



HAL
open science

A Fuzzy Modeling Approach to Cluster Validity

Hoel Le Capitaine, Carl Frelicot

► **To cite this version:**

Hoel Le Capitaine, Carl Frelicot. A Fuzzy Modeling Approach to Cluster Validity. 18th International Conference on Fuzzy Systems, FUZZIEEE'09, 2009, Jeju Island, South Korea. pp.462–467. hal-00441442

HAL Id: hal-00441442

<https://hal.science/hal-00441442v1>

Submitted on 16 Dec 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Fuzzy Modeling Approach to Cluster Validity

Hoel Le Capitaine and Carl Frélicot

Abstract—This paper presents a new approach to find the optimal number of clusters of a fuzzy partition. It is based on a fuzzy modeling approach which combines measures of clusters' separation and overlap. These measures are based on triangular norms and a discrete Sugeno integral. Results on artificial and real data sets prove its efficiency compared to indexes from the literature.

I. INTRODUCTION

Clustering aims at detecting natural groups (or clusters) in multidimensional data sets. The principle is that data points within a cluster are as similar as possible whereas data points of different clusters as dissimilar as possible. Since clusters may have different shapes and sizes, a partition resulting from this unsupervised classification process needs to be validated. A key point is the number of clusters and many Cluster Validity Indexes (CVIs) have been proposed, in particular for fuzzy clustering, see [3], [11], [19].

In this paper, we propose a new index following a fuzzy system modeling approach using measures of separation and overlap degree of the data points from the obtained clusters. These measures are based on dedicated operators, based on triangular norms and a discrete Sugeno integral, that aggregate the clusters fuzzy labels provided by the clustering algorithm.

II. CLUSTER VALIDITY FOR FUZZY CLUSTERING

A. The Fuzzy C-Means algorithm

Let $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ be a n points data set in a p -dimensional feature space, say \mathbb{R}^p , with the usual euclidean norm $\|\cdot\|$. The fuzzy c-means (FCM) algorithm partitions X into $c > 1$ clusters by minimizing the following objective function [2]:

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

where $u_{ik} \in [0, 1]$ is the membership degree of \mathbf{x}_k to the i^{th} cluster represented by its centroid $\mathbf{v}_i \in \mathbb{R}^p$. Centroids are gathered into a $(c \times p)$ matrix $V = [\mathbf{v}_1, \dots, \mathbf{v}_c]$. Degrees u_{ik} are subject to a normalization constraint $\sum_{i=1}^c u_{ik} = 1$ for all \mathbf{x}_k in X , and are elements of the fuzzy c -partition $U(c \times n)$ whose columns $\mathbf{u}_k = {}^t(u_{1k}, \dots, u_{ck})$ are the membership vectors of each \mathbf{x}_k . The so-called fuzzifier $m > 1$ is a weighting exponent which makes the resulting partition more or less fuzzy: the higher m , the softer the cluster boundaries are. Minimization of (1) is obtained by iteratively updating (U, V) as follows:

Hoel Le Capitaine/Carl Frélicot are with the MIA Laboratory, University of La Rochelle, Avenue Michel Crépeau, France (phone: +33 546 458 322/234; email: {hlecap01},{carl.frelicot}@univ-lr.fr).

$$u_{ik} = 1 / \sum_{j=1}^c \left(\frac{\|\mathbf{x}_k - \mathbf{v}_i\|}{\|\mathbf{x}_k - \mathbf{v}_j\|} \right)^{2/(m-1)} \quad (2)$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^n u_{ik}^m \mathbf{x}_k}{\sum_{k=1}^n u_{ik}^m} \quad (3)$$

The usual euclidian norm $\|\cdot\|$ induces hyper-spherical clusters, hence FCM can only detect clusters with the same shape and orientation. Thus, such clusters description is not well suited to every possible situation, e.g.: bridges, outliers, additional noisy points. Cluster Validity (CV) is then a more challenging problem using FCM instead of other algorithms that behave better in such situations.

B. Cluster validity procedure and classical fuzzy indexes

Validating the provided clustering (U, V) of X consists in assessing whether the resulting partition reflects the data structure or not. Since clustering is unsupervised, no prior knowledge on the data is taken into account, and the number c of clusters is a user-defined parameter for clustering algorithms such as FCM. Most of works on cluster validity focus on the number of clusters problem and many CVIs have been proposed, refer to [3], [11] for comparative studies. Given a CVI, the procedure to automatically select the optimal number of clusters c_{best} in a predefined range $[c_{min}, c_{max}]$ and therefore the best partition is as follows:

- (1) choose values c_{min} and c_{max}
- (2) for $c = c_{min}$ to c_{max}
 - run FCM
 - compute CVI(c) from (X, U, V)
- (3) select c_{best} such as $CVI(c_{best})$ is optimal and take the corresponding partition (U, V)

CVIs can be classified either according to the type of information they handle (only membership degrees to clusters vs additional information on the geometrical structure of clusters) or to cluster properties (compactness within each cluster and/or separation between clusters). Note these categories are not mutually exclusive and most indexes present advantages and drawbacks. Earliest CVIs only use partial membership degrees (U) . Let us cite the *Partition Coefficient* [2], taking values in $[\frac{1}{c}, 1]$:

$$PC(c) = \frac{1}{n} \sum_{k=1}^n \sum_{i=1}^c u_{ik}^2 \quad (4)$$

or the *Partition Entropy*, taking values in $[0, \log(c)]$:

$$PE(c) = -\frac{1}{n} \sum_{k=1}^n \sum_{i=1}^c u_{ik} \log(u_{ik}) \quad (5)$$

Both PC to be maximized and PE to be minimized are monotonic with c , as well as their bounds. Normalized versions have been proposed to overcome these drawbacks. For the experiments in section IV, we use NPC [18], [6] and NPE [7] defined by:

$$NPC(c) = \frac{cPC(c) - 1}{c - 1} \quad (6)$$

$$NPE(c) = \frac{nPE(c)}{n - 1} \quad (7)$$

The second category consists of indexes that use membership degrees but also some information about the geometrical structure of the data (U, V, X) , e.g. the Xie-Beni index [22]:

$$XB(c) = \frac{J_m(U, V) / n}{\min_{i,j=1,c;j \neq i} \|\mathbf{v}_i - \mathbf{v}_j\|^2} \quad (8)$$

or the Fukuyama-Sugeno index [8]:

$$FS(c) = J_m(U, V) - \sum_{k=1}^n \sum_{i=1}^c u_{ik}^m \|\mathbf{v}_i - \bar{\mathbf{v}}\|^2 \quad (9)$$

where $\bar{\mathbf{v}}$ is the mean of centroids. Both XB and FS combine the FCM objective function (1) which measures how much the clusters are *compact* and an additional term which measures how much they are *separated*. Combination indicates that both indexes are to be minimized. The more compact and separated the clusters are, the less fuzzy and the more crisp the partition is, therefore the more optimal c is. In [9], Gath and Geva propose the Fuzzy Hyper Volume (FHV) defined by:

$$FHV(c) = \sum_{i=1}^c \sqrt{\det(C_i)} \quad (10)$$

where $C_i = \frac{\sum_{k=1}^n u_{ik}^m (\mathbf{x}_k - \mathbf{v}_i)^t (\mathbf{x}_k - \mathbf{v}_i)}{\sum_{k=1}^n u_{ik}^m}$ is the fuzzy covariance matrix of the i th cluster. This index should be low when clusters are compact, so the optimal number of clusters can be found by minimization of (10). In order to decrease the tendency of XB (8) to monotonically decrease when c tends to n , Kwon add a penalty function [13], yielding to an index to be minimized:

$$K(c) = \frac{J_m(U, V) + \frac{1}{c} \sum_{i=1}^c \|\mathbf{v}_i - \bar{\mathbf{v}}\|^2}{\min_{i,j=1,c;j \neq i} \|\mathbf{v}_i - \mathbf{v}_j\|^2} \quad (11)$$

where $\bar{\mathbf{v}} = \frac{1}{n} \sum_{k=1}^n \mathbf{x}_k$. Wu and Yang [21] propose a validity index defined by:

$$WY(c) = \sum_{i=1}^c \sum_{k=1}^n \frac{u_{ik}^2}{u_M} - \sum_{i=1}^c \exp\left(-\min_{j \neq i} \frac{\|\mathbf{v}_i - \mathbf{v}_j\|^2}{\beta_T}\right) \quad (12)$$

where $u_M = \min_{1 \leq i \leq c} (\sum_{k=1}^n u_{ik}^2)$ is the compactness of the most compact cluster and $\beta_T = \frac{1}{c} \sum_{i=1}^c \|\mathbf{v}_i - \bar{\mathbf{v}}\|^2$ is the total average distance measure for all clusters. A large value of WY means that the c clusters are compact and separated, so the optimal number of clusters can be found by maximizing (12). As compactness for each cluster is computed relative to u_M , WY gives good results in presence of outliers.

III. THE PROPOSED CLUSTER VALIDITY INDEX

A. Basic aggregation operators

Aggregation functions aim at associating a typical value to a number of several numerical values which are generally defined on a finite real interval or on ordinal scales. They are used in many fields, e.g. decision-making and pattern recognition as we assume in this paper, with no loss of generality, that they come from the unit interval. If not, a simple transformation can be found to make this true. Given c values, an aggregation operator is then a mapping $\Phi: [0, 1]^c \rightarrow [0, 1]$, $\mathbf{a} = \{a_1, a_2, \dots, a_c\} \mapsto \Phi(\mathbf{a})$. One finds many families, e.g.: triangular norms, OWA (*Ordered Weighted Averaging*) operators, γ -operators, or fuzzy integrals. They are classified either by mathematical properties they share or by the way the values are aggregated: conjunctives, disjunctives, compensatory, and weighted operators. An aggregation operator Φ is said to be *conjunctive* if $\Phi(\mathbf{a}) \leq \min\{a_1, a_2, \dots, a_c\}$, *disjunctive* if $\Phi(\mathbf{a}) \geq \max\{a_1, a_2, \dots, a_c\}$, and *compensatory* if $\min\{a_1, a_2, \dots, a_c\} \leq \Phi(\mathbf{a}) \leq \max\{a_1, a_2, \dots, a_c\}$, refer to [5] for a large survey on aggregation operators.

Triangular norms (t-norms) and conorms (t-conorms) are interesting conjunctive and disjunctive operators that generalize the logical AND and OR crisp operators to fuzzy sets. Briefly, a t-norm is a binary operation $\top: [0, 1]^2 \rightarrow [0, 1]$ which is commutative, associative, non decreasing and has 1 for neutral element. Alternatively, a t-conorm is the dual binary operation $\perp: [0, 1]^2 \rightarrow [0, 1]$ having the same properties except that its neutral element is 0, see [12] for a survey. Numerical results we give in in section IV are obtained using the *Standard*, *Algebraic* and *Hamacher* ($\gamma \in [0, +\infty[$) triangular norms respectively defined by:

- $a_1 \top_S a_2 = \min(a_1, a_2)$ and $a_1 \perp_S a_2 = \max(a_1, a_2)$
- $a_1 \top_A a_2 = a_1 a_2$ and $a_1 \perp_A a_2 = a_1 + a_2 - a_1 a_2$
- $a_1 \top_H a_2 = \frac{a_1 a_2}{\gamma + (1-\gamma)(a_1 + a_2 - a_1 a_2)}$ and $a_1 \perp_H a_2 = \frac{a_1 + a_2 - a_1 a_2 - (1-\gamma)a_1 a_2}{1 - (1-\gamma)a_1 a_2}$.

The discrete Sugeno integral is a fuzzy aggregation operator which computes the mean value of a function with respect to a fuzzy measure m . It is a non-additive measure of uncertainty, i.e. more general than a possibility one and therefore a probability one. The integral of a function μ is defined by

$$\mathcal{S}_m = \bigvee_{i=1}^n \mu(x_i) \wedge m(A_{(i)}) \quad (13)$$

where $A_{(i)}$ is the fuzzy subset $\{x_{(i)}, \dots, x_{(n)}\}$ with respect to a permutation so that $\mu(x_{(i)}) \leq \dots \leq \mu(x_{(n)})$, and \bigvee, \wedge are the maximum and minimum operators, respectively. This integral is widely used in decision making, and in particular for pattern recognition [10] because of its ability to model some kind of interaction between features describing a pattern \mathbf{x} .

B. Separation of clusters

Most separation measures between clusters are based on the distances between cluster centroids, e.g. (8-9), but it is seldom sufficient to interpret the geometrical structure of the

data, see [11] for examples. We propose to define a rule for the separation of clusters, for each point \mathbf{x}_k , based on its membership degrees \mathbf{u}_k (k^{th} column of U), using the l -order fuzzy OR operator (fOR- l) and the fuzzy exclusive OR operator (fXOR), both defined in [15] for the supervised classification problem with reject options. Let \mathcal{P} be the power set of $C = \{1, 2, \dots, c\}$ and $\mathcal{P}_l = \{A \in \mathcal{P} : |A| = l\}$ where $|A|$ denotes the cardinality of subset A , then the fOR- l associates to \mathbf{u}_k a single value $\underline{l}(\mathbf{u}_k) \in [0, 1]$ defined by:

$$\underline{l}(\mathbf{u}_k) = \bigwedge_{i=1, \dots, c}^l u_{ik} = \bigcap_{A \in \mathcal{P}_{l-1}} \left(\bigwedge_{j \in C \setminus A} u_{jk} \right) \quad (14)$$

It must be viewed as some kind of generalization of the notion of " l^{th} highest" value, with $l \in C$. Using standard t-norms, $\underline{l}(\mathbf{u}_k)$ is exactly the " l^{th} highest" element of \mathbf{u}_k . Given a fuzzy complement, e.g. $\overline{a_1} = 1 - a_1$, the fXOR associates a single value $\underline{\perp}(\mathbf{u}_k)$ to \mathbf{u}_k defined by:

$$\underline{\perp}(\mathbf{u}_k) = \bigwedge_{i=1, c} u_{ik} = \underline{1}(\mathbf{u}_k) \top \overline{\underline{2}(\mathbf{u}_k) / \underline{1}(\mathbf{u}_k)} \quad (15)$$

The value of fXOR is "high" if the "highest" value is large enough compared to the second "highest" one. Therefore, we introduce the following rule for separation, with respect to each \mathbf{x}_k :

*if u_{1k} is high xor \dots xor u_{ck} is high,
then \mathbf{x}_k belongs to a well separated cluster*

Using (14-15), it can be formally expressed by the satisfaction level $\tau^{(s)}(\mathbf{x}_k) = \underline{\perp}(\mathbf{u}_k)$, denoted $\tau_k^{(s)}$ for short.

C. Overlap of clusters

Non monotonic CVIs succeed on well separated clusters but they generally fail when some clusters naturally overlap. We propose to define a rule for the degree of overlap of clusters, for each point \mathbf{x}_k , based on its membership degrees \mathbf{u}_k using the aggregation operator $\Phi_{i,j}$ defined by the ratio of two Sugeno integrals in [14] for the supervised classification problem with reject options. Assuming each \mathbf{u}_k to be sorted in descending order, i.e. $u_{1k} \geq u_{2k} \geq \dots \geq u_{ck}$, let the blockwise similarity operator $\Phi_{i,j}$ taking values in $[0, 1]$, which quantifies the similarity of the block of values $\{u_{ik}, \dots, u_{jk}\}$ in \mathbf{u}_k , defined by:

$$\Phi_{i,j}(\mathbf{u}_k) = \begin{cases} \frac{\bigwedge_{l=\frac{i+j}{2}}^j u_{lk} \top \mathcal{K}_\lambda(l, j)}{\bigwedge_{l=\frac{i+j}{2}}^j u_{lk} \top \mathcal{K}_\lambda(l, i)} & \text{if } j-i \text{ is even} \\ \frac{\bigwedge_{l=\frac{i+j+1}{2}}^j u_{lk} \top \mathcal{K}_\lambda(l, j)}{\bigwedge_{l=\frac{i+j-1}{2}}^j u_{lk} \top \mathcal{K}_\lambda(l, i)} & \text{if } j-i \text{ is odd} \end{cases} \quad (16)$$

where $\mathcal{K}_\lambda(l, j)$ is a symmetrical kernel at resolution level $\lambda \in \mathbb{R}^+$, e.g. a gaussian kernel $\mathcal{N}_\lambda(l, j)$, taking values in $[0, 1]$ such that values within the block more or less contribute.

The larger λ is, the larger the contribution is, so increasing λ makes two consecutive values more similar but can increase the similarity of blocks of larger size, see [14] for proofs and details. A high value of $\Phi_{1,j}(\mathbf{u}_k)$ reveals that the j "highest" values are similar. Therefore, we introduce the following rule for degree of overlap, with respect to \mathbf{x}_k :

*if $\Phi_{1,2}(\mathbf{u}_k)$ is high or \dots or $\Phi_{1,c}(\mathbf{u}_k)$ is high,
then \mathbf{x}_k belongs to overlapping clusters*

Formally, it can be expressed, using (16), by the satisfaction level $\tau^{(o)}(\mathbf{x}_k) = \bigwedge_{j=1, c}^1 \Phi_{1,j}(\mathbf{u}_k)$, denoted $\tau_k^{(o)}$ for short.

D. Fuzzy modeling and the new CVI

In order to build up a new CVI, we propose to follow the so-called *fuzzy system modeling technique* [23], issued from the fuzzy control community. It expresses in terms of a two rules knowledge base:

- 1) *if \mathbf{x}_k belongs to a well separated cluster, then use the most satisfied corresponding measure*
- 2) *if \mathbf{x}_k does not belong to a well separated cluster or belongs to overlapping clusters, then use the least satisfied corresponding measure.*

Introducing two concepts *high* and *low*, as well as a measure $SO_{\top}(\mathbf{u}_k)$ (for *Separation-Overlap*), with respect to each point \mathbf{x}_k , depending on the dual couple (t-norm, t-conorm) at hand the corresponding fuzzy model gives:

Rule 1: if $\tau_k^{(s)}$ is high and $\tau_k^{(o)}$ is low, then $SO_{\top}(\mathbf{u}_k)$ is B_1

Rule 2: if $\tau_k^{(s)}$ is low or $\tau_k^{(o)}$ is high, then $SO_{\top}(\mathbf{u}_k)$ is B_2 where B_1 and B_2 are the following singleton fuzzy subsets:

$$B_1 = \left\{ 1 / \left(\tau_k^{(s)} \perp \tau_k^{(o)} \right) \right\} \quad \text{and} \quad B_2 = \left\{ 1 / \left(\tau_k^{(s)} \top \tau_k^{(o)} \right) \right\}.$$

Since t-norms and t-conorms model the *and* and *or* connectives, respectively, we get for the firing strength of *Rule 1* and *Rule 2*:

$$\rho_1 = \tau_k^{(s)} \top \overline{\tau_k^{(o)}} \quad \text{and} \quad \rho_2 = \overline{\tau_k^{(s)}} \perp \tau_k^{(o)}.$$

Then, by aggregating the effective output of each rule, we obtain the following system output:

$$O = \left\{ \frac{\rho_1}{\tau_k^{(s)} \perp \tau_k^{(o)}}, \frac{\rho_2}{\tau_k^{(s)} \top \tau_k^{(o)}} \right\}.$$

Depending the defuzzification method, several operators can be obtained. Using *center of area* method, we get the measure:

$$SO_{\top}(\mathbf{u}_k) = \frac{\rho_1 \times \left(\tau_k^{(s)} \perp \tau_k^{(o)} \right) + \rho_2 \times \left(\tau_k^{(s)} \top \tau_k^{(o)} \right)}{\rho_1 + \rho_2}.$$

Finally, we define the new CVI, called SOI_{\top} (*Separation-Overlap Index*), by simply averaging the resulting measure on X , taking values in $[0, 1]$, as follows:

$$SOI_{\top}(c) = \frac{1}{n} \sum_{k=1}^n SO_{\top}(\mathbf{u}_k) \quad (17)$$

The more separated and not overlapping the clusters are, the more $SOI_{\top}(c)$ is. Maximization of (17) gives the optimal

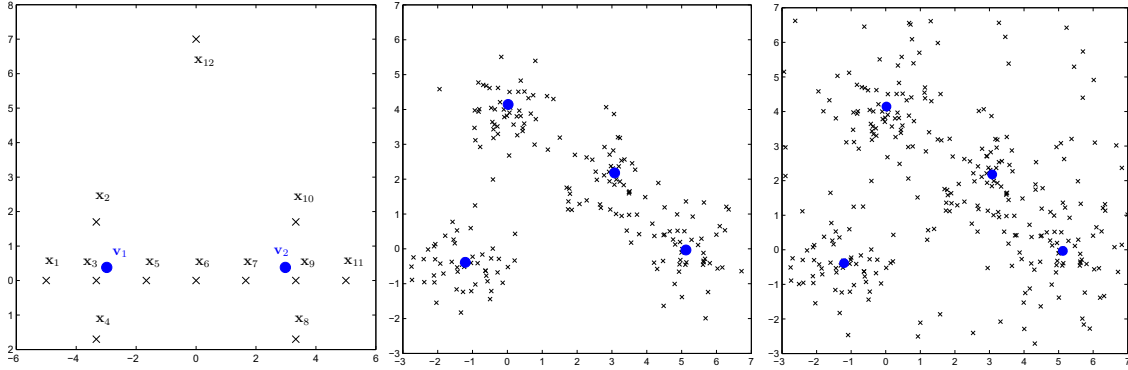
Fig. 1. Artificial data sets: *Diamond+* (left), D_1 (center) and D_2 (right), and their associated centroids represented in blue dots.

TABLE I

OVERLAP AND SEPARATION MEASURES FOR THE *Diamond+* DATA SET USING ALGEBRAIC T-NORMS AND T-CONORMS

	Point	\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}
$c = 2$	u_{1k}	.937	.957	.993	.908	.919	.500	.080	.091	.006	.042	.062	.500
	u_{2k}	.062	.042	.006	.091	.080	.500	.919	.908	.993	.957	.937	.500
	$\tau_k^{(s)}$.883	.917	.986	.833	.852	.500	.852	.833	.986	.917	.883	.500
	$\tau_k^{(o)}$.066	.044	.006	.100	.087	.999	.087	.100	.006	.044	.066	.999
	$SO_{\top_S}(\mathbf{u}_k, c)$.745	.812	.967	.658	.690	.499	.690	.658	.967	.812	.745	.499
$c = 3$	u_{1k}	.912	.871	.999	.912	.872	.450	.086	.056	.000	.057	.045	.000
	u_{2k}	.045	.057	.000	.056	.086	.450	.872	.912	.999	.871	.912	.000
	u_{3k}	.042	.070	.000	.030	.040	.098	.041	.030	.000	.070	.042	.998
	$\tau_k^{(s)}$.848	.791	.998	.847	.792	.550	.792	.847	.998	.791	.848	.997
	$\tau_k^{(o)}$.135	.206	.001	.147	.220	.999	.220	.147	.001	.206	.135	.001
$SO_{\top_S}(\mathbf{u}_k, c)$.668	.585	.994	.664	.585	.550	.585	.664	.994	.585	.668	.994	

number of clusters. Since each rule is defined with respect to a given dual couple (t-norm, t-conorm) the proposed cluster validity index is a family of indexes.

Proposition 1: if U is a hard c -partition, then $SOI_{\top}(c) = 1$ whatever c , for all couple (\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 $u_{1k} = 1$ and $u_{2k} = \dots = u_{ck} = 0$. In this case, we have $\tau_k^{(s)} = 1$: as 0 is the absorbing element of \top , it is easy to check in Eq. (14)

and $l = 2$ that $\perp = 0$. Therefore, replacing in Eq. (15), we have $\tau_k^{(o)} = 1 \top 1 = 1$, whatever c and (\top, \perp) .

We also have $\tau_k^{(o)} = 0$: here again, since 0 is the absorbing element of \top , we have $\Phi_{1,j} = \mathbf{0}/(\mathbf{0} \perp 1 \top 1) = 0$, for $j \geq 2$, whatever (\top, \perp) . Consequently, we obtain $\rho_1 = 1 \top 1 = 1$, and $\rho_2 = 0 \perp 0 = 0$.

Finally, we get $SO_{\top}(\mathbf{u}_k) = \frac{1 \times (1 \perp 0) + 0 \times (1 \top 0)}{1 + 0} = 1$, which gives $SOI_{\top}(c) = 1$, whatever c , for all (\top, \perp) . \square

Proposition 2: if U is a totally fuzzy c -partition, then $SOI_{\top_S}(c) = 0$ whatever c .

Proof: for all k , $u_{ik} = \frac{1}{c}$. By properties of $\Phi_{i,j}$, we have $\tau_k^{(o)} = 1$ whatever c and (\top, \perp) . Thus, we obtain $\rho_1 = 0 \top \tau_k^{(s)} = 0$, and $\rho_2 = 1 \perp \tau_k^{(s)} = 1$.

Finally, we get $SO_{\top}(\mathbf{u}_k) = \frac{0 \times (\tau_k^{(s)} \perp 1) + 1 \times (\tau_k^{(s)} \top 1)}{1 + 0} = \tau_k^{(s)}$.

Using $(\top, \perp)_S$, we have $\perp \mathbf{u}_k = \perp \mathbf{u}_k = \frac{1}{c}$ by (14) and

$\tau_k^{(s)} = \min\{\frac{1}{c}, 1 - 1\} = 0$, so that $SOI_{\top_S}(c) = 0$, whatever c . \square

Let us illustrate the ability of the proposed index to find the right number of clusters and the right partition on a small example, inspired by [17], that we call *Diamond+*. It consists of the eleven two-dimensional points first introduced by Windham [20] and an outlier, see Figure 1-left. Besides the outlier, the correct partition is composed of $c^* = 2$ clusters: the two touching diamond shapes. CVIs that only consider compactness and separation will select three clusters: the two touching diamond shapes and the outlier. Table I gives the detailed values of the satisfaction levels $\tau_k^{(s)}$ and $\tau_k^{(o)}$ as well as the measures $SO_{\top}(\mathbf{u}_k)$ for the twelve points, with $c = 2$ and $c = 3$. The algebraic triangular norms are used and the values of the new index are $SOI_{\top_A}(2) = 0.729$ and $SOI_{\top_A}(3) = 0.711$ showing that it recovers the natural partition in two clusters. In terms of separation and overlap, two points are of particular interest, namely \mathbf{x}_6 and \mathbf{x}_{12} .

For $c = 2$, both \mathbf{x}_6 and \mathbf{x}_{12} are the most overlapping points (0.999) and the most isolated points (0.500) as expected. The point \mathbf{x}_6 is equidistant to centroids \mathbf{v}_1 and \mathbf{v}_2 , so that it induces high overlapping and low separation. The point \mathbf{x}_{12} is equidistant to centroids \mathbf{v}_1 and \mathbf{v}_2 , but farther than \mathbf{x}_6 . Due to the sum constraint on \mathbf{u}_k , we have $u_{1,12} = u_{2,12} = 0.5$ so it has a high overlap and a low separation measure. For $c = 3$, \mathbf{x}_{12} becomes a well separated and not overlapping cluster, but

TABLE II
VALIDITY INDEXES ON ARTIFICIAL DATA D_1 (OPTIMAL VALUES ARE IN BOLD).

c	NPC	NPE	XB	FS $\times 10^{-3}$	K $\times 10^{-3}$	FHV	WY	SOI_{\top}		
								S	A	H_0
2	0.78	0.50	0.13	-1.48	0.32	3.97	1.64	0.67	0.58	0.58
3	0.80	0.54	0.07	-3.25	0.18	2.24	1.82	0.74	0.65	0.69
4	0.81	0.57	0.08	-2.95	0.20	1.83	2.20	0.79	0.69	0.72
5	0.73	0.78	0.29	-4.23	0.51	2.11	0.09	0.72	0.65	0.62
6	0.68	0.93	0.37	-1.02	0.54	2.35	-0.79	0.65	0.63	0.58
7	0.64	1.07	0.46	-3.75	0.61	2.28	-1.72	0.64	0.63	0.58
8	0.62	1.16	0.27	-0.88	0.57	2.25	0.99	0.57	0.65	0.58
9	0.59	1.25	0.32	-1.75	0.59	2.35	0.18	0.57	0.61	0.57
10	0.56	1.34	0.28	-1.17	0.55	2.32	-0.05	0.56	0.62	0.54

TABLE III
VALIDITY INDEXES ON NOISY DATA D_2 (OPTIMAL VALUES ARE IN BOLD).

c	NPC	NPE	XB	FS $\times 10^{-3}$	K $\times 10^{-3}$	FHV	WY	SOI_{\top}		
								S	A	H_0
2	0.74	0.58	0.18	-1.28	0.64	6.19	1.69	0.68	0.57	0.56
3	0.72	0.72	0.11	-4.10	0.45	4.48	2.24	0.71	0.63	0.58
4	0.71	0.83	0.13	-0.74	0.54	4.10	1.91	0.74	0.66	0.62
5	0.63	1.08	0.41	-3.44	0.94	5.12	0.84	0.68	0.63	0.60
6	0.59	1.21	0.40	-1.00	1.43	5.06	-0.23	0.59	0.59	0.59
7	0.56	1.34	0.42	-2.09	1.48	5.07	-0.95	0.60	0.58	0.57
8	0.55	1.41	0.40	-3.09	1.38	5.04	0.16	0.57	0.58	0.56
9	0.54	1.50	0.31	-0.99	1.37	5.25	-0.42	0.55	0.55	0.56
10	0.54	1.51	0.28	-1.01	1.27	5.29	-1.11	0.54	0.56	0.57

\mathbf{x}_6 remains the most overlapping point. The point \mathbf{x}_6 is still equidistant to \mathbf{v}_1 and \mathbf{v}_2 , but is far from the third centroid \mathbf{v}_3 , which is near \mathbf{x}_{12} , so that $u_{3,6}$ is low. Thus, resulting separation and overlap measures for \mathbf{x}_6 are almost the same than in case of $c = 2$. The point \mathbf{x}_{12} is close to \mathbf{v}_3 , leading to a high $u_{3,12}$ value, and low membership degree to the two previous clusters. This obviously gives a high separation and a low overlap value for \mathbf{x}_{12} . Averaging all $SO_{\top}(\mathbf{u}_k)$ values, in particular high values for points clearly belonging to clusters 1 and 2, makes the partition in $c = 3$ clusters less desirable than the one in $c = 2$ clusters, as shown by the proposed CVI values: $SOI_{\top A}(2) \geq SOI_{\top A}(3)$. This result clearly depends on the \mathbf{x}_{12} position. If it was farther, the membership degrees of the other points to the third cluster would become lower, resulting in a better partition in $c = 3$ clusters. Furthermore, if one adds one more point near \mathbf{x}_{12} , it would create a third cluster, and the mean value of $SO_{\top}(\mathbf{u}_k)$ would increase, leading to select $c = 3$ as the optimal number of clusters.

IV. NUMERICAL RESULTS

In this section, we evaluate the performance of the proposed CVI by conducting a comparison to the seven CVIs using the procedure described in section II for data sets of various structures (good separation, overlapping clusters, presence of outliers, additional noisy points) making the CV problem more or less easy. FCM is used with $m = 2$ and a termination parameter set to 10^{-3} . The optimal number of clusters is searched in the range $[c_{min} = 2, c_{max}]$ with $c_{max} = 10$ for data sets with high cardinality (n) and an

integer value close to \sqrt{n} for the others as usually done in the literature. For SOI_{\top} , we use a gaussian kernel $\mathcal{N}_{\lambda}(l, j) = \exp(-(l-j)^2/\lambda)$ in (16) with $\lambda = 5$ so that all the values in the blocks of size c have a high contribution to $\Phi_{1,c}$. Consequently, the choice of the t-norm has a little effect on the optimal value, otherwise the standard one tends to select a smaller number of clusters because it is not archimedean¹.

A. Artificial data sets

We generated an artificial data set D_1 containing $n = 200$ points composed of four bivariate gaussian distributions, two of them slightly overlapping, see Figure 1-center. Table II reports the obtained results for the tested CVIs, where bold values are the optimal ones indicating the selected number of clusters. We see that, for any t-norm, SOI_{\top} finds the optimal number of cluster while some classical indexes fail.

To emphasize the robustness to noisy data, we added 100 points drawn from a uniform distribution to generate another data set D_2 . This additional noisy points can make the FCM algorithm partitioning this second artificial data set into three clusters because the two least separated clusters tend to become one cluster with noise, see Figure 1-right. The corresponding results are given in Table III. Again, SOI_{\top} outperforms the classical indexes, none of them except FHV finding the optimal number of clusters (four).

B. Benchmark data sets

Additionally to data sets *Diamond+*, D_1 and D_2 , we compare SOI_{\top} on artificial and real benchmark data sets

¹a t-norm \top is archimedean if $\top(a, a) < a, \forall a \in [0, 1]$

TABLE IV

OPTIMAL NUMBER OF CLUSTERS SELECTED BY THE TESTED CVIS ON ARTIFICIAL AND BENCHMARK DATA SETS.

Data set	c_{max}	NPC	NPE	XB	FS	K	FHV	WY	SOI_{\top}			c^*
									S	A	H_0	
<i>Diamond+</i>	4	3	2	3	3	3	4	3	2	2	2	2
D_1	10	4	2	3	5	3	4	4	4	4	4	4
D_2	10	2	2	3	3	3	4	3	4	4	4	4
<i>X30</i>	5	3	3	3	4	2	3	3	3	3	3	3
<i>Bensaid</i>	7	3	2	3	7	3	3	6	3	3	3	3
<i>Iris</i>	10	2	2	2	3	2	3	2	2	2	2	2 or 3
<i>Starfield</i>	10	2	2	6	7	3	9	3	8	8	8	8 or 9
<i>Wine</i>	10	3	2	2	10	2	3	3	3	3	3	3
<i>Soybean</i>	7	3	2	3	4	3	2	3	3	4	4	4

commonly used in the literature:

- The artificial data set *X30* introduced in [3], consisting in 30 observations in \mathbb{R}^2 , for which 3 clusters are expected.
- The bi-dimensional artificial data set *Bensaid* [1] characterized by 3 classes of different cardinalities (6, 3 and 40).
- *Iris* [4], composed of 3 classes of 50 flowers each described by 4 physical attributes. Two classes have a substantial overlap in the feature space and the optimal number of clusters to be found is debatable: 2 or 3, see [3].
- The original *Starfield* [22], which contains the position and light intensity of 51 bright stars near Solaris. The expected number of clusters is 8 or 9, depending on the papers.
- *Wine* [4], which consists of 13 chemical attributes for 178 Italian wines belonging to 3 separable classes.
- *Soybean-small* [4], composed of four classes characterizing various diseases affecting soybean plants. Each of the 47 observations is described by 35 attributes.

The optimal number c^* of clusters for each data set and the one found by the different CVIs are given in Table IV. The proposed index always finds the optimal number of clusters except for the *Soybean-small* data set and standard t-norms because of the common value of λ , while the other do not.

V. CONCLUSION

We introduce a new family of cluster validity indexes for fuzzy partitions. They use new measures of separation and degree of overlap of the clusters based on triangular norms and a discrete Sugeno integral. Another novelty is that they lie on a fuzzy system modeling technique expressed in terms of a two rules knowledge base allowing to take into account the relative importance of each measure. Results on benchmark data sets of various structures show that the proposed indexes are more efficient than a large collection of indexes from the literature on cluster validity. Recommendations on the choice of the t-norms and the kernel function will be the subject of a longer forthcoming paper where we will show how their mathematical properties affect the indexes.

REFERENCES

- [1] A. M. Bensaid et. al., "Validity-Guided (Re)Clustering with Applications to Image Segmentation", *IEEE Trans. on Fuzzy Systems*, vol. 4, no. 2, pp. 112-123, 1996.
- [2] J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, 1981.
- [3] J.C. Bezdek and N.R. Pal, "Some new indexes of cluster validity", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 28, no. 3, pp. 301-315, 1998.
- [4] C. Blake and C. Merz, *UCI Repository of machine learning databases*, <http://www.ics.uci.edu/mllearn/MLRepository.html>, 1998.
- [5] T. Calvo and G. Mayor and R. Mesiar, *Aggregation Operators: New Trends and Applications*. Physica-Verlag, 2002.
- [6] R.N. Davé, "Validating fuzzy partitions obtained through c -shells clustering", *Pattern Recognition Letters*, vol. 17, no. 6, pp. 613-623, 1996.
- [7] J.C. Dunn, "Indices of partition fuzziness and the detection of clusters in large data sets", In: *Fuzzy Automata and Decision processes*, ed. by M.M. Gupta, Elsevier, 1977.
- [8] Y. Fukuyama and M. Sugeno, "A new method for choosing the number of clusters for the fuzzy c -means method", *Proc. 5th Fuzzy Systems Symposium*, pp. 247-250, 1989.
- [9] I. Gath and A.B. Geva, "Unsupervised optimal fuzzy clustering", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 773-780, 1989.
- [10] M. Grabisch, "Fuzzy pattern recognition by fuzzy integrals and fuzzy rules", In: *Pattern Recognition - From Classical to Modern Approaches*, ed. by S. Pal and P. Pal, pp. 257-280, World Scientific, 2002.
- [11] D-W. Kim, K.H. Lee and D. Lee, "On cluster validity index for estimating the optimal number of fuzzy clusters", *Pattern Recognition*, vol. 37, no. 3, pp. 2009-2025, 2004.
- [12] E.P. Klement and R. Mesiar (Eds), *Logical, algebraic, analytic and probabilistic aspects of triangular norms*. Elsevier, 2005.
- [13] S.H. Kwon, "Cluster validity index for fuzzy clustering", *Electronic Letters*, vol. 34, no. 22, pp. 2176-2177, 1998.
- [14] H. Le Capitaine and C. Frélicot, "A class-selective rejection scheme based on blockwise similarity of typicality degrees", *Proc. 19th Int. Conf on Pattern Recognition*, 2008.
- [15] L. Mascarilla, M. Berthier and C. Frélicot, "A k-order fuzzy OR operator for pattern classification with k-order ambiguity rejection", *Fuzzy Sets and Systems*, vol. 159, no. 2, pp. 2011-2029, 2008.
- [16] L. Mascarilla and C. Frélicot, "A class of reject-first possibilistic classifiers based on dual triples", *Proc. 9th Int. Fuzzy Systems Association World Congress*, pp. 743-747, 2001.
- [17] M-H. Masson and T. Denoeux, "ECM: An evidential version of the fuzzy c -means algorithm", *Pattern Recognition*, vol. 41, pp. 1384-1397, 2008.
- [18] M. Roubens, "Pattern classification problems and fuzzy sets", *Fuzzy Sets and Systems*, vol. 1, pp. 239-253, 1978.
- [19] W. Wang and Y. Zhang, "On fuzzy cluster validity indices", *Fuzzy Sets and Systems*, vol. 158, no. 19, pp. 2095-2117, 2007.
- [20] M.P. Windham, "Numerical classification of proximity data with assignment measure", *J. of Classification*, vol. 2, pp. 157-172, 1985.
- [21] K.L.Wu and M.S. Yang, "A cluster validity index for fuzzy clustering", *Pattern Recognition Letters*, vol. 26, no. 9, pp. 1275-1291, 2005.
- [22] X.L. Xie and G. Beni, "A validity measure for fuzzy clustering", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 841-847, 1991.
- [23] R.R. Yager and D.P. Filev, *Essentials of Fuzzy Modeling and Control*. Wiley, 1994.
- [24] H-J. Zimmerman and P. Zysno, "Quantifying vagueness in decision models", *European J. of Operational Research*, vol. 22, no. 2, pp. 148-158, 1985.