Blockwise similarity in [0,1] via triangular norms and Sugeno integrals – Application to cluster validity.

Hoel Le Capitaine, Thomas Batard, Carl Frélicot, Michel Berthier

Blockwise similarity in [0,1] via triangular norms and Sugeno integrals – Application to cluster validity

Hoel Le Capitaine, Thomas Batard, Carl Frélicot and Michel Berthier

***Abstract*— In many fields, e.g. decision-making, numerical values in [0,1] are available and one is often interested in detecting which are similar. In this paper, we propose an operator which is able to detect whether some values can be gathered by blocks with respect to their similarity or not. It combines the values and a kernel function using triangular norms and Sugeno integrals. This operator allows to estimate this blockwise similarity at different levels. For illustration purpose, we use it to define an index suitable for the cluster validity problem in pattern recognition.**

I. Introduction

The main topic of this paper is to define an indicator which measures, for a given c-tuple of values in $[0,1]$, whether some values can be gathered by blocks with respect to their similarity or not. For this purpose, we propose a new operator based on triangular norms and Sugeno integrals which combines the values and a kernel function. The resolution parameter of the kernel allows to view the induced similarity at different levels.

Such an operator can be used in many fields, in pattern recognition in particular and more specifically in supervised and unsupervised classification. Within this framework, the values to be aggregated generally express to which extent a pattern can either be associated to a specified class (supervised) or contribute to the definition of a particular cluster (unsupervised). Therefore, given a pattern, such a similarity operator is suitable for detecting, either ambiguities with respect to the classes at hand or a natural grouping tendancy.

The remainder of this paper is organized as follows. In section II, we briefly recall previous works that lead us to consider aggregation functions of a new kind. The blockwise similarity operator is proposed in section III. Properties and numerical examples are given. Next, in section IV, we use it to define an new index for cluster validity in the framework of fuzzy clustering. Results on artificial and real data show that the proposed index is performant thanks to the similarity operator.

II. Previous work

Aggregation functions or operators aim at combining (say c) numerical values. They are used in many fields, e.g. in multicriteria decision making and pattern recognition where values to be agregated are most often in $[0,1]$. Then, many families of functions have been defined that are formally mappings Φ: $[0,1]^c \to [0,1]$, $u = \{u_1, u_2, \cdots, u_c\} \mapsto \Phi(u)$, e.g. the triangular norms (briefly t-norm) and dual t-conorms, e.g. in subsection IV-B, refer to [10] for complete definitions and examples. Because of applications in pattern recognition we have in mind, we are interested in functions that qualify the similarity between some of the u_i's. For convenience, the u_i's are supposed to be sorted in decreasing order but of course operators do not need this assumption.

In [11], the $\overset{k}{\perp}$ operator has been defined as follows. Let $\mathcal{P}$ be the powerset of $C = \{1, 2, ..., c\}$ and $\mathcal{P}_k = \{A \in \mathcal{P} : |A| = k\}$ where $|A|$ denotes the cardinality of subset A, then

$$\overset{k}{\underset{i=1,c}{\perp}} u_i = \underset{A\in\mathcal{P}_{k-1}}{\top} \left(\underset{j\in C\setminus A}{\perp} u_j \right) \tag{1}$$

where $\top$ is a t-norm and $\perp$ is its dual t-conorm.

It must be viewed as some kind of generalization of the notion of "k^{th} bigger", with k in C. In particular, with standard triangular norms, $\overset{k}{\perp}$ is exactly the "k^{th} bigger". This operator satisfies nice mathematical properties (see [11] for details and applications).

By combining $\overset{k}{\perp}$, $k = 1, 2$, the authors built a fuzzy exclusive OR operator that extends the crisp XOR operator to the fuzzy context [6]:

$$\underset{i=1,c}{\overline{\perp}} u_i = \left(\overset{1}{\underset{i=1,c}{\perp}} u_i \right) \top \left(\overline{\overset{2}{\underset{i=1,c}{\perp}} u_i \Big/ \overset{1}{\underset{i=1,c}{\perp}} u_i} \right) \tag{2}$$

where $\overset{1}{\perp} = \perp$ and $\overline{(.)}$ is the fuzzy complement of $(.)$. Intuitively, it can be understood as u_1 being significantly high and u_2 significantly low while remaining greater than the other u_i's.

This leads us to investigate the question of how to define blockwise similarity and what kind of mathematical properties must be taken into account.

III. blockwise similarity

A. Definition

We define a blockwise similarity operator as a family of functions $\Phi_{j,k} : [0,1]^c \to [0,1]$, $(j,k) \in \{1, 2, \cdots, c\}^2$, $j < k$, satisfying the following four properties:

(P1) $\Phi_{j,k}(u) = 0$ whenever $u_j = 1$ and $u_k = 0$

(P2) $\Phi_{j,k} = 1 \Leftrightarrow u_j = u_k$

(P3) $\forall\ 0 \leq \varepsilon \leq u_{j-1} - u_j$, $\Phi_{j,k}(u_1, \cdots, u_j + \varepsilon, \cdots, u_c) \leq \Phi_{j,k}(u_1, \cdots, u_j, \cdots, u_c)$

All authors are with the Mathematics, Image and Applications (MIA) – Université de La Rochelle Avenue M. Crépeau, 17042 La Rochelle Cedex 1, FRANCE, email:{firstname.lastname}@univ-lr.fr

(P4) $\forall\ 0 \leq \varepsilon \leq u_{k-1} - u_k$, $\Phi_{j,k}(u_1, \cdots, u_k + \varepsilon, \cdots, u_c) \geq \Phi_{j,k}(u_1, \cdots, u_k, \cdots, u_c)$

The function $\Phi_{j,k}$ measures the similarity of u_i's belonging to the block bounded by j and k, and consequently $\Phi_{1,c}$ measures the similarity in the largest block, i.e. the total similarity in u. In the particular case when the sum of the u_i's is one, it follows from (P2) that $\Phi_{j,k} = 1$ whenever $u_1 = u_2 = \cdots = u_c$.
A straightforward solution to define blockwise similarity is to define $\Phi_{j,k}$ as the quotient $\overset{k}{\perp}/\overset{j}{\perp}$. However, it can be shown that this quotient does not satisfy the properties listed above.

B. The new operator

Since fuzzy integrals are able to model some kind of interaction between features, let us study the Sugeno integral. In the context described before the Sugeno integral (refer to [13] for complete definitions and examples) takes the following form:

$$\mathcal{S}_\mu(u) = \max_{i=1,c}[\min(u_i, \mu\{i, \cdots, c\})]$$

for a fuzzy measure μ on $\{1, 2, \cdots c\}$. This definition can be extended to any pair $(\top, \perp)$ of t-norm and t-conorm. $\mathcal{S}_\mu$ is then the fuzzy integral of u on subsets $\{i, \cdots c\}$ with respect to the fuzzy measure μ. and thus can be used to measure the ambiguity associated to u. Let $A_i = \{j, u_j \geq u_i\}$ and μ_k be defined by:

$$\mu_k(A_i) = \begin{cases} 0 & \text{if } Card(A_i) < k \\ 1 & \text{else} \end{cases} \tag{3}$$

we set

$$\overset{k}{\perp}(u) = \underset{i=1,\cdots,c}{\perp} u_i \top \mu_k(A_i) \tag{4}$$

(take care that these operators differ from those in II ; we adopt the same notations for simplicity).

It is easy to show that:

$$\overset{k}{\perp}(u) = \begin{cases} \underset{i=k,\cdots,c}{\perp} u_i & \text{if } u_{k-1} > u_k \\ \underset{i=j,\cdots,c}{\perp} u_i & j \text{ being defined by} \\ & u_{j-1} > u_j = \cdots = u_k \end{cases} \tag{5}$$

(with the convention: $u_0 > u_1$).

It can be proved that the operator $\Phi_{j,k} = \overset{k}{\perp}/\overset{j}{\perp}$ satisfies the four properties of subsection III-A for standard and algebraic t-norms. However, it is not fully convenient for measuring the similarity between u_j and u_k since the result depends on all the u_l's for $l \geq k+1$ while the u_l's for $l < j$ are not taken into account. To make the values between u_j and u_k meaningful, we introduce symmetrical kernel functions $\mathcal{K}_\lambda(i, l)$ centered at l and actually define:

$$\Phi_{j,k}(u) = \begin{cases} \dfrac{\underset{i=\frac{k+j}{2}}{\overset{k}{\perp}} u_i \top \mathcal{K}_\lambda(i,k)}{\underset{i=\frac{k+j}{2}}{\overset{j}{\perp}} u_i \top \mathcal{K}_\lambda(i,j)} & \text{if } k-j \text{ is even} \\ \dfrac{\underset{i=\frac{k+j+1}{2}}{\overset{k}{\perp}} u_i \top \mathcal{K}_\lambda(i,k)}{\underset{i=\frac{k+j-1}{2}}{\overset{j}{\perp}} u_i \top \mathcal{K}_\lambda(i,j)} & \text{if } k-j \text{ is odd} \end{cases} \tag{6}$$

Many kernel functions can be used, all being parametrized by a resolution parameter λ which controls its area of influence. Let us consider the particular case of a gaussian kernel $\mathcal{K}_\lambda(i, l) = \mathcal{N}_\lambda(i, l)$ defined by (see Figure 1):

$$\mathcal{N}_\lambda(i, l) = \exp \frac{-\pi(i-l)^2}{\lambda} \tag{7}$$

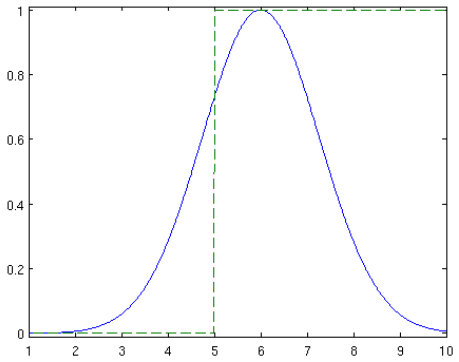

Fig. 1. Kernel $\mathcal{N}_{10}(i, l)$ and cardinal measure with $u_5 > u_6$ for $l = 6$

When λ tends to 0, this kernel becomes a dirac δ_k centered in k, the convergence being not uniform by continuity of $\mathcal{K}$ and discontinuity of δ_k. If the triangular norms used are continuous, then:

$$\lim_{\lambda \to 0} \Phi_{j,k}(u) = \begin{cases} 1 & \text{if } u_j = 0 \\ \frac{u_k}{u_j} & \text{otherwise} \end{cases} \tag{8}$$

and

$$\lim_{\lambda \to +\infty} \Phi_{j,k}(u) = \begin{cases} \dfrac{\underset{i=\frac{k+j}{2}}{\overset{k}{\perp}} u_i}{\underset{i=\frac{k+j}{2}}{\overset{j}{\perp}} u_i} & \text{if } k-j \text{ is even} \\ \dfrac{\underset{i=\frac{k+j+1}{2}}{\overset{k}{\perp}} u_i}{\underset{i=\frac{k+j-1}{2}}{\overset{j}{\perp}} u_i} & \text{if } k-j \text{ is odd} \\ 1 & \text{if } u_j = 0 \end{cases} \tag{9}$$

These results show how it is possible to adjust the weights that are given to the values between u_j and u_k: the contribution of u_{j+1}, ..., u_{k-1} is small if λ is close to zero and

increases with λ. Note however that $\Phi_{i-1,i} = u_{i-1}/u_i$ does not depend on λ in $\mathbb{R}^+$. This means that increasing λ will not make two consecutive u_i's more similar but may increase the similarity of blocks of larger size. Note also that if u_i is constant for all $i \in \{j, \ldots, k\}$, then $\Phi_{j,k}(u) = 1$ whatever (j, k), as expected in subsection III-A.

C. Numerical Examples

Given u, computation of $\Phi_{j,k}(u)$ for all $(i, j) \in C \times C$ result in a symmetrical table, see examples in Tables I and II for $u = \{0.9, 0.8, 0.68, 0.51, 0.5, 0.48, 0.32, 0.1\}$ obtained with kernels $\mathcal{N}_1$ and $\mathcal{N}_5$ respectively . Entries $\Phi_{j,k}(u)$ decrease as k increases for a fixed j ($k > j$). Detection of similarities then simply consists in exploring the upper triangular part of the table and compare the entries to a user-specified threshold s as follows:

- for $j = 1, c$ (row)
  - for $k = j + 1, c$ (column)
    if $\Phi_{j,k}(u) > s$, then $\{u_j, \ldots, u_k\}$ are similar.

TABLE I

$\Phi_{j,k}$ VALUES WITH $(\perp, \top)_S$ AND $\mathcal{K}_\lambda(i, j) = \mathcal{N}_1(i, j)$

$\Phi_{j,k}$	u	0.9	0.8	0.68	0.51	0.5	0.48	0.32	0.1
u	j, k	1	2	3	4	5	6	7	8
0.9	1	1.00	0.89	0.76	0.57	0.56	0.53	0.36	0.11
0.8	2	0.89	1.00	0.85	0.64	0.62	0.60	0.40	0.12
0.68	3	0.76	0.85	1.00	0.75	0.74	0.71	0.47	0.15
0.51	4	0.57	0.64	0.75	**1.00**	**0.98**	**0.94**	0.63	0.19
0.5	5	0.56	0.62	0.74	0.98	1.00	0.96	0.64	0.20
0.48	6	0.53	0.60	0.71	0.94	0.96	1.00	0.67	0.21
0.32	7	0.36	0.40	0.47	0.63	0.64	0.67	1.00	0.31
0.1	8	0.11	0.12	0.15	0.19	0.20	0.21	0.31	1.00

To understand the behaviour of the operator with respect to λ, let us compare the 4^{th} row of each table, i.e. $\Phi_{4,k}$ for $k = 4, \ldots, 8$:

- at resolution $\lambda = 1$, $\Phi_{4,4}$, $\Phi_{4,5}$ and $\Phi_{4,6}$ are greater than $s = 0.9$ and the corresponding u_k's $\{0.51, 0.5, 0.48\}$ are detected to be similar. Small value of $\Phi_{4,7}$ and $\Phi_{4,8}$ indicate that both $u_7 = 0.32$ and $u_8 = 0.1$ are not similar to the previous u_k's
- at resolution $\lambda = 5$, $\Phi_{4,7}$ becomes greater than s and $u_7 = 0.32$ is considered as being similar to $\{0.51, 0.5, 0.48\}$ while $u_8 = 0.1$ still not.

TABLE II

$\Phi_{j,k}$ VALUES WITH $(\perp, \top)_S$ AND $\mathcal{K}_\lambda(i, j) = \mathcal{N}_5(i, j)$

$\Phi_{j,k}$	u	0.9	0.8	0.68	0.51	0.5	0.48	0.32	0.1
u	j, k	1	2	3	4	5	6	7	8
0.9	1	1.00	0.89	0.76	0.59	0.57	0.56	0.53	0.36
0.8	2	0.89	1.00	0.85	0.67	0.64	0.63	0.60	0.40
0.68	3	0.76	0.85	1.00	0.75	0.75	0.74	0.71	0.47
0.51	4	0.59	0.67	0.75	**1.00**	**0.98**	**0.98**	**0.94**	0.63
0.5	5	0.57	0.64	0.75	0.98	1.00	0.96	0.96	0.64
0.48	6	0.56	0.63	0.74	0.98	0.96	1.00	0.67	0.67
0.32	7	0.53	0.60	0.71	0.94	0.96	0.67	1.00	0.31
0.1	8	0.36	0.40	0.47	0.19	0.64	0.67	0.31	1.00

IV. Application to cluster validity

A straightforward application of the proposed operator (6) to pattern recognition is supervised classification with ambiguity rejection. Let $\mathbf{x}$ be a pattern in a feature space, say $\mathbb{R}^p$, to be classified with respect to a set $\Omega = \{\omega_1, ..., \omega_c\}$ of c classes. Given a labelling function: $\mathbf{x} \mapsto u(\mathbf{x}) \in [0, 1]^c$ whose general term $u_i = u_i(\mathbf{x})$ is the posterior probability that $\mathbf{x}$ belongs to ω_i or a membership degree to a fuzzy set associated to ω_i, a decision rule is generally based on the aggregation of labels u_i ($i = 1, c$). By thresholding the values of $\Phi_{j,k}(u)$ for an incoming pattern $\mathbf{x}$, a reject option can easily be included. Table I gives a good example of how it could be done. A threshold $s = 0.9$ would result in rejecting $\mathbf{x}$ for ambiguity between the classes whose membership degrees are $\{0.51, 0.50, 0.48\}$. However, such reject option aims at reducing the misclassication risk, so it often focuses on subsets of degrees that include the higher one. Therefore, depending on the application, a particular attention to $\Phi_{1,k}(u)$ can be paid. This is clearly the case for cluster analysis, another task of major importance in pattern recognition we are interested in.

A. Cluster validity for fuzzy clustering and indexes

Clustering is an instance of unsupervised classification which aims at finding a structure of groups in set of n patterns $X = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$. In this framework, the label vectors $u_k = u(\mathbf{x}_k)$ are unknown and clustering algorithms can be used to obtain them from X. For instance, the *fuzzy c-means* (FCM) algorithm partitions X into $c > 1$ clusters by minimizing the following objective function [2]:

$$J_m(U, V) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^m \, ||\mathbf{x}_k - \mathbf{v}_i||^2 \tag{10}$$

where u_{ik} is the membership degree of $\mathbf{x}_k$ to the i^{th} cluster represented by its centroid $\mathbf{v}_i \in \Re^p$. Centroids are gathered into a $(c \times p)$ matrix $V = [\mathbf{v}_1, ..., \mathbf{v}_c]$. Degrees u_{ik} are subject to $\sum_{i=1}^{c} u_{ik} = 1$ for all x_k in X, to $0 < \sum_{k=1}^{n} u_{ik} < n$ ($\forall i = 1, c$), and are elements of the fuzzy $c-$partition matrix U ($c \times n$). The so-called *fuzzyfier* $m > 1$ is a weighting exponent which makes the resulting partition more or less fuzzy [12]. The higher m is, the softer the cluster boundaries are. Minimization of (10) is obtained by iteratively updating (U, V) as follows:

$$u_{ik} = 1 \Big/ \sum_{j=1}^{c} \left(\frac{||\mathbf{x}_k - \mathbf{v}_i||}{||\mathbf{x}_k - \mathbf{v}_j||} \right)^{2/(m-1)} \tag{11}$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^{n} u_{ik}^m \, \mathbf{x}_k}{\sum_{k=1}^{n} u_{ik}^m} \tag{12}$$

The usual euclidian norm $||.||$ induces hyperspherical clusters, hence FCM can only detect clusters with the same shape and orientation. In [8], a variant called FCM-GK has been proposed by extended FCM to cluster-dependent norms $||.||_{A_i}$ in order to detect clusters of different geometrical shapes. This results in modifying the objective function (10) as $J_m(U, V, A)$ where A is a $c-$tuple of norm-inducing

matrices A_i taking part in the minimization process, hence to be iteratively updated. To obtain a feasible solution, the determinant of these matrices are constrained allowing to optimize the clusters' shapes while their volumes remain constant (see [2], [8] for details).

Validating the provided clustering of X consists in assessing whether the resulting partition reflects the data structure or not. Since c is a user-defined parameter of clustering algorithms such as FCM, most of works on cluster validity focus on the number of clusters problem. Many validity indexes have been proposed for fuzzy clustering (refer to [4], [9], [14] for comparative studies). They can be classified in two main categories. The first one is composed of indexes that only use membership degrees (U). Let us cite the *Partition Coefficient* [2], taking values in $[\frac{1}{c}, 1]$:

$$PC(c) = \frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^2 \tag{13}$$

or the *Partition Entropy* [1], taking values in $[0, \log(c)]$:

$$PE(c) = -\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}\, log(u_{ik}) \tag{14}$$

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 reduce this tendency, *e.g.* in [5]. 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 [12], [15]:

$$XB(c) = \frac{J_m(U,V)\,/n}{\min_{i,j=1,c;j\neq i}\ ||\mathbf{v}_i - \mathbf{v}_j||^2} \tag{15}$$

or the Fukuyama-Sugeno index [7]:

$$FS(c) = J_m(U,V) - \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m\, ||\mathbf{v}_i - \overline{\mathbf{v}}||^2 \tag{16}$$

where $\overline{\mathbf{v}}$ is the mean of centroids. Both XB and FS combine the FCM objective function (10) which measures how much 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 more optimal c is.

B. A new index

Since the blockwise operator $\Phi_{j,k}$ (6) presents a special case ($j = 1, k = c$) which can reflect the overall similarity of u_k's components, it reflects the overall ambiguity of pattern $\mathbf{x}_k$ with respect to the c clusters at hand. Therefore, a very simple cluster validity index belonging to the first category can be derived by averaging $\Phi_{1,c}(u_k)$ over the columns of U. Given a c-partition matrix U resulting from a fuzzy clustering algorithm (FCM, FCM-GK, ...), we define the BwS (*BlockWise Similarity*) index by:

$$BwS(c) = \frac{1}{n}\sum_{k=1}^{n} \Phi_{1,c}(u_k) \tag{17}$$

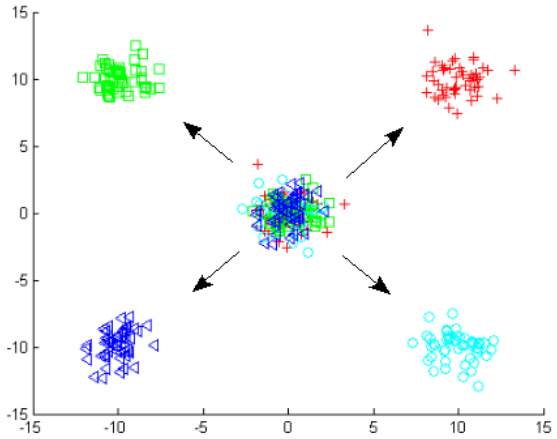

Fig. 2. α-separated data sets – $\alpha = 1$ and 10

The least valid c-partition arises when U is totally fuzzy, *i.e.* $u_{ik} = \frac{1}{c}$ for all $i = 1, c$. Then $\Phi_{1,c}(u_k) = 1$ by (P2) for all u_k in X and so $BwS(c)$ whatever c. On the other hand, the most valid c-partition arises when U is hard, *i.e.* $u_{ik} \in \{0, 1\}$. Then $\Phi_{1,c}(u_k) = 0$ by (P1) and $BwS(c) = 0$ whatever c. The more separated clusters are, the less BwS, and minimizing (17) gives the optimal number of clusters $c^\star$. In practice, $BwS(c)$ is computed for c varying from 2 up to c_{max} and $c^\star$ will correspond to a knee.

Recalling that $\Phi_{j,k}(u)$ defines a family of operators because of the many choices for the pair $(\top, \perp)$ and the kernel function $\mathcal{K}_\lambda$, therefore $BwS(c)$ is a family of validity indexes. In the remaining part of the paper, we present numerical results using the following basic norms:

- *Standard*: $a \top_S b = min(a, b)$ and $a \perp_S b = max(a, b)$
- *Algebraic*: $a \top_A b = a\, b$ and $a \perp_A b = a + b - a\, b$
- *Lukasiewicz*: $a \top_L b = max(a + b - 1, 0)$ and $a \perp_L b = min(a + b, 1)$

Among the possible kernel functions, we used the gaussian one (7). The resolution parameter λ must be set with great care, depending on the application and the magnitude of the u_i to be agregated. For instance, since U is fuzzy, degrees u_{ki} become as similar as c increases because of the normalization constraint. So, for the fuzzy cluster validity application, we recommend to chose a low λ in order to not take into account too many degrees that are similar only because of this constraint. In a further study, we will propose an upper bound for λ as a function of c which will probably result in modifying BwS.
In next subsections, we will use either the FCM algorithm or the FCM-GK one with the settings: $m = 2$, a threshold $\epsilon = 10^{-5}$ for termination criterion and a maximum of 100 iterations.

C. Artificial data sets

Experiment #1:
A series of 10 data sets was generated, each composed of 800 points drawn from a mixture of $c = 4$ bivariate normal

distributions. The covariance matrix of each component is the same $\Sigma_i = I$ $(i = 1, c)$ and the mean vectors are:

- $\mu_1 = (0 \quad 0)^t + \alpha\,(1 \quad 1)^t$,
- $\mu_2 = (0 \quad 0)^t + \alpha\,(1 \quad -1)^t$,
- $\mu_3 = (0 \quad 0)^t + \alpha\,(-1 \quad -1)^t$ and
- $\mu_4 = (0 \quad 0)^t + \alpha\,(-1 \quad 1)^t$

for increasing values of $\alpha = 1, 2, \ldots, 10$. This successively moves the clusters in opposite directions, creating less overlap as the clusters become more and more separated. The first and last data sets are shown in Figure 2. Each data set was then clustered using FCM with $c = 4$, providing a fuzzy partition matrix U_α. Corresponding values of BwS for the different basic norms are plotted in Figure 3 as a function of α. As expected, BwS decreases towards 0 as α increases whatever the norms.

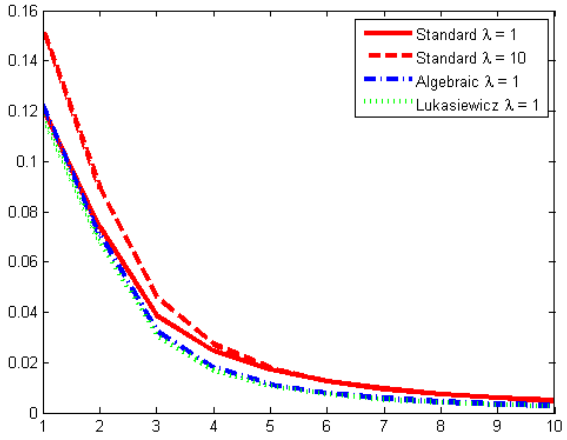

Fig. 3. BwS for α-separated data sets – $\alpha = 1$ to 10

Experiment #2:
In order to compare the proposed index to the classical ones recalled in subsection IV-A, we generated a data set containing $n = 200$ points consisting of 50 points each drawn from a mixture of $c = 4$ bivariate normal distributions with various ellipsoidal shapes. FCM-GK was used with $c_{max} = 10$ and an efficient index should find $c^\star = 4$. Table III reports the results obtained for the tested indexes. Optimal values are boldfaced and acceptable ones are italicized. We can see that BwS always gives the right number of clusters whatever λ while some classical indexes fail. The centroids (12) resulting from clustering with $c^\star = 4$ are represented by special symbols (•) in Figure 4.

Experiment #3:
The last artificial data set is similar to the previous one except that clusters are spherically shaped and 100 points drawn from a uniform distribution were added, as shown in Figure 5. These additional points act as noise and can make the FCM algorithm partitioning the data set in more than $c^\star = 4$ clusters. FCM was used with $c_{max} = 10$ and comparative results of the tested validity indexes are given in Table IV. None of the classical indexes was able to detect the right number of clusters while BsW succeed whatever $(\top, \perp)$.

TABLE III
VALIDITY INDEXES ON ELLIPSOIDAL CLUSTERS

c	PC	PE	XB	FS $\times 10^{-3}$	BwS with $(\top, \perp)_S$ and $\mathcal{N}_\lambda$ $\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$
2	0.790	**0.499**	0.132	-2.164	0.177	0.177	0.177
3	0.816	0.511	0.067	-0.438	0.057	0.061	0.089
4	**0.822**	*0.536*	**0.067**	-5.058	**0.027**	**0.033**	**0.044**
5	0.760	0.715	0.329	-3.391	0.021	0.028	0.033
6	0.721	0.841	0.259	-1.386	0.017	0.022	0.023
7	0.681	0.915	0.195	**-5.545**	0.013	0.016	0.016
8	0.651	1.059	0.336	-0.609	0.010	0.013	0.013
9	0.636	1.126	0.284	-0.398	0.009	0.012	0.012
10	0.624	1.176	0.265	-1.287	0.008	0.010	0.010

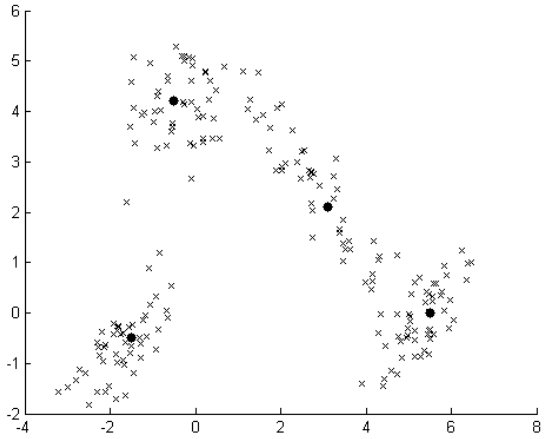

Fig. 4. Optimal $c^\star$ centroids for ellipsoidal clusters

Moreover, multiple runs showed us that it gives more stable results, showing its better robustness to noisy data.

TABLE IV
VALIDITY INDEXES ON NOISY DATA

c	PC	PE	XB	FS $\times 10^{-3}$	BwS with $(\top, \perp)$ and $\mathcal{N}_1$ S	A	L
2	**0.752**	**0.572**	0.188	-1.416	0.236	0.236	0.236
3	0.734	0.708	**0.109**	-4.186	0.102	0.110	0.102
4	0.731	0.782	0.129	-1.644	**0.050**	**0.053**	**0.050**
5	*0.691*	0.948	0.134	**-4.409**	0.040	0.040	0.038
6	0.596	1.202	0.543	-3.411	0.037	0.035	0.034
7	0.588	1.277	0.501	-2.508	0.031	0.031	0.030
8	0.565	1.386	0.432	-4.032	0.027	0.028	0.027
9	0.541	1.460	0.367	-1.795	0.022	0.023	0.022
10	0.525	1.552	0.398	-4.104	0.020	0.020	0.020

D. Real data sets [3]

Iris data:
The iris data set contains $n = 150$ observations from three 4-dimensional classes (iris species) of 50 points each. It is one of the most used benchmarks in pattern recognition, especially for cluster validity because two classes have a substancial overlap in the feature space. Therefore, the number of clusters to be found is debatable, *e.g.* in [4], some authors claiming that the right physical number $c = 3$ has to be detected while others say that the geometrical number is $c = 2$, so a good index should detect one of these two values as being $c^\star$. Indexes that only use U are a priori more prone to merge the two overlaping classes into a single cluster because they do not combine compactness and separation measures like the ones that use (U, V, X). As

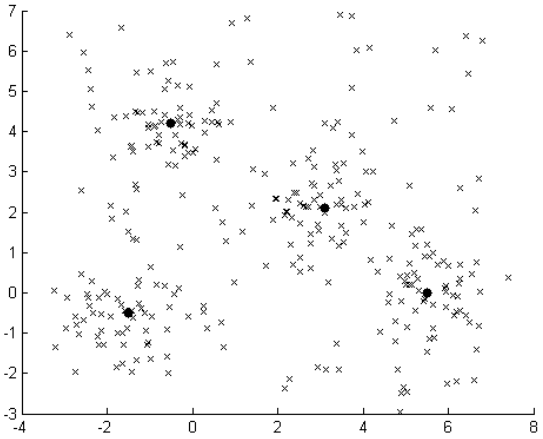

Fig. 5. Optimal $c^\star$ centroids for noisy clusters

the classes are known to have a hyperellipsoidal shape, we used FCM-GK with $c_{max} = 10$. It can be seen in Table V that all indexes exhibit one of the expected optimal numbers of clusters showing their ability in assessing the structure of the data and that the debate is not closed. However, it is worthnoting that BsW, despite it only uses U (like PC and PE), overcomes this limitation because small values of $\Phi_{1,c}$ and therefore the absence of similarity blocs (in average) clearly indicates that the clusters are well separated.

TABLE V

VALIDITY INDEXES ON IRIS DATA

c	PC	PE	XB	FS $\times 10^{-3}$	BwS with $(\top, \perp)_S$ and $\mathcal{N}_\lambda$ $\lambda = 0.5$	$\lambda = 1$	$\lambda = 2$
2	**0.738**	**0.589**	**0.027**	-4.761	0.289	0.289	0.289
3	0.727	0.671	0.192	**-4.845**	**0.022**	**0.051**	**0.179**
4	0.620	1.006	0.222	-3.030	0.015	0.054	0.140
5	0.534	1.291	0.264	-1.669	0.011	0.060	0.110
6	0.482	1.434	1.363	-2.744	0.009	0.051	0.073
7	0.458	1.583	1.172	-2.390	0.008	0.017	0.016
8	0.440	1.693	0.929	-1.357	0.007	0.011	0.012
9	0.432	1.789	0.983	-2.771	0.006	0.010	0.010
10	0.411	1.876	1.256	-1.223	0.002	0.003	0.004

Glass data:

This last set contains 214 observations of $c = 6$ types of glass that one can find in the scene of the crime (building window, vehicule window, container, headlamp, ...) described by 9 physical and chemical attributes. As shown in Table VI, BsW is the only index which was able to select the right number of clusters.

V. CONCLUSION

In this paper, we have proposed a new operator which estimates, given a c-tuple of values in [0,1], the similarity of some components. Based on triangular norms and Sugeno integrals, it combines the values and a kernel function. We have demonstrate how the definition of this operator makes it able to detect blockwise similarities at different levels of resolution via the kernel.

TABLE VI

VALIDITY INDEXES ON GLASS DATA

c	PC	PE	XB	FS $\times 10^{-3}$	BwS with $(\top, \perp)$ and $\mathcal{N}_1$ S	A	L
2	**0.807**	**0.457**	**0.224**	**-9.123**	0.189	0.189	0.189
3	0.666	0.853	0.489	-7.519	0.144	0.143	0.134
4	0.634	0.995	0.590	-7.157	*0.082*	*0.078*	*0.075*
5	0.499	1.367	2.988	-5.628	0.076	0.072	0.069
6	0.493	1.437	2.357	-5.561	**0.053**	**0.043**	**0.040**
7	0.467	1.592	1.973	-4.954	0.049	0.037	0.035
8	0.408	1.824	1.649	-4.603	0.047	0.035	0.033
9	0.380	1.985	2.211	-4.389	0.046	0.034	0.032
10	0.377	2.031	1.921	-4.297	0.042	0.031	0.029

Among the applications that can be considered, we have chosen to present a solution to the cluster validity problem in pattern recognition. For this purpose, we have proposed a new index based on the blockwise similarity operator. Given results show its performance when compared to classical indexes. Further works will concern the different choices the pratitioner must make (t-norms, kernel functions and their resolution parameter) to use the blockwise similarity operator as well as its application to selective ambiguity rejection in pattern classification.

REFERENCES

[1] J.C. Bezdek, "Cluster validity with fuzzy sets", *Journal of Cybernetics*, 3:5873, 1974.

[2] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New-York, 1981.

[3] C.L. Blake and C.J. Merz, UCI repository of machine learning databases, 1998.

[4] J.C. Bezdek and N.R. Pal, "Some new indexes of cluster validity", *IEEE Transactions on Systems, Man and Cybernetics*, 28(3):301-315, 1998.

[5] R.N. Davé, "Validating fuzzy partition obtained through c-shells clustering", *Pattern Recognition Letters*, 17:413-623, 1996.

[6] C. Frélicot, L. Mascarilla and M. Berthier, A new cluster validity index for fuzzy clustering based on Combination of dual triples, *Proc. IEEE International Conference on Fuzzy systems*, Vancouver, Canada, 2006.

[7] 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*, 247-250, July 1989.

[8] D.E. Gustafson and W.C. Kessel, "Fuzzy clustering with fuzzy covariance matrix", *Proc. IEEE Conference on Decision and Control*, 761-766, San Diego, California, 1979.

[9] D-W. Kim, K.H. Lee and D. Lee, "On cluster validity index for estimating the optimal number of fuzzy clusters", *Pattern Recognition*, 37(3):2009-2025, 2004.

[10] E.P. Klement and R. Mesiar (Eds), Logical, algebraic, analytic and probabilistic aspects of triangular norms. Elsevier, 2005.

[11] L. Mascarilla, M. Berthier and C. Frélicot, A K-order Fuzzy OR Operator-Application in Pattern Classification with k-order Ambiguity. *Proc. of 11th International Conference on Information Processing and Management of Uncertainty in Knowledge-based Systems*, Paris, France, 2006.

[12] N.R. Pal and J.C. Bezdek, "On cluster validity for fuzzy c-means model", *IEEE Transactions on Fuzzy Systems*, 3:370-379, 1995.

[13] T. Terano, K. Asai and M. Sugeno, Fuzzy Systems Theory and Its Applications. Academic Press, 1992.

[14] K-L. Wu and M-S Yang, "A cluster validity index for fuzzy clustering", *Pattern Recognition Letters*, 26(3):1275-1291, 2005.

[15] X.L. Xie and G. Beni, "A validity measure for fuzzy clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:841-847, 1991.