A Cluster-Validity Index Combining an Overlap Measure and a Separation Measure Based on Fuzzy-Aggregation Operators

Hoel Le Capitaine, Carl Frélicot

A cluster validity index combining an overlap measure and a separation measure based on fuzzy aggregation operators

Hoel Le Capitaine, Carl Frélicot
Mathematics, Image and Applications Laboratory, University of La Rochelle,
17000 La Rochelle, FRANCE
hoel.le_capitaine@univ-lr.fr & carl.frelicot@univ-lr.fr

***Abstract*—Since a clustering algorithm can produce as many partitions as desired, one needs to assess their quality in order to select the partition that most represents the structure in the data if there is any. This is the rationale for the cluster validity problem and indices. This paper presents a cluster validity index that helps to find the optimal number of clusters of data from partitions generated by a fuzzy clustering algorithm such as the Fuzzy C-Means (FCM) or its derivatives. Given a fuzzy partition, this new index uses a measure of multiple clusters overlap and a separation measure for each data point, both based on an aggregation operation of membership degrees. Experimental results on artificial and benchmark data sets are given to demonstrate the performance of the proposed index as compared to traditional and recent indices.**

***Index Terms*—Cluster validity, fuzzy cluster analysis, aggregation operators, triangular norms.**

I. Introduction

Clustering is the unsupervised classification of data points or *patterns* into groups or *clusters*, such that patterns in the same group are similar to each other while patterns in different groups are dissimilar [1]. An alternative to traditional crisp clustering methods that generate partitions where each pattern belongs to only one cluster is fuzzy clustering where each pattern is associated with every cluster to various membership degrees [2]. Such methods are less sensitive to local minimum problems than crisp ones because of the fuzzy updating at each iteration. Unfortunately, fuzzy methods as well as crisp ones are not robust to the existence of isolated points (*outliers*) and noisy data for which other approaches have been proposed such as possibilistic clustering [3]. In this work, we are interested in fuzzy clustering and we use the Fuzzy C-Means (FCM) partitioning method introduced by Bezdek [2] because it is the most widely used fuzzy clustering algorithm.

The main drawback of clustering methods is that they require the user to specify the number, c, of clusters which the user does not usually know or may not want to specify. A resulting fuzzy c-partition has to be validated because its quality highly depends on this parameter. A very challenging problem in cluster analysis, called the *Cluster Validity* (CV) problem in the literature, consists of finding the optimal value for c [4]. One may compare the resulting partition to a reference one obtained from background knowledge, using a different algorithm, or the same algorithm with a different specified c value. Such comparisons can be done using so-called *relative* indices, as in [5], [6], [7]. An alternative approach is to select the most appropriate number of clusters, given a particular clustering algorithm, based on so-called *internal* indices, referred hereafter to as *Cluster Validity Indices* (CVIs). In addition, clustering algorithms always produce a partition even if there is no cluster structure. This checking step, called *Cluster Tendency*, is done prior to CV, and is outside the scope of this paper. It has received more attention in recent years, *e.g.* in [8], [9], [10], mainly by use of re-ordering similarity/distance matrices.

Many CVIs have been proposed for this purpose in the last three decades and their number has been increasing in recent years, (see [11], [12] for reviews). Historically CVIs only use partial membership degrees, *e.g.* [13], [14]. They are easy to compute, well-adapted to situations where clusters overlap with each other, but suffer from a monotonic tendency with respect to c. Another widely reported problem is that such indices may not have any relation to the geometrical structure of the data, *e.g.* the distances between the patterns of similar and different clusters. More recently developed CVIs are based on compactness and/or separation measures that simultaneously use membership degrees and clusters' centroids, *e.g.* [15], [16], [17], [18], [19], [20], [21]. They are less monotonic with respect to c but are more difficult to compute and not as efficient in case of overlapping clusters. Furthermore, the way compactness and separation measures are computed does not allow for distinguishing numerous different situations [22]. Some reported problems of existing CVIs, whatever the category they belong to, are worsened by the inability of the underlying fuzzy clustering algorithm to deal with noisy points and outliers. Noisy points cause cluster-overlap by building bridges between separated clusters, and outliers may result in singleton clusters, such that the natural number of clusters can not be correctly assessed. Examples of such difficult situations are illustrated in Figure 2-(c) to -(f). Fuzzy modeling approaches have been proposed (see [23]), but we restrict ourselves to commonly used indices involving compactness, separability and/or overlap measures.

In this paper, we present a new CVI for fuzzy clustering which aims to overcome most of the well-known difficulties discussed above. It consists of the average value of the ratio of two measures: an *overlap* measure and a *separation* measure. Both measures are based on an aggregation operator which combines triangular norms applied to membership degrees.

The rest of the paper is organized as follows. Section II briefly describes the FCM fuzzy clustering algorithm, and recalls some traditional and recent CVIs that will be used for comparison. In section III, the necessary background on aggregation operators is given and the new CVI is presented. We discuss its properties and illustrate its behavior on a simple example. Experimental results on synthetic and real data sets are provided in section IV and concluding remarks are given in section V.

II. Cluster validity indices for fuzzy clustering

In this section, we first recall the well-known FCM fuzzy clustering algorithm. We then give nine previous CVIs that will be used in the experiments, described in section IV.

A. The fuzzy c-means algorithm

Let $X = \{x_1, \cdots, x_n\}$ be a n point data set in a p-dimensional feature space, $\mathbb{R}^p$, with the usual Euclidean norm $||.||$. The fuzzy c-means (FCM) algorithm partitions X into $c > 1$ clusters by minimizing the following objective function [2]:

$$J_m(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m \, ||\mathbf{x}_k - \mathbf{v}_i||^2 \qquad (1)$$

TABLE I
OTHER CLUSTER VALIDITY INDICES ACCORDING TO THE LITERATURE.

Name & Reference	Index	Search for
Normalized Partition Entropy [13]	$NPE(U,c) = \frac{-\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}\log(u_{ik})}{n-c}$	min
Normalized Partition Coefficient [14]	$NPC(U,c) = \frac{\frac{c}{n}\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^2 - 1}{c-1}$	max
Fukuyama & Sugeno [15]	$FS(U,V,X,c) = J_m(U,V) - \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m \, \|\|\mathbf{v}_i - \overline{\mathbf{v}}\|\|^2$	min
Fuzzy Hypervolume [16]	$FHV(U,V,X,c) = \sum_{i=1}^{c} \sqrt{\det\left(\frac{\sum_{k=1}^{n} u_{ik}^m (\mathbf{x}_k - \mathbf{v}_i)(\mathbf{x}_k - \mathbf{v}_i)^T}{\sum_{k=1}^{n} u_{ik}^m}\right)}$	min
Xie & Beni [17]	$XB(U,V,X,c) = \frac{J_m(U,V)/n}{\min_{i,j=1,c;j\neq i} \|\|\mathbf{v}_i - \mathbf{v}_j\|\|^2}$	min
Bensaid et al. [18]	$SC(U,V,X,c) = \sum_{i=1}^{c} \frac{\sum_{k=1}^{n} u_{ik}^m \|\|\mathbf{x}_k - \mathbf{v}_i\|\|_A^2}{\sum_{k=1}^{n} u_{ik} \cdot \sum_{j=1}^{c} \|\|\mathbf{v}_i - \mathbf{v}_j\|\|_A^2}$	min
Kwon [19]	$K(U,V,X,c) = \frac{J_m(U,V) + \frac{1}{c}\sum_{i=1}^{c} \|\|\mathbf{v}_i - \overline{\mathbf{v}}\|\|^2}{\min_{i,j=1,c;j\neq i} \|\|\mathbf{v}_i - \mathbf{v}_j\|\|^2}$	min
Pakhira et al. [20]	$PBM(U,V,X,c) = \left(\frac{1}{c} \times \frac{\sum_{k=1}^{n} \|\|\mathbf{x}_k - \overline{\mathbf{v}}\|\|}{J_m(U,V)} \times \max_{i,j=1}^{c} \|\|\mathbf{v}_j - \mathbf{v}_i\|\|\right)^2$	max
Wu & Yang [21]	$WY(U,V,c) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}^2 / \min_{1\leq i\leq c}\left(\sum_{k=1}^{n} u_{ik}^2\right)$ $- \sum_{i=1}^{c} \exp\left(-\min_{j\neq i}\left(\|\|\mathbf{v}_i - \mathbf{v}_j\|\|^2 / \sum_{i=1}^{c} \|\|\mathbf{v}_i - \overline{\mathbf{v}}\|\|^2/c\right)\right)$	max

where u_{ik} is the membership degree of $\mathbf{x}_k$ in the i^{th} cluster represented by its centroid $\mathbf{v}_i \in \mathbb{R}^p$. Centroids are gathered into a $(p \times c)$ matrix $V = [\mathbf{v}_1, ..., \mathbf{v}_c]$. Degrees u_{ik} are subject to $\sum_{i=1}^{c} u_{ik} = 1$ for all $\mathbf{x}_k$ in X and to $0 < \sum_{k=1}^{n} u_{ik} < n$ $(\forall i = 1, c)$, and are elements of the fuzzy $c-$partition matrix $U = [\mathbf{u}_1, ..., \mathbf{u}_n]$ of size $(c \times n)$. The so-called *fuzzifier* $m > 1$ is a weighting exponent which makes the resulting partition more or less fuzzy [11]. The higher m, the more fuzzy the clusters' boundaries. Given X and c, minimization of (1) is obtained by alternatively updating (U, V) (see [4] p.17 for details). The Euclidean distance used in FCM induces hyperspherical clusters with similar numbers of points. Thus, this partitioning is not well suited to every possible situation, *e.g.*: bridges, outliers, additional noisy points. CV is then a more challenging problem when using FCM instead of other algorithms that behave better in such situations.

B. Cluster validity and previous cluster validity indices

Validating the provided clustering (U, V) of X consists of assessing whether the resulting partition reflects the structure in the data or not. Due to the unsupervised aspect of the method, the user does not have any prior knowledge of the structure of the data and c is a user-defined parameter of clustering algorithms such as FCM. Most of the work on CV focuses on the "optimal number of clusters" problem and many CVIs have been proposed. Given a CVI, the procedure to automatically select the optimal number of clusters c_{best} consists of running the FCM algorithm with c varying in a user-defined range $[c_{min}, c_{max}]$, computing $CVI(c)$ for each partition produced, and selecting c_{best} such that CVI(c_{best}) is optimal within the predefined range.

Many CVIs have been proposed in the last three decades with their number increasing in recent years. It is not practical to review all of them. Nine indices are summarized in Table I, some of them being the most frequently referred to in the literature and some more recent that address the drawbacks of the former ones. We invite the interested reader to refer to review papers [24], [25], [4], [12], [26] and individual references for details on each CVI. They can be classified into possibly mixed categories according to:

$i)$ the type of information they handle: only membership degrees in clusters, *e.g.* [13], [14], vs additional information on the geometrical structure of clusters, *e.g.* [15], [16], [17], [18], [19], [20], [21]

$ii)$ cluster properties: compactness within each cluster, *e.g.* [13], [14], [16], *and/or* separation between clusters, *e.g.* [15], [17], [18], [19], [20], [21].

III. THE NEW CLUSTER VALIDITY INDEX

A. Background on aggregation operators and proposed approach

The aggregation of several input values into a single one is a fundamental step in many data analysis problems. In such problems, one has to represent a multidimensional vector by a single value; it may be a prototype, or a class, for clustering or pattern classification, or it may be an overall satisfaction degree for multi-criteria decision making. Generally speaking, an aggregation function is an operator that, with a number of input values, say c, will associate a typical value, representing as much as possible all the inputs. Since a rescaling operation is always possible, we restrict ourselves to the interval $I = [0, 1]$ for inputs and outputs.

Definition 1: A c-ary aggregation operator (AO for short) is a mapping $\mathcal{A} : [0,1]^c \to [0,1]$, $\{a_1, \cdots, a_c\} \mapsto \mathcal{A}(a_1, \cdots, a_c)$.
Among these operators, one finds a lot of commonly used functions such as arithmetic and geometric means, triangular norms, fuzzy integrals and OWA (*Ordered Weighted Averaging*) operators, (see [27], [28] for large surveys). These operators are divided into several categories, depending on the way the values are aggregated: conjunctive, disjunctive, compensative, or weighted operators.

Definition 2: An AO $\mathcal{A}$ is said to be conjunctive if $\mathcal{A}(a_1, \cdots, a_c) \leq min(a_1, \cdots, a_c)$.
If we add properties of non decreasingness, commutativity and associativity, we obtain the triangular norms (t-norms for short) family.

Definition 3: A t-norm is a commutative, associative and monotonic function $\top : [0,1]^2 \to [0,1]$, satisfying $\top(a, 1) = a$, *i.e.* 1 is the neutral element of t-norms.
It follows from these properties that $\top(a, b) \leq min(a, b)$. Since the minimum operator satisfies the above mentioned properties, it is a t-norm. Consequently, the minimum operator is the largest t-norm for all $[a, b] \in [0,1]^2$.

TABLE II
EXEMPLES OF BASIC AND PARAMETRIZED T-NORM AND T-CONORM COUPLES.

Name	$a \top b$	$a \perp b$
Standard (S)	$\min(a,b)$	$\max(a,b)$
Algebraic (A)	ab	$a+b-ab$
Łukasiewicz (L)	$\max(a+b-1,0)$	$\min(a+b,1)$
Hamacher (H_γ)	$\frac{ab}{\gamma+(1-\gamma)(a+b-ab)}$	$\frac{a+b-ab-(1-\gamma)ab}{1-(1-\gamma)ab}$
Dombi (D_γ)	$\frac{1}{1+\left((\frac{1-a}{a})^\gamma+(\frac{1-b}{b})^\gamma\right)^{1/\gamma}}$	$1-\frac{1}{1+\left((\frac{a}{1-a})^\gamma+(\frac{b}{1-b})^\gamma\right)^{1/\gamma}}$

***Definition** 4:* An AO $\mathcal{A}$ is said to be disjunctive if $\mathcal{A}(a_1,\cdots,a_c) \geq max(a_1,\cdots,a_c)$.
If we add the same properties of non decreasingness, commutativity and associativity, we obtain the family of triangular conorms (t-conorms for short).

***Definition** 5:* A t-conorm is a commutative, associative and monotonic function $\perp : [0,1]^2 \to [0,1]$ satisfying $\perp(a,0) = a$, *i.e.* 0 is the neutral element of t-conorms.
It follows from these properties that $\perp(a,b) \geq max(a,b)$. Since the maximum operator is a t-conorm, it is the smallest one. Besides the classical triangular norm couples, many parametric families have been introduced, *e.g.* the Hamacher or Dombi families (see Table II). Introducing parameters allows to control the way the values are aggregated, and special values of the parameter generally correspond to some basic couples, *e.g.* the Hamacher couple reduces to the algebraic one if $\gamma = 1$. A complete review of triangular norms can be found in [29]. Returning to the clustering problem, we assume that the values u_{ik} of a fuzzy c-partition matrix U to be aggregated represent the degree to which an object $\mathbf{x}_k$ satisfies the i^{th} group, *i.e.* its similarity to the prototypes describing each group. Using this knowledge contained in the membership vector $\mathbf{u}_k(\mathbf{x}_k) = [u_{1k},\cdots,u_{ck}]$, clustering consists of selecting the most appropriate group to which the object will be assigned. The maximum operator, or standard triangular conorm, is commonly used in this situation, but we may be interested in the lower values that interact with the largest value. The maximum operator does not allow the aggregated values to compensate each other, whereas other triangular conorms do (see [29]; especially the Archimedean ones for which $\perp(a,a) > a$). This property can be very useful, in particular in situations where objects satisfy more than one group description, making an exclusive partitioning inefficient. A fundamental issue is the determination of the overall degree of strict membership in a group or a cluster.

In [30], the authors define the l-order fuzzy OR operator (fOR-l for short) and use it in the context of supervised classification with reject options. This operator evaluates degrees of satisfaction at a given order by combination of triangular norms.

***Definition** 6:* Let $\mathcal{P}$ be the power set of $C = \{1,2,\cdots,c\}$ and $\mathcal{P}_l = \{A \in \mathcal{P} : |A| = l\}$ where $|A|$ denotes the cardinality of the subset A. The fOR-l operator is an aggregation operator that associates $\mathbf{u}_k$ with a single value $\overset{l}{\perp}(\mathbf{u}_k) \in [0,1]$ defined by:

$$\overset{l}{\underset{i=1,c}{\perp}} u_{ik} = \top_{A \in \mathcal{P}_{l-1}} \Big(\perp_{j \in C \setminus A} u_{jk} \Big). \tag{2}$$

It can be viewed as some kind of generalization of the notion of "l^{th} largest" value, with l in C. In particular, it is easy to show that in case of standard triangular norms, $\overset{l}{\perp}(\mathbf{u}_k)$ is exactly the "l^{th} largest" element of $\mathbf{u}_k$ (see proof in [30]). We use the ability of the fOR-l to evaluate membership degrees for various orders in a different context: unsupervised classification and in particular the selection of the optimal number of clusters. For convenience and without loss of generality, we will use $u_{(l)k}$ to denote the "l^{th} largest" value of $\mathbf{u}_k$.

B. The new CVI combining a separation measure and an overlap measure

A reliable validity index for the FCM algorithm must consider both compactness and separation within a fuzzy c-partition. If only a measure of compactness is considered, the best partition is obtained when each data point is considered as a separate (singleton) cluster. On the other hand, if only a separation measure is considered, the trivial solution corresponding to one cluster is obtained.

It is generally accepted that a CVI has to consider both separation and compactness measures (see [4], [12]), provided that such measures reflect the right data structure. This is not always true. CVIs that use the objective function (1) to quantify compactness, *e.g.* XB, FS, K and PBM, are not as efficient as one could expect. The reason is the multiplication of u_{ik} and $||\mathbf{x}_k - \mathbf{v}_i||^2$ that act in an opposite way. If one is increasing then the other is decreasing, and vice-versa. Furthermore, these indices tend to monotonically decrease when the number of clusters tends to the number of points in the data set, *i.e.* $\lim_{c \to n} ||\mathbf{x}_k - \mathbf{v}_i||^2 = 0$ (see [11], [19]). It is not anymore correct for CVIs that use distances between centroids to quantify separation, *e.g.* XB, SC, K, WY and PBM, namely $||\mathbf{v}_j - \mathbf{v}_{i \neq j}||^2$. Since these quantities do not take into account the shape and/or the scattering of the clusters, two close but not dispersed clusters can be more separated than two dispersed clusters that overlap despite the distance between the centroids being large. Furthermore, using only centroid information is not sufficient to interpret the geometrical structure of the data, and therefore not sufficient for the separation between clusters either (see [22] for examples).

We propose to use, for each point $\mathbf{x}_k$, two measures that overcome these drawbacks: a fuzzy overlap measure which evaluates the degree of overlap of a specified number, l, of fuzzy clusters and a fuzzy separation measure corresponding to the largest membership degree, with respect to the $c-1$ other ones. A low value of this latter measure will denote a large separation of the most probable cluster $\mathbf{x}_k$ from the others. In terms of fuzzy membership degrees, a high separation denotes how well a given point matches its supposedly true cluster description, while the overlap measure defines how much a given point satisfies several cluster descriptions. The ability to deal with overlapping clusters is now considered to be a major criterion when comparing indices [31]. Despite its importance, the majority of the existing work is based on an intuitive representation of *overlap*. An overlap measure between l fuzzy clusters for each point $\mathbf{x}_k$ in X described by its membership degrees can be obtained by (2) as illustrated in Figure 1 where overlap values of three fuzzy clusters are plotted for various orders and triangular norm couples. According to Figure 1, the l-order overlap value is null when $l-1$ clusters are overlapping, and increases as the clusters increasingly overlap. By successively computing $\overset{l}{\perp}(\mathbf{u}_k)$ for different values of l, we get a combination of l-order overlap degrees for $\mathbf{x}_k$. In order to determine the overall degree of overlap for a given point, we have to determine which order(s) induce(s) high overlap. The best order(s) is(are) obtained by the fuzzy disjunction of the l-order overlap measures ($l = 2, c$).

***Definition** 7:* We define the overall overlap measure for $\mathbf{x}_k$ as:

$$O_\perp(\mathbf{u}_k(\mathbf{x}_k), c) = \overset{1}{\underset{l=2,c}{\perp}} \Big(\overset{l}{\underset{i=1,c}{\perp}} u_{ik} \Big). \tag{3}$$

Several other CVIs use overlap measures between couples of clusters (see [22], [26]) that can be viewed as 2-order overlap measures.

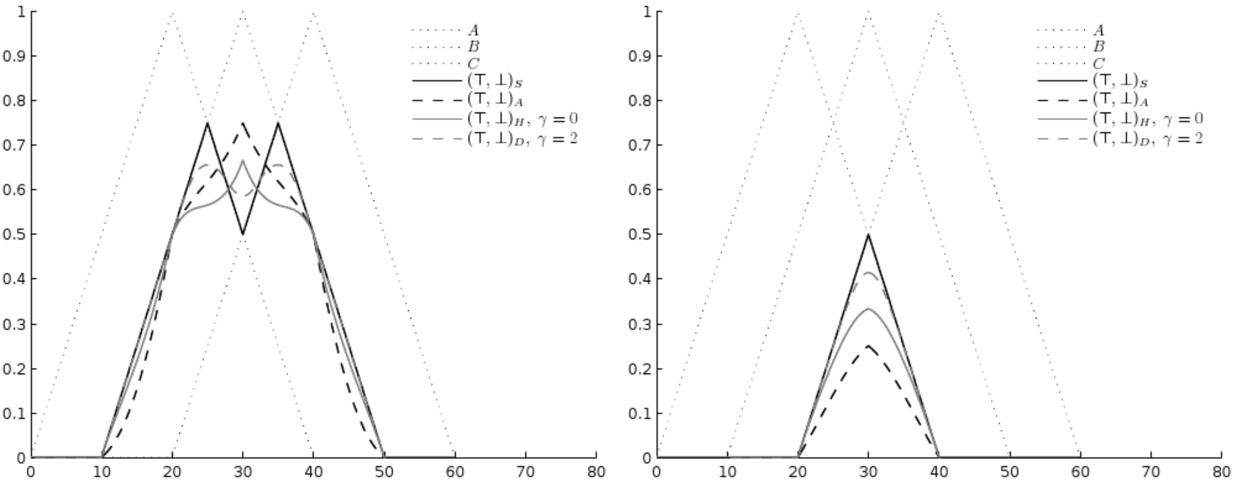


Fig. 1. $\overset{l}{\perp}(\mathbf{u}_k)$ for three fuzzy sets A, B and C with triangular membership functions at orders $l = 2$ (*left*) and $l = 3$ (*right*) using four different triangular norm couples: Standard, Algebraic, Hamacher ($\gamma = 0$) and Dombi ($\gamma = 2$).

In (3), not only couples, but also triplets of clusters up to a c-tuple of clusters combinations are taken into account. We will not compare the new CVI with these CVIs because it would require an exponential combination of them (to extend couples to all possible l-uples) resulting in a prohibitive computation time.

In [24], Bezdek and Pal show that inter cluster separation plays a more important role in CV than diameters. We propose to introduce such a measure by quantifying the fuzzy separation of each point $\mathbf{x}_k$ with $\overset{1}{\perp}(\mathbf{u}_k)$. This denotes how well $\mathbf{x}_k$ matches the cluster in which it has the largest membership; thus how well this cluster is separated from the others when only considering $\mathbf{x}_k$. Notice that, according to (3), this individual measure also corresponds to the overlap measure within one cluster, *i.e.* its separation from the other fuzzy clusters since $\mathbf{u}_k$ components sum up to one. For normalization purpose, and because there are $(c-1)$ other clusters, we use the fuzzy disjunction of the $(c-1)$ individual measures $\overset{1}{\perp}(\mathbf{u}_k)$ in order to select the most probable cluster.

***Definition** 8:* We define the fuzzy separation of $\mathbf{x}_k$ with respect to the c clusters as:

$$S_\perp(\mathbf{u}_k(\mathbf{x}_k), c) = \overset{1}{\perp}\Bigg(\underbrace{\overset{1}{\underset{i=1,c}{\perp}} u_{ik}, \cdots, \overset{1}{\underset{i=1,c}{\perp}} u_{ik}}_{c-1 \text{ times}}\Bigg). \qquad (4)$$

A small value of the overlapping degree $O_\perp(\mathbf{u}_k(\mathbf{x}_k), c)$ and a large value of the separation degree $S_\perp(\mathbf{u}_k(\mathbf{x}_k), c)$ indicates that $\mathbf{x}_k$ lies in a well separated and not overlapping part of a cluster.

***Definition** 9:* We define the overlap-separation measure from the perspective of $\mathbf{x}_k$ as:

$$OS_\perp(\mathbf{u}_k(\mathbf{x}_k), c) = \frac{O_\perp(\mathbf{u}_k(\mathbf{x}_k), c)}{S_\perp(\mathbf{u}_k(\mathbf{x}_k), c)}. \qquad (5)$$

This measure can be related to measures of fuzziness in the sense that it measures the amount of average ambiguity between fuzzy sets. However, $OS_\perp$ is not a measure of fuzziness H as defined in [32] because of the maximality property P2: $H(\mathbf{u})$ is maximum ⇔ $u_i = 0.5$ for all i. For $OS_\perp$, it holds only for (⇒): $OS_\perp(\mathbf{u})$ is maximal if $u_i = 0.5$ for all i. Nevertheless, $OS_\perp$ defines a measure of nonspecificity, (see [33] for a definition). We have $OS_\perp(\mathbf{u}) = 0$ iff $\mathbf{u}$ is a singleton, $OS_\perp(\emptyset)$ is undefined, but can be set to one by convention, and for two normal fuzzy sets $\mathbf{u}$, $\mathbf{v}$ such that $\mathbf{u} \subset \mathbf{v}$, then $OS_\perp(\mathbf{u}) \leq OS_\perp(\mathbf{v})$ for all t-norm couples, (see [32], [34] for a discussion on uncertainty measures).

***Definition** 10:* We define the *Overlap and Separation Index* (OSI) taking values in $[0, 1]$ as the average value of the individual overlap-separation measures:

$$OSI_\perp(U, c) = \frac{1}{n}\sum_{k=1}^{n} OS_\perp(\mathbf{u}_k(\mathbf{x}_k), c). \qquad (6)$$

Minimizing (6) over the range $[c_{min}, c_{max}]$ gives the local optimal number of clusters for the data in X. A short discussion about the range specification is provided in Section IV. Note that we already proposed to average the ratio of the overlap measure $O_\perp(\mathbf{u}_k, c)$ over another separation measure $S'_\perp(\mathbf{u}_k, c)$ based on the fOR-l operator in order to define a CVI in a recent paper [35]. Unfortunately, $S'_\perp(\mathbf{u}_k, c)$ can be lower than $O_\perp(\mathbf{u}_k, c)$, depending on the norm couple $(\top, \perp)$, so that the ratio is not always in [0,1]. A consequence is that such a large ratio value can significantly affect the average value in (6). This inconvenience does not hold for the new index (see Proposition 1 below).

C. Properties and example

Let us show some properties that the proposed family of indices OSI (6) satisfy. For all properties, c is supposed to be greater or equal to 2.

***Proposition** 1: For all $(\top, \perp)$ norm couples, we have $0 \leq OSI_\perp(U, c) \leq 1$.*

Proof: It is proved in [30] that $\overset{1}{\perp}(\mathbf{u}_k) \geq \overset{2}{\perp}(\mathbf{u}_k) \cdots \geq \overset{c}{\perp}(\mathbf{u}_k)$, for all $(\top, \perp)$. If $\mathbf{o}$ denotes the $(c-1)$-dimensional vector constructed by $\overset{l}{\perp}(\mathbf{u}_k)$ and $\mathbf{s}$ the one constructed with the $(c-1)$ values $\overset{1}{\perp}(\mathbf{u}_k)$, we have $o_i \leq s_i$ for all $i \in \{1, \cdots, c-1\}$. By monotony of $\top$, we finally have $O_\perp(\mathbf{u}_k, c) \leq S_\perp(\mathbf{u}_k, c)$. Q.E.D. ■

***Proposition** 2: If U is a crisp partition matrix, then $OSI_\perp(U, c) = 0$, for all $(\top, \perp)$.*

Proof: For all k, $u_{ik} \in \{0, 1\}$ and $\sum_{i=1}^{c} u_{ik} = 1$, then one value equals 1 while the others are 0, say $u_{(1)k} = 1$ and $u_{(2)k} = \cdots = u_{(c)k} = 0$. Since 0 is the absorbing element of $\top$ (*i.e.* $a\top 0 = 0$

for all $\top$), we easily verify that $O_{\perp}(\mathbf{u}_k, c) = 0$ for all $(\top, \perp)$:

$$\overset{1}{\underset{l=2,c}{\perp}}\Big(\overset{l}{\underset{i=1,c}{\perp}} u_{ik}\Big) = \Big(\overset{2}{\underset{i=1,c}{\perp}}(1,0,\cdots,0)\Big)\overset{1}{\perp}\cdots\overset{1}{\perp}\Big(\overset{c}{\underset{i=1,c}{\perp}}(1,0,\cdots,0)\Big)$$
$$= \overset{1}{\perp}(\underbrace{0,\cdots,0}_{c-1 \text{ times}}) = 0 \tag{7}$$

Since 1 is the absorbing element of $\perp$ (*i.e.* $a\perp 1 = 1$ for all $\top$), then $\overset{1}{\perp}(1,0,\cdots,0) = 1$. Hence, we have $OSI_{\perp}(U,c) = \frac{1}{n}\sum_{k=1}^{n} 0/1 = 0$. Q.E.D. ■

***Proposition** 3: If U is a totally fuzzy partition matrix, then $OSI_{\perp_S}(U,c) = 1$.*

Proof: For all k, $u_{ik} = \frac{1}{c}$. From (3), we can write $O_{\perp}(\mathbf{u}_k, c)$ as:

$$\overset{1}{\underset{l=2,c}{\perp}}\Big(\overset{l}{\underset{i=1,c}{\perp}} u_{ik}\Big) = \Big(\overset{2}{\underset{i=1,c}{\perp}}\Big(\frac{1}{c},\cdots,\frac{1}{c}\Big)\Big)\overset{1}{\perp}\cdots\overset{1}{\perp}\Big(\overset{c}{\underset{i=1,c}{\perp}}\Big(\frac{1}{c},\cdots,\frac{1}{c}\Big)\Big). \tag{8}$$

With standard triangular norms ($\top = min, \perp = max$), it becomes:

$$\overset{1}{\underset{l=2,c}{\perp}}\Big(\overset{l}{\underset{i=1,c}{\perp}} u_{ik}\Big) = \overset{1}{\perp}\Big(\underbrace{\frac{1}{c},\cdots,\frac{1}{c}}_{c-1 \text{ times}}\Big) = max\Big(\underbrace{\frac{1}{c},\cdots,\frac{1}{c}}_{c-1 \text{ times}}\Big) = \frac{1}{c}. \tag{9}$$

It is easy to check that it is the value of $S_{\perp_S}(\mathbf{u}_k, c)$ for all $\mathbf{u}_k$ such as $u_{ik} = 1/c$. Therefore, we have $OSI_{\perp_S}(U,c) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{c}/\frac{1}{c} = 1$. Q.E.D. ■

***Proposition** 4: If we use standard triangular norms, then $OSI_{\perp_S}(U,c) = \frac{1}{n}\sum_{k=1}^{n}\frac{u_{(2)k}}{u_{(1)k}}$.*

Proof: It is proved in [30] that, when using standard triangular norms, $\overset{l}{\perp}_S(\mathbf{u}_k) = u_{(l)k}$ and $\overset{1}{\perp}_S(\mathbf{u}_k) = max_{i=1,c}u_{ik}$. Then, we have $O_{\perp_S}(\mathbf{u}_k, c) = max_{l=2,c}u_{(l)k} = u_{(2)k}$ by (3) and $S_{\perp_S}(\mathbf{u}_k, c) = max(max_{i=1,c}u_{ik}, ..., max_{i=1,c}u_{ik}) = max(u_{(1)k}, ..., u_{(1)k}) = u_{(1)k}$ by (4). Therefore, we have $OSI_{\perp_S}(U,c) = \frac{1}{n}\sum_{k=1}^{n}\frac{u_{(2)k}}{u_{(1)k}}$. Q.E.D. ■

Let us illustrate the ability of the proposed index to find the right number of clusters and the right partition for a toy example, inspired by [36], that we call *Diamond+*. It consists of the eleven two-dimensional points first introduced by Windham [37] and an outlier, with coordinates $(6,6)$, (see Figure 2-(f)). Ignoring the outlier, the correct partition is composed of the $c^{\star} = 2$ touching clusters. Most of CVIs that only consider compactness and separation will select three clusters, (see Table V in section IV). The membership degrees of the twelve points (i.e. U) provided by the fuzzy c-means algorithm, for $c = 2$ and $c = 3$ clusters, are given in Table III, as well as the values of the overlap (3), separation (4) and overlap-separation (5) measures. The standard norms $\top = \min$ and $\perp = \max$ are used for simplicity, so the three measures are respectively the second largest value, the largest value and the ratio of both, regardless c (refer to Proposition 4). Since the membership degrees are given for both cases ($c = 2$ and $c = 3$), the values of the three measures are easy to compute: $O_{\perp_S}(\mathbf{u}_k, c) = u_{(2)k}$, $S_{\perp_S}(\mathbf{u}_k, c) = u_{(1)k}$ and $OS_{\perp_S}(\mathbf{u}_k, c) = u_{(2)k}/u_{(1)k}$, as well as their average value over the twelve points which gives the index values. The obtained index values are $OSI_{\perp_S}(U,2) = 0.132$ and $OSI_{\perp_S}(U,3) = 0.142$ showing that the proposed CVI recovers the natural partition. Note that for $c = 2$, the membership degrees of the eleven diamonds points, and therefore the measures, are not symmetrical because of the non central outlier's position. Let us focus on two particular points, compared to the ten others: the middle of the two diamonds $\mathbf{x}_6$ and the outlier $\mathbf{x}_{12}$. As expected, independently of c, $\mathbf{x}_6$ is the most ambiguous (0.387 and 0.466) and the less separated (0.613 and 0.477) point. For $c = 2$, $\mathbf{x}_{12}$ does not belong to an overlapping part of the cluster (0.222), and is more separated (0.778) than $\mathbf{x}_6$. However, $\mathbf{x}_{12}$ is less separated and more ambiguous than the ten other points so the resulting ratio, even if it is much greater than the others (except $\mathbf{x}_6$ of course), does not make the average value $OSI_{\perp_S}(U,2)$ be significantly high. For $c = 3$, $\mathbf{x}_{12}$ is the most (and only) representative point of the third cluster as expected, but also the most separated (0.999) and least ambiguous (0.001) point so that the resulting ratio is much lower. The overlap (respectively separation) measure for $\mathbf{x}_6$ increases (respectively decreases) significantly because the second cluster that it belongs to does not contain $\mathbf{x}_{12}$ anymore. For the ten other points, its depends on their relative position to the three centroids (from one side, between) and the ratios increase or decrease in a very compensatory way with respect to the average value because they belong to two well-balanced and symmetrical clusters. Therefore, the main changes in contributions to the index are due to both $\mathbf{x}_6$ and $\mathbf{x}_{12}$. When c increases from 2 to 3, the decrease of the ratio for $\mathbf{x}_{12}$ is not sufficient to compensate for the increase of the ratio for $\mathbf{x}_6$ when averaging, so that $OSI_{\perp_S}(U,3)$ does not become less than $OSI_{\perp_S}(U,2)$. However, as well as others, the proposed CVI is sensitive to the number of outliers, their relative position and their scattering (*i.e.* are they still outliers or noisy points?), because overlap and separation measures highly depend on the number and distribution of clusters. For instance, we find that $OSI_{\perp_S}(U,3) < OSI_{\perp_S}(U,2)$ if $\mathbf{x}_{12}$ is put far away from the diamonds or if there are a few additional points close to it. This is the reason why experiments on various mixed situations are presented in section IV - Figure 2. We experimentally found that when a data set contains about 50% (or more) of uniformly distributed and sufficiently scattered (noisy) points, all the CVIs fail in recovering the right number of clusters if a large range for c is tested.

IV. Experimental Results

We evaluate the performance of the proposed OSI by conducting an extensive comparison with the nine CVIs and the validation procedure described in section II-B in conjunction with the FCM algorithm. As in almost all papers dealing with fuzzy CV, the fuzzifier exponent m is set to 2, the termination parameter for the test for convergence is set to 10^{-3} and the Euclidean distance is used. The optimal number of clusters is sought in the range $[c_{min} = 2, c_{max}]$ with $c_{max} = 10$ for real data sets and $c_{\max} = \min(10, \lfloor\sqrt{n}\rfloor)$ for artificial data sets, where $\lfloor . \rfloor$ denotes the floor function, in order to ensure a good balance between the number of clusters and the number of points, (see [11]).

A. Data sets

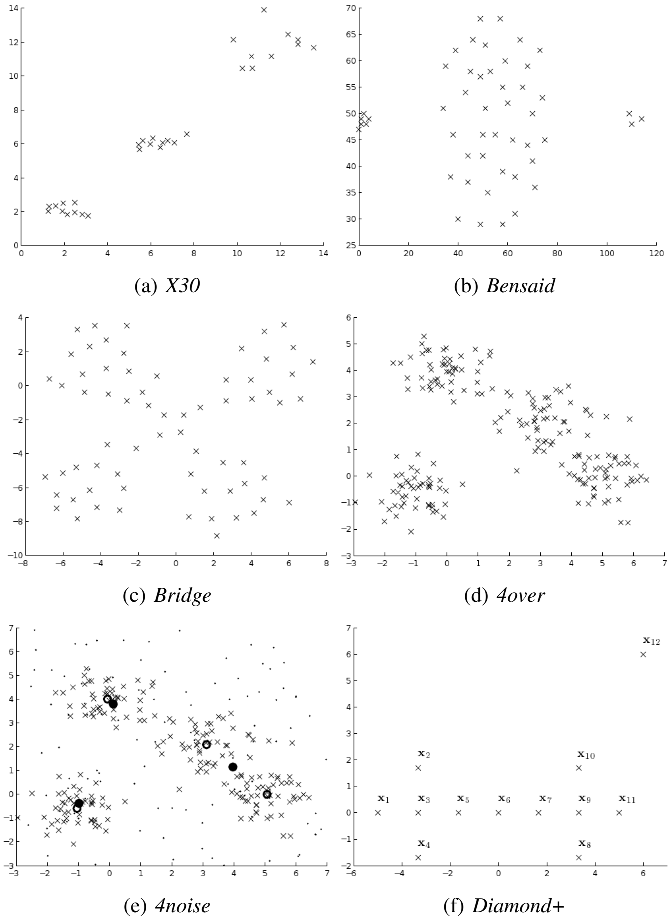
We make use of eleven data sets with varying properties such as good separation, overlapping clusters, presence of outliers, additional noisy points, making the CV problem more or less easy. Most of these data sets are described in the CV literature. The first six data sets are two-dimensional artificial data sets such that the true number of clusters can be visually assessed, the five others are real data sets from the public domain, (see Table IV). The data set *Bridge* is composed of four connected clusters, (see Figure 2-(c)). The data set *4over* contains 200 points drawn from a mixture of $c = 4$ bivariate normal distributions of 50 points each, with two of them slightly overlapping. The *4noise* data set is *4over* to which 100 points drawn from a uniform distribution are added to simulate a noisy environment, (Figure 2-(e)).

TABLE III
MEMBERSHIP DEGREES, OVERLAP AND SEPARATION MEASURES, AND $OSI_{\perp_S}$ VALUES FOR $c = 2, 3$ CLUSTERS ON THE *Diamond+* DATA SET, RESULTING IN CVI VALUES $OSI_{\perp_S}(2) = 0.132 < OSI_{\perp_S}(3) = 0.142$.

c		$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$	$\mathbf{x}_8$	$\mathbf{x}_9$	$\mathbf{x}_{10}$	$\mathbf{x}_{11}$	$\mathbf{x}_{12}$	$OSI_{\perp_S}$ (avg.)
2	$\mathbf{u}_k = \begin{pmatrix} u_{1k} \\ u_{2k} \end{pmatrix}$	.946 .054	.944 .056	.996 .004	.943 .057	.945 .055	.613 .387	.174 .826	.129 .871	.018 .982	.022 .978	.039 .961	.222 .778	
	$O_{\perp_S}(\mathbf{u}_k, c)$	.054	.056	.004	.057	.055	.387	.174	.129	.018	.022	.039	.222	
	$S_{\perp_S}(\mathbf{u}_k, c)$	.946	.944	.996	.943	.945	.613	.826	.871	.982	.978	.961	.778	
	$OS_{\perp_S}(\mathbf{u}_k, c)$	.057	.059	.004	.060	.058	.631	.210	.148	.018	.022	.040	.285	0.132
3	$\mathbf{u}_k = \begin{pmatrix} u_{1k} \\ u_{2k} \\ u_{3k} \end{pmatrix}$	.933 .047 .019	.913 .063 .024	.998 .001 .001	.921 .061 .018	.876 .100 .024	.466 .477 .057	.100 .874 .026	.055 .905 .039	.001 .998 .001	.058 .824 .119	.042 .866 .092	.000 .001 .999	
	$O_{\perp_S}(\mathbf{u}_k, c)$	.047	.063	.001	.061	.100	.466	.100	.055	.001	.119	.092	.001	
	$S_{\perp_S}(\mathbf{u}_k, c)$	.933	.913	.998	.921	.876	.477	.874	.905	.998	.824	.866	.999	
	$OS_{\perp_S}(\mathbf{u}_k, c)$	.050	.069	.001	.066	.114	.976	.114	.061	.001	.144	.106	.001	0.142

TABLE V
OPTIMAL NUMBER OF CLUSTERS c_{best} OBTAINED USING DIFFERENT CVIS ON ARTIFICIAL AND REAL DATA SETS.

data set	c_{max}	NPE	NPC	FS	FHV	XB	SC	K	PBM	WY	$OSI_\perp$				$c^\star$
											$\perp_S$	$\perp_{H_{10}}$	$\perp_A$	$\perp_{D_2}$	
X30	5	3	3	4	3	3	5	2	3	3	3	3	3	3	3
Bensaid	7	2	3	7	6	3	7	3	6	6	3	3	3	3	3
Bridge	8	2	4	6	5	5	8	4	4	5	4	4	4	4	4
4over	10	3	4	4	4	3	3	3	4	4	4	4	4	4	4
4noise	10	2	3	5	4	2	8	3	4	3	4	4	3	4	4
Diamond+	4	2	3	3	4	3	4	3	3	2	2	2	3	2	2
Iris	10	2	3	5	3	2	3	2	3	2	2	2	2	2	2 or 3
Wine	10	2	3	10	3	2	6	2	3	3	3	3	3	3	3
Starfield	10	2	2	7	9	6	8	3	4	3	9	9	9	9	8 or 9
Cancer	10	2	2	3	2	2	4	2	2	2	2	2	2	2	2
Pima	10	2	2	3	7	2	6	2	3	2	2	2	2	2	2



(a) *X30* (b) *Bensaid*

(c) *Bridge* (d) *4over*

(e) *4noise* (f) *Diamond+*

Fig. 2. The artificial data sets.

TABLE IV
ARTIFICIAL AND REAL DATA SETS USED IN THE EXPERIMENTS.

Data set	Cluster properties	$c^\star$
X30 [24]	well separated	3
Bensaid [18]	different dispersions and cardinalities	3
Bridge	connected	4
4over	overlapping and noisy points	4
4noise	overlapping and noisy points	4
Diamond+	touching and one outlier	2
Iris	overlapping	2 or 3
Wine	well separated	3
Starfield [17]	chain-like structure	8 or 9
Breast Cancer	overlapping	2
Pima	overlapping	2

B. Results

Table V summarizes the local optimal number of clusters obtained with the tested CVIs on artificial and real data sets. The $c^\star$ column gives the expected number of clusters which is either the physical number of clusters given by an expert (real data sets) or the (most) visually perceptive one (synthetic data sets). The structure of the *X30* data set is easy to recover, so most of the presented indices including the proposed one correctly identify three clusters, except for FS, SC and K. For the *Bensaid* data set, only NPC, XB, K and $OSI_\perp$ find the correct number of clusters, while the others have a tendency to overestimate the number of clusters, *e.g.* six and seven, by dividing the central cluster in order to obtain clusters with a similar number of points. The same problem arises with the linking points of the *Bridge* data set for most of the indices, except for NPC, PBM, K and $OSI_\perp$. For the *4over* data set, indices NPE, XB, SC and K fail by merging the two overlapping clusters while the others succeed. Only FHV, PBM and $OSI_{\perp=S,H_{10},D_2}$ identify the correct number of clusters for the *4noise* data set. The claimed ability of WY to deal well with noisy points is not true in all cases. However, it performs well in the presence of outliers, as shown for data set *Diamond+*, as well as NPC and the proposed $OSI_\perp$. Since it is generally accepted that the right number of clusters for the *Iris* data set is two or three, it is not surprising that all indices find it except FS. More surprising is the inability of NPE, XB, FS and SC to recover the structure of the *Wine* data set whose clusters are known to be linearly separable. The low number of examples compared to the high number of poorly separated clusters of the *Starfield* data set makes the problem of finding the number of clusters difficult. The proposed $OSI_{\perp}$, FHV and SC are the only indices that give an acceptable number of clusters. Almost all indices identify the right number of clusters for the *Cancer* data set. The two groups of the *Pima* data set present an overlap which makes FS, FHV, PBM and SC indices fail in correctly recognizing the true number of clusters. None of the previous existing indices correctly recognizes the expected number $c^{\star}$ for all the data sets. Some of them are very robust to outliers, *e.g.* WY, others are less adapted to a structure where clusters strongly overlap, *e.g.* PBM, but fail when faced with another case or with mixed ones. The new proposed index performs well, whatever the structure, for most of the norm couples. It only fails for the algebraic one for the $4noise$ and *Diamond+* data sets. This is due to the high compensatory behavior of the t-norm (product) which particularly arises in presence of isolated points. For illustration purpose, Figure 3 shows the $OSI_{\perp_S}(U, c)$ plot for c in the specified range for all the considered data sets. Let us focus on the respective results of OSI versus NPC and NPE, which are all based only on the fuzzy partition matrix U. First of all, NPE has a very strong tendency to select two clusters, making it fail most often when $c^{\star} > 2$ whenever the clusters are well separated, *e.g. Bensaid* data set. Although NPC outperforms NPE, it fails in presence of noisy points or outliers (*e.g. 4noise, Diamond+*). The main reason is that the information measure provided by the squared membership degrees is not adapted to combined situations, whereas the OSI index permits balancing of the presence of isolated and/or ambiguous points. Moreover, when the data set consist of a high number of poorly separated clusters, *e.g. Starfield*, OSI still succeeds while NPC fails because of its higher monotonic tendency in spite of the normalization factor. More generally, NPC and NPE are outperformed because they do not exploit the relationship between membership degrees. Both compute a measure of uncertainty where each value u_{ik} is multiplied with itself or a function of itself, respectively $u_{ik} \times u_{ik}$ and $u_{ik} \times \log(u_{ik})$. The indices then sum up these individual measures over the data set. In contrast, OSI evaluates the underlying uncertainty by combining different degrees u_{ik} and u_{jk} where $i \neq j$. Therefore, the relationship between degrees is taken into account if any, hence the data structure is better reflected.

C. Sensitivity to fuzzifier m

As mentioned in section I, the fuzzy c-partition matrix U resulting from the FCM algorithm depends on the fuzzy exponent m. A CVI should be robust in the presence of changes in this user-defined parameter. Since it has been shown in [24] that FCM provides best results for m lying in $[1.5, 2.5]$, we test the different CVIs in this range of values. The *4over* data set and even more the *4noise* one are chosen because two close clusters strongly overlap (see Figure 2-(d) and -(e)). This could make the indices favor three fuzzy clusters instead of four as the partition matrix U becomes more fuzzy with m. The optimal number of clusters selected by the different CVIs on both data sets are reported in Table VI for c in the range $[2, 10]$. One must distinguish robust or quite robust CVIs that fail in selecting the correct number of clusters, *e.g.* XB and K, and the others. The results show that the proposed index OSI is at least as robust as FHV, WY and PBM, and even more robust for the most cautious norms (Standard and Dombi). It only fails for the most compensatory norms (Algebraic and Hamacher) for high values of m (2.3 and 2.5) because the induced clusters' boundaries are less crisp. As compared to the other CVIs that only use U, NPC is more sensitive to m than OSI, while NPE is less but does not provide the correct number of clusters for both data sets despite m lying in a favorable range.

TABLE VI
OPTIMAL NUMBER OF CLUSTERS SELECTED BY DIFFERENT CVIS ON ARTIFICIAL *4over* AND *4noise* DATA SETS FOR VARIOUS VALUES OF m.

data set	m	NPE	NPC	FS	FHV	XB	SC	K	PBM	WY	$OSI_{\perp}$			
											$\perp_S$	$\perp_{H_{10}}$	$\perp_A$	$\perp_{D_2}$
4over	1.5	3	4	4	4	3	3	3	4	4	4	4	4	4
	1.7	3	4	4	4	3	3	3	4	4	4	4	4	4
	1.9	3	4	4	4	3	3	3	4	4	4	4	4	4
	2.1	3	4	4	4	3	3	3	4	4	4	4	4	4
	2.3	3	4	4	4	3	7	3	4	4	4	4	4	4
	2.5	3	4	5	4	3	7	3	4	4	4	4	4	4
4noise	1.5	2	4	6	4	3	7	3	5	4	4	4	4	4
	1.7	2	4	5	4	3	8	3	5	3	4	4	4	4
	1.9	2	3	5	4	2	8	3	4	3	4	4	4	4
	2.1	2	3	5	4	3	10	3	4	4	4	4	4	4
	2.3	2	3	5	3	3	5	3	4	4	4	3	3	4
	2.5	2	3	4	3	3	7	3	3	4	4	3	3	4

V. Conclusion

A new family of CVIs for the fuzzy c-means algorithm has been proposed in this paper. It is simply defined by the average value of two new overlap and separation measures that do not use clusters' centroids but only the membership degrees. Both measures are defined for each point of the data set to be clustered through combinations of triangular norms applied to its membership degrees so that the relative importance of the degrees is taken into account in spite of the fuzzy context that makes their sum constant. A low value of the overlap measure for a given point means that it does not belong to an overlapping part of the most probable cluster, while a high separation measure implies that it is located near the cluster's prototype. An extensive comparison to the most frequently referred to indices in the literature and more recent CVIs, when determined for a number of artificial and real data sets, shows that the new indices outperform the existing ones. It is worth noting that the proposed family of indices behaves well when faced with particularly difficult (mixed) data properties such as overlapping clusters, bridges, outliers and additional noisy points. This is all the more noticeable since indices that only use membership values are generally unable to handle such difficult structures.

The selection of triangular norm couples is not an easy task and remains an open problem from a theoretical point of view. It requires further study on how their properties make the induced index behave with respect to the number of clusters (monotonic tendency) and to membership values for a given clustering situation. For instance, we observed that t-norms having a high compensative property, *e.g.* the algebraic ones, are not well suited to very noisy environments. Other combinations of overlap-separation measures than the simple taking of the average are possible to define new CVIs, *e.g.*: median, OWA or fuzzy integrals.

References

[1] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comp. Surv.*, vol. 31, pp. 264–323, 1999.

[2] J. C. Bezdek, *Pattern Recognition with fuzzy objective function algorithm.* Plenum Press, 1981.

[3] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol. 1, no. 2, pp. 98–110, 1993.

[4] J. C. Bezdek, J. M. Keller, R. Krishnapuram, and N. R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing.* Kluwer Academic, 1999.

[5] C. Borgelt and R. Kruse, "Finding the number of fuzzy clusters by resampling," in *FUZZ-IEEE*, 2006, pp. 48–54.

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

[7] R. J. G. B. Campello, "Generalized external indexes for comparing data partitions with overlapping categories," *Pattern Recognit. Lett.*, vol. 31, no. 9, pp. 966–975, 2010.

[8] R. J. Hathaway and J. C. Bezdek, "Visual cluster validity for prototype generator clustering models," *Pattern Recognit. Lett.*, vol. 24, no. 9-10, pp. 1563–1569, 2003.

[9] J. C. Bezdek, R. J. Hathaway, and J. M. Huband, "Visual assesment of clustering tendency for rectangular dissimilarity matrices," *IEEE Trans. Fuzzy Syst.*, vol. 15, no. 5, pp. 890–903, 2007.

[10] R. K. Brouwer, "Permutation clustering using the proximity matrix," in *FUZZ-IEEE*, 2009, pp. 441–446.

[11] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 370–379, 1995.

[12] W. Wang and Y. Zhang, "On fuzzy cluster validity indices," *Fuzzy Sets Syst.*, vol. 158, no. 19, pp. 2095–2117, 2007.

[13] J. C. Dunn, "Indices of partition fuzziness and the detection of clusters in large data sets," in *Fuzzy Automata and Decision processes*. Elsevier, NY, 1977.

[14] M. Roubens, "Pattern classification problems and fuzzy sets," *Fuzzy Sets Syst.*, vol. 1, no. 4, pp. 239–253, 1978.

[15] Y. Fukuyama and M. Sugeno, "A new method for choosing the number of clusters for the fuzzy c-means method," in *Proc. 5th Fuzzy Systems Symposium*, 1989, pp. 247–250.

[16] I. Gath and A. B. Geva, "Unsupervised optimal fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 773–780, 1989.

[17] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, 1991.

[18] A. M. Bensaid, L. O. Hall, J. C. Bezdek, L. P. Clarke, M. L. Silbiger, J. A. Arrington, and R. F. Murtagh, "Validity-guided (re)clustering with applications to image segmentation," *IEEE Trans. Fuzzy Syst.*, vol. 4, no. 2, pp. 112–123, 1996.

[19] S. Kwon, "Cluster validity index for fuzzy clustering," *Electronic Letters*, vol. 34, no. 22, pp. 2176–2177, 1998.

[20] M. Pakhira, S. Bandyopadhyay, and U. Maulik, "Validity index for crisp and fuzzy clusters," *Pattern Recognit.*, vol. 37, no. 3, pp. 487–501, 2004.

[21] K. Wu and M. Yang, "A cluster validity index for fuzzy clustering," *Pattern Recognit. Lett.*, vol. 26, no. 9, pp. 1275–1291, 2005.

[22] D. Kim, K. Lee, and D. Lee, "On cluster validity index for estimation of the optimal number of fuzzy clusters," *Pattern Recognit.*, vol. 37, no. 10, pp. 2009–2025, 2004.

[23] H. Le Capitaine and C. Frélicot, "A fuzzy modeling approach to cluster validity," in *FUZZ-IEEE*, 2009, pp. 462–467.

[24] J. C. Bezdek and N. R. Pal, "Some new indexes of cluster validity," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 23, no. 3, pp. 301–315, 1998.

[25] M. R. Rezaee, B. Lelieveldt, and J. Reiber, "A new cluster validity index for the fuzzy c-mean," *Pattern Recognit. Lett.*, vol. 19, no. 3-4, pp. 237–246, 1998.

[26] M. H. F. Zarandi, E. Neshat, and I. B. Türksen, "A new cluster validity index for fuzzy clustering based on similarity measure," in *Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, 11th Int. Conf.*, 2007, pp. 127–135.

[27] T. Calvo, A. Kolesárová, M. Komorníková, and R. Mesiar, "Aggregation operators: properties, classes and construction methods." Heidelberg, Germany, Germany: Physica-Verlag, 2002, pp. 3–104.

[28] M. Grabisch, J. Marichal, R. Mesiar, and E. Pap, *Aggregation Functions*, ser. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2009, no. 127.

[29] E. P. Klement and R. Mesiar, *Logical, Algebraic, Analytic, and Probabilistic Aspects of Triangular Norms*. Elsevier, 2005.

[30] L. Mascarilla, M. Berthier, and C. Frélicot, "A k-order fuzzy or operator for pattern classification with k-order ambiguity rejection," *Fuzzy Sets Syst.*, vol. 159, no. 15, pp. 2011–2029, 2008.

[31] M. Bouguessa, S. Wang, and H. Sun, "An objective approach to cluster validation," *Pattern Recognit. Lett.*, vol. 27, no. 13, pp. 1419–1430, 2006.

[32] N. R. Pal, "On quantification of different facets of uncertainty," *Fuzzy Sets Syst.*, vol. 107, no. 1, pp. 81–91, 1999.

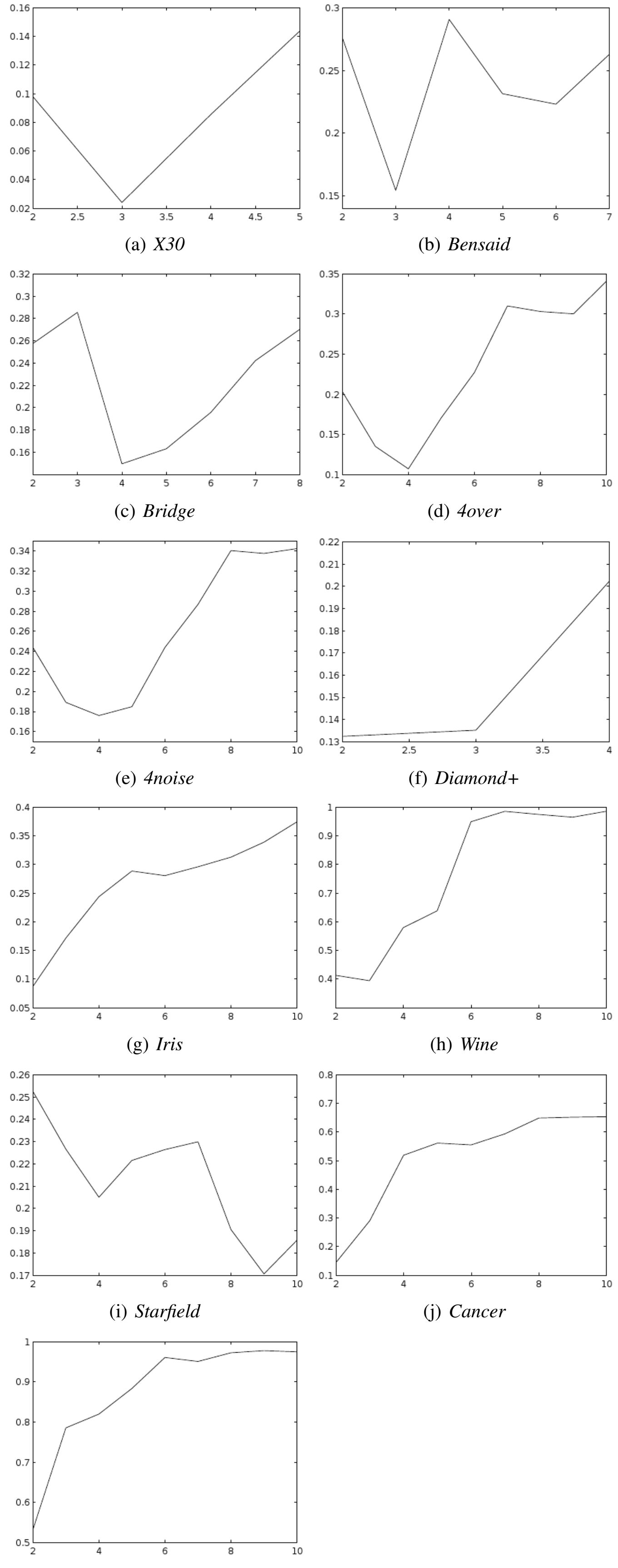
[33] L. Garmendia, R. Yager, E. Trillas, and A. Salvador, "On t-norms based measures of specificity," *Fuzzy Sets Syst.*, vol. 133, no. 2, pp. 237–248, 2003.

[34] G. J. Klir, *Uncertainty and information: foundations of generalized information theory*. John Wiley, 2005.

[35] H. Le Capitaine and C. Frélicot, "A family of cluster validity indexes based on a l-order fuzzy or operator," in *LNCS 5342*, 2008, pp. 622–631.

[36] M. Masson and T. Denoeux, "Ecm: An evidential version of the fuzzy c-means algorithm," *Pattern Recognit.*, vol. 41, no. 4, pp. 1384–1397, 2008.

[37] M. Windham, "Numerical classification of proximity data with assignment measure," *J. of Classification*, vol. 2, pp. 157–172, 1985.



(a) *X30*

(b) *Bensaid*

(c) *Bridge*

(d) *4over*

(e) *4noise*

(f) *Diamond+*

(g) *Iris*

(h) *Wine*

(i) *Starfield*

(j) *Cancer*

(k) *Pima*

Fig. 3. $OSI_{\perp_S}(U,c)$ curves for the different data sets, c in the specified range.