Available online at www.sciencedirect.com
ScienceDirect
Fuzzy Sets and Systems 305 (2016) 70–94
www.elsevier.com/locate/fss

Capacities and overlap indexes with an application in fuzzy rule-based classification systems

D. Paternain [a], H. Bustince [a,b,*], M. Pagola [a], P. Sussner [c], A. Kolesárová [d], R. Mesiar [e,f]

[a] *Departamento de Automática y Computación, Universidad Pública de Navarra, Campus Arrosadia s/n, 31006 Pamplona, Spain*
[b] *Institute of Smart Cities, Universidad Publica de Navarra, Campus Arrosadia s/n, 31006 Pamplona, Spain*
[c] *Department of Applied Mathematics, University of Campinas, 13081-970 Campinas, SP, Brazil*
[d] *Institute of Information Engineering, Automation and Mathematics, Slovak University of Technology, Radlinského 9, 81237 Bratislava, Slovakia*
[e] *Slovak University of Technology, Radlinského 11, 81005 Bratislava, Slovakia*
[f] *Slovakia and Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 18208 Prague, Czech Republic*

Received 4 March 2014; received in revised form 8 September 2015; accepted 30 December 2015

Available online 6 January 2016

Abstract

In this work, we introduce a method for constructing capacities using overlap indexes between the fuzzy sets which are generated from the inputs of the considered problem. We also use these capacities to aggregate information by means of the Choquet integral in a fuzzy rule-based classifier. We observe that with these capacities the obtained results are better than those obtained with other measures.

Keywords: Capacity; Overlap index; Overlap function; Choquet integral; Fuzzy rule-based classification systems

1. Introduction

For some problems, the solution may be improved in two different ways:

(a) employing representations of information that take the inherent imprecision in data into account [19,22,39], and
(b) using information fusion mechanisms adapted to such representations [6,7,11].

These strategies were, for instance, applied to the calculation of the volume of clinically significant regions in the brain [50] and to the biometric recognition from digital fingerprints [40].

* Corresponding author.
E-mail addresses: daniel.paternain@unavarra.es (D. Paternain), bustince@unavarra.es (H. Bustince), miguel.pagola@unavarra.es (M. Pagola), sussner@ime.unicamp.br (P. Sussner), anna.kolesarova@stuba.sk (A. Kolesárová), mesiar@math.sk (R. Mesiar).

http://dx.doi.org/10.1016/j.fss.2015.12.021
0165-0114/

In many problems, it is crucial to find a relation between groups of data. Such relation can be expressed, for instance, in terms of an appropriate fuzzy measure or capacity that is capable to detect the links between data [43].

In this work, we focus on the latter problem. In particular, our main goal is to describe a method to build capacities [43,49] from the data (inputs) of a given problem.

The main idea is as follows: after constructing fuzzy sets from the data (inputs) of the problem, we establish links between these fuzzy sets in terms of overlap indexes [8,19,25] that are derived from overlap functions [10,7,34]. Finally, capacities are defined on the basis of these overlaps.

The notion of an overlap function appears in image processing and classification settings as a way to measure to what extent a given element belongs to two (in the original bidimensional case) or three or more (in the multidimensional case) considered classes. The interest of using bidimensional overlap functions for classification was made clear in [7], where they were used to define an analog of preference structures in those situations where associativity is not required or even natural. Moreover, they can also be used for approximate reasoning, as shown in [25]. Basically, a bidimensional overlap function is a continuous aggregation function which vanishes whenever any of the inputs is equal to zero and which equals one only when both inputs are 1. In particular, continuous t-norms without divisors of zero are examples of overlap functions [10].

These capacities are used for the aggregation of information through the use of Choquet integrals. Note that, in this way, the resulting aggregation function can be adapted to the specific problem at hand. To illustrate our approach, we apply interpretable Fuzzy Rule-Based Classification Systems (FRBCSs) [32] to some benchmark classification problems [20]. In this setting, the aggregation of the information plays a key role since it determines the character of the Fuzzy Reasoning Method (FRM) [15].

Specifically, we apply the Choquet integral to aggregate the local information given by each fuzzy rule of the system. The fuzzy sets used to construct the associated fuzzy measure are the rule weights and consequently, it expresses the interaction among the rules of the different classes. Furthermore, we propose an evolutionary method to learn a different capacity for each class of the problem. The quality of the proposal is tested using 40 datasets selected from the KEEL dataset repository [1]. We compare our approach with 6 classifiers, namely SGERD [41], SLAVE [28], C4.5 [44], CART [5], RIPPER [14] and FURIA [30], and we use statistical tests to support our conclusions.

This paper is organized as follows: after providing some preliminaries, we analyse some properties of overlap functions and indexes. In Sections 4 and 5, we present a method for constructing capacities from overlap indexes. In Sections 6 and 7, we study the main properties of our constructions. Section 8 presents the FRM in which we apply Choquet integrals defined in terms of our measures to FRBCSs. Section 9 exhibits both the results obtained by our approach in a set of well-known classification problems and the statistical analysis. Finally, we present some conclusions, future research directions and references.

2. Preliminaries

Given a set U, we denote its cardinality by $card(U)$.

Recall that, given a referential set (or universe) U, a fuzzy set A over U corresponds to a function μ_A:

$$\mu_A : U \to [0,1] .$$

More precisely, a fuzzy set A is given by the graph of the function μ_A, called the membership function of A:

$$\{(i, \mu_A(i)) \mid i \in U\} .$$

For simplicity, we write $A(i)$ instead of $\mu_A(i)$ in this work.

Given a referential set U, we denote by $FS(U)$ the space of all fuzzy sets defined over U. In this work, we only deal with finite referential sets. Henceforth, let $U = \{1, \ldots, n\}$.

$FS(U)$ can be endowed with a partial order $\subseteq$ as follows. For $A, B \in FS(U)$, $A \subseteq B$ if and only if $A(i) \le B(i)$ for every $i \in U$. Similarly, a partial order on $[0,1]^n$ is given as follows: $(x_1, \ldots, x_n) \le (y_1, \ldots, y_n)$ if and only if $x_i \le y_i$ for every $i \in U$. Together with these partial orders, $FS(U)$ and $[0,1]^n$ constitute lattices, that is, all pair-wise infima and suprema exist in $FS(U)$ and $[0,1]^n$, respectively.

Let L_1 and L_2 be lattices. A function $f : L_1 \to L_2$ is said to be increasing if $x \leq y$ implies that $f(x) \leq f(y)$ for all $x, y \in L_1$. The function $f : L_1 \to L_2$ is a (lattice-)isomorphism if f is bijective and f as well as f^{-1} are increasing [17]. In this case, L_1 and L_2 are called (lattice-)isomorphic and one can identify L_1 with L_2. For example, $FS(U)$ and $[0, 1]^n$ are isomorphic.

By abuse of notation, we denote by $\emptyset$ empty set (that is, the fuzzy set where the membership values of all the elements are equal to 0), and by U the fuzzy set with all its memberships equal to 1.

The support of a fuzzy set $A \in FS(U)$ is given by:

$$supp(A) = \{i \in U \mid A(i) \neq 0\}.$$

We say that A is a full fuzzy set if $supp(A) = U$. To distinguish fuzzy sets from the classical subsets of U, we will use the notation $\widetilde{A}$ in the case of the later class.

Let $\widetilde{A} \subseteq U$ and $t \in [0, 1]$. The symbol $t\widetilde{A}$ denotes the fuzzy set given by:

$$t\widetilde{A}(i) = \begin{cases} t & \text{if } i \in \widetilde{A}; \\ 0 & \text{otherwise.} \end{cases}$$

In particular, and by abuse of notation, we use the symbol $1_{\widetilde{A}}$ to denote $1\widetilde{A}$ for every $\widetilde{A} \subseteq U$, since $1\widetilde{A}$ equals the characteristic function of the set $\widetilde{A}$.

Besides, note that this definition corresponds to the basic function $b(\widetilde{A}, t)$ introduced by Benvenuti et al. in 2002 [3].

Given a fuzzy set $A \in FS(U)$ and a mapping $f : [0, 1] \to [0, 1]$, the symbol $f(A)$ denotes the following the fuzzy set:

$$f(A) = \{(i, f(A(i)))|i \in U\}.$$

Let $A, B \in FS(U)$. Note that

$$A \cup B = \{(i, \max(A(i), B(i))) \mid i \in U\}$$

and

$$A \cap B = \{(i, \min(A(i), B(i))) \mid i \in U\}.$$

Given a function $F : [0, 1]^k \to [0, 1]$ (with $k \in \mathbb{N}$) and k fuzzy sets $A_k \in FS(U)$, the symbol $F(A_1, \ldots, A_k)$ denotes the fuzzy set over U whose membership function is given by:

$$F(A_1, \ldots, A_k)(i) = F(A_1(i), \ldots, A_k(i)).$$

Definition 1. An aggregation function of dimension n [9,11,29,35,39] is a mapping M from the (complete) lattice $[0, 1]^n$ to the (complete) lattice $[0, 1]$ such that (see also [2,16]):

(1) $M(0, \cdots, 0) = 0$ and $M(1, \ldots, 1) = 1$;
(2) M is increasing.

3. Overlap functions and overlap indexes

The concept of overlap function was extensively studied in [10]. Let us recall some definitions and results from this article.

Definition 2. An overlap function is a mapping $G_O : [0, 1]^2 \to [0, 1]$ such that:

(1) $G_O(x, y) = G_O(y, x)$ for every $x, y \in [0, 1]$;
(2) $G_O(x, y) = 0$ if and only if $xy = 0$;
(3) $G_O(x, y) = 1$ if and only if $xy = 1$;
(4) G_O is increasing;
(5) G_O is continuous.

Example 1.

(1) The minimum and the product are examples of overlap functions which are also t-norms.
(2) If $f : [0, 1] \to [0, 1]$ is an increasing bijection, then

$$f(xy) \text{ and } f(\min(x, y))$$

are examples of overlap functions.

On the one hand, overlap functions generalize binary intersection operators such as the minimum, or, in general, t-norms. On the other hand, observe that an overlap function is a particular instance of an aggregation function without divisors of zero or divisors of one.

The class of all overlap functions is convex. It is closed under any continuous aggregation of its members, provided that the applied aggregation function has no divisors of zero and no divisors of one. In particular, it is closed under aggregation by means of any of its members.

Overlap functions can be used to build overlap indexes by aggregating them. We start by recalling some basic notions about the idea of an overlap index and we will formalize the construction method in Theorem 1.

Definition 3. An overlap index is a mapping $O : FS(U) \times FS(U) \to [0, 1]$ such that

(O1) $O(A, B) = 0$ if and only if A and B have disjoint supports; that is, $A(i)B(i) = 0$ for every $i \in U$;
(O3) $O(A, B) = O(B, A)$;
(O4) If $B \subseteqq C$, then $O(A, B) \leq O(A, C)$.

An overlap index such that

(O2′) $O(A, B) = 1$ if there exists $i \in U$ such that $A(i) = B(i) = 1$

is called a normal overlap index.

Remark 1. In the original definition of overlap index [19], condition (O2) states that

$$O(A, B) = 1 \text{ if } A(i) = 0 \text{ or } B(i) = 1 \text{ or } A(i)B(i) = 0$$

for all $i \in U$. For $A = \emptyset$ we obtain the following contradiction: (O1) implies that $O(A, A) = 0$ whereas (O2) implies $O(A, A) = 1$. Therefore we removed condition (O2) from the definition of an overlap index.

Example 2.

(1) The first example of overlap index in the literature is Zadeh's consistency index [52]:

$$O_Z(A, B) = \max_{i=1}^{n}(\min(A(i), B(i))).$$

Note that O_Z is normal.
(2) Let $M : [0, 1]^2 \to [0, 1]$ be a symmetric aggregation function such that $M(x, y) = 0$ if and only if $xy = 0$. We have that

$$O_{M,Z}(A, B) = \max_{i=1}^{n}(M(A(i), B(i)))$$

is a normal overlap index that generalizes Zadeh's index.
(3) If in the previous example, we consider a symmetric, increasing function $M : [0, 1]^2 \to [0, 1]$ such that $M(1, 1) < 1$ and $M(x, y) = 0$ if and only if $xy = 0$, then we obtain an overlap index which is not normal. For instance, when taking $M(x, y) = \frac{(xy)^p}{2}$ with $p > 0$, we arrive at the overlap index:

$$O(A, B) = \max_{i=1}^{n}\left(\frac{(A(i)B(i))^p}{2}\right).$$

(4) The following is also an example of overlap index:

$$O_\pi(A, B) = \frac{1}{n}\sum_{i=1}^{n} A(i)B(i).$$

Remark 2. For each overlap index $O : FS(U) \times FS(U) \to [0, 1]$, the function $M_O : [0, 1]^n \to [0, 1]$ given by

$$M_O(E) = \frac{O(E, U)}{O(U, U)}$$

with $E \in [0, 1]^n$, is an aggregation function with no zero divisors.

Remark 3. Formally, overlap indexes can be seen as generalized measures of fuzzy intersections of considered fuzzy sets A and B. Consider, for example, a full probability measure P on U (i.e., $P(\widetilde{A}) > 0$ whenever $\widetilde{A}$ is non-empty). Then $O_P(A, B) = P(suppA \cap suppB)$ defines an overlap index. Following the Zadeh idea of fuzzy probability measures [52], also $O^P(A, B) = \sum_{i=1}^{n} P(i)\min(A(i), B(i))$ defines an overlap index. For the uniform probability measure (counting measure in Example 5) m_c we obtain $O^{m_c} = O_\pi$. Similarly, O_Z can be seen as the Sugeno integral with respect to the top capacity m^*, see Example 6, applied to the fuzzy intersection $A \cap B$.

3.1. Modularity of overlap indexes

The main goal of this work is to show how overlap indexes can be used to build capacities. To do so, we first introduce a set of properties of overlap indexes that are going to be linked to similar properties of capacities.

We first introduce the idea of symmetry for overlap indexes.

Definition 4. Let $O : FS(U) \times FS(U) \to [0, 1]$ be an overlap index and let $E \in FS(U)$. O is E-symmetric if for every $A, B \in FS(U)$ such that $card(supp(A)) = card(supp(B))$ it holds that:

$$O(A, E) = O(B, E).$$

Example 3.

(1) Every overlap index O is E-symmetric if $E = \emptyset \in FS(U)$.
(2) Consider the strongest overlap index:

$$O_s(A, B) = \begin{cases} 0 & \text{if } A, B \text{ are disjoint fuzzy sets;} \\ 1 & \text{otherwise.} \end{cases}$$

We have that O_s is E-symmetric for every full set E.

Note that an overlap index cannot be E-symmetric unless E is a full fuzzy set, as the next result shows.

Proposition 1. *If O is an overlap index which is E-symmetric with respect to some fuzzy set $E \in FS(U)$, $E \neq \emptyset$, then E is a full fuzzy set.*

Proof. Assume that E is not a full fuzzy set and that

$$k = \min(card(supp(E)), n - card(supp(E))) > 0 .$$

Let $\widetilde{A} \subseteq supp(E)$ and $\widetilde{B} \subseteq U \backslash supp(E)$ with $card(\widetilde{A}) = card(\widetilde{B}) = k$. Consider the fuzzy sets

$$A = 1_{\widetilde{A}} \text{ and } B = 1_{\widetilde{B}} .$$

We have that $O(E, A) > 0$ (since E and A are not mutually disjoint) whereas $O(E, B) = 0$. Therefore, O cannot be E-symmetric. □

Evidently, the concept of E-symmetry of overlap indexes only applies to full fuzzy sets E and $E = \emptyset$.

Corollary 1. *If O is an E-symmetric overlap index, then*

$$O(E, E) = O(E, U).$$

Now we consider the notion of modularity.

Definition 5. Let $O : FS(U) \times FS(U) \to [0, 1]$ be an overlap index and let $E \in FS(U)$.

(1) O is called E-supermodular if $O(E, A \cap B) + O(E, A \cup B) \geq O(E, A) + O(E, B)$ holds for all $A, B \in FS(U)$. Similarly, O is called E-submodular if $O(E, A \cap B) + O(E, A \cup B) \leq O(E, A) + O(E, B)$ for all $A, B \in FS(U)$.
(2) If O is E-submodular and E-supermodular, then O is simply called E-modular.

Example 4.

(1) Every overlap index O is E-modular for $E = \emptyset \in FS(U)$.
(2) The overlap index O_π is E-modular for every fuzzy set E.

Remark 4. Observe that O_Z is E-submodular but not E-modular. Moreover, O_Z is E-symmetric if and only if $E = \emptyset$.

As mentioned before, overlap indexes can be built by means of overlap functions. In particular, the following construction method by means of aggregation functions can be found in [25].

Theorem 1. *Let $M : [0, 1]^n \to [0, 1]$ be an aggregation function such that $M(x_1, \dots, x_n) = 0$ if and only if $x_1 = \dots = x_n = 0$ and let $G_O : [0, 1]^2 \to [0, 1]$ be an overlap function. The mapping $O : FS(U) \times FS(U) \to [0, 1]$ given by*

$$O(A, B) = M(G_O(A(1), B(1)), \dots, G_O(A(n), B(n))) \tag{1}$$

is a normal overlap index in the sense of Definition 3.

Conversely, if G_O is an overlap function and $M : [0, 1]^n \to [0, 1]$ is an aggregation function such that O defined by Equation 1 is an overlap index, then $M(x_1, \dots, x_n) = 0$ if and only if $x_1 = \dots = x_n = 0$.

4. Capacities and overlap indexes

In recent years, capacities have attracted a growing interest due to their applicability in many different areas, ranging from decision making to image processing. In the following, we recall some basic notions concerning capacities [49].

Definition 6. Let $U = \{1, 2, \dots, n\}$. A capacity (or non-additive measure) over U is a mapping $m : 2^U \to [0, 1]$ such that

(1) $m(\emptyset) = 0$ and $m(U) = 1$;
(2) If $\widetilde{A} \subset \widetilde{B}$ then $m(\widetilde{A}) \leq m(\widetilde{B})$.

Example 5.

(1) Any probability measure yields an example of a capacity.
(2) The bottom capacity is defined by

$$m_*(\widetilde{A}) = \begin{cases} 1 & \text{if } \widetilde{A} = U; \\ 0 & \text{otherwise.} \end{cases}$$

It is called the bottom capacity because, for any other capacity m over U, it holds that $m_*(\widetilde{A}) \leq m(\widetilde{A})$ for every $\widetilde{A} \subseteq U$.

(3) The top capacity is defined by

$$m^*(\widetilde{A}) = \begin{cases} 0 & \text{if } \widetilde{A} = \emptyset; \\ 1 & \text{otherwise.} \end{cases}$$

It is called the top capacity because, for any other capacity m over U, it holds that $m^*(\widetilde{A}) \geq m(\widetilde{A})$ for every $\widetilde{A} \subseteq U$.

Next we recall some additional properties that capacities may fulfil.

Definition 7. If m is a capacity over $U = \{1, \dots, n\}$, then:

(1) m is called additive if $m(\widetilde{A} \cup \widetilde{B}) = m(\widetilde{A}) + m(\widetilde{B})$ whenever $\widetilde{A} \cap \widetilde{B} = \emptyset$.
(2) m is called symmetric if $m(\widetilde{A}) = m(\widetilde{B})$ whenever $card(\widetilde{A}) = card(\widetilde{B})$.
(3) m is called supermodular (submodular) if $m(\widetilde{A} \cup \widetilde{B}) + m(\widetilde{A} \cap \widetilde{B}) \geq m(\widetilde{A}) + m(\widetilde{B})$ $(m(\widetilde{A} \cup \widetilde{B}) + m(\widetilde{A} \cap \widetilde{B}) \leq m(\widetilde{A}) + m(\widetilde{B}))$ for every $\widetilde{A}, \widetilde{B} \in 2^U$.
(4) m is called modular if it is supermodular and submodular.

Remark 5. Since we have $m(\emptyset) = 0$ for every capacity m, additivity and modularity are equivalent properties of capacities.

Example 6.

(1) The bottom capacity m_* is symmetric but not additive if $n > 1$. Moreover, this measure is also supermodular.
(2) The top capacity m^* is submodular as well as symmetric, non-additive, and non-supermodular for $n > 1$.
(3) The counting capacity (uniform probability measure) m_c is the capacity defined as:

$$m_c(\widetilde{A}) = \frac{1}{n} card(\widetilde{A}).$$

This capacity is additive and symmetric. Moreover, the identity $card(\widetilde{A} \cup \widetilde{B}) + card(\widetilde{A} \cap \widetilde{B}) = card(\widetilde{A}) + card(\widetilde{B})$ implies that m_c is modular.
However, if we consider $p > 0$ and define

$$m_{c,p}(\widetilde{A}) = (m_c(\widetilde{A}))^p$$

then $m_{c,p}$ is a capacity which is submodular if $p \in]0, 1[$ and supermodular if $p > 1$.

Example 7. The Dirac delta

$$m_{d,j}(\widetilde{A}) = \begin{cases} 1 & \text{if } j \in \widetilde{A}; \\ 0 & \text{otherwise} \end{cases} = 1_{\widetilde{A}}(j)$$

for every $\widetilde{A} \subseteq U$, is an additive measure which is not symmetric.

Note that a capacity can always be defined by means of the following result.

Proposition 2. *(See [4,49].) Let $m : 2^U \to [0, 1]$ be a set function. The following items are equivalent.*

(1) *m is a capacity.*
(2) *There exists an aggregation function $M : [0, 1]^n \to [0, 1]$ such that, for every $\widetilde{A} \in 2^U$*

$$m(\widetilde{A}) = M(1_{\widetilde{A}}) .$$

5. Some relationships between capacities, overlap indexes, and overlap functions

One of the most powerful tools in the probabilistic framework is the Bayesian approach based on conditional probabilities $P_{\widetilde{B}}$, i.e., on new probability measures derived from a given probability measure P such that $P(\widetilde{B}) > 0$. In this case, $P_{\widetilde{B}} : 2^U \to [0, 1]$ is given by $P_{\widetilde{B}}(\widetilde{A}) = \frac{P(\widetilde{A}\cap\widetilde{B})}{P(\widetilde{B})} = \frac{P(\widetilde{A}\cap\widetilde{B})}{P(\widetilde{B}\cap\widetilde{B})}$.

In this section, we generalize the construction of Bayesian conditional probabilities by showing how to build a capacity from an overlap function.

Let $E \in FS(U)$ be a fixed non-empty fuzzy set (that is, with at least one membership different from zero). Given $\widetilde{A} \in 2^U$, let us define a fuzzy set $E_{\widetilde{A}}$ induced by E as follows:

$$E_{\widetilde{A}}(i) = \begin{cases} E(i) & \text{if } i \in \widetilde{A}; \\ 0 & \text{otherwise.} \end{cases}$$

Observe that $E_{\widetilde{A}}$ is the fuzzy intersection of the fuzzy set E and the crisp set $\widetilde{A}$, since

$$E_{\widetilde{A}}(i) = \min(1_{\widetilde{A}}(i), E(i)) .$$

Therefore, any aggregation function with no zero divisors could also be used instead of the minimum in this definition for the subsequent developments.

Example 8.

(1) If $\widetilde{A} = U$, then $E_{\widetilde{A}} = E$ for every $E \in FS(U)$.
(2) If $\widetilde{A} = \emptyset$, then $E_{\widetilde{A}} = \emptyset$ for every $E \in FS(U)$.

We have the following straightforward results.

Proposition 3. *Let E be a fixed fuzzy set over U and $\widetilde{A} \in 2^U$. Then:*

(1) $E_{\widetilde{A}} \subseteq E$;
(2) $E_{\widetilde{A}} = E$ *if and only if* $supp(E) \subseteq \widetilde{A}$.
(3) *If* $\widetilde{B} \in 2^U$ *and* $\widetilde{A} \subseteq \widetilde{B}$ *then* $E_{\widetilde{A}} \subseteq E_{\widetilde{B}}$.
(4) $supp(E_{\widetilde{A}}) = \widetilde{A} \cap supp(E)$.

Remark 6. Note that from (4) in the previous proposition, we have that $E_{\widetilde{A}} = \emptyset$ if and only if $E(i) = 0$ for every $i \in \widetilde{A}$; that is, if and only if $\widetilde{A} \cap supp(E) = \emptyset$.

Now we are ready to introduce the definition of a measure in terms of a fixed fuzzy set and an overlap index.

Theorem 2. *If $E \in FS(U)$ is a fixed, non-empty fuzzy set, then the mapping $m_{O,E} : 2^U \to [0, 1]$ given by*

$$m_{O,E}(\widetilde{A}) = \frac{1}{O(E, E)} O(E, E_{\widetilde{A}})$$

is a capacity for every overlap index O.

Proof. First of all observe that $m_{O,E}$ is well defined since $O(E, E) \neq 0$ and $O(E, E_{\widetilde{A}}) \leq O(E, E)$.

If $\widetilde{A} = U$, then it follows that $E_{\widetilde{A}} = E$, so we have that $m_{O,E}(\widetilde{A}) = 1$. Moreover, if $\widetilde{A} = \emptyset$, then $E_{\widetilde{A}}(i) = 0$ for every $i \in U$. So, in particular, $O(E_{\widetilde{A}}, E) = 0$.

Finally, if $\widetilde{A} \subset \widetilde{B}$, then it follows that $E_{\widetilde{A}} \subseteq E_{\widetilde{B}}$, so, in particular, $O(E, E_{\widetilde{A}}) \leq O(E, E_{\widetilde{B}})$ and hence $m_{O,E}(\widetilde{A}) \leq m_{O,E}(\widetilde{B})$. □

Recall that Benvenuti et al. [4] defined for each aggregation function $M : [0, 1]^n \to [0, 1]$ and $e \in]0, 1]$ such that $M(E) > 0$, where $E = (e, \ldots, e)$, a capacity $m_{M,e} : 2^U \to [0, 1]$ given by

$$m_{M,e}(\widetilde{A}) = \frac{M(e\widetilde{A})}{M(E)}$$

(for $e = 1$ see also Proposition 2). Obviously, in the terms of Theorem 2, $m_{O,E} = m_{M_O,e}$. Here M_O was defined in Remark 2.

Remark 7. Observe that, for a fixed full probability measure P on U, if we consider the overlap index O_P introduced in Remark 3, we recover the definition of Bayesian conditional probabilities, i.e., $m_{O_P,E} = P_{suppE}$.

Example 9. Let $U = \{1, 2, 3\}$ and $E = \{(1, 0.2), (2, 0), (3, 0)\}$. Consider the overlap index $O_Z(A, B) = \max_{1 \le i \le 3}(\min(A(i), B(i))$. We have that $O_Z(E, E) = 0.2$ and $m_{O_Z,E}(\{1\}) = \frac{1}{0.2}0.2 = 1$, whereas $m_{O_Z,E}(\{2\}) = m_{O_Z,E}(\{3\}) = 0$ and therefore $m_{O_Z,E}$ is not symmetric. Observe that in this example $m_{O_Z,E}$ corresponds to the Dirac measure concentrated at $\{1\}$.

Let us continue with some comments regarding $m_{O,E}$.

Proposition 4. *Let E be a non-empty fuzzy set and $\widetilde{A} \in 2^U$. For any overlap index O we have:*

(1) $m_{O,E}(\widetilde{A}) = 0$ *if and only if* $supp(E) \cap \widetilde{A} = \emptyset$;
(2) *If* $O(D, E) < O(E, E)$ *for every* $D \subset E$, *then* $m_{O,E}(\widetilde{A}) = 1$ *if and only if* $supp(E) \subseteq \widetilde{A}$.

Proof.

(1) $m_{O,E}(\widetilde{A}) = 0$ if and only if $O(E, E_{\widetilde{A}}) = 0$. From (O1) in the definition of overlap index, this may happen if and only if $supp(E) \cap supp(E_{\widetilde{A}}) = \emptyset$. From (4) in Proposition 3, the result follows.
(2) This claim follows immediately from (2) in Proposition 3. □

With respect to the comparison of the measures, that we obtain for different choices of the full fuzzy set E, we can state the following.

Proposition 5. *Let O be an overlap index. For all full fuzzy sets $E_1, E_2 \in FS(U)$, the following statements are equivalent:*

(1) $m_{O,E_1}(\widetilde{A}) \le m_{O,E_2}(\widetilde{A})$ *for every* $\widetilde{A} \in 2^U$.
(2) $\min_{\widetilde{A} \in 2^U} \frac{O(E_2, \widetilde{A}_{E_2})}{O(E_1, \widetilde{A}_{E_1})} = \frac{O(E_2, E_2)}{O(E_1, E_1)}$

Proof. The inequality $m_{O,E_1}(\widetilde{A}) \le m_{O,E_2}(\widetilde{A})$ implies that

$$\frac{O(E_2, E_2)}{O(E_1, E_1)} \le \frac{O(E_2, \widetilde{A}_{E_2})}{O(E_1, \widetilde{A}_{E_1})} \tag{2}$$

for every $\widetilde{A} \in 2^U$, so (2) holds. Conversely, if (2) holds, then Equation 2 is satisfied as well and we obtain (1). □

Corollary 2. *Let $E_1, E_2 \in FS(U)$ such that $O(E_1, E_1) = 1$. If $E_1 \subseteq E_2$, then $m_{O,E_1}(\widetilde{A}) \le m_{O,E_2}(\widetilde{A})$ for every $\widetilde{A} \in 2^U$.*

Proof. If $E_1 \subseteq E_2$, then we have that $1 \ge O(E_2, E_2) \ge O(E_1, E_1) = 1$. Consequently,

$$1 = \frac{O(E_2, E_2)}{O(E_1, E_1)} \le \frac{O(E_2, \widetilde{A}_{E_2})}{O(E_1, \widetilde{A}_{E_1})}$$

for every $\widetilde{A} \in 2^U$, which implies that $m_{O,E_1}(\widetilde{A}) \le m_{O,E_2}(\widetilde{A})$ for every $\widetilde{A} \in 2^U$. □

Now, let us show how to build overlaps from capacities. To this end, let us first introduce the concept of contraction.

Definition 8. Let $E \in FS(U)$. The contraction to E (or E-contraction) is the mapping $C_E : FS(U) \to FS(U)$ defined by:

$$C_E(A) = \{(i, E(i)A(i)) \mid i \in U\}.$$

The following result is now straightforward.

Proposition 6. *For every $E \in FS(U)$ and for all $A, B \in FS(U)$:*

(1) *$C_E(A) \subseteq A$;*
(2) *If $A \subseteq B$ then $C_E(A) \subseteq C_E(B)$;*
(3) *$C_E(A) \subseteq E$ and, if E is a full fuzzy set, $C_E(A) = E$ if and only if $A = U$.*
(4) *If E is a full fuzzy set, then C_E is a one-to-one mapping.*
(5) *If E is a full fuzzy set, then $Im(C_E) = \{A \in FS(U) \mid A \subseteq E\}$, where $Im(C_E)$ denotes the image of the mapping C_E.*

Remark 8. The definition of contraction can be generalized by substituting the product with any other t-norm or even an overlap function. We postpone the analysis of the resulting operators to future works.

Let us continue by introducing some notations. For a fixed fuzzy set $E \in FS(U)$ and for $\widetilde{A} \in 2^U$, we define

$$Cl(E, \widetilde{A}) = \{A \in FS(U) \mid A \subseteq E_{\widetilde{A}} \text{ and } A \nsubseteq E_{\widetilde{B}} \text{ for every } \widetilde{B} \subset \widetilde{A}\}.$$

The proof of the following lemma is straightforward.

Lemma 1. *Let $E \in FS(U)$. Then $Cl(E, \widetilde{A}) = \emptyset$ for every $\widetilde{A} \in 2^U$ such that $\widetilde{A} \cap supp(E) = \emptyset$.*

Then we can state the following.

Proposition 7. *Let E be a full fuzzy set. The family $(Cl(E, \widetilde{A}))_{\widetilde{A} \in 2^U}$ is a partition of the set $\{A \in FS(U) \mid A \subseteq E\}$.*

Proof. For any $A \subseteq E$ let's take $\widetilde{A} = supp(A)$. Then, and only then $A \subseteq E_{\widetilde{A}}$ and for any $\widetilde{B}$ which is a proper subset of $\widetilde{A}$, $A \nsubseteq E_{\widetilde{B}}$, i.e., $A \in Cl(E, \widetilde{A})$. The fact that $Cl(E, \widetilde{A}) \cap Cl(E, \widetilde{B}) = \emptyset$ for $\widetilde{A} \neq \widetilde{B}$ is trivial. Hence $\{Cl(E, \widetilde{A}) | \widetilde{A} \in 2^U\}$ is a partition of $\{A \in FS(U) \mid A \subseteq E\}$. □

Now we can show how to recover overlap indexes from capacities.

Theorem 3. *Let m be a capacity such that $m(\widetilde{A}) = 0$ if and only if $\widetilde{A} = \emptyset$. If E is a full fuzzy set, then the function $O_{E,m} : FS(U) \times FS(U) \to [0, 1]$ defined by:*

$$O_{E,m}(A, B) = \begin{cases} m(\widetilde{A}) & \text{if } A \cap B \in Cl(E, \widetilde{A}); \\ 1 & \text{otherwise} \end{cases}$$

is a normal overlap index such that the capacity induced by $O_{E,m}$ is equal to m.

Proof. First of all, due to Proposition 7, $O_{E,m}$ is well defined. Let us prove that $O_{E,m}$ is an overlap index.

(O1) Assume that $O_{E,m}(A, B) = 0$. Since $m(\widetilde{A}) \neq 0$ for every $\widetilde{A} \neq \emptyset$, this happens if and only if $A \cap B \in Cl(E, \emptyset)$, i.e., if and only if A and B have disjoint supports.
(O3) Symmetry is obvious from the definition.
(O4) Let $A \in FS(U)$ be arbitrary, but fixed and let $B \subseteq C$. If $A \cap C \nsubseteq E$, then $O_{E,m}(A, C) = 1 \geq O_{E,m}(A, B)$. Now let us assume that $A \cap C \subseteq E$. From Proposition 7 and the fact that $A \cap B \subseteq A \cap C$, it follows that there exist $\widetilde{A}, \widetilde{B} \in 2^U$ with $\widetilde{B} \subseteq \widetilde{A}$ such that $A \cap C \in Cl(E, \widetilde{A})$ and $A \cap B \in Cl(E, \widetilde{B})$. Since m is a capacity, we have that $m(\widetilde{A}) \leq m(\widetilde{B})$ and therefore $O_{E,m}(A, B) \leq O_{E,m}(A, C)$.

(O2′) Note that $U = U \cap U$. So if $E \neq U$ it follows that $O_{E,m}(U, U) = 1$, whereas if $E = U$, we have that $U \in Cl(U, U)$ and $O_{E,m}(U, U) = m(U) = 1$.

Finally, note that $E_{\widetilde{A}} = E_{\widetilde{A}} \cap E \in Cl(E, \widetilde{A})$ for every $\widetilde{A} \in 2^U$, which concludes the proof of the theorem since

$$m_{O_{E,m},E}(\widetilde{A}) = \frac{1}{O_{E,m}(E, E)} O_{E,m}(E, E_{\widetilde{A}}) = m(\widetilde{A}) .$$

This completes the proof. □

Remark 9. Note that the construction of the previous overlap function relies on considering the quotient space for the following equivalence relation:

A is related to B if and only if there exists $\widetilde{A} \in 2^U$ such that $A, B \in Cl(E, \widetilde{A})$ or $A, B \notin Cl(E, \widetilde{A})$ for every $\widetilde{A} \in 2^U$.

In this way, all the sets that are greater than or not comparable to E, which are irrelevant for the construction of the capacity, are compressed into a single equivalence class to which the value 1 is assigned.

Remark 10. As previously remarked, the use of the minimum in the construction of the capacity in Theorem 3 can be generalized by any other symmetric aggregation function M such that $M(x, 0) = 0$ for every $x \in [0, 1]$. In particular, the definition could be rewritten in terms of overlap functions.

Example 10. Consider the bottom capacity m_*. Let $E \in FS(U)$ be a full fuzzy set. An application of Theorem 3 yields the following function:

$$O_{E,m_*}(A, B) = \begin{cases} 0 & \text{if } A \cap B \in Cl(E, \widetilde{A}) \text{ with } \widetilde{A} \neq U; \\ 1 & \text{otherwise.} \end{cases}$$

In particular, if $E = U$, then we obtain

$$O(A, B) = \begin{cases} 0 & \text{if } \min(A(i), B(i)) = 0 \text{ for some } i \in U\,; \\ 1 & \text{otherwise,} \end{cases}$$

which is not an overlap index, since condition (O1) is violated. This shows that the condition of strictness for the capacity m, that is, $m(\widetilde{A}) > 0$ for all $\widetilde{A} \neq \emptyset$ is crucial in Theorem 3.

Theorem 3 can be extended to include non-strict measures as follows.

Corollary 3. *Consider a capacity m. Let $\widetilde{A}_0 = \{i \in U \mid m(\{i\}) = 0\}$. Suppose that E is a fuzzy set such that $E(i) \neq 0$ for $i \notin \widetilde{A}_0$. The function $O : FS(U) \times FS(U) \to [0, 1]$ given by*

$$O_{E,m,\widetilde{A}_0}(A, B) = \begin{cases} m(\widetilde{A}) & \text{if } A \cap B \in Cl(E, \widetilde{A} \backslash \widetilde{A}_0); \\ 1 & \text{otherwise} \end{cases}$$

is an overlap index.

Proof. Symmetry and monotonicity are clear. We only need to check that (O1) holds. To see (O1), note that $O(A, B) = 0$ if and only if $A \cap B \in Cl(E, \widetilde{A} \backslash \widetilde{A}_0)$ for some $\widetilde{A} \in 2^U$ such that $m(\widetilde{A}) = 0$. But, $m(\widetilde{A}) = 0$ if and only if $\widetilde{A} \subseteq \widetilde{A}_0$, due to the monotonicity of capacities. So, $\min(A(i), B(i)) = 0$ if $i \in \widetilde{A}_0$. □

Example 11. For the bottom capacity m_*, we obtain $\widetilde{A}_0 = \{i \in U \mid m_*(\{i\}) = 0\} = U$. Thus, setting $E = \emptyset$ yields the following overlap function:

$$O_{E,m_*,U}(A, B) = \begin{cases} 0 & \text{if } A \cap B = \emptyset; \\ 1 & \text{otherwise,} \end{cases}$$

which is the strongest overlap index.

Remark 11. Note that Corollary 3 yields Theorem 3 if $\widetilde{A}_0 = \emptyset$.

Corollary 4. *For every capacity m there exists a fuzzy set E and a continuous overlap index $O_{E,m}$ such that the measure induced by $O_{E,m}$ is equal to m.*

Proof. It is just a matter of using the overlap index defined by means of Theorem 3 and Corollary 3. □

6. Some properties of capacities constructed from overlap indexes

Note that we can characterize which of these capacities are also classical measures.

Recall that $m_{O,E}$ is a (probability) measure on U if and only if there exist constants $k_1, \dots, k_n \in [0,1]$ such that $k_1 + \dots + k_n = 1$ and

$$m_{O,E}(\widetilde{A}) = \sum_{j \in \widetilde{A}} k_j \,.$$

This fact leads directly to the following two corollaries:

Corollary 5. *The measure $m_{O,E}$ is additive if and only if*

$$m_{O,E}(\widetilde{A}) = \sum_{j \in \widetilde{A}} O(E(j)\{j\}, E)\,.$$

Corollary 6. *If E is a fuzzy set and O is an E-modular overlap index, then $m_{O,E}$ is additive and hence a classical measure.*

Regarding symmetry, we can state the following.

Theorem 4. *Let E be a fixed fuzzy set and O an overlap index. The capacity $m_{O,E}$ is symmetric if and only if $O(E_{\widetilde{A}}, E) = O(E_{\widetilde{B}}, E)$ for every $\widetilde{A}, \widetilde{B} \in 2^U$ such that $card(\widetilde{A}) = card(\widetilde{B})$.*

Proof. This theorem follows from the definition of a symmetric measure. □

Corollary 7. *The capacity $m_{O,E}$ is symmetric if and only if the overlap index O is E-symmetric.*

Regarding possible applications, it is important to take the following expressions into account. Recall that, given a capacity m, the interaction index I_{ij} between $i, j \in U$ is defined as:

$$\sum_{\widetilde{A} \subseteq U \setminus \{i,j\}} \frac{(n - card(\widetilde{A}) - 2)!\,card(\widetilde{A})!}{(n-1)!} \left(m(\widetilde{A} \cup \{i,j\} - m(\widetilde{A} \cup \{i\}) - m(\widetilde{A} \cup \{j\}) + m(\widetilde{A})\right).$$

We obtain the following straightforward result.

Proposition 8. *Let E be a fuzzy set and O an overlap index. Then, the interaction index for the measure $m_{O,E}$ is given by:*

$$I_{ij} = \frac{1}{O(E,E)} \sum_{\widetilde{A} \subseteq U \setminus \{i,j\}} \frac{(n - card(\widetilde{A}) - 2)!\,card(\widetilde{A})!}{(n-1)!} F, \text{ where}$$

$$F = \left(O((\widetilde{A} \cup \{i,j\})_E, E) - O((\widetilde{A} \cup \{i\})_E, E) - O((\widetilde{A} \cup \{j\})_E, E) + O(E_{\widetilde{A}}, E)\right).$$

7. Construction of capacities from overlap indexes based on overlap functions

Using Theorem 1, we can also derive measures from aggregation and overlap functions as follows.

Proposition 9. *Let $M : [0,1]^n \to [0,1]$ be an aggregation function such that $M(x_1, \dots, x_n) = 0$ if and only if $x_1 = \dots = x_n = 0$, let $G_O : [0,1]^2 \to [0,1]$ be an overlap function and let $E \in FS(U)$ be a non-empty fuzzy set. The mapping $m_{E,M,G_O} : 2^U \to [0,1]$ given by*

$$m_{E,M,G_O}(\widetilde{A}) = \frac{1}{M(G_O(E))} M(G_O(E(1), E_{\widetilde{A}}(1)), \dots, G_O(E(n), E_{\widetilde{A}}(n))),$$

where $M(G_O(E)) = M(G_O(E(1), E(1)), \dots, G_O(E(n), E(n))$ is a capacity.

Note that if we take $E = U$, then we have $m_{U,M,G_O}(\widetilde{A}) = M(1_{\widetilde{A}})$.
Some properties of these types of capacities are the following.

Proposition 10. *Let M be an aggregation function as in Proposition 9. For any non-empty fuzzy set E, we have:*

(1) *$m_{E,M,G_O}(\widetilde{A}) = 0$ if and only if $E(i) = 0$ for every $i \in \widetilde{A}$;*
(2) *$m_{E,M,G_O}(\widetilde{A}) = 1$ whenever $E(i) \neq 0$ for every $i \in \widetilde{A}$.*

Proof.

(1) If $m_{E,M,G_O}(\widetilde{A}) = 0$, then

$$\frac{1}{M(G_O(E))} M(G_O(E(1), E_{\widetilde{A}}(1)), \dots, G_O(E(n), E_{\widetilde{A}}(n))) = 0\,.$$

Since M is an aggregation function, this implies that $G_O(E(i), E_{\widetilde{A}}(i)) = 0$ for every $i = 1, \dots, n$. From the definition of an overlap function, this happens only if $E(i)E_{\widetilde{A}}(i) = 0$ for every $i = 1, \dots, n$. If $E_{\widetilde{A}}(i) \neq 0$ it follows that $E(i) = 0$, which is impossible due to the definition of $E_{\widetilde{A}}$. Therefore, we infer that $E_{\widetilde{A}}(i) = 0$ for every $i \in \widetilde{A}$, that is, if $i \in \widetilde{A}$ then $E(i) = 0$.
The other direction follows from the fact that $E(i) = 0$ for every $i \in \widetilde{A}$ implies that $E_{\widetilde{A}}(i) = 0$ for every $i \in U$.
(2) If $\widetilde{A}$ is as in the statement of this property, then we obtain $E_{\widetilde{A}} = E$, and the result follows from the monotonicity of aggregation and overlap functions. □

The following corollary is a straightforward consequence of the previous result.

Corollary 8. *Let M be an aggregation function as in Proposition 9. For any non-empty fuzzy set E, we have:*

(1) *m_{E,M,G_O} satisfies the property*

$$m_{E,M,G_O}(\widetilde{A}) = 0 \text{ if and only if } \widetilde{A} = \emptyset$$

if and only if E is a full fuzzy set.
(2) *$m_{E,M,G_O}(supp(E)) = 1$.*

Theorem 5. *For a fixed overlap function G_O and an n-ary aggregation function M as in Proposition 9, the following claims are equivalent:*

(1) *For each non-empty fuzzy subset $E \in FS(U)$, the measure m_{E,M,G_O} is additive.*
(2) *M is modular, i.e., $M(\max(x, y)) + M(\min(x, y)) = M(x) + M(y)$ for all $x, y \in [0,1]^n$.*

Proof. Observe first that any modular aggregation function M such that $M(x_1, .., x_n) = 0$ only if $x_1 = \dots = x_n = 0$ has the form $M(x_1, \dots, x_n) = \sum_{i=1}^n f_i(x_i)$, where each $f_i : [0,1] \to [0,1]$ is increasing, $f_i(x) = 0$ only if $x = 0$, and $\sum f_i(1) = 1$, see, e.g., [42].

To see the necessity, observe that the additivity of m_{E,M,G_O} implies that

$$M(G_O(E(1), E_{\widetilde{A}}(1)), \ldots, G_O(E(n), E_{\widetilde{A}}(n))) =$$
$$M(G_O((E(1), E_{\widetilde{A}}(1)), 0, \ldots, 0)) + M(0, \ldots, 0, G_O(E(n), E_{\widetilde{A}}(n))).$$

In view of the mean value theorem for overlap functions, this equality is equivalent to

$$M(x_1, \ldots, x_n) = M(x_1, 0, \ldots, 0) + M(0, x_2, 0, \ldots, 0) + \cdots + M(0, \ldots, 0, x_n)$$

for every $x_1, \ldots, x_n \in [0, 1]$. Defining $f_i(x) = M(0, 0, \ldots, 0, x, 0, \ldots, 0)$, where the x is in the i-th position, the result follows. To see the converse, observe that $M(x_1, \ldots, x_n) = \sum_{i=1}^{n} f_i(x_i)$ is an aggregation function such that $M(x_1, \ldots, x_n) = 0$ if and only if $x_1 = \cdots = x_n = 0$. $\square$

Observe that if M satisfies the requirements of the previous theorem, then m_{E,M,G_O} is additive for all G_O and all $E \neq \emptyset$.

8. An approach towards classification using capacities based on overlap indexes

In this section, we provide an example of an application of capacities based on overlap indexes in classification. To this end, we first provide an appropriate background about Fuzzy Rule-Based Classification Systems (FRBCSs) [32,46], which is the technique used to tackle classification problems. Then we describe in detail the proposed Fuzzy Reasoning Method (FRM) using capacities (Section 8.2) and finally, we introduce the evolutionary approach used to learn the capacities (Section 8.3).

8.1. Fuzzy rule-based classification systems

A classification problem consists in learning a mapping function called *classifier* from a set of training examples, named *training set*, that allows one to classify previously unknown examples. Let $x_p = (x_{p1}, \ldots, x_{pn})$ be the p-th example of the training set which is composed of P examples, where x_{pi} is the value of the i-th attribute ($i = 1, 2, \ldots, n$) of the p-th training example. Each example belongs to a class $y_p \in \mathbb{C} = \{C_1, C_2, \ldots, C_m\}$, where M is the number of classes of the problem.

FRBCSs [32] are widely used to solve classification problems since they allow the inclusion of all available information in system modelling, that is, the information coming from expert knowledge and the one coming from empirical measures or mathematical models. Furthermore, FRBCSs provide a good balance between accuracy and interpretability through the use of linguistic labels [47]. Specifically, the fuzzy rules used in this paper have the following form:

$$\text{Rule } R_j : \text{ If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \text{ then Class} = C_j \text{ with } RW_j \tag{3}$$

where R_j is the label of the jth rule, $x = (x_1, \ldots, x_n)$ is an n-dimensional example vector, A_{ji} is an antecedent fuzzy set representing a linguistic term, C_j is a class label, and $RW_j \in [0, 1]$ is the rule weight [31].

In this paper, we generated the initial knowledge base by means of the fuzzy rule learning method of Chi et al. [12]. This method is one of the most commonly used learning algorithms in the specialized literature due to the simplicity of the fuzzy rule generation method.

8.2. Fuzzy reasoning method based on capacities

The FRM [15] is the mechanism that using the information stored in the knowledge base assigns a class label to the new examples to be classified. In this paper, we propose a modification of the classical FRM using capacities. Specifically, we consider to perform the aggregation of the local information given by the fuzzy rules having the same class label in their consequent part using the Choquet integral [13]. To do so, we create a capacity for each class of the classification problem using the rule weights of those rules belonging to the same class for the sake of modelling their interaction.

Let $x_q = (x_{q1}, \ldots, x_{qn})$ be a new pattern to be classified, let L be the number of rules in the rule base, and let M be the number of classes of the problem. The steps of the proposed FRM, in which we have modified the first and third step of the classical FRM [15], used to classify the pattern x_q are:

Table 1
Names and equations of the overlap indexes used to construct the capacities.

Name	Equation
O_1	$O_1(A,B)=\max_{i=1}^{n}\min(A(i),B(i))$
O_2	$O_2(A,B)=\frac{1}{n}\sum_{i=1}^{n}\min(A(i),B(i))$
O_3	$O_3(A,B)=\frac{1}{n}\sum_{i=1}^{n}(A(i)B(i))^p$, with $p=1$
O_4	$O_4(A,B)=\frac{1}{n}\sum_{i=1}^{n}\sqrt{(A(i)B(i))}$
O_5	$O_5(A,B)=\frac{1}{n}\sum_{i=1}^{n}\min\left(A(i)^pB(i),A(i)B(i)^p\right)$, with $p=0.5$
O_6	$O_6(A,B)=\frac{1}{n}\sum_{i=1}^{n}\frac{\sqrt{(A(i)B(i))}}{\sqrt{(A(i)B(i))}+1-(A(i)B(i))}$
O_7	$O_7(A,B)=\max_{i=1}^{n}(A(i)B(i))^p$, with $p=1$
O_8	$O_8(A,B)=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(A(i)B(i))^2}$

(1) *Matching degree*, that is, *the strength of activation of the if-part for all rules in the rule base with the example* x_q: for each variable we compute the overlap index among each value of the example modelled using a singleton and the triangular membership functions used to model each linguistic term A_{jn}. Then, we apply a t-norm over these values.

$$A_j(x_q)=T(O(Ant_1(x_{q1}),A_{j1}),\ldots,O(Ant_n(x_{qn},A_{jn}))),\qquad j=1,\ldots,L \tag{4}$$

where $Ant_i=1_{x_{pi}}(1)$ with $i=1,\ldots,n$.

(2) *Association degree*, i.e., the *association degree of the example* x_q *with the class of each rule in the rule base*:

$$b_j^k(x_q)=A_j(x_q)\times RW_j^k \qquad k=Class(R_j),\quad j=1,\ldots,L \tag{5}$$

(3) *Example classification soundness degree for all classes*: we aggregate the positive association degrees calculated in the previous step using the Choquet integral whose capacity is obtained in terms of the rule weights as explained below.

$$Y_k=C_{m_{O,E^k}}(b_j^k(x_q)\mid j=1,\ldots,L \text{ and } b_j^k(x_q)>0),\qquad k=1,\ldots,M \tag{6}$$

where m_{O,E^k} is the capacity considered for the k-th class of the problem computed as:

$$m_{O,E^k}(\widetilde{A})=\frac{1}{O(E^k,E^k)}O(E^k,\widetilde{A}_{E^k}) \tag{7}$$

where

$$E^k=\{(R_j,RW_j^k)|R_j \text{ such that } Class(R_j)=k \text{ and } b_j^k(x_q)>0\}$$

(4) *Classification*: we apply a decision function F over the example classification soundness degree for all classes. This function determines the class corresponding to the maximum value.

$$F(Y_1,\ldots,Y_M)=\underset{k=1,\ldots,M}{arg\ max}(Y_k) \tag{8}$$

As we can observe, Equation (7) is applied to obtain the capacities associated with the different classes. Specifically, we propose to apply eight overlap indexes (introduced in Table 1) to compute them, that is, we obtain 8 different ways of constructing the capacities. Furthermore, we also consider the usage of the well-known power capacity shown in Equation (9), where we would like to point out that if we use $p=1$, then we will obtain the classical cardinality capacity.

$$m(A)=\left(\frac{|A|}{n}\right)^p \tag{9}$$

8.3. Evolutionary approach to learning capacities

From the overlap indexes shown in Table 1 we can observe that three of them as well as the power capacity shown in Equation (9) are parametrized. If we use the same value for the parameter p we create the same capacity for all the classes of the problem. However, the interaction among the rules of the different classes can be different. Therefore, in order to model this situation we can consider a different value for the parameter p related to each class of the problem. Consequently, we have an optimization problem in which we have to search for the best set of values that lead to the best possible performance of the system.

To accomplish the optimization process, we propose to use an evolutionary algorithm. Specifically, we apply the CHC evolutionary algorithm [21] due to both its good properties and the good results shown in the specialized literature [23,24,45]. The specific features of the evolutionary algorithm are the following ones:

- **Coding Scheme**. We use real coded chromosomes composed of M genes whose structure is as follows:

  $$Cr = \{g_1, \dots, g_M\},$$

  where $g_i \in (0, 2)$ with $i = \{1, \dots, M\}$ represents the value of the parameter p associated with the i-th class of the problem. In order to achieve a wider range for each value of the parameter p_i, we translate the value of each gene g_i from $(0, 2)$ to $(0, 100)$. To do so, we apply the following equation

  $$p_i = \begin{cases} g_i, & \text{if } g_i \le 1, \\ \frac{1}{2-g_i}, & \text{if } g_i > 1 \end{cases}$$
- **Initial Gene Pool**. To consider the initial situation, we include a chromosome having all genes with value one. The remaining chromosomes will be randomly initialized in the range $(0, 2)$.
- **Chromosome Evaluation**. We use the most common metric for classification, i.e. the accuracy rate.
- **Crossover Operator**. We apply the Parent Centrix BLX operator [36] (which is based on the BLX-α).
- **Restarting Approach**. To try to avoid local optima, we consider a restarting approach. In each restart of the population we include the best global solution found so far, as in the elitist scheme, and we randomly create the remaining individuals.

9. Showing the behaviour of the fuzzy reasoning method based on capacities in classification

This section is aimed at showing the results achieved by our proposals as well as the corresponding analysis. In Section 9.1, we introduce the framework considered for the study. Section 9.2 provides some experimental results as well as a statistical analysis of these results.

9.1. Experimental framework

We have selected forty data-sets selected from the KEEL repository [1]. Table 2 summarizes the following characteristics of each data-set: number of examples (#Ex.), number of attributes (#Atts.) and number of classes (#Class.). We would like to point out that the *Magic, Page-blocks, Penbased, Shuttle* and *Twonorm* data-sets are stratified-sampled to 10% to improve the learning process efficiency. We removed the missing values of *Breast, Cleveland, Crx, Housevotes, Mammographic and Wisconsin* before partitioning them.

We have applied a *5-fold cross-validation model* to test the performance of the approaches. To do so, we first split the data-set into 5 random partitions of data, each one with 20% of the examples maintaining the original distribution of the classes. Then, we join 4 of them to train (optimize) the system (80% of the examples) and the remaining one is used to test the learned (optimized) classifier. We repeat the process 5 times by changing the testing partition in each run. Therefore, after the 5 runs all the examples are treated as testing (unseen) examples and the reported final result is the average among the 5 testing results. As performance measure we have used the accuracy rate, which is computed dividing the number of correctly classified examples by the total number of examples.

In order to give statistical support to the analysis of the results, we use some hypothesis validation techniques [26,27,48]. We use non-parametric tests because the initial conditions that guarantee the reliability of the parametric tests cannot be fulfilled, which implies that the statistical analysis loses credibility with these parametric tests [18].

Table 2
Summary description for the considered data-sets.

Id.	Data-set	#Ex.	#Atts.	#Class.	Id.	Data-set	#Ex.	#Atts.	#Class.
app	Appendicitis	106	7	2	mag	Magic	1902	10	2
aus	Australian	690	14	2	mam	Mammographic	830	5	2
bal	Balance	625	4	3	mon	Monk-2	432	6	2
ban	Banana	5300	2	2	new	New-Thyroid	215	5	3
bre	Breast	277	9	2	pag	Page-blocks	548	10	5
bup	Bupa	345	6	2	pen	Penbased	1992	16	10
car	Car	1728	6	4	pho	Phoneme	5404	5	2
cle	Cleveland	297	13	5	pim	Pima	768	8	2
crx	Crx	653	15	2	rin	Ring	7400	20	2
eco	Ecoli	336	7	8	sat	Satimage	6435	36	7
fla	Flare	1066	11	6	seg	Segment	2310	19	7
ger	German	1000	20	2	shu	Shuttle	5800	9	7
gla	Glass	214	9	7	tae	Tae	151	5	3
hab	Haberman	306	3	2	tic	Tictactoe	958	9	2
hay	Hayes–Roth	160	4	3	tit	Titanic	2201	3	2
hea	Heart	270	13	2	two	Twonorm	740	20	2
hou	Housevotes	232	16	2	veh	Vehicle	846	18	4
ion	Ionosphere	351	33	10	win	Wine	178	13	3
iri	Iris	150	4	3	wis	Wisconsin	683	9	2
led	Led7digit	500	7	10	zoo	Zoo	101	16	7

Specifically, we use the aligned Friedman ranks test [37] to detect statistical differences among a group of results and we graphically show the obtained ranks of each method in order to compare this method with the other methods. We also consider the usage of the Holm post-hoc test [38] to find the algorithms that reject the equality hypothesis with respect to the best method according to the aligned Friedman ranks test. Furthermore, we compute the adjusted p-value (APV) in order to take into account the fact that multiple tests are conducted. In this manner, we can directly compare the APV with respect to the level of significance α in order to be able to reject the null hypothesis. Finally, we use the Wilcoxon signed-ranks test [51] to perform pairwise comparisons.

The configuration for the algorithm of Chi et al. is as follows: we use 3 linguistic labels per variable, which are modelled using triangular shaped membership functions (using Ruspini's fuzzy partitions). Discrete attributes are modelled using singleton fuzzy sets (one for each value of the attribute). We have to point out that this criterion implies that the study is conducted only for fuzzy partitions using a grid structure. We compute the rule weights using the penalized certainty factor [33]. Regarding the evolutionary process, we use populations composed of 50 individuals, 20,000 evaluations and 30 bits per gene for the *Gray* codification.

Finally, we have selected 6 classifiers[1] for the sake of comparing our results. We have used two FRBCSs that include a genetic algorithm in their learning (SLAVE [28] and SGERD [41]), two decision trees (C4.5 [44] and CART [5]) and RIPPER [14] besides its fuzzy extension, that is, FURIA [30] as a representative of a state-of-the-art fuzzy classifier. Recall that FURIA is not restricted to a fixed grid structure. In our opinion, this feature implies that it is more flexible and, consequently, it is less interpretable than our FRM approach based on capacities. The set-up of all these methods is included in Table 3.

9.2. Experimental results and analysis

Table 4 contains the results in testing that were produced using:

- Classical aggregation functions such as maximum (Max.) and mean, instead of the Choquet integral in the third step of the FRM.
- The proposed FRM when using the capacities generated with the eight different overlap indexes (O_1–O_8) considered in this work.

[1] We have executed the KEEL software over the datasets used in this paper.

Table 3
Set-up of the parameters of the classifiers used in the comparison.

Algorithm	Parameters
SLAVE	Number of linguistic labels per variable: 5 Number of individuals: 100, Mutation probability: 0.01 Number of allowed iterations without change: 500
SGERD	Enhanced compatibility degree Heuristic computation of the number of Q rules
C4.5	Pruning: true, Confidence level: 0.25 Minimum number of examples per leaf: 2
CART	Maximum depth: 90
RIPPER	k: 2, Growing set: $\frac{2}{3}$ of the examples
FURIA	Number of optimizations: 2, Number of linguistic labels per variable: 3

- The classical capacity using the cardinality of the set (Card.).
- The proposed FRMs in which we learn a different capacity for each class using an evolutionary method. GA_Card is the proposal in which we have learnt the parameter p of the power capacity whereas GA_Ov is the approach where we have optimized the parameter p of the O_7 overlap index.[2]

The last column (#FiredRules) shows the average number of fired rules in each data-set, that is, we compute the average number of fired rules per class within the five folders and then we average those classes which have fired some rule. The data-sets in this table are sorted according to the average number of fired rules so as to analyse the importance of the new proposal depending on the number of elements to be aggregated. The four last rows are the mean (Mean) ans averaged standard deviations (Mean Std.) for all the datasets considered in the study and the average result for those dataset having an averaged number of fired rules larger than 3. For each data-set the best results is highlighted in **bold-face**.

From the results presented in Table 4 it can be observed three main facts:

- All the proposals based on capacities that do not use an evolutionary method to learn a different capacity for each class exhibit a similar behaviour.
- The proposals based on capacities as well as the FRMs using the classical aggregation operators (Max and Mean) achieve similar results.
- When the number of elements to be aggregated (#FiredRules) is less than 3, the result obtained by all the approaches are quite similar. Moreover, the construction method of capacities becomes more appropriate when the number of elements to be aggregated is larger, since the number of elements to construct the capacity is also larger.

Due to the latter fact, we focus on the analysis of the results of those datasets in which the averaged number of rules is greater than 3. First, we investigate whether there are statistical differences among the eleven approaches that do not use an optimization process. To do so, we have applied the aligned Friedman ranks test, whose obtained p-value is 0.034 and its ranks are shown in Fig. 1. Note that there are five methods (O_1, O_4, O_7, O_8 and Max.) yielding ranks of approximately one hundred or less, whereas the remaining six approaches lead to worse rankings. Finally, we applied Holm's post-hoc test [38] to verify if the best ranking method, which is the classical FRM of the winning rule (Max.), statistically outperforms the other ones. The obtained statistical results are presented in Table 5 which is in agreement with the graphical representation of the rankings.

Regarding these 11 approaches, the observations above indicate that the best way to aggregate the information is using the Choquet integral whose capacities are computed in terms of the overlap indexes O_1, O_4, O_7 and O_8 as well as the maximum as an aggregation function.

[2] O_7 was selected because among the three parametrized overlap indexes is the one providing the best mean result according to Table 4.

Table 4
Results achieved in testing by the different approaches.

Dataset	Max	Mean	O_1	O_2	O_3	O_4	O_5	O_6	O_7	O_8	Card	GA_Ov	GA_Card	#FiredRules
bre	**71.43**	**71.43**	**71.43**	**71.43**	**71.43**	**71.43**	**71.43**	**71.43**	**71.43**	**71.43**	**71.43**	**71.43**	**71.43**	1.00
car	**70.06**	**70.06**	**70.06**	**70.06**	**70.06**	**70.06**	**70.06**	**70.06**	**70.06**	**70.06**	**70.06**	**70.06**	**70.06**	1.00
fla	**63.08**	**63.08**	**63.08**	**63.08**	**63.08**	**63.08**	**63.08**	**63.08**	**63.08**	**63.08**	**63.08**	**63.08**	**63.08**	1.00
hou	**64.26**	**64.26**	**64.26**	**64.26**	**64.26**	**64.26**	**64.26**	**64.26**	**64.26**	**64.26**	**64.26**	**64.26**	**64.26**	1.00
led	**64.20**	**64.20**	**64.20**	**64.20**	**64.20**	**64.20**	**64.20**	**64.20**	**64.20**	**64.20**	**64.20**	**64.20**	**64.20**	1.00
tic	**65.42**	**65.42**	**65.42**	**65.42**	**65.42**	**65.42**	**65.42**	**65.42**	**65.42**	**65.42**	**65.42**	**65.42**	**65.42**	1.00
zoo	**72.38**	**72.38**	**72.38**	**72.38**	**72.38**	**72.38**	**72.38**	**72.38**	**72.38**	**72.38**	**72.38**	**72.38**	**72.38**	1.00
ger	**70.40**	**70.40**	**70.40**	**70.40**	**70.40**	**70.40**	**70.40**	**70.40**	**70.40**	**70.40**	**70.40**	**70.40**	**70.40**	1.05
hay	65.93	67.41	65.93	65.93	65.93	65.93	65.93	65.19	65.93	65.93	67.41	**68.15**	65.93	1.26
mon	**75.40**	**75.40**	**75.40**	**75.40**	**75.40**	**75.40**	**75.40**	**75.40**	**75.40**	**75.40**	**75.40**	**75.40**	**75.40**	1.29
crx	**69.31**	69.16	69.16	69.16	69.16	69.16	69.16	69.16	69.16	69.16	69.16	69.01	69.16	1.42
tit	78.32	78.32	78.32	78.32	78.32	78.32	78.32	78.32	78.32	78.32	78.32	78.32	**78.57**	1.46
gla	57.67	58.60	58.14	58.60	57.67	58.14	58.14	54.42	58.14	58.14	58.60	**59.53**	58.14	1.74
cle	**55.00**	53.67	54.33	53.67	54.00	54.33	54.00	54.00	54.33	54.33	53.67	54.00	54.33	1.78
tae	55.48	56.13	56.13	56.13	56.13	56.13	56.13	55.48	56.13	56.13	56.13	57.42	**58.06**	1.84
ban	60.32	60.47	60.30	61.04	61.36	60.83	61.25	63.23	60.83	60.42	60.47	**69.60**	62.62	1.94
bal	**90.56**	86.72	**90.56**	88.32	89.12	**90.56**	88.80	88.32	**90.56**	89.60	86.72	90.40	**90.56**	2.52
pag	91.45	91.82	91.45	91.82	**92.00**	91.45	91.82	**92.00**	91.45	91.82	91.82	**92.00**	91.59	2.76
shu	80.23	80.23	80.23	80.23	80.23	80.23	80.23	80.23	80.23	80.23	80.23	**83.59**	80.40	2.76
hab	72.26	72.26	72.26	**72.58**	**72.58**	71.94	**72.58**	72.26	71.94	71.94	72.26	72.26	71.94	3.03
hea	**71.48**	**71.48**	71.11	71.11	**71.48**	71.11	**71.48**	71.11	71.11	**71.48**	**71.48**	**71.48**	71.11	3.07
iri	92.67	87.33	**93.33**	90.00	91.33	**93.33**	91.33	92.00	**93.33**	92.67	87.33	**93.33**	**93.33**	3.32
new	85.12	86.05	86.05	86.05	86.05	86.05	86.05	87.44	86.05	85.58	86.05	**91.16**	88.37	3.44
eco	71.76	69.71	71.18	70.00	70.29	71.47	70.88	70.00	71.47	70.59	69.71	**77.94**	72.06	3.62
aus	**83.33**	82.17	82.75	81.88	81.74	82.75	81.59	81.74	82.75	82.46	82.17	82.46	82.61	4.10
mam	**80.99**	80.41	80.41	79.94	79.82	79.58	79.94	80.40	79.58	80.64	80.41	80.64	80.88	5.32
wis	95.62	**96.06**	95.77	**96.06**	95.91	95.62	95.91	95.91	95.62	95.91	**96.06**	95.04	94.89	5.42
app	**87.27**	85.45	**87.27**	84.55	85.45	**87.27**	85.45	83.64	**87.27**	**87.27**	85.45	86.36	**87.27**	6.54
ion	**92.11**	91.55	**92.11**	91.55	91.55	**92.11**	91.55	91.27	**92.11**	91.83	91.55	**92.11**	**92.11**	7.26
sat	47.29	50.39	47.29	50.39	50.39	46.98	50.39	50.23	46.98	48.37	50.39	**79.22**	56.20	7.42
rin	52.70	51.08	51.76	51.08	51.08	51.49	51.08	51.08	51.49	51.49	51.08	**89.46**	56.25	7.81
pho	71.91	71.23	71.60	70.92	71.17	71.64	71.12	72.45	71.64	71.30	71.23	**78.04**	71.58	8.67
seg	85.02	83.51	**85.2**8	83.68	83.64	84.98	83.81	84.29	84.98	84.46	83.51	84.72	85.06	8.84
pen	**94.36**	93.09	94.18	93.09	93.00	**94.36**	92.91	93.00	**94.36**	94.00	93.09	94.09	94.20	10.64
bup	57.97	59.42	58.84	59.71	59.42	59.13	60.00	57.68	59.13	58.26	59.42	61.16	**62.61**	11.37
win	92.78	91.11	92.78	91.67	91.67	92.78	91.67	91.11	92.78	91.67	91.11	92.78	**94.44**	11.68
veh	61.41	60.94	60.71	60.59	60.35	61.06	60.24	58.82	61.06	60.94	60.94	61.29	**61.76**	12.01
pim	72.99	73.77	72.99	74.29	74.16	73.25	74.29	73.90	73.25	73.51	73.77	74.68	**75.00**	16.70
mag	74.75	72.65	74.38	72.49	72.55	74.17	72.60	73.23	74.17	73.07	72.65	**77.90**	77.11	21.39
two	83.78	70.41	84.86	70.41	71.08	84.05	70.54	69.59	84.05	78.24	70.41	**91.62**	87.50	26.68
Mean	73.71	72.98	73.70	73.05	73.13	73.67	73.14	72.95	73.67	73.41	72.98	**76.66**	74.56	
Mean Std.	2.49	2.73	2.68	2.84	2.89	2.74	2.90	2.87	2.74	2.72	2.73	2.95	3.07	
Mean (FR > 3)	77.50	76.19	77.47	76.29	76.41	77.39	76.45	76.25	77.39	76.94	76.19	**82.27**	78.87	
Mean Std. (FR > 3)	2.72	2.96	3.00	3.21	3.25	3.03	3.27	3.27	3.03	3.05	2.96	3.41	3.75	

However, according to the results shown in Table 4, when learning a different capacity for each class of the problem a notable enhancement of the obtained results is found, which is especially remarkable when applying the GA_Ov approach. It is remarkable that both approaches using genetic algorithms (especially GA_Ov) outperform the remaining ones when dealing with data-sets having a number of elements to be aggregated greater than 3 (number of fired rules).

In order to check if one of these optimization proposals leads to improvements in the results given by the classical aggregation functions (Max. and Mean), we apply the same statistical study that has been previously explained. The p-value computed with the aligned Friedman ranks test is 7.20E-4, which means that there are statistical differences among these four approaches. The results of the Holm's post-hoc process are shown in Table 6. The GA_Ov, which

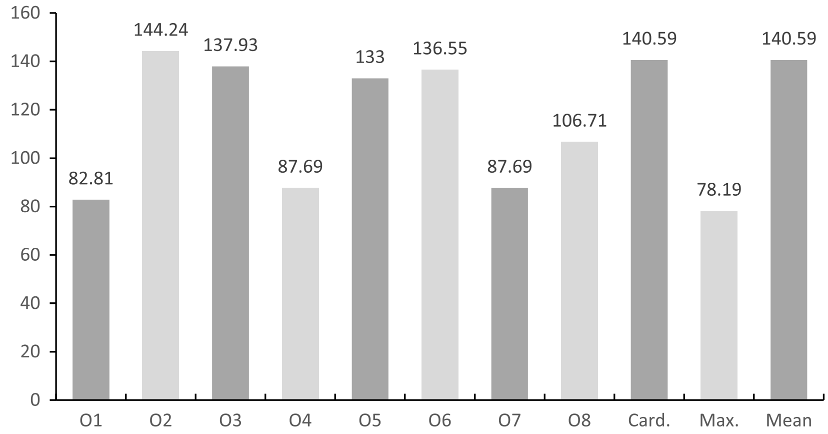

Fig. 1. Rankings of the approaches that do not learn capacities.

Table 5
Holm test to compare the approaches that do not learn a capacity. Max is used as a reference.

i	Algorithm	APV
1	O_2	0.014
4	Mean	0.022
5	Card.	0.022
2	O_3	0.026
3	O_6	0.028
6	O_5	0.039
7	O_8	0.666
8	O_4	1.00
9	O_7	1.00
10	O_1	1.00

Table 6
Holm test to compare the approaches that learn capacities and classical ones. GA_Ov is used as a reference.

i	Algorithm	APV
1	Mean	8.38E-7
2	Max.	0.007
3	GA_Card	0.139

achieved the best ranking according to Fig. 2, statistically outperforms the approaches using classical aggregation operators with a high level of confidence. Regarding GA_Card, there are not statistical differences in favour to GA_Ov. However, a closer look at the results provided by these two methods, we can observe that GA_Ov provides the best results in 18 out of the 21 datasets, there is 1 draw and GA_Card achieves the best result in the remainder 2 datasets. In light of all these facts, we can confirm that the best FRM among the ones considered in this paper is the one based on capacities constructed from overlap indexes using an evolutionary method to learn a different capacity for each class of the problem, that is, GA_Ov.

Finally, we want to compare the best approach among the ones based on capacities constructed from overlap indexes (GA_Ov) versus several classifiers that are published in the specialized literature, namely, SGERD, SLAVE,[3] C4.5, CART, RIPPER and FURIA. The results of these 6 classifiers and the ones provided by GA_Ov are reported

[3] Let us remark that the SLAVE algorithm does not learn any rule when all the attributes of the datasets are nominal and the number of classes is greater than 2. However, this fact does not affect our analysis since the average number of fired rules in all these datasets is one.

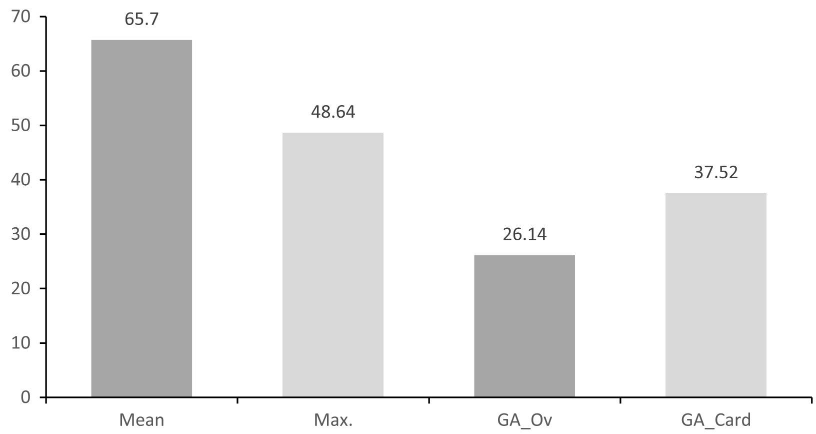

Fig. 2. Rankings of the approaches that learn capacities and the classical ones.

in Table 7, which is also sorted according to the number of fired rules (elements to be aggregated using the GA_Ov approach). The best result for each dataset is highlighted in **bold-face**.

The results shown in Table 7 reveal that when considering all the datasets in the study the results of GA_Ov are worse than those of four classifiers. However, GA_Ov works well providing the third best mean result (just worse than FURIA and the C4.5 decision tree) with respect to the datasets whose averaged number of fired rules is larger than 3 (where our proposal becomes more important). Furthermore, we can observe that the difference in average is reduced from 7.16%, using all datasets, to 2.88%, in the latter scenario, with FURIA and from 5.62% to 1.05% versus C4.5. This reduction shows the benefits of our new FRM in this setting.

Following the previously explained scheme, we carry out a statistical test to check whether there are statistical differences among these seven approaches. The p-value computed with the aligned Friedman test is 0.001 and the obtained ranks are graphically depicted in Fig. 3, where it is shown that the best ranking method is FURIA. Finally, the results of Holm's post-hoc test are reported in Table 8, where it can be observed that FURIA outperforms all the classifiers as it was expected according to the results shown in Table 7

Finally, we analyse the performance of GA_Ov by carrying out a set of pairwise comparisons of our approach with the other four classifiers from the literature using the Wilcoxon test, whose results are shown in Table 9. These results indicate that GA_Ov yields competitive results with respect all the classifiers in this study (since there are not statistical differences) except FURIA and even it outperforms SGERD as FURIA does. The greater flexibility of the fuzzy partition used by FURIA implies obtaining a leap in the obtained performance and thus the improvement of all the methods including GA_Ov. Anyway, we would like to emphasize the large improvement provided by our novel FRM when the number of elements to be aggregated is large enough, which makes it an alternative for combining fuzzy rules in FRBCSs.

10. Conclusions

In this paper, we presented a method for constructing capacities from the data of a given problem. This method has the following steps:

1. Build fuzzy sets from the inputs;
2. Build as many subsets of the considered fuzzy set as the cardinal of the parts of the considered referential set;
3. Compute the overlaps of the fuzzy set and each of its subsets.
4. Build the capacity as a quotient of overlap indexes.

The advantage of this method is that it makes use of the data of the considered problem for constructing the measures. In this way, aggregation by means of the Choquet integral based on these measures is much more conditioned by the data than in the case where interactions between data are not taken into account.

Table 7
Results achieved in testing by the different approaches.

Dataset	GA_Ov	SGERD	SLAVE	RIPPER	C4.5	CART	FURIA	#FiredRules
bre	71.43	70.79	70.76	63.56	74.37	70.76	**76.54**	1.00
car	70.06	67.19	70.02	89.01	90.80	80.85	**91.26**	1.00
fla	63.08	66.89	0.00	68.38	74.48	**75.23**	73.73	1.00
hou	64.26	87.95	53.45	**96.98**	96.56	93.52	96.56	1.00
led	64.20	38.00	0.00	47.80	70.60	40.20	**71.80**	1.00
tic	65.42	68.89	65.34	97.49	85.80	72.65	**98.22**	1.00
zoo	72.38	83.19	0.00	89.14	94.10	86.14	**95.05**	1.00
ger	70.40	67.50	70.30	67.80	71.80	70.10	**73.30**	1.05
hay	68.15	49.26	80.28	81.74	**83.30**	76.50	81.00	1.26
mon	75.40	80.55	97.45	**100.00**	**100.00**	**100.00**	**100.00**	1.29
crx	69.01	86.22	68.45	82.54	85.30	84.38	**86.37**	1.42
tit	78.32	77.60	**78.87**	70.47	77.78	77.10	78.51	1.46
gla	59.53	61.23	57.04	64.01	68.73	68.70	**72.91**	1.74
cle	54.00	48.13	**57.23**	42.44	51.82	49.17	56.57	1.78
tae	**57.42**	49.68	53.61	45.01	54.99	47.76	45.61	1.84
ban	69.60	66.89	68.85	61.17	**89.00**	75.66	88.57	1.94
bal	**90.40**	72.96	77.76	48.96	77.28	62.24	83.68	2.52
pag	92.00	90.33	93.61	94.70	95.07	94.34	**95.25**	2.76
shu	83.59	78.48	85.70	99.49	99.54	**99.72**	99.68	2.76
hab	72.26	**73.86**	72.88	52.61	72.22	69.61	72.55	3.03
hea	71.48	75.19	76.67	76.67	79.26	72.59	**80.37**	3.07
iri	93.33	93.33	**96.00**	94.00	93.33	95.33	94.00	3.32
new	91.16	88.84	89.30	93.02	91.16	94.42	**94.88**	3.44
eco	77.94	73.82	**84.52**	70.56	78.28	73.82	80.06	3.62
aus	82.46	85.51	70.29	82.03	84.20	84.06	**86.09**	4.10
mam	80.64	78.68	78.90	76.82	**83.61**	83.24	83.57	5.32
wis	95.04	93.85	95.75	95.61	95.03	93.27	**96.63**	5.42
app	86.36	**87.75**	85.93	80.04	84.98	82.08	87.71	6.54
ion	**92.11**	74.65	87.75	88.89	88.03	85.77	89.75	7.26
sat	79.22	63.47	66.72	78.54	80.09	79.62	**82.27**	7.42
rin	**89.46**	73.51	86.22	82.43	82.70	83.38	85.54	7.81
pho	78.04	74.67	76.17	82.22	85.57	80.03	**85.90**	8.67
seg	84.72	78.01	87.66	94.20	96.32	95.93	**97.32**	8.84
pen	**94.09**	66.55	89.91	85.27	89.36	62.00	92.45	10.64
bup	61.16	56.81	60.00	63.48	66.09	68.99	**70.14**	11.37
win	92.78	92.14	93.79	89.86	**94.90**	92.11	93.78	11.68
veh	61.29	52.60	64.66	69.39	**71.87**	65.36	70.21	12.01
pim	74.68	73.30	73.04	69.92	74.09	71.87	**76.17**	16.70
mag	77.90	71.50	78.55	77.60	79.81	74.76	**80.65**	21.39
two	**91.62**	71.76	82.16	84.19	78.78	80.41	88.11	26.68
Mean	76.66	72.79	71.14	77.45	82.28	77.84	**83.82**	
Mean (FR > 3)	82.27	76.18	80.80	80.35	83.32	80.41	**85.15**	

Finally, we performed experiments in classification by using our measures in FRBCSs. The experimental results produced by our approach were better than those obtained with classical aggregation functions such as the maximum (FRM of the wining rule) or the mean. Our proposal is suitable for situations where the number of elements to be aggregated is larger than 3, since the construction method of the capacities makes sense in this scenario. Among all the FRMs using capacities, the best one is the FRM based on capacities whose fuzzy measure has been genetically learnt to model the interaction among the rules belonging to the different classes.

In future research, we intend to analyse aggregations that make use of our measures, such as Sugeno, Shilkret, Choquet or copula-based integrals, and compare the results obtained using other aggregations in different problems, such as, for instance, digital fingerprint recognition, decision making, etc. Moreover, we intend to generalize the proposed algorithm to fuzzy rule-based classification systems using different aggregation functions in each of their steps as well as with different structures for the fuzzy partitions.

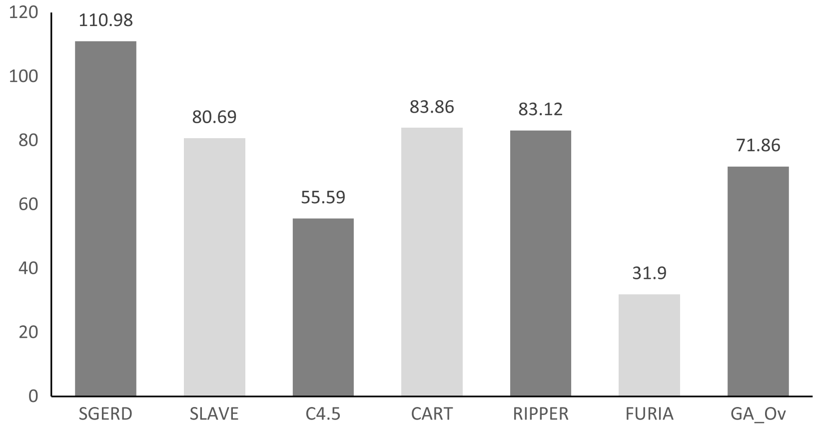

Fig. 3. Rankings of GA_Ov and the classifiers published in the specialized literature.

Table 8
Holm test to compare GA_Ov and the six classifiers selected in this paper. FURIA is used as the control method.

i	Algorithm	APV
1	SGERD	1.06E-8
2	CART	3.85E-4
3	RIPPER	3.89E-4
4	SLAVE	6.15E-4
5	GA_Ov	0.005
6	C4.5	0.071

Table 9
Wilcoxon test to compare GA_Ov (R+) with the classifiers selected in this study (R−).

Comparative	R+	R−	p-value	Hypothesis
GA_Ov vs. SGERD	199.5	31.5	0.004	Rej. GA_Ov 95%
GA_Ov vs. SLAVE	145.0	86.0	0.305	Accepted
GA_Ov vs. C4.5	86.5	144.5	0.295	Accepted
GA_Ov vs. CART	169.5	91.5	0.404	Accepted
GA_Ov vs. RIPPER	149.0	82.0	0.244	Accepted
GA_Ov vs. FURIA	45.0	186.0	0.014	Rej. FURIA 95%

Acknowledgements

The work has been supported by projects TIN2013-40765-P of the Spanish Ministry of Science and by grants VEGA 1/0420/15 and VEGA 1/0419/13.

References

[1] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, KEEL data-mining software tool: data set repository, integration of algorithms and experimental analysis framework, J. Mult.-Valued Log. Soft Comput. 17 (2–3) (2011) 255–287.
[2] A. Amo, J. Montero, D. Gómez, E. Molina, Representation of consistent recursive rules, Eur. J. Oper. Res. 130 (2001) 29–53.
[3] P. Benvenuti, R. Mesiar, Pseudo-additive measures and triangular-norm-based conditioning, Ann. Math. Artif. Intell. 35 (2002) 63–69.
[4] P. Benvenuti, D. Vivona, M. Divari, Aggregation operators and associated fuzzy measures, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 9 (2001) 197–204.
[5] L. Breiman, J. Friedman, C.J. Stone, R.A. Olshen, Classification and Regression Trees, CRC press, Boca Raton, 1984.
[6] H. Bustince, E. Barrenechea, T. Calvo, S. James, G. Beliakov, Consensus in multi-expert decision making problems using penalty functions defined over a Cartesian product of lattices, Inf. Fusion 17 (2014) 56–64.
[7] H. Bustince, M. Pagola, R. Mesiar, E. Hüllermeier, F. Herrera, Grouping, overlap, and generalized bientropic functions for fuzzy modeling of pairwise comparisons, IEEE Trans. Fuzzy Syst. 20 (3) (2012) 405–415.

[8] H. Bustince, E. Barrenechea, M. Pagola, F. Soria, Weak fuzzy S-subsethood measures: overlap index, Int. J. Uncertain. Fuzziness Knowl.-Based Syst. 14 (5) (2006) 537–560.
[9] H. Bustince, J. Montero, E. Barrenechea, M. Pagola, Semiautoduality in a restricted family of aggregation operators, Fuzzy Sets Syst. 158 (2007) 1360–1377.
[10] H. Bustince, J. Fernandez, R. Mesiar, J. Montero, R. Orduna, Overlap functions, Nonlinear Anal. 72 (3–4) (2010) 1488–1499.
[11] T. Calvo, A. Kolesárová, M. Komorníková, R. Mesiar, Aggregation operators: properties, classes and construction methods, in: T. Calvo, G. Mayor, R. Mesiar (Eds.), Aggregation Operators. New Trends and Applications, Physica-Verlag, Heidelberg, 2002, pp. 3–104.
[12] Z. Chi, H. Yan, T. Pham, Fuzzy Algorithms with Applications to Image Processing and Pattern Recognition, World Scientific, 1996.
[13] G. Choquet, Theory of capacities, Ann. Inst. Fourier 5 (1953–1954) 131–295.
[14] W.W. Cohen, Fast effective rule induction, in: 12th Int. Conf. Mach. Learn., Lake Tahoe, CA, USA, 1995.
[15] O. Cordón, M.J. del Jesus, F. Herrera, A proposal on reasoning methods in fuzzy rule-based classification systems, Int. J. Approx. Reason. 20 (1) (1999) 21–45.
[16] V. Cutello, J. Montero, Recursive connective rules, Int. J. Intell. Syst. 14 (1999) 3–20.
[17] B.A. Davey, H.A. Priestley, Introduction to Lattices and Order, Cambridge University Press, 2002.
[18] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[19] D. Dubois, W. Ostasiewicz, H. Prade, Fuzzy sets: history and basic notions, in: D. Dubois, H. Prade (Eds.), Fundamentals of Fuzzy Sets, Kluwer, Boston, MA, 2000, pp. 21–124.
[20] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley, 2001.
[21] L.J. Eshelman, The CHC adaptive search algorithm: how to have safe search when engaging in nontraditional genetic recombination, in: G.J.E. Rawlins (Ed.), Foundations of Genetic Algorithms, Morgan Kaufmann, San Mateo, 1991, pp. 265–283.
[22] J. Fodor, M. Roubens, Fuzzy Preference Modelling and Multicriteria Decision Support, Kluwer Academic Publishers, Dordrecht, 1994.
[23] M. Galar, A. Fernández, E. Barrenechea, F. Herrera, EUSBoost: enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling, Pattern Recognit. 46 (12) (2013) 3460–3471.
[24] M. Galar, A. Fernández, E. Barrenechea, F. Herrera, Empowering difficult classes with a similarity-based aggregation in multi-class classification problems, Inf. Sci. 264 (2014) 135–157.
[25] S. Garcia-Jimenez, H. Bustince, E. Hüllermeier, R. Mesiar, N.R. Pal, A. Pradera, Overlap Indices: construction of and application to interpolative fuzzy systems, IEEE Trans. Fuzzy Syst. 23 (4) (2015) 1259–1273.
[26] S. García, A. Fernández, J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability, Soft Comput. 13 (10) (2009) 959–977.
[27] S. García, A. Fernández, J. Luengo, F. Herrera, Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power, Inf. Sci. 180 (10) (2010) 2044–2064.
[28] A. González, R. Perez, SLAVE: a genetic learning system based on an iterative approach, IEEE Trans. Fuzzy Syst. 7 (2) (1999) 176–191.
[29] M. Grabisch, J.-L. Marichal, R. Mesiar, E. Pap, Aggregation Functions, Cambridge University Press, Cambridge, 2009.
[30] J. Hühn, E. Hüllermeier, FURIA: an algorithm for unordered fuzzy rule induction, Data Min. Knowl. Discov. 19 (3) (2009) 293–319.
[31] H. Ishibuchi, T. Nakashima, Effect of rule weights in fuzzy rule-based classification systems, IEEE Trans. Fuzzy Syst. 9 (4) (2001) 506–515.
[32] H. Ishibuchi, T. Nakashima, M. Nii, Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining, Springer-Verlag, Berlin, 2004.
[33] H. Ishibuchi, T. Yamamoto, Rule weight specification in fuzzy rule-based classification systems, IEEE Trans. Fuzzy Syst. 13 (2001) 428–435.
[34] A. Jurio, H. Bustince, M. Pagola, A. Pradera, R.R. Yager, Some properties of overlap and grouping functions and their application to image thresholding, Fuzzy Sets Syst. 229 (2013) 69–90.
[35] D. Gómez, J. Montero, A discussion on aggregation operators, Kybernetika 40 (2004) 107–120.
[36] F. Herrera, M. Lozano, A.M. Sánchez, A taxonomy for the crossover operator for real-coded genetic algorithms: an experimental study, Int. J. Intell. Syst. 18 (2003) 309–338.
[37] J.L. Hodges, E.L. Lehmann, Ranks methods for combination of independent experiments in analysis of variance, Ann. Math. Stat. 33 (1962) 482–497.
[38] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (1979) 65–70.
[39] G.J. Klir, T.A. Folger, Fuzzy Sets, Uncertainty and Information, Prentice Hall, Englewood Cliffs, NJ, 1988.
[40] D. Maltoni, D. Maio, A. Jain, S. Prabhakar, Handbook of Fingerprint Recognition, Springer, New York, 2009.
[41] E.G. Mansoori, M.J. Zolghadri, S.D. Katebi, SGERD: a steady-state genetic algorithm for extracting fuzzy classification rules from data, IEEE Trans. Fuzzy Syst. 16 (4) (2008) 1061–1071.
[42] R. Mesiar, A. Mesiarová-Zemánková, The ordered modular averages, IEEE Trans. Fuzzy Syst. 19 (2011) 42–50.
[43] E. Pap, Handbook of Measure Theory, Part I, Part II, Elsevier, Amsterdam, 2002.
[44] J. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo–California, 1993.
[45] J. Sanz, A. Fernández, H. Bustince, F. Herrera, IVTURS: a linguistic fuzzy rule-based classification system based on a new Interval-Valued fuzzy reasoning method with TUning and Rule Selection, IEEE Trans. Fuzzy Syst. 21 (3) (2013) 399–411.
[46] J. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, H. Bustince, Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system, Appl. Soft Comput. 20 (2014) 103–111.
[47] J. Sanz, D. Bernardo, F. Herrera, H. Bustince, H. Hagras, A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data, IEEE Trans. Fuzzy Syst. 23 (4) (2015) 973–990.
[48] D. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, second edition, Chapman & Hall/CRC, Boca Raton, 2006.
[49] V. Torra, Y. Narukawa, M. Sugeno (Eds.), Non-Additive Measures, Theory and Applications, Studies in Fuzziness and Soft Computing, vol. 310, 2014.

[50] Z. Tu, S. Zheng, A.L. Yuille, A. Reiss, E. Dutton, A. Lee, A. Galaburda, I. Dinov, P. Thompson, A. Toga, Automated extraction of the cortical sulci based on a supervised learning approach, IEEE Trans. Med. Imaging 25 (2007) 541–552.
[51] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics 1 (1945) 80–83.
[52] L.A. Zadeh, Fuzzy sets, Inf. Control 8 (1965) 338–353.