selecting the focal set with the highest mass [9], or by using a decision rule such as the interval dominance rule [27].

B. Rand index and fuzzy extensions

In this section, we represent a c-partition of n objects by a binary matrix $\boldsymbol{U} = (u_{ik}) \in \{0,1\}^{n\times c}$ such that $u_{ik} = 1$ if object i is assigned to cluster k, and $u_{ik} = 0$ otherwise. Similarly, a fuzzy or possibilistic partition can be represented by a matrix $\widetilde{\boldsymbol{U}} = (u_{ik}) \in [0,1]^{n\times c}$. Without any risk of confusion, we will identify a partition (crisp or fuzzy) with its matrix representation $\boldsymbol{U}$.

The Rand index [16] is a measure of similarity between hard partitions. Given two such partitions $\boldsymbol{U}$ and $\boldsymbol{U}'$ (with possibly different numbers of clusters), let us denote by a the number of object pairs that are clustered together in both $\boldsymbol{U}$ and $\boldsymbol{U}'$, and let b denote the number of object pairs that are in different clusters in both $\boldsymbol{U}$ and $\boldsymbol{U}'$. Then, the Rand index is defined by

$$\rho(\boldsymbol{U},\boldsymbol{U}') = \frac{a+b}{n(n-1)/2}. \tag{5}$$

As noted by Rand [16], ρ is a measure of similarity, with $\rho(\boldsymbol{U},\boldsymbol{U}')=1$ if $\boldsymbol{U}=\boldsymbol{U}'$, and $\rho(\boldsymbol{U},\boldsymbol{U}')=0$ when $\boldsymbol{U}$ and $\boldsymbol{U}'$ have no similarities, i.e., when one partition consists of a single cluster, and the other one consists of n clusters containing single points. Also, $1-\rho$ is a metric in the space of partitions of the n objects.

As mentioned in the introduction, several authors [18], [21], [22], [28] have proposed extensions of index ρ to measure the similarity between fuzzy partitions. Most of these proposals consist in starting with some formal expression for ρ, and fuzzifying it. For instance, Anderson et al. [18] observe that a and b in (5) can be written as

$$a = \frac{1}{2}\sum_{k,\ell} n_{k\ell}(n_{k\ell}-1) \tag{6a}$$

and

$$b = \frac{1}{2}\left(n^2 + \sum_{k,\ell} n_{k\ell}^2 - \sum_k n_{k\cdot}^2 - \sum_\ell n_{\cdot\ell}^2\right), \tag{6b}$$

where $n_{k\ell}$ is the number of objects i such that $u_{ik} = u'_{i\ell} = 1$, $n_{k\cdot} = \sum_\ell n_{ij}$ and $n_{\cdot\ell} = \sum_k n_{k\ell}$. Now, the confusion matrix $\boldsymbol{N} = (n_{k\ell})$ can be obtained as

$$\boldsymbol{N} = \boldsymbol{U}\boldsymbol{U}'^T, \tag{7}$$

Equations (7) can still be used together with (6a)-(6b) when matrices $\boldsymbol{U}$ and $\boldsymbol{U}'$ are replaced by $\widetilde{\boldsymbol{U}}$ and $\widetilde{\boldsymbol{U}}'$ with elements in $[0,1]$ and represent fuzzy or possibilistic partitions, which allows one to obtained a fuzzy version ρ_F of the Rand index, or of any index based on the confusion matrix $\boldsymbol{N} = (n_{k\ell})$. One advantage of this approach is that ρ_F can be computed in time proportional to the number n of objects, in contrast with previous methods whose complexity was $O(n^2)$. However, as noticed by Hüllermeier [17], this method for fuzzifying the Rand index also has shortcomings. It is partially arbitrary, because a and b can be expressed in different ways as a function of the $n_{k\ell}$, resulting in different values in the fuzzy case. Also, the semantics of the index is not clear in the fuzzy case; for instance, the quantity $n_{k\ell}(n_{k\ell}-1)$ can be negative when $0 < n_{k\ell} < 1$, and so can a. Finally, and maybe more importantly, $1-\rho_F$ is no longer a metric in the space of fuzzy partitions; in particular, we have in general $\rho_F(\widetilde{\boldsymbol{U}},\widetilde{\boldsymbol{U}}) < 1$, which means that a fuzzy partition is not fully similar to itself; this issues makes index ρ_F unsuitable to some applications, such as comparing the fuzzy clusterings produced by two algorithms.

A different approach, proposed by Hüllermeier et al. [17], is to consider the Rand index as a measure of similarity between equivalence relations. Let $\boldsymbol{R} = (r_{ij})$ be the $n \times n$ binary matrix representing the equivalence relation associated to $\boldsymbol{U}$, defined by $r_{ij} = 1$ if objects i and j and clustered together in $\boldsymbol{U}$, and $r_{ij} = 0$ otherwise. In the sequel, $\boldsymbol{R}$ will be called the *relational representation* of $\boldsymbol{U}$. Similarly, let $\boldsymbol{R}' = (r'_{ij})$ be the relational representation of $\boldsymbol{U}'$. Then, the Rand index can be written as

$$\rho(\boldsymbol{U},\boldsymbol{U}') = 1 - \frac{\sum_{i<j} |r_{ij} - r'_{ij}|}{n(n-1)/2}. \tag{8}$$

When $\widetilde{\boldsymbol{U}} \in [0,1]^{n\times c}$ represents a fuzzy partition, we can still define a relation representation $\widetilde{\boldsymbol{R}} = (\widetilde{r}_{ij}) \in [0,1]^{n\times n}$ as follows. Let $\boldsymbol{u}_i = (u_{i1},\ldots,u_{ic})$ be the vector of membership degrees for object i. Then,

$$\widetilde{r}_{ij} = 1 - \delta(\boldsymbol{u}_i,\boldsymbol{u}_j), \tag{9}$$

where δ is a metric in $[0,1]^c$ taking values in $[0,1]$. Equation (8) and (9) together define a comparison index ρ_H between fuzzy partitions,

$$\rho_F(\widetilde{\boldsymbol{U}},\widetilde{\boldsymbol{U}}') = 1 - \frac{\sum_{i<j} |\widetilde{r}_{ij} - \widetilde{r}'_{ij}|}{n(n-1)/2}, \tag{10}$$

which boils down to the Rand index when the partitions are hard. As shown in [17], $d_H = 1-\rho_H$ is a pseudometric, i.e., it verifies reflexivity ($\forall \widetilde{\boldsymbol{U}}, d_H(\widetilde{\boldsymbol{U}},\widetilde{\boldsymbol{U}}) = 0$), symmetry ($\forall \widetilde{\boldsymbol{U}},\widetilde{\boldsymbol{U}}', d_H(\widetilde{\boldsymbol{U}},\widetilde{\boldsymbol{U}}') = d_H(\widetilde{\boldsymbol{U}}',\widetilde{\boldsymbol{U}})$) and the triangular inequality ($\forall \widetilde{\boldsymbol{U}},\widetilde{\boldsymbol{U}}',\widetilde{\boldsymbol{U}}'', d_H(\widetilde{\boldsymbol{U}},\widetilde{\boldsymbol{U}}'') \leq d_H(\widetilde{\boldsymbol{U}},\widetilde{\boldsymbol{U}}') + d_H(\widetilde{\boldsymbol{U}}',\widetilde{\boldsymbol{U}}'')$). It is not a metric in general, because it does not satisfy separability: we may have $d_H(\widetilde{\boldsymbol{U}},\widetilde{\boldsymbol{U}}') = 0$ for $\widetilde{\boldsymbol{U}} \neq \widetilde{\boldsymbol{U}}'$. The reason for this property is that have some information is lost when transforming $\widetilde{\boldsymbol{U}}$ into its relational representation $\widetilde{\boldsymbol{R}}$, and two different fuzzy partitions may share the same relational representation (except under some restricting conditions analyzed in [17]).

Index ρ_H has several advantages over ρ_F: its meaning is clearer, as it is simply a measure of similarity between relational representations. Even more importantly, its pseudometric properties make it more suitable for comparing fuzzy partitions produced by different clustering algorithms, or for sensitivity analysis. However, it obviously cannot be applied to general evidential partitions, and even its application to possibilitic partitions seems problematic. Consider, for instance, two objects i and j, and their possibility degree vectors computed by two clustering algorithms:

- Algorithm 1: $\boldsymbol{u}_i = (1,0,0)$, $\boldsymbol{u}_j = (0,1,0)$;
- Algorithm 2: $\boldsymbol{u}'_i = (1,0,0)$, $\boldsymbol{u}'_j = (1,1,1)$.

Defining δ as $\delta(\boldsymbol{u}_i,\boldsymbol{u}_j) = \frac{1}{c}\sum_{k=1}^{c} \|u_{ik} - u_{jk}\|$ to have a metric taking values in $[0,1]$, we get $\delta(\boldsymbol{u}_i,\boldsymbol{u}_j) = \delta(\boldsymbol{u}'_i,\boldsymbol{u}'_j) = 2/3$ and, consequently, $r_{ij} = r'_{ij} = 1/3$. We thus consider that the two algorithms agree on this object pair. However, Algorithm 1 assigns the two objects to different clusters, whereas Algorithm 2 is undecided about object j. These two situations are arguably very different, yet this difference is not reflected in the relational representations.

III. Evidential extensions of the Rand index

In this section, we introduce two new extensions of the Rand index for comparing evidential partitions. As in [17], our extensions will be based on the comparison of relational representations. To this aim, we first need to define the relational representation of an evidential partition, which will be the subject of Section III-A. The new definitions will then be introduced in Sections III-B and III-C.

A. Relational representation of an evidential partition

Let $\Omega = \{\omega_1,\ldots,\omega_c\}$ be the set of clusters, and let m_i and m_j be two mass functions quantifying our beliefs about the cluster-membership of objects i and j. We wish to express our beliefs on the frame $\Theta = \{s, \neg s\}$, where s and $\neg s$ mean, respectively, "the two objects belong to the same cluster", and "the two objects belong to different clusters". The set Θ is a coarsening of Ω^2, characterized by the following refining,

$$f(s) = S \tag{11a}$$

$$f(\neg s) = \overline{S}, \tag{11b}$$

where $S = \{(\omega_k,\omega_k), k = 1\ldots,c\}$ and $\overline{S}$ denotes the complement of S (see Figure 3). To express our belief on Θ, we need to compute the extensions of m_i and m_j in Ω^2, combine them using Dempster's rule (31), and then compute the restriction of the combined mass function on Θ. The result is expressed by the following proposition.

Proposition 1: Let m_i and m_j be two mass functions on Ω, and let $\Theta = \{s,\neg s\}$ be the coarsening of Ω^2 defined by (11). The mass function on Θ obtained by combined mass function m_i and m_j by Dempster's rule is

$$m_{ij}(\emptyset) = m_i(\emptyset) + m_j(\emptyset) - m_i(\emptyset)m_j(\emptyset) \tag{12a}$$

$$m_{ij}(\{s\}) = \sum_{k=1}^{c} m_i(\{\omega_k\})m_j(\{\omega_k\}) \tag{12b}$$

$$m_{ij}(\{\neg s\}) = \sum_{A\cap B=\emptyset} m_i(A)m_j(B) - m_{ij}(\emptyset) \tag{12c}$$

$$m_{ij}(\Theta) = \sum_{A\cap B\neq\emptyset} m_i(A)m_j(B) - m_{ij}(s). \tag{12d}$$

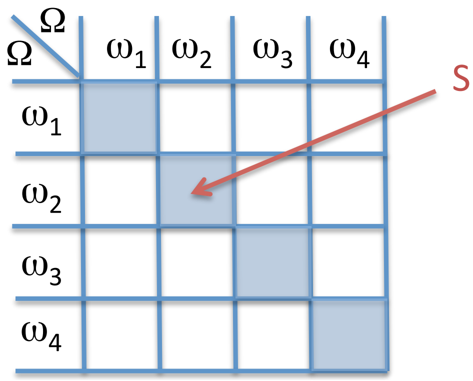

Fig. 3. Product space Ω^2, and subset S corresponding to the hypothesis that two objects belong to the same cluster.

Proof. Let A be a focal set of m_i and B be a focal set of m_j. After extension on Ω^2 and combination by Dempster's rule, the product $m_i(A)m_j(B)$ is assigned to the focal set

$$C = (A \times \Omega) \cap (\Omega \times B) = A \times B.$$

We can distinguish four cases:

1) If $A = \emptyset$ or $B = \emptyset$, then $C = \emptyset$;
2) If $A = B = \{\omega_k\}$, then $C = \{(\omega_k, \omega_k)\} \subset S$ hence, the mass $m_i(A)m_j(B)$ is transferred to s;
3) If $A \cap B = \emptyset$, $A \neq \emptyset$ and $B \neq \emptyset$, then $C \subset \overline{S}$ and $C \neq \emptyset$; hence, the mass $m_i(A)m_j(B)$ is transferred to $\neg s$;
4) If $A \cap B \neq \emptyset$ and ($|A| > 1$ or $|B| > 1$), then $C \cap S \neq \emptyset$ and $C \cap \overline{S} \neq \emptyset$; hence, the mass $m_i(A)m_j(B)$ is transferred to Θ.

By considering all focal sets A and B and summing up the masses, we get (12). □

Given an evidential partition $M = (m_1, \ldots, m_2)$, the tuple $R = (m_{ij})_{1 \leq i < j \leq n}$ will be called the *relational representation* of M. To better understand the nature of this representation, let us consider the following special cases:

- If M corresponds to a hard partition, i.e., if the mass functions m_i are certain, then

$$m_{ij}(\{s\}) = r_{ij}, \quad m_{ij}(\{\neg s\}) = 1 - r_{ij}, \tag{13}$$

with $r_{ij} = 1$ if objects i and j are in the same cluster, and $r_{ij} = 0$ otherwise; consequently, R contains the same information as the binary matrix $\boldsymbol{R} = (r_{ij})$ representing the equivalence relation corresponding to the hard partition.

- If M corresponds to a fuzzy partition, i.e., if the mass functions m_i are Bayesian, then m_{ij} is also Bayesian:

$$m_{ij}(\{s\}) = \sum_{k=1}^{c} u_{ik}u_{jk},$$
$$m_{ij}(\{\neg s\}) = 1 - \sum_{k=1}^{c} u_{ik}u_{jk};$$

- If M corresponds to a rough partition, i.e., if each mass functions m_i has a single focal set $A_i \subseteq \Omega$, then

$$m_{ij}(\{s\}) = 1 \quad \text{if } A_i = A_j = \{\omega_k\}$$
$$m_{ij}(\{\neg s\}) = 1 \quad \text{if } A_i \cap A_j = \emptyset$$
$$m_{ij}(\Theta) = 1 \quad \text{otherwise.}$$

Example 2: As an illustrative example, consider the evidential partition $M = (m_1, m_2, m_3)$ with $n = 3$ objects and $c = 2$ clusters shown in Table I. Both objects 1 and 2 seem to belong to cluster ω_1, while object 3 more likely belongs to cluster ω_2. There is also some evidence that object 1 might belong to none of the two clusters (as shown by the mass $m_1(\emptyset)$), as well as some ignorance about the cluster membership of objects 1 and 2 (as shown by the masses $m_i(\{\omega_1, \omega_2\})$ for $i = 1, 2$). The corresponding relational representation R is shown in Figure II. We can see that objects 1 and 2 seem to belong to the same cluster ($m_{12}(s) = 0.43$), while the pairs (1,3) and (2,3) seem to belong to different clusters ($m_{13}(\neg s) = 0.37$ and $m_{23}(\neg s) = 0.43$). The belief that object 1 might belong to none of the two clusters translates into the masses $m_{12}(\emptyset) = m_{13}(\emptyset) = 0.3$ being assigned the empty set.

TABLE I
EXAMPLE OF AN EVIDENTIAL PARTITION WITH $n = 3$ OBJECTS AND $c = 2$ CLUSTERS.

A	$\emptyset$	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$
$m_1(A)$	0.3	0.6	0.1	0.0
$m_2(A)$	0.0	0.7	0.1	0.2
$m_3(A)$	0.0	0.1	0.6	0.3

TABLE II
RELATIONAL REPRESENTATION OF THE EVIDENTIAL PARTITION SHOWN IN TABLE I.

A	$\emptyset$	$\{s\}$	$\{\neg s\}$	$\{s, \neg s\}$
$m_{12}(A)$	0.30	0.43	0.13	0.14
$m_{13}(A)$	0.30	0.12	0.37	0.21
$m_{23}(A)$	0.00	0.13	0.43	0.44

□

Remark 1: Given a relational representation R, can we uniquely recover the evidential partition M? For $n = 2$, it is easy to find counterexamples in which several choices of m_1 and m_2 correspond to the same mass function $m_{1,2}$. Consider, for instance, the following pair of mass functions on $\Omega = \{\omega_1, \omega_2, \omega_3\}$:

$$m_1(\{\omega_1, \omega_2\}) = 0.5, \quad m_1(\Omega) = 0.5$$

$$m_2(\{\omega_3\}) = 0.5, \quad m_2(\Omega) = 0.5.$$

We have $m_{12}(\emptyset) = m_{12}(\{s\}) = 0$, $m_{12}(\{\neg s\}) = 0.5 \times 0.5 = 0.25$ and $m_{12}(\Theta) = 1 - 0.25 = 0.75$. We can see that the pairwise mass function m_{12} depends only, in this case, on the product $m_1(\{\omega_1, \omega_2\}) \times m_2(\{\omega_3\})$. Consequently, all pairs of mass functions m'_1 and m'_2 such that

$$m'_1(\{\omega_1, \omega_2\}) = u, \quad m'_1(\Omega) = 1 - u$$

$$m'_2(\{\omega_3\}) = 0.25/u, \quad m'_2(\Omega) = 1 - 0.25/u$$

for some $u \in [0.25, 1]$ yield the same pairwise mass function m_{12}. We conjecture that the relation between M and R becomes one-to-one (up to a permutation of cluster labels) for large n, subject to some condition on n and c. However, proving this conjecture is an open problem. □

B. Similarity index between evidential partitions

Having computed the relational representations R and R' of two credal partitions M and M', we can compare M and M' from their relational representations. A first idea, somewhat similar to the one developed in [17], is to define a similarity between relational representations. Given a distance δ between mass functions taking values in $[0,1]$, we could define a generalized Rand index as

$$\rho_S(M, M') = 1 - \frac{\sum_{i<j} \delta(m_{ij}, m'_{ij})}{n(n-1)/2}. \quad (14)$$

Many distances between mass functions have been proposed: the extensive survey presented in [29] lists more than 60 of them. Here, we will consider two of the most widely used belief function distances, which also have the property of being full metrics: Jousselme's distance [30] and the belief distance [31], defined as follows.

1) Let us arrange the masses $m_{ij}(A)$ for $A \subseteq \Theta$ in a vector $\boldsymbol{m}_{ij} = (m_{ij}(\emptyset), m_{ij}(\{s\}), m_{ij}(\{ns\}), m_{ij}(\Theta))^T$. Jousselme's distance between m_{ij} and m'_{ij} is

$$\delta_J(m_{ij}, m'_{ij}) = \left(\frac{1}{2}(\boldsymbol{m}_{ij} - \boldsymbol{m}'_{ij})^T \boldsymbol{J} (\boldsymbol{m}_{ij} - \boldsymbol{m}'_{ij})\right)^{1/2} \quad (15)$$

where $\boldsymbol{J}$ is the Jaccard matrix

$$\boldsymbol{J} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & 1/2 \\ 0 & 1/2 & 1/2 & 1 \end{pmatrix}.$$

2) The belief distance between between m_{ij} and m'_{ij} is

$$\delta_B(m_{ij}, m'_{ij}) = \frac{1}{2} \sum_{A \subseteq \Theta} \mid Bel_{ij}(A) - Bel'_{ij}(A) \mid, \quad (16)$$

where Bel_{ij} and Bel'_{ij} are, respectively, the belief functions (30) associated to m_{ij} and m'_{ij}, and the multiplicative constant $1/2$ ensures that $\delta_B(m_{ij}, m'_{ij}) \in [0,1]$. This distance can be expressed as a function of the masses as

$$\begin{aligned} \delta_B(m_{ij}, m'_{ij}) = \frac{1}{2} \{ & \mid m(\{s\}) - m'(\{s\}) \mid \\ & + \mid m(\{\neg s\}) - m'(\{\neg s\}) \mid \\ & + \mid m(\emptyset) - m'(\emptyset) \mid \}. \end{aligned} \quad (17)$$

In the following, we will denote by ρ_S^J and ρ_S^B the similarity indices ρ_S defined by (14)-(15), and (14)-(17), respectively. Both indices take values in $[0,1]$ and generalize the Rand index (they become the Rand index when M and M' correspond to hard partitions). Furthermore, Jousselme's distance δ_J was shown in [32] to be a metric, as a consequence of the positive-definiteness of $\boldsymbol{J}$, and so is δ_B. Consequently, $1 - \rho_S^J$ and $1 - \rho_S^B$ are pseudometric (they are reflexive, symmetric and they verifiy the triangular inequality).

Example 3: As a continuation of Example 2, assume that we now have a second evidential partition $M' = (m'_1, m'_2, m'_3)$ of the same three objects, shown in Table III, with the corresponding relational representation R' shown in Table IV. The distances between the pairwise mass functions are

$$\delta_J(m_{12}, m'_{12}) = 0.261, \ \delta_J(m_{13}, m'_{13}) = 0.263$$
$$\delta_J(m_{23}, m'_{23}) = 0.124,$$

and

$$\delta_B(m_{12}, m'_{12}) = 0.32, \ \delta_B(m_{13}, m'_{13}) = 0.265$$
$$\delta_B(m_{23}, m'_{23}) = 0.120.$$

The corresponding similarity indices between evidential partitions M and M' are thus

$$\rho_S^J(M, M') = 1 - (0.261 + 0.263 + 0.124)/3 = 0.784$$

and

$$\rho_S^B(M, M') = 1 - (0.32 + 0.265 + 0.120)/3 = 0.765.$$

TABLE III
EVIDENTIAL PARTITION M' OF EXAMPLE 3.

A	$\emptyset$	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$
$m'_1(A)$	0	0.2	0.7	0.1
$m'_2(A)$	0	0.2	0.8	0.0
$m'_3(A)$	0	0.6	0.2	0.2

TABLE IV
RELATIONAL REPRESENTATION OF THE EVIDENTIAL PARTITION M' SHOWN IN TABLE III.

A	$\emptyset$	$\{s\}$	$\{\neg s\}$	$\{s, \neg s\}$
$m'_{12}(A)$	0.0	0.6	0.3	0.1
$m'_{13}(A)$	0.00	0.26	0.46	0.28
$m'_{23}(A)$	0.00	0.28	0.52	0.20

□

Similarity indices such as ρ_S^J and ρ_S^B thus allow us to measure the closeness of two soft (fuzzy, possibilistic, rough or, more generally, evidential) partitions. The corresponding distances being pseudometrics, they are suitable, for instance, for comparing the results of several clustering algorithms, or for analyzing the sensitivity of the clustering to perturbations of the data. However, when comparing a ground truth hard partition with an evidential partition, the notion of distance may not always be relevant. Assume, for instance, that two objects i and j are known to belong to the same group, which we denote by $m^*_{ij}(\{s\}) = 1$. Consider two clustering algorithms, producing the following results about the pair (i,j):

- Algorithm 1: $m_{ij}(\{s\}) = 0.5$, $m_{ij}(\{\neg s\}) = 0.5$;
- Algorithm 2: $m'_{ij}(\Theta) = 1$.

The result of Algorithm 1 is partially in conflict with the truth, as it assigns some mass to the singleton $\{\neg s\}$. In contrast, Algorithm 2 is more cautious, and produces a mass function m'_{ij} that is imprecise, but consistent with the truth. Yet, these two results are not distinguished by criterion ρ_S^B, as $\delta_B(m^*_{ij}, m_{ij}) = \delta_B(m^*_{ij}, m'_{ij}) = 0.5$. With Jousselme's distance, we find $\delta_J(m^*_{ij}, m_{ij}) = 0.5$ and $\delta_J(m^*_{ij}, m_{ij}) = 0.71$, but these values do not seem to indicate that m_{ij} is preferable to m'_{ij} in any way. To address this issue, a new criterion will be introduced in the next section.

C. Consistency index between evidential partitions

To measure the consistency between an evidential partition and a true hard partition, another notion than that of distance might thus be suitable. We propose to use the notion of *degree of conflict* (32), which leads to another generalized Rand index:

$$\rho_C(M, M') = 1 - \frac{2\sum_{i<j} \kappa(m_{ij}, m'_{ij})}{n(n-1)}. \quad (18)$$

where

$$\kappa(m_{ij}, m'_{ij}) = \sum_{A\cap B=\emptyset} m_{ij}(A) m'_{ij}(B) = \boldsymbol{m}_{ij}^T \boldsymbol{C}\, \boldsymbol{m}'_{ij}, \quad (19)$$

with

$$\boldsymbol{C} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

This is a proper generalization of the Rand index, as in the case of hard partitions $\kappa(m_{ij}, m'_{ij}) = 0$ if $r_{ij} = r'_{ij}$ and $\kappa(m_{ij}, m'_{ij}) = 1$ otherwise. In the special case where we compare an evidential partition M with a true hard partition M^*, index ρ_C has a simpler expression. Let $m^*_{ij}(\{s\}) = r^*_{ij}$ and $m^*_{ij}(\{\neg s\}) = 1 - r^*_{ij}$, where $r^*_{ij} = 1$ if objects i and j truly belong to the same cluster, and $r^*_{ij} = 0$ otherwise. Then, the degree of conflict between m_{ij} and m^*_{ij} is

$$\kappa(m_{ij}, m^*_{ij}) = \begin{cases} m_{ij}(\emptyset) + m_{ij}(\{\neg s\}) & \text{if } r^*_{ij} = 1, \\ m_{ij}(\emptyset) + m_{ij}(\{s\}) & \text{if } r^*_{ij} = 0 \end{cases} \quad (20)$$

$$= \begin{cases} 1 - pl_{ij}(s) & \text{if } r^*_{ij} = 1, \\ 1 - pl_{ij}(\neg s) & \text{if } r^*_{ij} = 0. \end{cases} \quad (21)$$

Consequently,

$$\rho_C(M, M^*) = 1 - \frac{\sum_{i<j, r^*_{ij}=1}(1 - pl_{ij}(s)) + \sum_{i<j, r^*_{ij}=0}(1 - pl_{ij}(\neg s))}{n(n-1)/2}, \quad (22)$$

which can be simplified to

$$\rho_C(M, M^*) = \frac{\sum_{i<j} pl_{ij}(s)^{r^*_{ij}}\, pl_{ij}(\neg s)^{1-r^*_{ij}}}{n(n-1)/2}. \quad (23)$$

The meaning of (23) is clear: the consistency $\rho_C(M, M^*)$ between M and the true partition M^* is high when $pl_{ij}(s)$ is high whenever $r^*_{ij} = 1$, and $pl_{ij}(\neg s)$ is high whenever $r^*_{ij} = 0$.

If the evidential partition M is vacuous, i.e., if $m_{ij}(\Theta) = 1$ for all $i < j$, meaning that we are in a state of total ignorance regarding the membership of objects to clusters, then $\rho_C(M, M^*) = 1$: since M encodes total ignorance, it is perfectly consistent with the truth, whatever it is. It is obvious from this special case that $1 - \rho_C$ is not a distance: it is a measure of conflict, and ρ_C can be seen as a measure of consistency between two evidential partitions.

As $\rho_C(M, M)$ is, in general, strictly less than one, index ρ_C is not suitable to compare, for instance, two evidential partitions generated by two algorithms. But it does make sense to use $\rho_C(M, M^*)$ to measure how consistent the evidential partition M is, with respect to the true hard partition M^*. A cautious clustering algorithm, which produces imprecise evidential partitions, will tend to have high ρ_C scores. When comparing the performances of two clustering algorithms generating evidential partitions M and M', we should, therefore, also take into account the imprecision of the evidential partitions. As mentioned in Section A, imprecision can be measured by nonspecificity (29). The imprecision of a credal partition can thus be defined from the average nonspecificity of the m_{ij}'s,

$$N(M) = \frac{\sum_{i<j} N(m_{ij})}{n(n-1)/2} = \frac{\sum_{i<j}[m_{ij}(\emptyset) + m_{ij}(\Theta)]}{n(n-1)/2}, \quad (24)$$

which ranges in $[0, 1]$. We note that, for crisp and fuzzy partitions, we have $N(M) = 0$.

The quality of an evidential partition M can thus be described by two numbers: $\rho_C(M, M^*)$ and $N(M)$, both taking values in $[0, 1]$. Given two evidential partitions M and M', M can be considered to be preferable to M' if it is both more consistent with the truth, and more precise, i.e., if

$$\rho_C(M, M^*) \geq \rho_C(M', M^*) \text{ and } N(M) \leq N(M'), \quad (25)$$

and M is strictly more preferable than M' if at least one of the inequalities on (25) is strict.

Example 4: Consider again the evidential partitions M and M' of Examples 2 and 3. Assume that the true hard partition M^* is defined by $m^*_{12}(\{s\}) = r^*_{12} = 1$, $m^*_{13}(\{\neg s\}) = 1 - r^*_{13} = 1$ and $m^*_{23}(\{\neg s\}) = 1 - r^*_{23} = 1$, i.e., objects 1 and 2 are in the same cluster, and object 3 forms another cluster. For evidential partition M (see Table II), the degrees of conflict are

$$\kappa(m_{12}, m^*_{12}) = 1 - pl_{12}(s) = 1 - (0.43 + 0.14) = 0.43$$
$$\kappa(m_{13}, m^*_{13}) = 1 - pl_{13}(\neg s) = 1 - (0.37 + 0.21) = 0.42$$
$$\kappa(m_{23}, m^*_{23}) = 1 - pl_{23}(\neg s) = 1 - (0.43 + 0.44) = 0.13.$$

Consequently, we have $\rho_C(M, M^*) = 1 - (0.43 + 0.42 + 0.13)/3 = 0.673$. Similar calculation yields $\rho_C(M', M^*) = 0.72$: M' is thus strictly more consistent than M with the true partition. Now, the nonspecificities of M and M' are, respectively, $N(M) = 0.46$ and $N(M') = 0.19$. We can conclude that M' is also strictly more precise than M. Consequently M' is strictly preferable to M. □

IV. Numerical experiments

As noted by Hüllermeier et al. [17], there is no gold standard to which an external validity index such as a generalized Rand index can be compared. The best we can do is to show that the indices introduced in Sections III-B and III-C above provide useful information for assessing and comparing

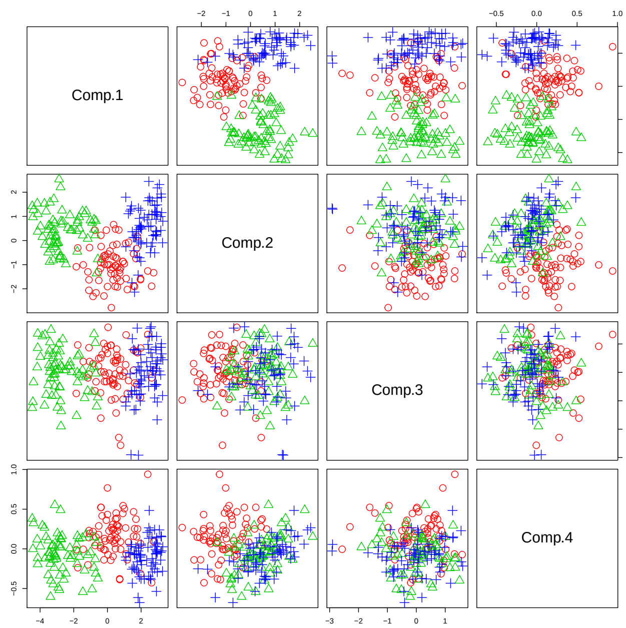

Fig. 4. The seeds data, in the space spanned by the first four components of Principal Component Analysis.

TABLE V
CLUSTERING ALGORITHMS USED IN EXPERIMENT 1.

algorithm	R function	R package
HCM	`kmeans`	`stats` [34]
Hierarchical clust. (Ward distance)	`hclust`	`stats`
ECM [9]	`ecm`	`evclust` [26]
EVCLUS [10]	`kevclus`	`evclust` [26]
FCM [35]	`FKM`	`fclust` [36]
Fuzzy k-medoids [37]	`FKM.med`	`fclust` [36]
π-Rough k-means [38]	`RoughKMeans_PI`	`SoftClustering` [39]
EM [2]	`Mclust`	`mclust` [40]

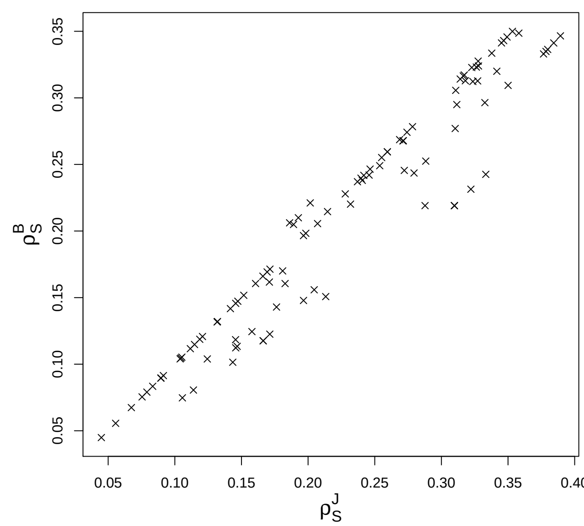

Fig. 5. Similarity indices ρ_S^B vs. ρ_S^J for the Seeds data.

soft partitions produced by a variety of clustering algorithms. This is our objective in the section. Two experiments will be presented. In Section IV-A, we will show that similarity index ρ_S can be used to compare hard and soft partitions produced by various algorithms and to generate informative graphical representations. In Section IV-B, we will use consistency index ρ_C together with the nonspecificity measure to compare evidential and rough clustering algorithms for different values of their tuning parameters.

A. Experiment 1

As an example, we consider the Seeds data from the UCI Repository of machine learning databases [33]. This dataset consists of 7 attributes observed for 210 kernels belonging to three different varieties of wheat, with 70 elements each. Figure 4 shows the three clusters in the space spanned by the first four principal components of the data.

The clustering algorithms used in this experiment are shown in Table V. The simulations were done using the R environment [34]. Each algorithm was run using the default parameter values. For the EM algorithm, we used the Gaussian mixture model with spherical classes of equal volume, which is the most similar to the assumptions of FCM. For hierarchical clustering, we used the Ward distance and we cut the dendrogram to obtain three clusters. For ECM, we used the version with masses assigned to all subsets of Ω, except Ω itself. For EVCLUS, we specified the focal sets to be the empty set, the singletons and Ω. The evidential partitions generated by ECM and EVCLUS were transformed into hard (ECMh, EVCh), fuzzy (ECMf, EVCf) and rough (ECMr, EVCr) partitions, as explained in Section II-A. We thus obtained a total of 15 hard and soft partitions (including the true one). These partitions and the corresponding notations are recapitulated in Table VI.

The similarity indices ρ_S^J between these partitions are shown in Table VII. As shown in Figure 5, indices ρ_S^J and ρ_B^J are highly correlated. Figures 6(a) and 6(b) display Multi-dimensional Scaling (MDS) configurations of, respectively, the distances $1-\rho_S^J$ and $1-\rho_B^J$ between the 15 partitions, computed using Kruskal's nonmetric method [41]. We recall that MDS finds a configuration of points in a low-dimensional space, such that Euclidean distances between the points match the observed distances or dissimilarities between the objects. Comparing Figures 6(a) and 6(b), we can see that the MDS configurations computed from similarity indices ρ_S^J and ρ_B^J are quite similar, which seems to indicate that these two indices can be used interchangeably.

From Figure 6, we can see that partitions of the same type (evidential, fuzzy, rough, hard) are mapped together in the MDS representation, i.e., they are similar according to index ρ_S, whereas partitions of different types are mapped in different regions of the configuration space. (The partition derived from the EM algorithm is clustered with hard partitions because most of the estimated cluster membership probabil-

TABLE VI
DESCRIPTION OF THE 15 HARD AND SOFT PARTITIONS IN EXPERIMENT 1.

Acronym	Method	Acronym	Method
TRUE	True partition	EVC	EVCLUS
HCM	Hard c-means	EVCh	Hard partition derived from EVCLUS
Ward	Ward hierarchical clustering	EVCf	Fuzzy partition derived from EVCLUS
EM	EM algorithm with Gaussian Mixture Model	EVCr	Rough partition from EVCLUS
FCM	Fuzzy c-means	ECM	Evidential c-means
FKMm	Fuzzy k-medoids	ECMh	Hard partition derived from ECM
RCM	Rough c-means	ECMf	Fuzzy partition derived from ECM
		ECMr	Rough partition derived from ECM

TABLE VII
SIMILARITY INDICES ρ_S^J BETWEEN THE 15 HARD AND SOFT PARTITIONS OF EXPERIMENT 1 (SEEDS DATASET).

	True	HCM	Ward	ECM	ECMh	ECMf	ECMr	EVC	EVCh	EVCf	EVCr	FCM	FKMm	EM
HCM	0.90													
Ward	0.85	0.89												
ECM	0.62	0.62	0.61											
ECMh	0.89	0.91	0.84	0.62										
ECMf	0.68	0.68	0.67	0.81	0.69									
ECMr	0.68	0.69	0.67	0.73	0.69	0.64								
EVC	0.65	0.66	0.65	0.82	0.65	0.8	0 0.66							
EVCh	0.88	0.92	0.91	0.62	0.92	0.68	0.69	0.66						
EVCf	0.75	0.76	0.76	0.77	0.76	0.90	0.69	0.81	0.77					
EVCr	0.82	0.88	0.84	0.67	0.85	0.67	0.79	0.68	0.89	0.76				
FCM	0.73	0.75	0.73	0.79	0.74	0.93	0.67	0.81	0.74	0.92	0.73			
FKMm	0.83	0.86	0.83	0.71	0.83	0.80	0.72	0.75	0.85	0.88	0.83	0.87		
EM	0.89	0.96	0.87	0.65	0.92	0.72	0.72	0.69	0.91	0.80	0.88	0.79	0.89	
RCM	0.80	0.83	0.80	0.69	0.83	0.67	0.86	0.68	0.83	0.75	0.89	0.73	0.82	0.85

ities are close to 0 or 1.) Evidential partitions (ECM, EVC) are the most dissimilar to the true one, but hard partitions derived from them (ECMh, EVCh) are quite similar to the hard partition generated by the HCM algorithm, and have similar values of the Rand index. In the same way, fuzzy and rough partitions derived from evidential partitions are similar to those generated by the FCM and π-RCM algorithms, respectively. This finding confirms the idea that evidential clustering extracts rich information from the data, which can be summarized into different clustering structures (hard, fuzzy, rough). A related finding is that the similarity to the true partition according to ρ_S^B or ρ_S^J is not a good criterion to rank clustering algorithms of different types: the similarity between a soft partition and the true one may be low, even through the true partition can be approximated quite accurately by summarizing the soft partition.

Figure 7 shows another representation of the 15 hard and soft partitions, according to nonspecificity (x-axis) and consistency with the true partition (y-axis). Hard and fuzzy partitions all have zero nonspecificity: consequently, they are located on the line $x = 0$. Once again we can see that rough, fuzzy and hard partitions extracted from ECM and EVCLUS are quite similar to partitions generated by the RCM, FCM and HCM algorithms, respectively. The rough partition extracted from ECM (ECMr) has the highest degree of consistency, but it is also the most imprecise. Consequently, it is not comparable to any of the other soft partitions. Similarly, the evidential partitions generated by ECM and EVCLUS are not comparable: the evidential partition generated by ECM is slightly more consistent than that produced by EVC, but it is also less precise.

Figure 7 also reveals that hard partitions generally have a higher degree of consistency with the true partition, as compared to fuzzy partitions. In a similar way, rough partitions are more consistent than evidential partitions. This effect can be easily explained: assume that two objects i and j actually belong to the same class, i.e., $m_{ij}^*(\{s\}) = 1$. Consider the following pairwise mass function: $m_{ij}(\{s\}) = 0.5$, $m_{ij}(\{\neg s\}) = 0.2$, $m_{ij}(\Theta) = 0.3$. The degree of conflict between m_{ij}^* and m_{ij} is $\kappa_{ij} = 1 - pl_{ij}(s) = 0.2$. Now, for the derived rough partition, we will have $m'_{ij}(\Theta) = 1$, and the degree of conflict will be $\kappa'_{ij} = 1 - pl'_{ij}(s) = 0$. This observation suggest that preference relation (25) might be specially relevant to compare evidential partitions (with graded membership values) with fuzzy partitions, and rough partitions (with 0-1 membership values) with hard partitions.

B. Experiment 2

When comparing imprecise predictions of some quantities, we usually need to take into account both precision and accuracy. For instance, when predicting a real number x by an interval I, we need to consider not only whether I contains x, but also the length of I. If two intervals I_1 and I_2 both contain x, I_1 may be considered a better prediction than I_2 if it is smaller. A similar issue arises when comparing evidential partitions. A more imprecise evidential partition will usually be more consistent with the true partition; consequently, nonspecificity and consistency both need to be taken into account when assessing the quality of an evidential partition. Evidential clustering algorithms such as ECM [9] and EVCLUS [10]

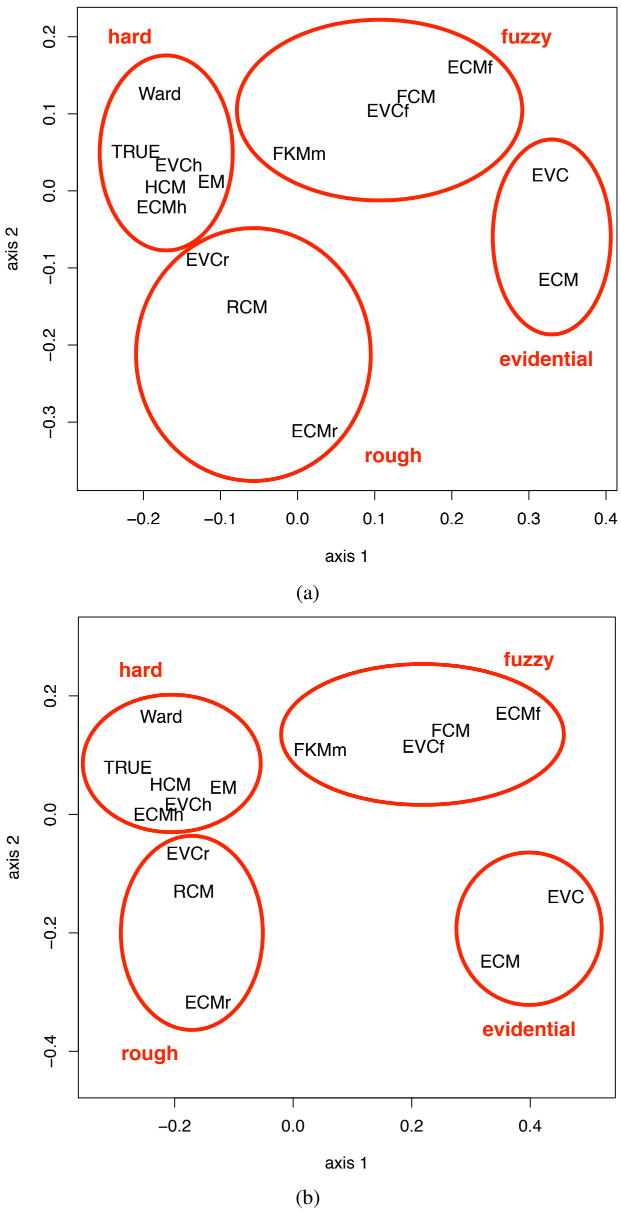

Fig. 6. MDS representations of similarities ρ_S^J (a) and ρ_S^B (b) for the Seeds data. Distances in the MDS representation match dissimilarities $1-\rho_S$ between the evidential partitions. The axes are arbitrary, as the representation is defined up to an isometric transformation.

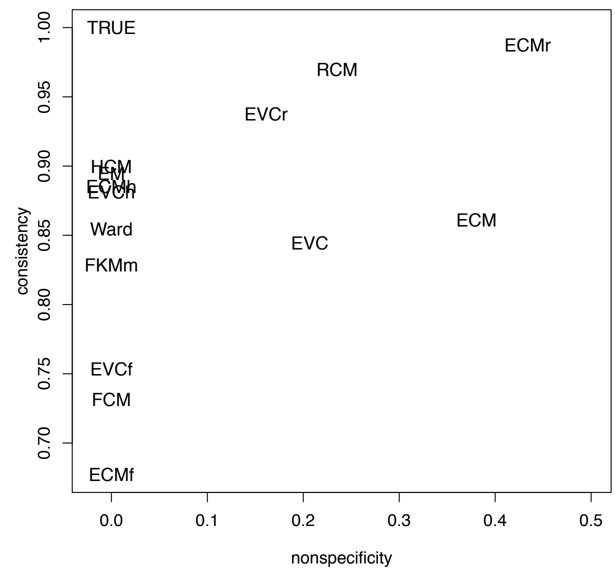

Fig. 7. Nonspecificity $N(M)$ (x-axis) vs. consistency with the true partition $\rho_C(M, M^*)$ (y-axis) for the 15 hard and soft partitions of Experiment 1 (Seeds dataset).

have a parameter that allows one to tune the nonspecificity of the evidential partition. Larger nonspecificity results in smaller conflict. When comparing two algorithms, we thus need to compute the pairs $(N(M), \rho_C(M, M^*))$ for a range of parameter values.

In this second experiment, we considered the ECM and EVCLUS algorithms. ECM [9] minimizes the following cost function:

$$J_{\text{ECM}}(M,V) = \sum_{i=1}^{n}\sum_{j=1}^{f} |A_j|^\alpha [m_i(A_j)]^\beta d_{ij}^2 + \sum_{i=1}^{n} \delta^2 [m_i(\emptyset)]^\beta, \quad (26)$$

where $A_1, \ldots, A_f$ are f nonempty subsets of Ω, d_{ij} denotes the Euclidean distance between attribute vector $\boldsymbol{x}_i$ and the prototype $\boldsymbol{v}_j$ representing the set of clusters A_j, and $V = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_f)$. In (26), coefficient α controls the imprecision of the evidential partition: a larger value of α penalizes focal sets A_j with a large cardinality, resulting in a more specific evidential partition. In this experiment, the focal sets A_j were defined as subsets of Ω with cardinality one or two. Parameters β and δ were set to 1 and to the maximum of distances $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$, respectively.

EVCLUS [10] minimizes the stress function

$$J(M) = \sum_{i<j} (\kappa_{ij} - \delta_{ij})^2, \quad (27)$$

where κ_{ij} is the degree of conflict (32) between mass functions m_i and m_j, and δ_{ij} is the transformed dissimilarity between objects o_i and o_j. Usually, δ_{ij} is defined from distances d_{ij} through the transformation $\delta_{ij} = 1 - \exp(-\gamma d_{ij}^2)$, where γ is determined in such a way that $\delta_{ij} = 0.95$ when d_{ij} equals some predefined value d_0. Parameter d_0 is interpreted as follows: two objects o_i and o_j such that $d_{ij} \geq d_0$ have a plausibility at least 0.95 of belonging to different clusters. Parameter d_0 influences nonspecificity, larger values of d_0 resulting in less specific evidential partitions. Here, d_0 was set to the q-quantile of the distances d_{ij}, with $q \in [0.9, 1]$.

As an illustration, three datasets from the UCI Repository of machine learning databases [33] were used in this experiment: Seed (see Section IV-A), Ecoli and Wine. The Ecoli dataset contains data about protein localization sites in E. coli bacteria. We used only the quantitative attributes (2, 3, 6, 7, and 8) and the four most frequent classes: ‘im’, ‘pp’, ‘imU’ and ‘cp’, resulting in a dataset with 307 objects and 5 attributes. The Wine data are results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. The dataset contains 13 attributes (the quantities of 13 constituents) and 178 objects (wines).

We applied ECM and EVCLUS to each of these three datasets. For ECM, parameter α was varied between 0 and 3. For EVCLUS, we varied q between 0.9 and 1. The results are shown in Figures 8, 9 and 10. Each of these figures contains two graphs. The upper graphs (Figures 8(a), 9(a) and 10(a)) show the nonspecificity and consistency indices of the evidential partitions obtained by ECM and EVCLUS with different parameters. The lower graphs (Figures 8(b), 9(b) and 10(b)) show the same information for the rough partitions derived from the evidential partitions generated by ECM and EVCLUS. We also show the fuzzy partition generated by the Fuzzy c-means algorithm with the evidential partitions, and the hard partition generated by the Hard c-means with the rough partitions. (As remarked in Section IV-A, evidential partitions with graded membership values are comparable to fuzzy partitions, while rough partitions with 0-1 membership values are comparable to hard partitions.) Hard and fuzzy partitions have zero nonspecificity: they are thus located on the vertical axes in Figures 8-10. We can see that evidential (respectively, rough) partitions outperform fuzzy (respectively, hard) partitions in terms of consistency, and this advantage can be further enhanced by increasing nonspecificity.

Figures 8-10 reveal similar patterns. For each of the two algorithms ECM and EVCLUS, it possible to tune parameters so as to increase the consistency $\rho_C(M, M^*)$ with the true partition, at the price of also increasing the nonspecificity (imprecision) of the evidential partition. For the three datasets, EVCLUS outperforms ECM in terms of the evidential partitions: for comparable nonspecificity, evidential partitions generated by EVCLUS are more consistent with the true partition than those generated by ECM (Figures 8(a)-10(a)). However, the opposite effect if observed for the rough partitions: those derived from ECM are more consistent than those derived from EVCLUS (Figures 8(b)-10(b)). Overall, the two algorithms thus appear to be roughly equivalent for these three datasets. We can remark that these nonspecificity-consistency curve provide a rich description of the quality of evidential partitions generated by a clustering algorithm with different parameter tunings.

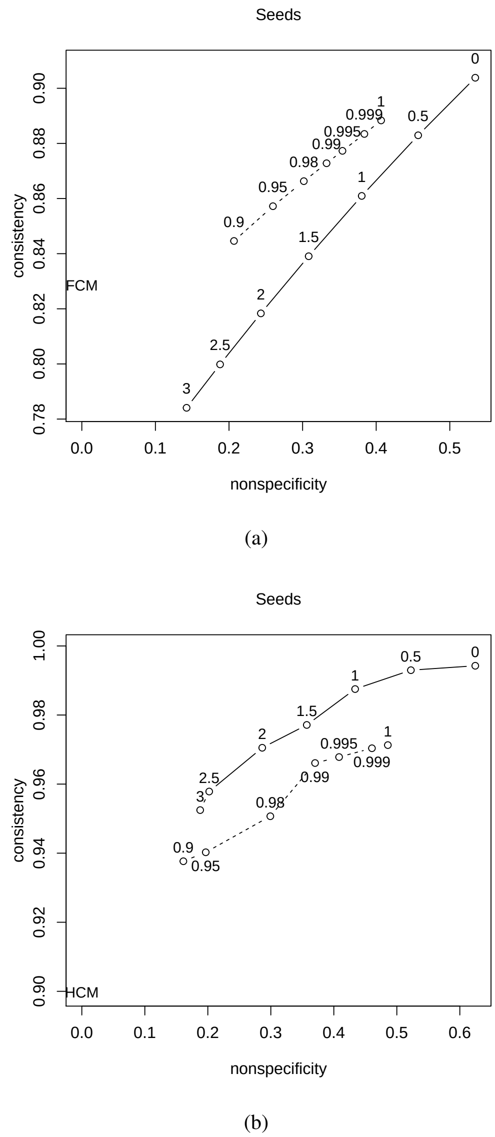

Fig. 8. Top: nonspecificity $N(M)$ (x-axis) vs. consistency with the true partition $\rho_C(M, M^*)$ (y-axis) for ECM (solid line) and EVCLUS (dashed line) for different values of parameters α and q, respectively (Seeds dataset). Bottom: same graph for the rough partitions derived from the evidential partitions.

V. Conclusions

The Dempster-Shafer theory of belief functions [11] provides a very general framework for representing and reasoning with uncertainty. In evidential clustering, this framework is exploited to represent uncertain assignment of objects to clusters. In previous work, this approach has allowed us to devise new clustering procedures such as ECM [9], EVCLUS [8], [10] and Ek-NNclus [24]. In this paper, we have shown that the Dempster-Shafer framework also makes it possible to evaluate and compare soft partitions produced by a wide range of methods, including probabilistic, fuzzy, possibilistic and rough clustering algorithms.

The Rand index is probably the most widely used external cluster validity measure in a hard clustering context. It can be seen both as a similarity measure, allowing one to measure the *closeness* of two hard partitions, and as a measure of *agreement* or *consistency* between a hard partition computed by some algorithm and the true one. When moving from hard to evidential partitions, the two notions need to be distinguished: by making an evidential partition more

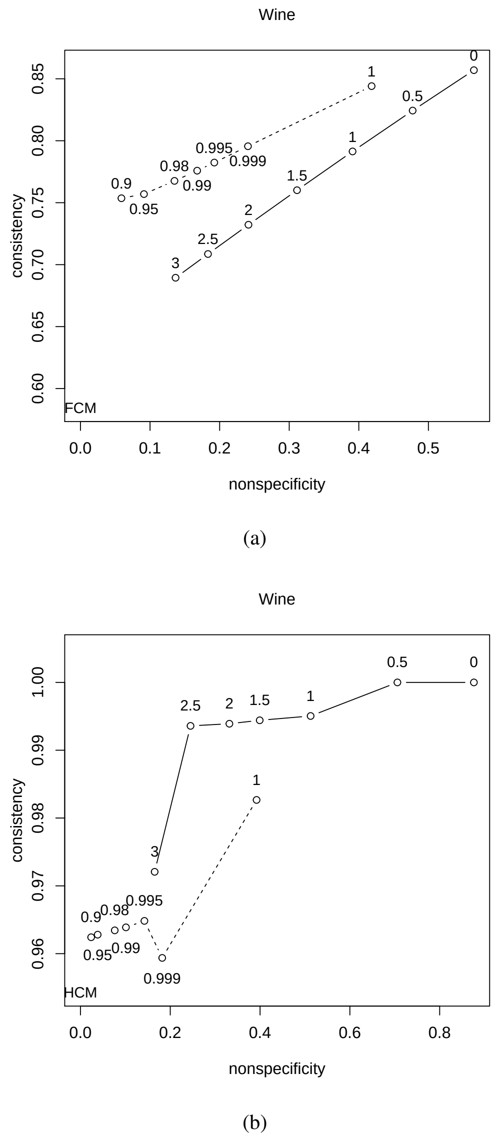

Fig. 9. Top: Nonspecificity $N(M)$ (x-axis) vs. consistency with the true partition $\rho_C(M, M^*)$ (y-axis) for ECM (solid line) and EVCLUS (dashed line) for different values of parameters α and q, respectively (Wine dataset). Bottom: same graph for the rough partitions derived from the evidential partitions.

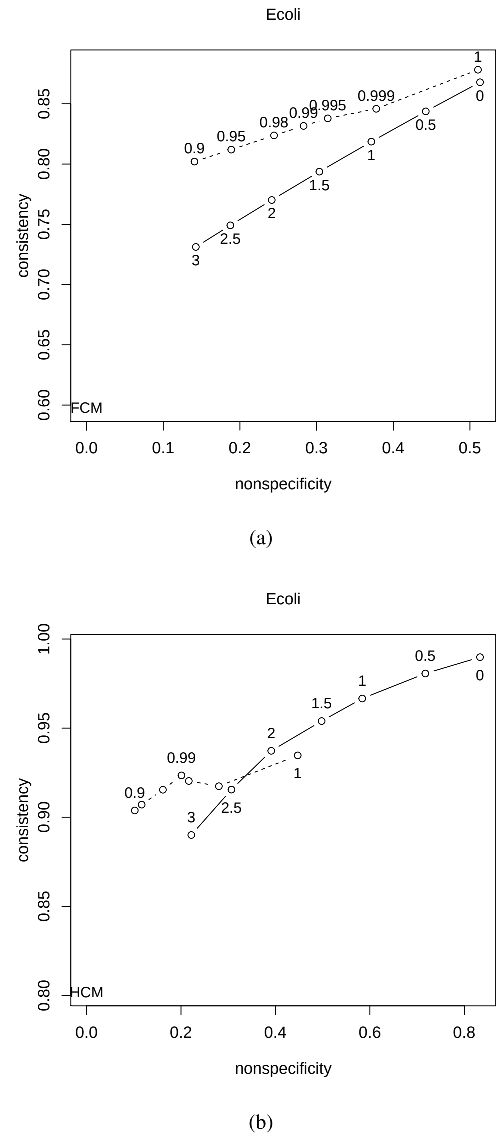

Fig. 10. Top: Nonspecificity $N(M)$ (x-axis) vs. consistency with the true partition $\rho_C(M, M^*)$ (y-axis) for ECM (solid line) and EVCLUS (dashed line) for different values of parameters α and q, respectively (Ecoli dataset). Bottom: same graph for the rough partitions derived from the evidential partitions.

imprecise, we can make it more consistent with the true partition, while increasing the distance to it. We thus need two indices: a similarity index and a consistency index. To define such indices, we have introduced the notion of relational representation of an evidential partition, defined as a set of pairwise mass functions indicating whether any two objects belong to the same class or not, with some uncertainty. The mean distance between such pairwise mass functions for two evidential partitions allows one to define a similarity index, while the mean degree of conflict yields a consistency index. The former is suitable for comparing partitions produced by different soft clustering algorithms, or by the same algorithm with different initial conditions or parameter values, while the latter makes it possible to compare a soft clustering to the ground truth partition, when available.

We hope that the unified approach introduced in this paper will stimulate research on the comparison of clustering methods based on different frameworks, including the fuzzy, possibilistic and rough settings. Another potential application concerns the combination of heterogeneous soft partitions computed by various clustering procedures. This topic is left for further study.

APPENDIX
DEMPSTER-SHAFER THEORY

Let Ω be a finite set. A *mass function* [11] is a mapping m from the power set of Ω, denoted as 2^Ω, to the interval $[0,1]$, such that $\sum_{A\subseteq\Omega} m(A)=1$. The subsets A of Ω such that $m(A) > 0$ are called the *focal sets* of m. Typically, Ω is a set of possible answers to some question, and $m(A)$ is interpreted as a share of a unit mass of belief allocated to the hypothesis that the truth is in A based on some evidence, and which cannot be allocated to any strict subset of A. If $m(\emptyset) = 0$, the mass function is said to be normalized. This condition will not be imposed in this paper. A mass function m is said to be *logical* if it has only one focal set; *Bayesian* if its focal sets are singletons; *certain* if it is both logical and Bayesian, i.e., if it has only one focal set, and this focal set is a singleton; *consonant* if its focal sets are nested.

The imprecision of a mass function can be measured by its *nonspecificity* [42], defined for a normalized mass function as

$$N(m) = \sum_{A\subseteq\Omega} m(A) \log_2 |A|, \tag{28}$$

where $|A|$ denotes the cardinality of A. This function was shown by Ramer [43] to be the only one satisfying some axioms. Klir and Wierman [44, page 51] proposed to extend it to unnormalized mass function as

$$N(m) = \sum_{A\subseteq\Omega, A\neq\emptyset} m(A) \log_2 |A| + m(\emptyset) \log_2 |\Omega|. \tag{29}$$

The rationale of this extension is that the mass assigned to the empty set expresses inconsistency of the evidence, which is a form of uncertainty. Consequently, the coefficient of $m(\emptyset)$ in (29) should be as high as that of $m(\Omega)$, i.e., it should be equal to $\log_2 |\Omega|$.

Given a mass function m, *belief* and *plausibility* functions are defined as follows:

$$Bel(A) = \sum_{\emptyset\neq B\subseteq A} m(B) \quad Pl(A) = \sum_{B\cap A\neq\emptyset} m(B), \tag{30}$$

for all $A \subseteq \Omega$. The quantity $Bel(A)$ can be interpreted as a degree of support in A, while $Pl(A)$ can be seen as a degree to which hypothesis A is consistent with the evidence [11]. The *contour function* $pl : \Omega \to [0, 1]$ is the restriction of the plausibility function Pl to singletons, i.e., $pl(\omega) = Pl(\{\omega\})$, for all $\omega \in \Omega$.

Two mass functions m_1 and m_2 representing independent items of evidence can be combined using the unnormalized Dempster's rule [45] as follows,

$$(m_1 \cap m_2)(A) = \sum_{B\cap C=A} m_1(B)m_2(C), \tag{31}$$

for all $A \subseteq \Omega$. The quantity

$$\kappa = (m_1 \cap m_2)(\emptyset) = \sum_{B\cap C=\emptyset} m_1(B)m_2(C) \tag{32}$$

is called the *degree of conflict* between m_1 and m_2.

Much of the power of Dempster-Shafer theory resides in the possibility to change the granularity of the set on which beliefs are expressed [11]. Let Ω and Θ be two finite sets. We say that Ω is a *refinement* of Θ (or Θ is a *coarsening* of Ω) if there is a mapping $f : \Theta \to 2^\Omega$ (called a *refining*) such that $f(\theta) \cap f(\theta') = \emptyset$ for any $\theta \neq \theta'$, and $\bigcup_{\theta\in\Theta} f(\theta) = \Omega$ (see Figure 11). Each element $\theta \in \Theta$ is thus split into a subset $f(\theta)$ of Ω, resulting in a finer representation. If Ω is a refinement of Θ, a mass function m^Θ on Θ can be carried to Ω by transferring each mass $m^\Theta(A)$ to the set $\bigcup_{\theta\in A} f(\theta)$. The resulting mass function m^Ω is called the *extension* of m^Θ in Ω. Conversely, a mass function m^Ω on Ω can be carried to Θ (possibly with some loss of information) by transferring each mass $m^\Omega(B)$ to the set $\{\theta \in \Theta | f(\theta) \cap B \neq \emptyset\}$. For instance, in Figure 11, the mass $m^\Omega(B)$ would be transferred to $\{\theta_2, \theta_3\}$. The resulting mass function m^Θ is called the *restriction* of m^Ω in Θ.

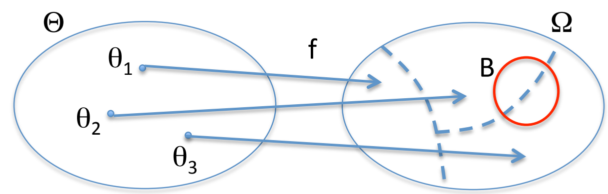

Fig. 11. Refining of a set.

ACKNOWLEDGMENT

This research was supported by a grant from the Beijing Government as part of the Overseas Talents program, and by the Puey Ungphakorn Center of Excellence in Econometrics at Chiang Mai University.

REFERENCES

[1] 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*, vol. 54, no. 2, pp. 307–322, 2013.
[2] C. Fraley and A. E. Raftery, "Model-based clustering, discriminant analysis and density estimation," *Journal of the American Statistical Association*, vol. 97, pp. 611–631, 2002.
[3] J. Bezdek, *Pattern Recognition with fuzzy objective function algorithm*. New-York: Plenum Press, 1981.
[4] R. Krishnapuram and J. Keller, "A possibilistic approach to clustering," *IEEE Trans. on Fuzzy Systems*, vol. 1, pp. 98–111, 1993.
[5] M.-S.Yang and C.-Y. Lai, "A robust automatic merging possibilistic clustering method," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 1, pp. 26–41, 2011.
[6] P. Lingras and G. Peters, "Applying rough set concepts to clustering," in *Rough Sets: Selected Methods and Applications in Management and Engineering*, G. Peters, P. Lingras, D. Ślezak, and Y. Yao, Eds. London, UK: Springer-Verlag, 2012, pp. 23–37.
[7] G. Peters, "Is there any need for rough clustering?" *Pattern Recognition Letters*, vol. 53, pp. 31–37, 2015.
[8] T. Denœux and M.-H. Masson, "EVCLUS: Evidential clustering of proximity data," *IEEE Trans. on Systems, Man and Cybernetics B*, vol. 34, no. 1, pp. 95–109, 2004.
[9] M.-H. Masson and T. Denoeux, "ECM: an evidential version of the fuzzy c-means algorithm," *Pattern Recognition*, vol. 41, no. 4, pp. 1384–1397, 2008.
[10] T. Denœux, S. Sriboonchitta, and O. Kanjanatarakul, "Evidential clustering of large dissimilarity data," *Knowledge-based Systems*, vol. 106, pp. 179–195, 2016.
[11] G. Shafer, *A mathematical theory of evidence*. Princeton, N.J.: Princeton University Press, 1976.
[12] L. Serir, E. Ramasso, and N. Zerhouni, "Evidential evolving Gustafson-Kessel algorithm for online data streams partitioning using belief function theory," *International Journal of Approximate Reasoning*, vol. 53, no. 5, pp. 747–768, 2012.
[13] N. Makni, N. Betrouni, and O. Colot, "Introducing spatial neighbourhood in evidential c-means for segmentation of multi-source images: Application to prostate multi-parametric MRI," *Information Fusion*, vol. 19, pp. 61–72, 2014.

[14] B. Lelandais, S. Ruan, T. Denœux, P. Vera, and I. Gardin, "Fusion of multi-tracer PET images for dose painting," *Medical Image Analysis*, vol. 18, no. 7, pp. 1247–1259, 2014.

[15] K. Zhou, A. Martin, Q. Pan, and Z.-G. Liu, "Median evidential c-means algorithm and its application to community detection," *Knowledge-Based Systems*, vol. 74, no. 0, pp. 69–88, 2015.

[16] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical Association*, vol. 66, no. 336, pp. 846–850, 1971.

[17] E. Hüllermeier, M. Rifqi, S. Henzgen, and R. Senge, "Comparing fuzzy partitions: A generalization of the rand index and related measures," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 3, pp. 546–556, 2012.

[18] D. T. Anderson, J. C. Bezdek, M. Popescu, and J. M. Keller, "Comparing fuzzy, probabilistic, and possibilistic partitions," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 5, pp. 906–918, 2010.

[19] A. Thalamuthu, I. Mukhopadhyay, X. Zheng, and G. Tseng, "Evaluation and comparison of gene clustering methods in microarray analysis," *Bioinformatics*, vol. 22, no. 19, pp. 2405–2412, 2006.

[20] D. S. V. Wong, F. K. Wong, and G. R. Wood, "A multi-stage approach to clustering and imputation of gene expression profiles," *Bioinformatics*, vol. 23, no. 8, pp. 998–1005, 2007.

[21] R. Campello, "A fuzzy extension of the rand index and other related indexes for clustering and classification assessment," *Pattern Recognition Letters*, vol. 28, no. 7, pp. 833–841, 2007.

[22] R. K. Brouwer, "Extending the rand, adjusted rand and jaccard indices to fuzzy partitions," *Journal of Intelligent Information Systems*, vol. 32, no. 3, pp. 213–235, 2009.

[23] M.-H. Masson and T. Denœux, "RECM: relational evidential c-means algorithm," *Pattern Recognition Letters*, vol. 30, pp. 1015–1026, 2009.

[24] T. Denœux, O. Kanjanatarakul, and S. Sriboonchitta, "EK-NNclus: a clustering procedure based on the evidential k-nearest neighbor rule," *Knowledge-based Systems*, vol. 88, pp. 57–69, 2015.

[25] T. Denœux, "A k-nearest neighbor classification rule based on Dempster-Shafer theory," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 25, no. 05, pp. 804–813, 1995.

[26] ——, *evclust: Evidential Clustering*, 2016, R package version 1.0.3. [Online]. Available: https://CRAN.R-project.org/package=evclust

[27] T. Denoeux and O. Kanjanatarakul, "Beyond fuzzy, possibilistic and rough: An investigation of belief functions in clustering," in *Soft Methods for Data Science (Proc. of the 8th International Conference on Soft Methods in Probability and Statistics SMPS 2016)*, ser. Advances in Intelligent and Soft Computing, vol. AISC 456. Rome, Italy: Springer-Verlag, September 2016, pp. 157–164.

[28] H. Frigui, C. Hwang, and F. C.-H. Rhee, "Clustering and aggregation of relational data with applications to image database categorization," *Pattern Recognition*, vol. 40, no. 11, pp. 3053–3068, 2007.

[29] A.-L. Jousselme and P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations," *International Journal of Approximate Reasoning*, vol. 53, no. 2, pp. 118–145, 2012.

[30] A.-L. Jousselme, D. Grenier, and E. Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, no. 2, pp. 91–101, 2001.

[31] T. Denœux, "Inner and outer approximation of belief structures using a hierarchical clustering approach," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 4, pp. 437–460, 2001.

[32] M. Bouchard, A.-L. Jousselme, and P.-E. Doré, "A proof for the positive definiteness of the jaccard index matrix," *International Journal of Approximate Reasoning*, vol. 54, no. 5, pp. 615–626, 2013.

[33] P. M. Murphy and D. W. Aha, *UCI Repository of machine learning databases [Machine-readable data repository]*. Irvine, CA: University of California, Department of Information and Computer Science., 1994.

[34] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2016. [Online]. Available: https://www.R-project.org/

[35] J. Bezdek, "Numerical taxonomy with fuzzy sets," *Journal of mathematical biology*, vol. 1, pp. 57–71, 1974.

[36] M. Ferraro and P. Giordani, "A toolbox for fuzzy clustering using the R programming language," *Fuzzy Sets and Systems*, vol. 279, pp. 1–16, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.fss.2015.05.001

[37] R. Krishnapuram, A. Joshi, O. Nasraoui, and L. Yi, "Low-complexity fuzzy relational clustering algorithms for web mining," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 595–607, 2001.

[38] G. Peters, "Rough clustering utilizing the principle of indifference," *Information Sciences*, vol. 277, pp. 358 – 374, 2014.

[39] ——, *SoftClustering: Soft Clustering Algorithms*, 2015, R package version 1.1502. [Online]. Available: https://CRAN.R-project.org/package=SoftClustering

[40] C. Fraley, A. E. Raftery, T. B. Murphy, and L. Scrucca, "mclust version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation," Department of Statistics, University of Washington, Tech. Rep. 597, 2012.

[41] T. F. Cox and M. A. Cox, *Multidimensional scaling*. London: Chapman and Hall, 1994.

[42] D. Dubois and H. Prade, "A note on measures of specificity for fuzzy sets," *International Journal of General Systems*, vol. 10, no. 4, pp. 279–283, 1985.

[43] A. Ramer, "Uniqueness of information measure in the theory of evidence," *Fuzzy Sets and Systems*, vol. 24, pp. 183–196, 1987.

[44] G. J. Klir and M. J. Wierman, *Uncertainty-Based Information. Elements of Generalized Information Theory*. New-York: Springer-Verlag, 1999.

[45] P. Smets, "The combination of evidence in the Transferable Belief Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 5, pp. 447–458, 1990.

Thierry Denœux Thierry Denoeux graduated in 1985 as an engineer from the Ecole des Ponts Paris-Tech in Paris, and received a doctorate from the same institution in 1989. Currently, he is Full Professor (Exceptional Class) with the Department of Information Processing Engineering at the Université de Technologie de Compiègne (UTC), France, and the scientific coordinator of the Laboratory of Excellence ontrol of Technological Systems of Systems (Labex MS2T). His research interests concern the management of uncertainty in intelligent systems. His main contributions are in the theory of belief functions with applications to pattern recognition, data mining and information fusion. He is the Editor-in-Chief of the *International Journal of Approximate Reasoning*, and an Associate Editor of several journals including *Fuzzy Sets and Systems* and *International Journal of Uncertainty, Fuziness and Knowledge-Based Systems*.

Shoumei Li Shoumei Li is a full professor at College of Applied Sciences, Beijing University of Technology (BJUT), China. She received her M.S. in Mathematics at Hebei University, China, in 1988, and then became an assistant professor at the same university. From 1991 to 1994, she was a lecture at BJUT. She got her Ph.D. in Probability at Saga University, Japan, in 1998 and then worked there as an associated professor. In 2000, she was back to BJUT to work as a professor. Her research interests focus on the theory of set-valued random variables and stochastic processes, nonadditive measures, nonlinear expectations, option pricing and portfolio.

Songsak Sriboonchitta Songsak Sriboonchitta received Bachelor and Master Degrees in Economics from Thammasat University, Thailand in 1972 and 1975 respectively and received Ph.D from University of Illinois at Urbana-Champaign in 1983. He has been working at the Faculty of Economics,Chiang Mai University since 1976. He was the dean of the Faculty of Economics,Chiang Mai University during 2004-2008. Presently he is Professor of Economics, Chiang Mai University, President of Thailand Econometric Society and also Director of Centre of Excellence in Econometrics. He has published one book and five edited books and more than 100 papers. His current research interest is econometrics.