A Distributional Approach for Soft Clustering Comparison and Evaluation

Andrea Campagner\textsuperscript{1}, Davide Ciucci\textsuperscript{1}, and Thierry Denœux\textsuperscript{2,3}

\textsuperscript{1} University of Milano-Bicocca, viale Sarca 336 – 20126 Milano, Italy
\textsuperscript{2} Université de technologie de Compiègne, CNRS, UMR 7253 Heudiasyc, Compiègne, France
\textsuperscript{3} Institut universitaire de France, Paris, France

Abstract. The development of external evaluation criteria for soft clustering (SC) has received limited attention: existing methods do not provide a general approach to extend comparison measures to SC, and are unable to account for the uncertainty represented in the results of SC algorithms. In this article, we propose a general method to address these limitations, grounding on a novel interpretation of SC as distributions over hard clusterings, which we call distributional measures. We provide an in-depth study of complexity- and metric-theoretic properties of the proposed approach, and we describe approximation techniques that can make the calculations tractable. Finally, we illustrate our approach through a simple but illustrative experiment.

Keywords: Soft clustering, Evidential clustering, External validation

1 Introduction

External clustering evaluation, defined as the act of objectively assessing the quality of a clustering result by means of a comparison between two or more clusterings (one of which is usually assumed to be the correct one), is one of the most relevant steps in clustering analysis [19]. In the case of hard clustering (HC), where each object is unambiguously assigned to a single cluster, several criteria have been considered in the literature [5, 17, 18].

By contrast, how to properly evaluate the results of a clustering analysis is much less clear in the case of soft clustering (SC) methods. Several SC methods have been proposed, including rough clustering (RC) [14], fuzzy clustering (FC) [2], possibilistic clustering (PC) [13] and evidential clustering (EC) [9, 7]. The development of evaluation measures for SC has largely focused on the extension of common measures [1], notably the Rand index, to the setting of FC [4, 11, 12], while only recently a formulation of this approach has been introduced for the more general case of EC [8]. Nonetheless, a general approach to extend other comparison measures to SC is still lacking. Furthermore, existing measures fail to properly distinguish and quantify different types of uncertainty that can arise in SC [8], namely: ambiguity, i.e., the inability to uniquely assign an object to a
single clustering (typical of RC); and partial assignment, i.e., the assignment of objects to multiple clusters (typical of FC and PC).

In this article, we propose a general method to address these limitations, which makes it possible to extend any clustering evaluation to the case of SC. This approach allows us providing a full account of the uncertainty in the two clusterings to be compared. It relies on a novel interpretation of SC as representing distributions over HCs, referred to as distributional measures. We provide an in-depth study of the proposed approach, with respect to both computational complexity and metric properties. Furthermore, we describe approximation techniques that can make the approach tractable. Finally, we illustrate the application of the proposed method through a simple but illustrative example involving commonly adopted SC algorithms.

2 Background and Related Work

In the following section we provide basic background on clustering (Section 2.1) and evaluation measures for SC (Section 2.2).

2.1 Background on Clustering

Let \( X = \{x_1, ..., x_n\} \) be a set of objects. A HC is an unique assignment of objects in \( X \) to clusters. Formally, a HC can be represented as a mapping \( C : X \rightarrow \Omega \), where \( \Omega = \{\omega_1, ..., \omega_k\} \) is a set of clusters. This representation is called object-based. An equivalent representation, called relational representation, can be obtained by defining a clustering as an equivalence relation \( \mathcal{C} \subseteq X \times X \). Clearly, for the case of HC the two representations are equivalent. Given a clustering \( C \), \( [C] \) denotes its relational representation. Given two clusterings \( C_1, C_2 \) we say that they are equivalent iff \( [C_1] = [C_2] \); we then write \( C_1 \sim C_2 \).

As mentioned in the introduction, in SC the unique assignment assumption is relaxed: the intuition is that we allow uncertainty in the cluster assignments. In the most general framework of EC, the uncertainty about cluster assignment is represented as a Dempster-Shafer mass function. Formally, using the object-based representation, an EC is a set \( M = \{m_x\}_{x \in X} \), where each \( m_x \) is a mass function; i.e., a function \( m_x : 2^\Omega \rightarrow [0, 1] \) such that \( \sum_{A \subseteq \Omega} m_x(A) = 1 \). If the mass functions \( m_x \) are logical, then the collection \( R = \{m_x\}_{x \in X} \) is said to be a RC. A RC can be seen equivalently as a set of HCs [3]. Namely, a HC \( C \) is compatible with \( R \) if \( \forall x \in X, C(x) \in R(x) \). Then, \( R \) can be represented by \( C(R) = \{C : C \text{ is compatible with } R\} \). If all \( m_x \) are Bayesian, then the collection \( F = \{m_x\}_{x \in X} \) is a FC. Finally, if all \( m_x \) are consonant, then the collection \( P = \{m_x\}_{x \in X} \) is a PC. Both FC and PC can be alternatively represented as a collection of cluster membership vectors \( F = \{\mu_x\}_{x \in X} \). In PC it is assumed that \( \forall x \in X, \max_{\omega \in \Omega} \mu_x(\omega) \leq 1 \), while in FC that \( \forall x \in X, \sum_{\omega \in \Omega} \mu_x(\omega) = 1 \).

\[4\] Due to space constraints, the complete version of all proofs appears online at https://arxiv.org/abs/2206.09827.
A relational representation can also be defined for the case of EC. Let \( \Theta = \{s, \neg s\} \) be the frame where \( s \) denotes that two objects are in the same cluster, and \( \neg s \) the opposite event. Given an EC \( M \), the corresponding relational representation can be obtained, for any two objects \( x, y \in X \), by combining \( m_x, m_y \) using Dempster’s rule of combination and then computing the restriction \( m^{xy} \) of the resulting mass function to \( \Theta \) [8].

Finally, we note that, if we interpret a SC as describing the uncertainty in regard to an underlying (unknown) HC, then two types of uncertainty can be distinguished. First, partial assignment, i.e., the fact that for \( \omega_1, \omega_2 \in \Omega \) it may happen that \( m_x(\omega_1), m_x(\omega_2) > 0 \). Second, ambiguity, i.e., the assignment of some mass to non-singleton events, describing our inability to exactly determine to which cluster an object belongs. It is easy to observe that in a FC only partial assignment is relevant, since all the mass is assigned to the singletons, while in the case of RC, only ambiguity is present. By contrast, the EC formalism is flexible enough to represent both types of uncertainty.

### 2.2 Clustering Comparison Measures

Several measures have been defined to compare clusterings. Given two HCs \( C_1, C_2 \), a commonly adopted approach to compare them is to evaluate the number of object pairs \( x, y \in X \) on which they agree. Formally, the Rand index can be defined as:

\[
\text{Rand}(C_1, C_2) = \frac{|\{(x, y) \in X^2 : (x, y) \in ([C_1] \cap [C_2]) \cup ([C_1]^c \cap [C_2]^c)\}|}{|X|^2}.
\]

(1)

It is easy to show that the Rand index is a similarity on HCs.

Several extensions of the Rand index to the setting of SC have been considered. In the case of FC and PC, Campello et al. [4] and Frigui et al. [11] proposed to use a t-norm \( \land \) and a t-conorm \( \lor \) in place of classical set operators. This setting was later generalized to the setting of RC [10], by means of a transformation from RC into FC. A further generalization of the Rand index was proposed by Hüllermeier et al. [12], in the setting of FC, and Denœux et al. [8], in the more general setting of EC. This approach is based on the use of a normalized metric \( d_M \) on mass functions, that is then extended to compare ECs as

\[
\text{Rand}_E(M_1, M_2) = \frac{2}{n(n - 1)} \sum_{x \neq y \in X} 1 - d_M(m_1^{xy}, m_2^{xy}).
\]

(2)

The authors of [8, 12] note that, while the measure they propose is a pseudo-similarity, it is not completely satisfactory for the comparison of ECs as it does not distinguish between partial assignment and ambiguity.

Another commonly adopted approach for the definition of clustering evaluation measure grounds on information theory. Let \( C_1, C_2 \) two HCs. The mutual information is defined as

\[
MI(C_1, C_2) = \sum_{\omega_1 \in \Omega_1} \sum_{\omega_2 \in \Omega_2} p_{ij} \log \frac{p_{ij}}{p_i \cdot p_j},
\]

(3)
where $p^1_i = |\{x \in X : C_1(x) = \omega^1_i\}|/|X|$, and similarly for each $\omega^2_i \in \Omega_2$, while $p_{ij} = |\{x \in X : C_1(x) = \omega^1_i \text{ and } C_2(x) = \omega^2_j\}|/|X|$. The mutual information is a similarity over HCs. This measure has been extended to the case of RC in [3], by representing RCs as collections of compatible HCs. To our knowledge, no proposal to extend these metrics to the more general case of EC can be found in the literature.

Finally, another comparison approach grounds on the notion of edit distance between partitions. The partition distance [5] which, for two HCs $C_1, C_2$, is defined as the minimum number of objects to be moved to transform $C_1$ into $C_2$ (or, equivalently, $C_2$ into $C_1$), can be computed as

$$d_\pi(C_1, C_2) = \frac{1}{|X| - 1} \min_w \frac{1}{2} \sum_i |\omega^1_i \Delta \omega^2_{w(i)}|, \quad (4)$$

where $w$ is a permutation function, and $\Delta$ is the symmetric difference operator. It is easy to note that $d_\pi$ is a normalized metric. An extension of the partition distance to the case of FC was proposed by Zhou [20]. The obtained measure is a proper generalization of the partition distance and is a metric. As for the mutual information, to our knowledge, the extension of the partition distance to rough and EC has not been considered in the literature.

3 A General Framework for Soft Clustering Evaluation Measures

As shown in the previous section, most of the research on clustering comparison measures for SC has focused on the analysis of some specific indices, while a general methodology for obtaining such measures is still missing. Furthermore, as noted in [12, 8], most of the existing methods fail to satisfy reasonable metric properties and can thus hardly be used for the objective comparison of SCs. Notably, also the more principled approaches [8] can have some drawbacks, such as the inability to properly distinguish between different types of uncertainty arising in SC. In this section, we propose an approach that attempts to address these limitations, based on the representation of a SC as a distribution over HCs.

3.1 Distribution-based Representation of Soft Clustering

As shown in Section 2.1, a RC $R$ can be represented as a set $C(R)$ of HCs $C$. Based on this observation, we extend this representation to general SCs. Formally, given an EC $M$, we consider the following probability distribution over RCs:

$$m_M(R) = \prod_{x \in X} m_x(R(x)), \quad (5)$$

which can also be seen as a Dempster-Shafer mass function over HCs. Given an EC $M$ and its distribution-based representation $m_M$, we denote with $\mathcal{F}(M)$ the collection of focal RCs of $m_M$, that is $\mathcal{F}(M) = \{R : m_M(R) > 0\}$. 
The distribution-based representation for RC, FC and PC can then be obtained as a special case of Eq (5). Indeed, in the case of RC, \( m_M \) is logical (i.e. \( \mathcal{F}(m_M) = 1 \)). In the case of a FC \( F = \{ \mu_x \}_{x} \), where \( \mu_x : C \rightarrow [0, 1] \) is a probability distribution, the focal RCs are all singletons (i.e., HCs), thus we can define \( \text{Pr}_F(C) = \prod_{x \in X} \mu_x(C(x)) \). Finally, given a PC and a t-norm \( \wedge \), we can view \( P \) as a possibility distribution over HCs \( \text{Poss}_P(C) = \bigwedge_{x \in X} \mu_x(C(x)) \). The possibility distribution \( \text{Poss}_P \) can equivalently be represented as a consonant mass function, i.e., a mass function for which the focal RCs are nested. Note that, when \( \wedge \) is the product t-norm, we recover the case of FC.

### 3.2 Distributional Measures

Let \( d \) be a normalized metric on HCs. Since, as shown in the previous section, any SC can be seen as a distribution over HCs, an intuitive approach would be to extend \( d \) to a distribution-valued function, providing a quantification of the belief about the real value of the evaluation measure. The intuition behind this idea is based on the definition of SC as representing a clustering with some uncertainty affecting our knowledge with respect to the assignment of objects to clusters. Thus, it is natural to require that an evaluation measure for SC should transfer this uncertainty to the possible outcomes of the evaluation.

Therefore, intuitively, a measure over RCs would provide, given two RCs \( R_1, R_2 \), a set of values, representing all possible distances between HCs compatible with \( R_1, R_2 \). More generally, a measure over ECs would provide a mass function over possible values of \( d \). Formally, we define the distributional measure, based on \( d \), between two RCs (resp., PCs, ECs) as, respectively:

\[
\forall v \in \mathbb{R}, \ d_P(P_1, P_2)(v) = \bigvee_{C_1, C_2 : d(C_1, C_2) = v} \text{Poss}_{P_1}(C_1) \wedge \text{Poss}_{P_2}(C_2)
\]

\[
\forall V \in \mathbb{R}^2, \ d_E(M_1, M_2)(V) = \sum_{R_1, R_2 : d_R(R_1, R_2) = V} m_{M_1}(R_1) \cdot m_{M_2}(R_2)
\]

where \( \wedge, \bigvee \) are a t-norm and the corresponding dual s-conorm. It is easy to observe that \( d_E \) is a generalization of \( d_R \) and \( d_P \). For ease of notation, we distinguish the case of \( d_P \) where \( \wedge = \otimes_P, \bigvee = \oplus_P \) (resp., the product t-norm and the corresponding t-conorm) and we denote it as \( d_E \), since it can be applied directly to the case of FC. Intuitively, \( d_E(M_1, M_2) = V \) can be interpreted as the evidence supporting the statement “The distance between the two real HCs underlying \( M_1, M_2 \) is within \( V \)”.

Therefore, \( d_E \) provides a complete representation of the possible distance values that arise when comparing HCs compatible with \( M_1, M_2 \). Though, clearly, \( d_E \) is not a metric, we can see that it satisfies the properties stated in the following theorem.

**Theorem 1.** Function \( d_E \) satisfies (M3) (see Appendix A\(^5\)). \( d_E(M_1, M_2)(0) = 1 \) iff there exists a HC \( C \) s.t. for \( i = 1, 2 \), \( R \in \mathcal{F}(M_i) \implies \forall C' \in C(R), C' \sim C. \)

\(^5\) https://arxiv.org/abs/2206.09827
As a consequence of the previous result, the value $d_E(M_1, M_2)(0)$ can be interpreted as the evidence that the unknown HC$s$ corresponding to $M_1$ and $M_2$ are the same (have a distance equal to 0). Indeed, $d_E$ assigns full evidence to value 0, if and only if $M_1, M_2$ are totally compatible. In particular, simple equality between $M_1, M_2$ does not suffice to obtain $d_E(M_1, M_2)(0) = 1$, unless $M_1, M_2$ are HC$s$.

In regard to computational complexity, it is easy to show that computing $d_R$ (resp., $d_F, d_P, d_E$) is computationally easy w.r.t. the size of the distribution-based representation introduced in the previous section, while it is intractable w.r.t. the size of the object-based representation:

**Theorem 2.** The problem of computing $d_R$ (resp., $d_F, d_P, d_E$) has complexity $O(k^m)$, where $m = |\{x \in X : |R(x)| \neq 1\}|$ and $k = |\Omega|$ is the number of clusters. More precisely, $d_R$ can be computed in constant amortized time, while $d_F, d_P, d_E$ can be computed in at most linear amortized time.

**Interval representation** A possible solution to the intractability of computing the distributional measures would be to consider a compact representation of the latter. For the case of RC, $d_R$ could be summarized as the interval:

$$\langle d_R^l, d_R^u \rangle(R_1, R_2) = \langle \min\{v \in \mathbb{R} : v \in d_R(R_1, R_2)\}, \max\{v \in \mathbb{R} : v \in d_R(R_1, R_2)\}\rangle.$$ 

We note that this definition satisfies the following properties:

**Proposition 1.** Let $R_1, R_2$ be two RCs. Then, $1 - d_R^l$ is a consistency: in particular $d_R^l = 0$ iff $C(R_1) \cap C(R_2) \neq \emptyset$. By contrast, $d_R^u$ is a meta-metric: in particular, $d_R^u = 0$ iff $R_1 = R_2$ and $|C(R_1)| = 1$, that is iff $R_1 = R_2$ is a HC.

**Corollary 1.** $d_R^u$ is a metric iff either $R_1, R_2$ is a HC.

As a result of the previous corollary, in the special case where the aim is to evaluate a RC $R$ with respect to a HC $C$ representing the ground truth, then $d_R^u$ is guaranteed to be a metric. Nonetheless, it is easy to observe that computing $\langle d_R^l, d_R^u \rangle$ is still computationally hard:

**Theorem 3.** Let $R_1, R_2$ be two RCs, represented through the object-based representation. Then, the problem of computing $\langle d_R^l, d_R^u \rangle$ is NP-HARD\(^6\).

For the case of FC, PC and, more generally, EC, a possible approach to obtain a similar summarization would be to apply a decision rule to transform the distribution-valued $d_F, d_P, d_E$ into simpler indices \([6]\). An example of this approach would be to compute the following lower and upper expectations:

$$E(d_E)(M_1, M_2) = \sum_{V \subseteq 2^\mathbb{R}} d_E(M_1, M_2)(V) \min_{d \in V} d = E(d_R^l), \quad (9)$$

$$E(d_E)(M_1, M_2) = \sum_{V \subseteq 2^\mathbb{R}} d_E(M_1, M_2)(V) \max_{d \in V} d = E(d_R^u). \quad (10)$$

\(^6\) The problem is trivially in P w.r.t. the distribution-based representation of $R_1, R_2$.  

6
If $M_1, M_2$ are two FCs we obtain that $E(d_E) = E(d_E) = E(d_F)$. Similarly to the case of RC, it is easy to show that the following properties hold:

**Theorem 4.** Let $M_1, M_2$ be two ECs. Then $E(d_E)$ is a meta-metric. Furthermore, $E(d_E)$ satisfies only (M1b) and (M2) (see Appendix A\(^7\)). In particular:
- If $M_1, M_2$ is a HC, then $1 - E(d_E)$ is a consistency and $E(d_E)$ is a metric;
- If $F_1, F_2$ are two FCs, then $E(d_F)$ is a meta-metric.

From the computational complexity point of view, computing $E(d_E), E(d_E)$ is at least as hard as computing $⟨d\_R, d\_R⟩$. However, for the case of FC, it is easy to show that for certain base distances $d$, $E(d_F)$ can be computed efficiently:

**Proposition 2.** Let $d = 1−Rand$. Then, $E(d_F)$ can be computed in time $O(n^2)$.

We leave the problem of characterizing the general complexity of computing $E(d_E), E(d_E)$ as open problem.

### 3.3 Approximation Methods

In the previous section we proposed distributional measures as a general approach to extend any HC comparison measure to a SC comparison measure. Nonetheless, the computation of these distributional measures is, in general, intractable. For this reason, in this section, we introduce some approximation methods and algorithms, based on a sampling approach, which can be applied to any base distance between HCs.

We start with the case of the summarized representation of $d_R$, that is with $d\_R, d\_R\_R$. Given two RCs $R_1, R_2$, we draw $s$ samples $(C^1, C^2), \ldots, (C^1, C^2)$ uniformly from $C(R_1), C(R_2)$. Then, we can approximate $d\_R$ as, respectively, $d\_R = \min_{i \in \{1, \ldots, s\}} d(C^i, C^i)$ and $d\_R = \max_{i \in \{1, \ldots, s\}} d(C^i, C^i)$. Clearly, the following result holds:

**Proposition 3.** The following bounds hold for any $\epsilon > 0$:

\[
Pr(d\_R - d\_R > \epsilon) \leq F(d\_R - \epsilon), \quad Pr(d\_R - d\_R > \epsilon) \leq 1 - (1 - F(\epsilon - d\_R))^s
\]

(11)

where $F$ is the cumulative distribution function (CDF) of the probability distribution $p_R$ defined as $p_R(t) = \frac{|\{C_1 \in C(R_1), C_2 \in C(R_2) : d(C_1, C_2) = t\}|}{|d(R_1, R_2)|}$.

Since for each $\epsilon$, the quantity $F(d\_R - \epsilon)$ (resp., $F(\epsilon - d\_R)$) is strictly less than 1, it holds that $Pr(d\_R - d\_R > \epsilon)$ (resp., $Pr(d\_R - d\_R > \epsilon)$) decays exponentially w.r.t. the size of the sample $s$. However, we note that the quality of the previously described approximation method largely depends on $d_R$. In particular, the convergence in Eq (11) is influenced by the **tailedness** of $p_R$: the heavier the tails of $p_R$, the lower the approximation error.

For the case of FC, if we use the expected value $E(d_F)$ to summarize $d_F$ and we use a sampling procedure to estimate $E(d_F)$ as $d_F$ then we can obtain a tail bound by applying Hoeffding’s inequality:

\[7\] https://arxiv.org/abs/2206.09827
Proposition 4. Assume that $d$ is a normalized metric on HCs. Then:

$$Pr(|\hat{d}_F - E(d_F)| \geq \epsilon) \leq 2e^{-2s\epsilon^2}$$

(12)

Hence, the deviation between the empirical mean $\hat{d}_F$ and $E(d_F)$ has exponential decay in the size of the sample $s$.

Combining Eqs (11) and (12), a similar result holds also for $d_E$:

Proposition 5. Assume that $d$ is a normalized metric on HCs. Let $\hat{d}_E^l, \hat{d}_E^u$ be the sample estimates of $E(d_E), E(d_E)$. Then:

$$Pr(|\hat{d}_E^l - E(d_E)| \geq \epsilon) \leq 2e^{-2s\epsilon^2}, \quad Pr(|\hat{d}_E^u - E(d_E)| \geq \epsilon) \leq 2e^{-2s\epsilon^2}$$

(13)

Given two ECs $M_1, M_2$, the previous estimate requires that $\hat{d}_E^l, \hat{d}_E^u$ are computed by sampling pairs $R_1, R_2$ of RCs from the distributions $m_{M_1}, m_{M_2}$ and then computing the exact values of $d_R^l(R_1, R_2), d_R^u(R_1, R_2)$. As a consequence of Proposition 3, this may not be feasible when $|X|$ is large. In such cases, a possible solution would be to compute $\hat{d}_E^l, \hat{d}_E^u$ by means of nested sampling (i.e, first we sample a RC $R$ from $m_M$, then we sample a HC $C$ from $C(R)$). In this case, however, one should expect a larger approximation error. Finally, we note that all the above mentioned sampling-based approximation methods can easily be implemented in time complexity $O(n^2s)$.

4 Illustrative Experiment

In this section, we illustrate the use of the proposed metrics using the Anderson’s Iris dataset. This latter is a small-scale benchmark problem comprising 150 objects, four numerical features and three perfectly balanced classes. We selected this dataset as, the three classes being approximately linearly separable, it can be expected that any SC algorithm would give as output a clustering close to being an HC. As a consequence of Theorem 3, this is a necessary condition for efficient computation of the exact versions of the distributional measures.

We considered five different clustering algorithms, namely: k-means (KM), rough k-means (RKM) [16], fuzzy c-means (FCM) [2], possibilistic c-means (PCM) [13] and evidential c-means (ECM) [15]. In order to reduce the complexity of computing the distributional measures, we set the algorithm hyper-parameters to obtain clusterings close to being hard. In particular, for RKM we set $\epsilon = 1.1$, for FCM and PCM we set $m = 5$, and for ECM we set $\delta = 10, \beta = 5, \alpha = 5$. The output of each algorithm was compared with the ground truth labeling of the iris dataset. We considered, in particular, the distributional generalizations of the Rand index (D-RI) and the partition distance (D-PD), as well as their sampling-based approximations (S-RI, S-PD). Code was implemented in Python (v. 3.8.8), using scikit-learn (v. 0.24.1), numpy (v. 1.20.1) and scipy (v. 1.6.2).
The results of the experiment are reported in Table 1, in terms of the metrics values as well as running time (in seconds). As far as running time is concerned, we can observe that the cost of computing the exact versions of the proposed measures sharply increases when considering more general SC algorithms. Indeed, the running times of D-RI and D-PD for ECM were approximately twice those of FCM and PCM. On the other hand, the differences in running time for the approximation algorithms were much smaller, and indeed the running times for FCM, PCM and ECM were similar.

In terms of approximation quality, even though for RKM and FCM there were no differences between the approximated and exact results, this was not the case for PCM and ECM. In particular, we note that the sampling-based approximation algorithm systematically underestimated the uncertainty in clustering comparison results, by producing intervals that were narrower than the exact ones. Nonetheless, we note that the approximation methods provided results that were aligned with the exact ones, with smaller values according to one method associated to smaller values according to the other one.

5 Conclusion

In this article we proposed a general framework for extending clustering comparison measures from HC to EC (hence, as special cases, also to RC, FC and PC), that we called distributional measures. We studied the metric- and complexity-theoretic properties of the proposed measures and, since a major limitation of the proposed approaches lies in their high computational complexity, we also proposed some strategies for approximation based on sampling. Finally, we illustrated the application of the proposed methods through a simple experiment.

We believe that this article could provide a first step toward the development of general and principled approaches for the comparison of SC algorithms. For this reason, we deem the following problems to be worthy of further investigation: 1) Generalizing Proposition 2 to other base distance measures, and determining whether this result can be extended to EC; 2) Designing more refined sampling approaches that can be used to correct the uncertainty underestimation that we observed in the experiments.

Table 1: Results of the experiment. For the Rand index higher is better, while for the partition distance lower is better.

| Metric | KM | RKM | FCM | PCM | ECM |
|--------|----|-----|-----|-----|-----|
| D-RI   | 0.877 (0.034s) | (0.874, 0.886) (0.802s) | 0.876 (784.388s) | (0.839, 0.941) (979.053s) | (0.781, 0.944) (1394.69s) |
| S-RI   | - | (0.874, 0.886) (0.429s) | 0.876 (11.266s) | (0.860, 0.927) (19.848s) | (0.681, 0.819) (19.845s) |
| D-PD   | 0.111 (0.031s) | (0.099, 0.113) (0.803s) | 0.112 (184.707s) | (0.033, 0.122) (224.31s) | (0.041, 0.229) (431.57s) |
| S-PD   | - | (0.100, 0.113) (0.202s) | 0.112 (11.391s) | (0.072, 0.103) (13.739) | (0.154, 0.209) (16.424s) |
References

1. Anderson, D.T., Bezdek, J.C., Popescu, M., et al.: Comparing fuzzy, probabilistic, and possibilistic partitions. IEEE Trans Fuzzy Syst 18(5), 906–918 (2010)
2. Bezdek, J.C.: Pattern recognition with fuzzy objective function algorithms. Springer (1981)
3. Campagner, A., Ciucci, D.: Orthopartitions and soft clustering: soft mutual information measures for clustering validation. Knowl Based Syst 180, 51–61 (2019)
4. Campello, R.J.: A fuzzy extension of the rand index and other related indexes for clustering and classification assessment. Pattern Recognit Lett 28(7), 833–841 (2007)
5. Day, W.H.: The complexity of computing metric distances between partitions. Math Soc Sci 1(3), 269–287 (1981)
6. Denoeux, T.: Decision-making with belief functions: A review. Int J Approx Reason 109, 87–110 (2019)
7. Denoeux, T., Kanjanatarakul, O.: Evidential clustering: A review. In: Huynh, V.N., Imiguchi, M., Le, B., Le, B.N., Denoeux, T. (eds.) Integrated Uncertainty in Knowledge Modelling and Decision Making. pp. 24–35. Springer International Publishing, Cham (2016)
8. Denoeux, T., Li, S., Sriboonchitta, S.: Evaluating and comparing soft partitions: An approach based on Dempster–Shafer theory. IEEE Trans Fuzzy Syst 26(3), 1231–1244 (2017)
9. Denoeux, T., Masson, M.H.: EVCLUS: evidential clustering of proximity data. IEEE Trans Syst Man Cybern B Cybern 34(1), 95–109 (2004)
10. Depaolini, M.R., Ciucci, D., Calegari, S., et al.: External indices for rough clustering. In: Lect Notes Comput Sci, vol 11103. pp. 378–391. Springer (2018)
11. Frigui, H., Hwang, C., Rhee, F.C.H.: Clustering and aggregation of relational data with applications to image database categorization. Pattern Recognit 40(11), 3053–3068 (2007)
12. Hüllermeier, E., Rifqi, M., Henzgen, S., et al.: Comparing fuzzy partitions: A generalization of the Rand index and related measures. IEEE Trans Fuzzy Syst 20(3), 546–556 (2011)
13. Krishnapuram, R., Keller, J.M.: A possibilistic approach to clustering. IEEE Trans Fuzzy Syst 1(2), 98–110 (1993)
14. Lingras, P., West, C.: Interval set clustering of web users with rough k-means. J Intell Inform Syst 23(1), 5–16 (2004)
15. Masson, M.H., Denoeux, T.: ECM: an evidential version of the fuzzy c-means algorithm. Pattern Recognition 41(4), 1384–1397 (2008)
16. Peters, G.: Rough clustering utilizing the principle of indifference. Information Sciences 277, 358–374 (2014)
17. Rand, W.M.: Objective criteria for the evaluation of clustering methods. J Am Stat Assoc 66(336), 846–850 (1971)
18. Vinh, N.X., Epps, J., Bailey, J.: Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. J Mach Learn Res 11, 2837–2854 (2010)
19. Xiong, H., Li, Z.: Clustering validation measures. In: Data Clustering, pp. 571–606. Chapman and Hall/CRC (2018)
20. Zhou, D., Li, J., Zhu, H.: A new Mallows distance based metric for comparing clusterings. In: Proceeding of ICML 2005. pp. 1028–1035 (2005)