Tatra Mt. Math. Publ. **21** (2001), 93–104

TATRA MOUNTAINS
Mathematical Publications

T_∞ DIVERGENCE FUNCTIONS: THE MULTIDIMENSIONAL CASE

Emilio Torres — Pedro Miranda — Pedro Gil

ABSTRACT. In this paper, we characterize divergence measures among n fuzzy sets when Łukasiewicz t-norm is used.

1. Introduction and conclusions

The comparison of description of objects is a usual operation in main domains: psychology, analogy,… This comparison is frequently achieved through a measure intended to determine to which extent descriptions bear resemblance to another one or differ from others one, and consequently, measures of comparison have various forms, depending on their use. Generally they can be divided in two classes ([2]): measures of similitude and measures of dissimilarity.

On the one hand, measures of similitude are used when we consider a reference object or class and we decide if a new object is compatible with it or correspond to the reference object. This situation is typical in prototype based reasoning, when the references are prototypes and a new object must be associated with one of them.

On the other hand, the dissimilarity between object evaluates to which extent they are different. This quantity may be useful when, in the retrieved step of a case-based reasoning system, no case is sufficiently similar to the new case. It is then interesting to be able to establish comparison with respect to difference between descriptions, and to choose the least different case from the new one.

Many measures of similitude have been proposed and studied (e.g., [9], [4]), generally in a given framework of application. A more formal approximation to measure of dissimilarity based on fuzzy equivalence relation was first proposed in [1] and an additional research can be found in [5]. In [2], there is a general overview about general measures of comparison, considering whether the

2000 Mathematics Subject Classification: 03E02, 62H30, 68T10.
Keywords: divergence measure, partitions, real functions.

Partially supported by PB97-1282 and PB97-1286.

elements which belong to a class and not to another class, or viceversa, or to both of them.

In this sense, [7] has proposed a measure of dissimilarity that tries to maintain the properties of classical divergence measures between two probability distributions which appear in Information Theory. It compares imperfect descriptions, affected with imprecisions and inaccuracies, represented by fuzzy sets, and it is based on an axiomatic approach through union and intersection of fuzzy sets.

Originally, it only measures the difference between two fuzzy sets, and it was only developed with union and intersection of fuzzy sets based on the minimum t-norm and on the maximum t-conorm, respectively. But in some cases, it is more interesting to use Łukasiewicz t-norm, specially when we are dealing with partitions ([6]).

In such cases, we must consider not only the divergence between two fuzzy sets, but the divergence existing among n fuzzy sets. In this paper we present a first step of this study, extending the original definition, so we are now able to measure global divergence among n fuzzy sets. It is showed that aggregation operators are very useful to quantify the global divergence, specially when they are based on the differences of the membership functions. Also, we may apply this measure in the context of hierarchical clustering analysis.

2. Preliminaries

2.1 Aggregation operators

Let us start with the most general definition of an aggregation operator, as given by [3], for instance:

Definition 1. An n ary aggregation operator is a $[0,1]^n \to [0,1]$ mapping s that satisfies:

1. boundary conditions: $s(0,\dots,0)=0$ and $s(1,\dots,1)=1$;
2. monotonicity: for any $(x_1,\dots,x_n)$ and $(y_1,\dots,y_n)$in $[0,1]^n$ it holds that if $x_i \le y_i$, $\forall i \in \{1,\dots,n\}$, then $s(x_1,\dots,x_n) \le s(y_1,\dots,y_n)$.

From here on, the arity n will be fixed and we will simply talk about aggregation operators. Of course, in practice the above definition is far too general, and additional properties have to be imposed, depending upon the context. One could consider, for instance, continuity (so it guarantees that an infinitesimal variation in any argument of s does not produce a noticeable change in the aggregate), symmetry, idempotency, etc. Recall that an aggregation operator s is called symmetric if for any permutation σ of $\{1,\dots,n\}$ and any

$(x_1, \ldots, x_n) \in [0,1]^n$ it holds that

$$s(x_1, \ldots, x_n) = s(x_{\sigma(1)}, \ldots, x_{\sigma(n)}) .$$

An aggregation operator is called idempotent if $\forall x \in [0,1]$, $s(x, \ldots, x) = x$. It is well known that an aggregation operator is idempotent if and only if for any $(x_1, \ldots, x_n)$ in $[0,1]^n$ it holds that

$$\min(x_1, \ldots, x_n) \leq s(x_1, \ldots, x_n) \leq \max(x_1, \ldots, x_n) .$$

A *weighting vector* is a n tuple $(w_1, \ldots, w_n)$ in $[0,1]^n$ such that $\sum_{i=1}^{n} w_i = 1$.

We recall some of well-known aggregation operators' definitions:

Definition 2. Consider a continuous, strictly monotone $[0,1] \to \overline{\mathbb{R}}$ mapping f such that $\{f(0), f(1)\} \neq \{-\infty, +\infty\}$, and a weighting vector $(w_1, \ldots, w_n)$ such that $0 < w_i < 1$, for any $i \in \{1, \ldots, n\}$. A *quasi linear weighted mean* is a $[0,1]^n \to [0,1]$ mapping s defined by

$$s(x_1, \ldots, x_n) = f^{-1}\left(\sum_{i=1}^{n} w_i f(x_i)\right) .$$

Clearly, a quasi-linear weighted mean is symmetric only if all weights are identical. Symmetric quasi-linear weighted means are usually called quasi-arithmetic means. The most important ones are the following:

1. The arithmetic mean: $s(x_1, \ldots, x_n) = \frac{1}{n} \sum_{i=1}^{n} x_i$.
2. The geometric mean: $s(x_1, \ldots, x_n) = \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}$.
3. The harmonic mean: $s(x_1, \ldots, x_n) = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} \frac{1}{x_i}}$.
4. The quadratic mean: $s(x_1, \ldots, x_n) = \left(\frac{1}{n} \sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$.

Definition 3. An n ary *ordered weighted averaging* (OWA) operator is a $[0,1]^n \to [0,1]$ mapping s defined by

$$s(x_1, \ldots, x_n) = \sum_{i=1}^{n} w_i x_{(i)} ,$$

where $\cdot_{(i)}$ indicates that the indices have been permuted so that $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(n)}$.

Note that the arithmetic mean is an OWA operator with the weighted vector $\left(\frac{1}{n}, \ldots, \frac{1}{n}\right)$.

2.2 Divergence Measure

Let us give the definition of divergence measure.

Definition 4. [7] Let Ω be the universal set and let $\widetilde{\mathcal{P}}(\Omega)$ be the set of all fuzzy subsets. An application $D\colon \widetilde{\mathcal{P}}(\Omega) \times \widetilde{\mathcal{P}}(\Omega) \to \overline{\mathbb{R}}$ is said to be a *divergence measure* between fuzzy subsets if and only if the following three axioms hold:

d′1. $D(A, B) = D(B, A)$.

d′2. $D(A, A) = 0$,

d′3. $\max\{D(A \cup C, B \cup C),\, D(A \cap C, B \cap C)\} \le D(A, B)$, $\forall C \in \widetilde{\mathcal{P}}(\Omega)$,

for any $A, B, C \in \widetilde{\mathcal{P}}(\Omega)$.

The three axioms used to define divergence measure are based on the reflexivity, symmetry, and that the divergence between any common handling of two fuzzy sets are always smaller than the divergence between them.

In the case of crisp sets, it is easy to check that this measure compares just the different part of both sets, not their similarity, so it enables us to evaluate the weight of the part of the universe not common to A and B, being true the following relation

$$D(A, B) = D(A - B, B - A),$$

with difference of crisp sets, $A - B$, defined as the intersection of A with the complement of B in Ω.

This property was studied using different continuous t-norms and arbitrary fuzzy sets, and it is not difficult to show ([8]) that the preceding property ($D(A, B) = D(A - B, B - A)$) holds only when Łukasiewicz t-norm is used.

This reason leads us to consider only the union and intersection based on the Łukasiewicz t-norm and t-conorm, so we restrict our research to this case.

3. Multi-dimensional case

Until now, all measures of comparison were based on the similitude or dissimilarity of two fuzzy sets. But what happens when we want to compare three or more objects? Is it possible to calculate the relation among various prototypes simultaneously?

Based on this motivation, we extend the original bi-dimensional divergence definition to a multi-dimensional case, so we can compare two or more fuzzy sets, quantifying their differences in a global context. First, we propose a generalized definition of the original one. Based on the three previous axioms, it is the following:

Definition 5. Let Ω be the universal set and let $\widetilde{\mathcal{P}}(\Omega)$ be the set of all fuzzy subsets. An application $D\colon \widetilde{\mathcal{P}}(\Omega) \times \widetilde{\mathcal{P}}(\Omega) \times \cdots \times \widetilde{\mathcal{P}}(\Omega) \to \overline{\mathbb{R}}$ is said to be a *general divergence measure* among fuzzy subsets if and only if the following three axioms hold:

- **d1.** $D(A_1, A_2, \dots, A_n) = D(A_{\sigma(1)},\ A_{\sigma(2)}, \dots, A_{\sigma(n)})$, for any permutation σ of $\{1, \dots, n\}$,
- **d2.** $D(A, A, \dots, A) = 0$,
- **d3.** $\max\{D(A_1 \cup C, A_2 \cup C, \dots, A_n \cup C),\ D(A_1 \cap C, A_2 \cap C, \dots, A_n \cap C)\} \le D(A_1, A_2, \dots, A_n)$, $\forall C, A_1, \dots, A_n \in \widetilde{\mathcal{P}}(\Omega)$,

for any $C, A_1, \dots, A_n \in \widetilde{\mathcal{P}}(\Omega)$.

In this paper, we begin supposing that the universal set consist of just one element, $\Omega = \{q\}$, so every fuzzy set can be identified with its membership function and we can characterize a divergence measure D through a function h

$$D(A_1, A_2, \dots, A_n) = h\big(A_1(q), A_2(q), \dots, A_n(q)\big)\,.$$

This function must fulfill the following properties:

- **p1.** $h(a, a, \dots, a) = 0$, $\forall a \in [0,1]$ (reflexivity),
- **p2.** $h(a_1, a_2, \dots, a_n) = h\big(a_{\sigma(1)}, \dots, a_{\sigma(n)}\big)$, $\forall a_i \in [0,1]$ (symmetry),
- **p3.** $h\big(T_\infty(a_1, c), \dots, T_\infty(a_n, c)\big) \le h(a_1, \dots, a_n)$, $\forall a_1, \dots,\ a_n,\ c \in [0,1]$,
- **p4.** $h\big(S_\infty(a_1, c), \dots, S_\infty(a_n, c)\big) \le h(a_1, \dots, a_n)$, $\forall a_1,\ \dots,\ a_n,\ c \in [0,1]$,

where $T_\infty(a,b) = \max(a + b - 1, 0)$ is the Łukasiewicz t-norm and S_∞ its dual t-conorm. In the following subsection we will study such functions. Finally, we extend this results to a universal finite set.

3.1 Real divergence functions

In this section, we suppose that the universal set is a single point, and let be $h_n\colon [0,1]^n \to \overline{\mathbb{R}}$, $(n \ge 2)$, a real function fulfilling **p1** and **p2**. By the symmetry of h and by **p1**, we suppose that $a_1 \ge a_2 \ge \cdots \ge a_n$ and $x_1 \ge x_2 \ge \cdots \ge x_n$ till the end, with a_i, c, x_i, y and so on in $[0,1]$.

First of all, we will review the characterization of h_n when $n = 2$, which only depends on the difference between its coordinates.

Proposition 1. *The following two statements are equivalent:*

1. $h_2(x_1, x_2) = g\big(|x_1 - x_2|\big)$ *with* $g(0) = 0$ *and* $g(\cdot)$ *monotonous non decreasing in* $[0,1]$,
2. h_2 *fulfills properties* **p3** *and* **p4**.

P r o o f . (1)$\Rightarrow$(2): Since $x_1 > x_2$, there are three cases: i) $x_1 + c \geq 1$ and $x_2 + c \geq 1$; ii) $x_1 + c \geq 1$ and $x_2 + c < 1$; iii) $x_1 + c < 1$ and $x_2 + c < 1$. The cases i) and iii) are trivial, and if $x_1 + c \geq 1 > x_2 + c$ then by the monotony of g, $h\big(\min(x_1 + c, 1), \min(x_2 + c, 1)\big) = g\big(|1 - (x_2 + c)|\big) \leq g\big(|(x_1 + c) - (x_2 + c)|\big) = g\big(|x_1 - x_2|\big) = h(x_1, x_2)$, and $h\big(\max(x_1 + c - 1, 0), \max(x_c + c - 1, 0)\big) = g\big(|(x_1 + c - 1) - 0|\big) \leq g\big(|x_1 + c - 1 + (1 - x_2 - c)|\big) = h(x_1, x_2)$.

(2)$\Rightarrow$(1): Let $0 \leq x_2 < x_1 \leq 1$, and $c = 1 - x_1 + x_2$, then by **p3** it holds that $h(0, \cdot)$ is monotonous nondecreasing in $[0, 1]$.

Let $t \in [-1, 1]$, $x_1 + t$, $x_2 + t \in [0, 1]$. If $t \leq 0$ and $x_1 + t$, $x_2 + t \in [0, 1]$, let $c = 1 - t$. Then $h(x_1 + t, x_2 + t) \leq h(x_1, x_2)$ by **p3**. If $t > 0$ and $x_1 + t$, $x_2 + t \in [0, 1]$, let $c = t$. Then $h(x_1 + t, x_2 + t) \leq h(x_1, x_2)$ by **p4**.

Since x_1, x_2 and t are arbitrary, then it follows that $h(x_1 + t, x_2 + t) = h(x_1, x_2)$, so $h(x_1, x_2) = h\big(|x_1 - x_2|, 0\big)$. Let $g(x) = h(x, 0)$. Then $g(0) = 0$ and $g(\cdot)$ is monotonous nondecreasing in $[0, 1]$. ☐

Now, we characterize such functions.

Proposition 2. *If $t \in [0, 1]$, then*

1. (**o3**) $h_n(x_1, x_2, \dots, x_n) \geq h_n\big(\max(x_1 - t, 0), \max(x_2 - t, 0), \dots, \max(x_n - t, 0)\big)$ *if and only if h_n fulfills property* **p3**,
2. (**o4**) $h_n(x_1, x_2, \dots, x_n) \geq h_n\big(\min(x_1 + t, 1), \min(x_2 + t, 1), \dots, \min(x_n + t, 1)\big)$ *if and only if h_n fulfills property* **p4**.

P r o o f . Since $T_\infty(x_i, 1 - t) = \max(x_i + 1 - t - 1, 0) = \max(x_i - t, 0)$ and $S_\infty(x_i, t) = \min(x_i + t, 1)$, the proposition follows. ☐

In particular, we apply this property on the border of $[0, 1]^n$.

Proposition 3. *Let $h_n(\cdot)$ be a function such that fulfills properties* **p3** *and* **p4**. *Then, if $\forall i$, $x_i + r \in [0, 1]$, with $r \in [-1, 1]$, then*

$$\begin{aligned} h_n(x_1, x_2, \dots, x_n) &= h_n(x_1 + r,\ x_2 + r, \dots, x_n + r) \\ &= h_n(x_1 - x_n,\ x_2 - x_n, \dots, x_{n-1} - x_n,\ 0) \\ &= h_n(1,\ x_2 + 1 - x_1, \dots, x_n + 1 - x_1). \end{aligned}$$

P r o o f . Proposition 2 guarantees the fulfillment of **o3** and **o4**. If $r \geq 0$, let $t = r$ and by **o4**, $h_n(x_1, x_2, \dots, x_n) \geq h_n(x_1 + r, x_2 + r, \dots, x_n + r)$. Now, if $r < 0$, then by **o3** with $t = -r$, $h_n(x_1, x_2, \dots, x_n) \geq h_n(x_1 + r, x_2 + r, \dots, x_n + r)$. It is easy to see that $h_n(x_1, x_2, \dots, x_n) = h_n(x_1 + r, x_2 + r, \dots, x_n + r)$ and the subsequent equalities, with $r = -x_n$ and $r = 1 - x_1$. ☐

In the case $n = 3$, we may apply this property to discover the region where h is monotonous non decreasing.

Proposition 4. *Let* $n = 3$, $(a_1, a_2, a_3) \in [0,1]^3$ *and* $R_{(a_1,a_2,a_3)} = \{(x_1, x_2, x_3) \in [0,1]^3 : 0 \le x_{(2)} - x_{(3)} \le a_{(2)} - a_{(3)},\ 0 \le x_{(1)} - x_{(2)} \le a_{(1)} - a_{(2)}\}$, *where* $\cdot_{(i)}$ *indicates that the indices have been permuted so that* $x_{(1)} \ge x_{(2)} \ge x_{(3)}$. *If* h_3 *is a divergence function, then*

$$\forall (x_1, x_2, x_3) \in R_{(a_1,a_2,a_3)}, \quad h(x_1, x_2, x_3) \le h_3(a_1, a_2, a_3).$$

P r o o f. Let $(x_1, x_2, x_3) \in R_{(a_1,a_2,a_3)}$. By the symmetry of h_3, we may suppose that $x_1 \ge x_2 \ge x_3$ and $a_1 \ge a_2 \ge a_3$. Let be $1 - c_1 = x_1 - x_2 + a_2 - a_3$ and $1 - c_2 = a_2 - a_3 - (x_2 - x_3)$. The c_1 and c_2 are well defined, since $a_2 - a_3 \le 1 - c_1 \le a_1 - a_3$ and $0 \le 1 - c_2 \le a_2 - a_3$.

Then, by Proposition 2 and 3, $h(a_1, a_2, a_3) = h(a_1 - a_3,\, a_2 - a_3,\, 0) \ge h(1,\, a_2 - a_3 + c_1,\, c_1) = h(1 - c_1,\, a_2 - a_3,\, 0) \ge h\big((1 - c_1) - (1 - c_2),\, (a_2 - a_3) - (1 - c_2),\, 0\big) = h(x_1 - x_3,\, x_2 - x_3,\, 0) = h(x_1, x_2, x_3)$. □

Now, we give an easy way to construct divergence functions using aggregation operators.

Proposition 5. *Let* $s(\cdot)$ *be a* $n-1$ *ary aggregation operator. Then*

$$h_n(x_1, \dots, x_n) = s(x_1 - x_2,\, x_2 - x_3,\, \dots, x_{n-1} - x_n)$$

is a n *multi-dimensional divergence function.*

P r o o f. Let be $t \in [0,1]$. Then

$$\max(x_i - t, 0) - \max(x_j - t, 0) = \begin{cases} 0 & \text{if } x_j \le x_i \le t, \\ x_i - t & \text{if } x_j \le t < x_i, \\ x_i - x_j & \text{if } t < x_j \le x_i \end{cases}$$

and

$$\min(x_i + t, 1) - \min(x_j + t, 1) = \begin{cases} x_i - x_j & \text{if } x_j \le x_i \le 1 - t, \\ 1 - t - x_j & \text{if } x_j \le 1 - t < x_i, \\ 0 & \text{if } 1 - t < x_j \le x_i \end{cases}$$

for any $i \in \{1, \dots, n\}$. Thus

$$\big|\max(x_i - t, 0) - \max(x_j - t, 0)\big| \le |x_i - x_j|$$

Table 1. Car 1.

Car	A_1	A_2	B_1	B_2	C
Car 1	0.6	0.1	0.5	0.2	0.8

and

$$\left|\min(x_i + t, 1) - \min(x_j + t, 1)\right| \le |x_i - x_j|.$$

Now, by the monotonicity of s, it is clear that h_n fulfills **o3** and **o4**, and by Proposition 2, h_n is a divergence function. □

For example, if $s_n(\cdot)$ is the arithmetic mean, the corresponding divergence function measures only the difference between the extreme values, ignoring the intermediate values, due to their alternative sign.

$$h_{n+1}(x_1, \dots, x_{n+1}) = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_{i+1}) = \frac{1}{n}(x_1 - x_{n+1}).$$

Using the quadratic mean, the function h is able to quantify the difference among all values,

$$h_{n+1}(x_1, \dots, x_{n+1}) = \left(\frac{1}{n}\sum_{i=1}^{n}(x_i - x_{i+1})^2\right)^{1/2},$$

and we present a brief example of this type of measure of dissimilarity.

Example 1. A car is being analyzed, and there are five characteristics evaluated (Table 1.). Two are about external view (shape A_1 and brightness A_2) and the two following are concerned with the auto mechanic (speed B_1 and acceleration B_2). The respective values are $A_1 = 0.6$, $A_2 = 0.1$, $B_1 = 0.5$, $B_2 = 0.2$. And a global evaluation, considering different aspects of the car, $C = 0.8$. Is this global evaluation more relationed with the external look or with the mechanic?

Applying the quadratic mean, since $D(A_1, A_2, C) = h_3(0.6,\ 0.1,\ 0.8) = \left(\frac{1}{2}(0.8 - 0.6)^2 + (0.6 - 0.1)^2\right)^{1/2} = 0.38$, and $D(B_1, B_2, C) = h_3(0.5,\ 0.2,\ 0.8) = \left(\frac{1}{2}(0.8 - 0.5)^2 + (0.5 - 0.2)^2\right)^{1/2} = 0.3$, it may be concluded that the global evaluation is more relationed with the mechanic.

3.2 Universal finite set

In this subsection we extend the previous results to a universal finite set $\Omega = \{q_1, \dots, q_m\}$. First, we show that the divergence among n fuzzy sets is equal to the divergence when the common part is eliminated.

Proposition 6. *Let C be a fuzzy set such that $C(x) = \min\left(A_1(x), \dots, A_n(x)\right)$. Then*

$$D(A_1 - C, \dots, A_n - C) = D(A_1, \dots, A_n),$$

where $(A_i - C)(x) = \max\left(A_i(x) - C(x), 0\right)$, $i = 1, \dots, n$.

Proof. Since

$$(A_i \cap C^c)(x) = \max\left(A_i(x) + \left(1 - C(x)\right) - 1, 0\right) = (A_i - C)(x)$$

and

$$\big((A_i - C) \cup C\big)(x) = \min\big(A_i(x) - C(x) + C(x), 1\big) = A_i(x),$$

then

$$\begin{aligned} D(A_1 - C, \dots, A_n - C) &= D(A_1 \cap C^c, \dots, A_n \cap C^c) \le D(A_1, \dots, A_n) \\ &= D\big((A_1 - C) \cup C, \dots, (A_n - C) \cup C\big) \\ &\le D(A_1 - C, \dots, A_n - C), \end{aligned}$$

by **d3**. □

Now, if we consider the previous subsection, we can define a global divergence among the fuzzy sets through the aggregation of divergence from each element, using the aggregation operators.

Proposition 7. *Let Ω be a universal finite set with m elements, $\{q_1, \dots, q_m\}$, and $A_1, \dots, A_n$ fuzzy sets. Then if $s_m(\cdot)$ is an operation of aggregation and $h^j(\cdot)$ is a divergence function, then*

$$D(A_1, \dots, A_n) = s_m(h^1, \dots, h^m)$$

is a divergence measure with $h^j = h_n\big(A_1(q_j), \dots, A_n(q_j)\big)$, $j = 1, \dots, m$.

Proof. By the reflexivity and symmetry of h, and the boundary conditions of s, **d1** and **d2** hold. And by **p3** and **p4** and the monotony of s, **d3** holds. □

Example 2. Following the Example 1, suppose now that five cars are analysed, and their evaluations are listed in the Table 2. If we consider that all the cars are equally important, we use the arithmetic mean, $s_5(\cdot)$ and the same divergence function than in Example 1 to calculate

$$D(A_1, A_2, C) = \frac{1}{5} \sum_{j=1}^{5} h_3\big(A_1(q_j),\ A_2(q_j),\ C(q_j)\big) = 0.144,$$

and

$$D(B_1, B_2, C) = \frac{1}{5} \sum_{j=1}^{5} h_3\big(B_1(q_j),\ B_2(q_j),\ C(q_j)\big) = 0.151,$$

so in this case the global evaluation C is more relationed with the external look.

Table 2. Evaluations for five cars.

Car	A_1	A_2	B_1	B_2	C
Car 1	0.6	0.1	0.5	0.2	0.8
Car 2	0.5	0.4	0.4	0.3	0.3
Car 3	0.4	0.5	0.5	0.2	0.3
Car 4	0.9	0.9	0.9	0.6	0.7
Car 5	0.8	0.8	0.8	0.9	0.8

4. Application to hierarchical cluster analysis

We can use divergence measure in Cluster Analysis. The object of clustering techniques is a partition of a set of objects into groups in such a way that the profiles of objects in the same cluster are very similar, whereas the profile of objects in different clusters are distinct. The most common approach to clustering analysis is the hierarchical method. In this method, the groups at each step are sequentially nested with respect to previous groups, beginning with n clusters, one for each object, and ending with one cluster containing all objects. Subsequentially, at each step the two groups less divergent are nested, and the new divergence in the new class of groups are calculated.

When a person want to buy a car, usually he or she looks at four or five cars before the final decision. Normally, these cars are very similar in global, similar price, engines, size, and so on, and only little differences may be detected. On the one hand, if the cars are quite similar, we must not use measure of similitude to compare them, but measure of dissimilarity, in order to point the idiosyncrasy of each car.

With the next theorem, we can compare three or more cars simultaneously, building up the divergence among n cars from the bidimensional divergence, calculating recurrent divergence measures. It only calculates the minimal divergence among different groups of sets and no new divergence measure is smaller than the smallest previous one.

Theorem 1. *Let $s_3(\cdot)$ be idempotent aggregation operators, and let D_2 be a 2-divergence measure in arbitrary partition of Ω, $\{A_1, \dots, A_n\}$. We subsequently define*

$$H^0 : G_1, \dots, G_n \text{ with } G_i = A_i\,, \qquad i = 1, \dots, n\,.$$
$$D^0(G_i^0, G_j^0) = D_2(G_i^0, G_j^0)\,, \qquad i, j = 1, \dots, n\,.$$

Let the partition in the step r be

$$H^{r-1} : G_1,\, G_2, \dots, G_p\,.$$

Choose G_i and G_j such that

$$D^{r-1}(G_i, G_j) \le D^{r-1}(G_k, G_s)\,,$$
$$k \ne s\,, \qquad k, s \in \{1, 2 \dots, p\}\,.$$

Let be

$$H^r : G_1, \dots, \{G_i, G_j\}, \dots, G_p$$

and

$$D^r(G_i, G_j, G_k) = s_3\big(D^{r-1}(G_i, G_j)\,, D^{r-1}(G_i, G_k), D^{r-1}(G_j, G_k)\big)\,;$$
$$D^r(G_r, G_s) = D^{r-1}(G_r, G_s)\,, \qquad r, s \notin \{i, j\}\,.$$

Then

1. D^r *is a divergence measure,* $r = 2, \dots, n$.
2. $\min(D^r) \leq \min(D^{r+1})$.

Proof. By the idempotent of s_3, the properties of divergence measure follow and $\min(D^r) \leq \min(D^{r+1})$. $\square$

EXAMPLE 3. Following the Example 2, the cars are now considered like fuzzy sets, and the five characteristics are evaluated for each car. Initially, we first begin with five different groups and calculate the divergence between each pair of them, using the quadratic mean. The cars less different are car 2 and car 3, so in the next class of groups, we consider four groups, merging the cars according to their divergence:

$$H^0 : \{1\}, \{2\}, \{3\}, \{4\}, \{5\},$$

$$H^1 : \{1\}, \{2, 3\}, \{4\}, \{5\},$$

$$H^2 : \{1\}, \{2, 3\}, \{4, 5\},$$

$$H^3 : \{1, 2, 3\}, \{4, 5\},$$

$$H^4 : \{1, 2, 3, 4, 5\}.$$

REFERENCES

[1] BEZDEK, J.—HARRIS, J.: *Fuzzy partitions and relations, an axiomatic basis for clustering*, Fuzzy Sets and Systems **1** (1978), 111–127.

[2] BOUCHON-MEUNIER, B.—RIFQI, M.—BOTHOREL, S.: *Towards general measures of comparison of objects*, Fuzzy Sets and Systems **84** (1996), 143–153.

[3] CALVO, T.—DE BAETS, B.—MESIAR, R.: *Weighted sums of aggregation operators*, Mathware Soft Comput. **6** (1999), 33–47.

[4] CHEN, S. M.—YEH, M. S.—HSIAO, P. Y.: *A comparison of similarity measures of fuzzy values*, Fuzzy Sets and Systems **72** (1995), 79–89.

[5] DE BAETS, B.—MESIAR, R.: *Pseudo-Metrics and T-equivalences*, J. Fuzzy Math. **5** (1997), 471–481.

[6] DUMITRESCU, D.: *Fuzzy partitions with the connectives T_∞, S_∞*, Fuzzy Sets and Systems **47** (1992), 193–195.

[7] MONTES, S.: *Particiones y medidas de divergencia en modelos difusos*, Ph. Thesis. University of Oviedo, 1998.

[8] TORRES, E.—MIRANDA, P.—GIL, P.: *Real divergence functions with continuous t-norm*, in: IPMU 2000, Madrid.

[9] WANG, X.—DE BAETS, B.—KERRE, E.: *A comparative study of similarity measures*, Fuzzy Sets and Systems **73** (1995), 259–268.

Received April 28, 2000

Dpto. de Estadística, I. O. y D. M.
Universidad de Oviedo
E.U.E. Empresariales Jovellanos
C./ Tomás y Valiente 1
E–33201 Gijón
SPAIN

E-mail: torres@etsiig.uniovi.es