A proposal for tuning the α parameter in $C_\alpha C$-integrals for application in fuzzy rule-based classification systems

Giancarlo Lucca · José A. Sanz · Graçaliz P. Dimuro · Benjamín Bedregal · Humberto Bustince

Abstract In this paper, we consider the concept of extended Choquet integral generalized by a copula, called CC-integral. In particular, we adopt a CC-integral that uses a copula defined by a parameter α, which behavior was tested in a previous work using different fixed values. In this contribution, we propose an extension of this method by learning the best value for the parameter α using a genetic algorithm. This new proposal is applied in the fuzzy reasoning method of fuzzy rule-based classification systems in such a way that, for each class, the most suitable value of the parameter α is obtained, which can lead to an improvement on the system's performance. In the experimental study, we test the performance of 4 different so called $C_\alpha C$-integrals, comparing the results obtained when using fixed values for the parameter α against the results provided by our new evolutionary approach. From the obtained results, it is possible to conclude that the genetic learning of the parameter α is statistically superior than the fixed one for two copulas. Moreover, in general, the accuracy achieved in test is superior than that of the fixed approach in all functions. We also compare the quality of this approach with related approaches, showing that the methodology proposed in this work provides competitive results. Therefore, we demonstrate that $C_\alpha C$-integrals with α learned genetically can be considered as a good alternative to be used in fuzzy rule-based classification systems.

Giancarlo Lucca
Dpt. de Automática y Computación
Universidad Publica de Navarra
Pamplona, Spain
E-mail: lucca.112793@e.unavarra.es

José Antonio Sanz
Dpt. de Automática y Computación and
Institute of Smart Cities
Universidad Publica de Navarra
Pamplona, Spain
E-mail: joseantonio.sanz@unavarra.es

Graçaliz P. Dimuro
Institute of Smart Cities
Universidad Publica de Navarra
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
E-mail: gracaliz@gmail.com

Benjamín Bedregal
Departamento de Informatica e
Matemática Aplicada,
Universidade Federal do Rio Grande do Norte
Natal, Brazil
E-mail: bedregal@dimap.ufrn.br

Humberto Bustince
Dpt. de Automática y Computación and
Institute of Smart Cities
Universidad Publica de Navarra
Pamplona, Spain
E-mail: bustince@unavarra.es

1 Introduction

Fuzzy Rule-Based Classification Systems (FRBCSs) are widely studied (Ishibuchi et al. (2005)) and also a powerful tool commonly used to tackle classifications problems (Duda et al. (2001)). Since this kind of systems have a good classification rate and provide interpretable models by using linguist labels in their antecedent rules, FRBCSs have been applied in severals real world problems, including industry (Samantaray et al. (2010)), health (Sanz et al. (2014)), economy (Sanz et al. (2015)) and many others.

The Fuzzy Reasoning Method (FRM) (Cordón et al. (1999)) is applied to classify new examples using the fuzzy rules of Fuzzy Rule-Based Classification Systems (FRBCSs). A key component of the FRM is the aggregation stage, where

different aggregation functions have been considered in the specialized literature. In Lucca et al. (2016c), an aggregation function generalizing the expanded form of the Choquet integral was proposed. Specifically, the product between the value to be aggregated and the difference of fuzzy measures is distributed and it is generalized by a copula. Copulas link (two-dimensional) probability distribution functions to their one-dimensional margins, playing an important role in the theory of probabilistic metric spaces and statistics (Alsina et al. (2006)). As a consequence of this generalization, the authors introduced a family of Choquet-based non-associative aggregation functions named CC-Integrals.

In Lucca et al. (2015), CC-integrals were applied in the FRM of FRBCSs, in order to improve the quality of the system. In that work, a copula that is directly related to a parameter α was proposed. The authors tested in this function 5 different fixed values for this parameter, showing that the copula using the value 0.1 for the parameter α presented a good classification rate.

In the work presented at BRACIS 2016, Lucca et al. (2016a), instead of considering fixed values for the parameter α, presented a method to learn the most suitable value for this parameter in each class, using an evolutionary algorithm.

In this present paper, we extend the work in Lucca et al. (2016a), by conducting a deeper study on the use of copulas considering the parameter α, defining the family of $C_\alpha C$-integrals. Specifically, we study the performance of the family of copulas related to parameters α defined in Alsina et al. (2006). Furthermore, we also consider the application of the genetic approach to learn the most suitable value of the parameter α. Regarding the work presented in Lucca et al. (2016a), we broaden the study as follows:

- We use four different functions that consider the parameter α in the family $C_\alpha C$-integrals.
- Since $C_\alpha C$-integrals are related to a fuzzy measure (Murofushi et al. (1994)), we adopt a method for building fuzzy measures based un fuzzy sets derived form FRBCS, which was introduced recently by Paternain et al. (2016).
- We analyse the results obtained with the values learned by the evolutionary method for the parameters α among the considered $C_\alpha C$-integrals, also performing a comparative study including other methods found in the literature.

In order to demonstrate the quality of the approach, 30 datasets, that are available in the KEEL [1] database repository (Alcalá-Fdez et al. (2009)), are selected. The FARC-HD (Alcalá-Fdez et al. (2011)) is applied to accomplish the fuzzy rule learning process. Finally, our conclusions are supported by the statistical Wilcoxon signed-rank test (Wilcoxon (1945)).

The paper is organized as follows. Section 2 presents some preliminary concepts that are necessary to develop the paper, also explaining the process to build fuzzy measures based on fuzzy sets, which we adopt in this work. The definition of the $C_\alpha C$-integral is presented in Section 3. In Section 4 the proposed FRM is presented, including the methodology to construct the fuzzy measure and the approach to genetically learn the value for the parameter α. We describe the experimental framework, introduce the results achieved in testing by the application of the generalized Choquet integral in FRBCSs besides the quality analysis of the best function in Section 5. The main conclusions are drawn in Section 6.

2 Preliminaries

This section aims at introducing the background necessary to understand the paper.

Definition 1 (Zadeh (1965)) A fuzzy set F defined on a finite and non-empty referential set (or universe) U is given by:

$$F = \{(u, \mu_F(u)) \mid u \in U\}$$

where $\mu_F : U \to [0,1]$ is the membership function.

The family of fuzzy sets on U is denoted by $FS(U)$.

Let $F_1, F_2 \in FS(U)$. Then $B \subseteq C$ if and only if, for all $u \in U$, it holds that $\mu_{F_1}(u) \leq \mu_{F_2}(u)$.

2.1 Aggregation and Pre-aggregation Functions

One important class of fuzzy operators are the *aggregation operators* (Beliakov et al. (2007); Mayor and Trillas (1986)).

Definition 2 A function $A : [0,1]^n \to [0,1]$ is said to be an n-ary aggregation operator if the following conditions hold:

(A1) A is increasing [2] in each argument: for each $i \in \{1, \ldots, n\}$, if $x_i \leq y$, then
$A(x_1, \ldots, x_n) \leq A(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n)$;
(A2) A satisfies the Boundary conditions: $A(0, \ldots, 0) = 0$ and $A(1, \ldots, 1) = 1$.

Let $\mathbf{r} = (r_1, \ldots, r_n)$ be a real n-dimensional vector, $\mathbf{r} \neq \mathbf{0}$. A function $F : [0,1]^n \to [0,1]$ is directionally increasing (Bustince et al. (2015)) with respect to $\mathbf{r}$ ($\mathbf{r}$-increasing, for short) if for all $(x_1, \ldots, x_n) \in [0,1]^n$ and $c > 0$ such that $(x_1 + cr_1, \ldots, x_n + cr_n) \in [0,1]^n$ it holds that

$$F(x_1 + cr_1, \ldots, x_n + cr_n) \geq F(x_1, \ldots, x_n). \qquad (1)$$

Similarly, one defines an $\mathbf{r}$-decreasing function.

---

[2] In this paper, a increasing (decreasing) function does not need to be strictly increasing (decreasing).

Definition 3 (Lucca et al. (2016b)) A function $F : [0, 1]^n \to [0, 1]$ is said to be an n-ary pre-aggregation function if the following conditions hold:

(PA1) F is $\mathbf{r}$-increasing, for some $\mathbf{r} \in \mathbb{R}^n$, $\mathbf{r} \neq \mathbf{0}$;
(PA2) F satisfies the boundary conditions: $F(0, \ldots, 0) = 0$, $F(1, \ldots, 1) = 1$.

If F is a pre-aggregation function with respect to a vector $\mathbf{r}$ we just say that F is an $\mathbf{r}$-pre-aggregation function. See also the works by Mesiar et al. (2016) and Dimuro et al. (2016b).

Definition 4 An aggregation function $T : [0, 1]^2 \to [0, 1]$ is a t-norm if, for all $x, y, z \in [0, 1]$, it satisfies the following properties:

(T1) Commutativity: $T(x, y) = T(y, x)$;
(T2) Associativity: $T(x, T(y, z)) = T(T(x, y), z)$;
(T3) Boundary condition: $T(x, 1) = x$.

If T satisfies only the property *(T3)* of a t-norm (but also its symmetric $T(1, x) = x$), then it is called a semi-copula. Some examples of t-norms are presented in Table 1.

Table 1 Examples of t-norms

Name	Definition
Minimum	$T_M(x, y) = \min\{x, y\}$
Algebraic Product	$T_P(x, y) = xy$
Lukasiewicz	$T_L(x, y) = \max\{0, x + y - 1\}$
Hamacher Product	$T_{HP}(x, y) = \begin{cases} 0 & \text{if } x = y = 0 \\ \frac{xy}{x+y-xy} & \text{otherwise} \end{cases}$

Definition 5 (Alsina et al. (2006)) A bivariate function $C : [0, 1]^2 \to [0, 1]$ is a copula if it satisfies the following conditions, for all $x, x', y, y' \in [0, 1]$ with $x \leq x'$ and $y \leq y'$:

(C1) $C(x, y) + C(x', y') \geq C(x, y') + C(x', y)$;
(C2) $C(x, 0) = C(0, x) = 0$;
(C3) $C(x, 1) = C(1, x) = x$.

In this paper, we consider 4 different commutative, non-associative copulas, that were defined at. Alsina et al. (2006) (appendix A (A.2.1)). These four functions are shown in Table 2, where we have to take into account the following notation:

- Max = $\max\{x, y\}$
- Min = $\min\{x, y\}$
- W = $\max\{0, x + y - 1\}$
- P = xy

2.2 Fuzzy measures and the Choquet integral

For the following, consider $N = \{0, \ldots, n\}$.

Definition 6 (Choquet (1953–1954); Murofushi et al. (1994)) A function $\mathfrak{m} : 2^N \to [0, 1]$ is a discrete fuzzy measure if, for all $X, Y \subseteq N$, it satisfies properties:

(m1) Increasing: if $X \subseteq Y$, then $\mathfrak{m}(X) \leq \mathfrak{m}(Y)$;
(m2) Boundary conditions: $\mathfrak{m}(\emptyset) = 0$ and $\mathfrak{m}(N) = 1$.

Definition 7 (Choquet (1953–1954)) Let $\mathfrak{m} : 2^N \to [0, 1]$ be a discrete fuzzy measure. The discrete Choquet integral for $\mathfrak{m}$ is defined as a function $\mathfrak{C}_\mathfrak{m} : [0, 1]^n \to [0, 1]$, given by

$$\mathfrak{C}_\mathfrak{m}(\mathbf{x}) = \sum_{i=1}^n \left(x_{(i)} - x_{(i-1)} \right) \cdot \mathfrak{m}\left(A_{(i)} \right), \tag{2}$$

where $(x_{(i)}, \ldots, x_{(n)})$ is an increasing permutation on the input x, that is, $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$, with the convention that $x_{(0)} = 0$, and $A_{(i)} = \{(i), \ldots, (n)\}$ is the subset of index of $n - i + 1$ largest components of $\mathbf{x}$.

Observe that the Equation (2) can be also written as:

$$\mathfrak{C}_\mathfrak{m}(\mathbf{x}) = \sum_{i=1}^n \left(x_{(i)} \cdot \mathfrak{m}\left(A_{(i)} \right) - x_{(i-1)} \cdot \mathfrak{m}\left(A_{(i)} \right) \right), \tag{3}$$

which we call the Choquet integral in its expanded form.

The Choquet integral combines the inputs in such a way that not only the importance of individual inputs or of their magnitude are taken into account, but also the importance of their groups (or coalitions in which it takes part), allowing to assign importance to all possible groups of criteria (Grabisch and Labreuche (2010)).

2.3 Overlap functions, overlap indexes and the construction of fuzzy measures

Overlap functions (Bedregal et al. (2013); Bustince et al. (2010); Dimuro and Bedregal (2014, 2015); Dimuro et al. (2016a,c)) are special aggregation functions proposed to deal with the overlap problem and/or when the associativity property is not strongly required, as in image processing (Jurio et al. (2013)) and decision making based on fuzzy preference relations (Bustince et al. (2012)). In particular, they play an important role in classification problems (Elkano et al. (2015, 2016); Lucca et al. (2015, 2016c)).

In this paper, overlap functions are used to derived overlap indexes (Bustince et al. (2009); Garcia-Jimenez et al. (2015)), which are used to build links among the fuzzy sets defining the input data. In the following, we show how to use overlap indexes to define fuzzy measures, as proposed by Paternain et al. (2016).

Definition 8 (Bustince et al. (2010)) An overlap function is a mapping $G_O : [0, 1]^2 \to [0, 1]$ such that:

(GO1) $G_O(x, y) = G_O(y, x)$ for every $x, y \in [0, 1]$;

Table 2 Example of copulas used in this paper

Copula ID	Functions	Property
(a)	$C_\alpha(x,y) = xy[1 + \alpha(1-x)(1-y)]$	$-1 \le \alpha \le 1 \,(\alpha \ne 0)$
(b)	$C_\alpha(x,y) = \frac{1}{1+\alpha} Max[x + y - 1 + \alpha - \alpha|x - y|, 0]$	$0 < \alpha < 1$
(c)	$C_\alpha = (1-\alpha)W + \alpha Min$	$0 < \alpha < 1$
(d)	$C_\alpha = \frac{\alpha^2(1-\alpha)}{2}W + (1-\alpha^2)P + \frac{\alpha^2(1+\alpha)}{2}Min$	$-1 < \alpha < 1 \,(\alpha \ne 0)$

(GO2) $G_O(x,y) = 0$ if and only if $xy = 0$;

(GO3) $G_O(x,y) = 1$ if and only if $xy = 1$;

(GO4) G_O is increasing;

[(GO5) G_O is continuous.

Definition 9 (Sola et al. (2006)) 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 $B \subseteq C$, then $O(A,B) \le O(A,C)$.

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.

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 $\widetilde{A}$ (the membership degree of all the elements to the fuzzy set A is always 1), 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.

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

$$m_{O,E}(\widetilde{A}) = \frac{1}{O(E,E)} O(E, E_{\widetilde{A}}) \tag{4}$$

is a capacity or fuzzy measure for every overlap index O.

3 Defining the Class of $C_\alpha C$-integrals

Definition 10 Let $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure and $C_\alpha : [0,1]^2 \to [0,1]$ be a family of copulas indexed by α. The family of discrete $C_\alpha C$-integrals with respect to $\mathfrak{m}$ is

defined as the function $\mathfrak{C}_{\mathfrak{m}}^{C_\alpha} : [0,1]^n \to [0,1]$, given, for all $x \in [0,1]^n$, by

$$\mathfrak{C}_{\mathfrak{m}}^{C_\alpha}(\mathbf{x}) = \sum_{i=1}^{n} C_\alpha\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - C_\alpha\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right),$$

$$\tag{5}$$

where $(x_{(i)}, \ldots, x_{(n)})$ is an increasing permutation on the input x and $A_{(i)} = \{(i), \ldots, (n)\}$ is the subset of indexes of $n - i + 1$ largest components of $\mathbf{x}$, and α having different ranges, according to adopted the function.

Theorem 1 *For any* $\alpha \in [-1,1]$, *bivariate copula* $C : [0,1]^2 \to [0,1]$ *and fuzzy measure* $\mathfrak{m} : 2^N \to [0,1]$, $\mathfrak{C}_{\mathfrak{m}}^{C_\alpha}$ *is an average aggregation function.*

Proof It follows directly from Lucca et al. (2016c, Thorem 4).

Applying the functions proposed in Table 2 in Equation (5), the definition of the family of discrete $C_\alpha C$-integrals with respect to a fuzzy measure $\mathfrak{m}$ (in this paper we build a measure based on overlap indexes (see Section 2.3)) is presented in Table 3, also considering the the notation introduced before (Min, W and P).

4 A proposal of a fuzzy reasoning method using the Choquet integral generalized by copulas and genetic algorithms

This section aims at introducing in detail the proposal of this paper. We start by describing the main components of FRBCSs. Then we present the methodology to build the fuzzy measure (Section 4.1), the FRM using the generalized Choquet integral (Section 4.2) and, finally, we present the genetic algorithm for determining the value of the parameter α (Section 4.3).

In what follows, consider that a classification problem consists of m training examples $x_p = (x_{p1}, \ldots, x_{pn}, y_p)$, with $p = 1, \ldots, m$, where x_{pi}, with $i = 1, \ldots, n$, is the value of the i-th variable and $y_p \in \mathbb{Y} = \{Y_1, \ldots, Y_M\}$ is the label of the class of the p-th training example. Among the different existing approaches to deal with classification problems, we adopt FRBCSs. These kind of system has two main components:

Table 3 Choquet-α Copula-based integral ($C_\alpha C - integral$)

ID	Functions	Range
(a)	$C_m^\alpha(x,y) = (x_i y[1 + \alpha(1-x_i)(1-y)]) - (x_{i-1}y[1 + \alpha(1-x_{i-1})(1-y)])$	$-1 \leq \alpha \leq 1\,(\alpha \neq 0)$
(b)	$C_m^\alpha(x,y) = (\frac{1}{1+\alpha}Max[x_i + y - 1 + \alpha - \alpha|x_i - y|, 0]) - (\frac{1}{1+\alpha}Max[x_{i-1} + y - 1 + \alpha - \alpha|[x_{i-1} - y|, 0])$	$0 < \alpha < 1$
(c)	$C_m^\alpha = ((1-\alpha)W_i + \alpha Min_i) - ((1-\alpha)W_{i-1} + \alpha Min_{i-1})$	$0 < \alpha < 1$
(d)	$C_m^\alpha = (\frac{\alpha^2(1-\alpha)}{2}W_i + (1-\alpha^2)P_i + \frac{\alpha^2(1+\alpha)}{2}Min_i) - (\frac{\alpha^2(1-\alpha)}{2}W_{i-1} + (1-\alpha^2)P_{i-1} + \frac{\alpha^2(1+\alpha)}{2}Min_{i-1})$	$-1 < \alpha < 1\,(\alpha \neq 0)$

1. *Knowledge Base*: it contains the Rule Base and the Data Base. The fuzzy rules used in this work have the following form:

   Rule R_j : If x_1 is A_{j1} and ... and x_n is A_{jn}

   then Class is Y_j with RW_j,

   where R_j is the label of the jth rule, A_{ji} is an antecedent fuzzy set modeling a linguistic term, Y_j is the class label and $RW_j \in [0,1]$ is the rule weight (Ishibuchi and Nakashima (2001)).

2. *Fuzzy Reasoning Method*: it is the mechanism used to classify examples using the information available in the knowledge base.

The Fuzzy Association Rule-based Classification model for High Dimensional problems (FARC-HD – Alcalá-Fdez et al. (2011)) is the fuzzy classifier considered in this paper. As mentioned before, we adopt this classifier since it is the same fuzzy classifier considered in the previous work (Lucca et al. (2016a)). The main steps of this classifier to build the fuzzy rules are:

- Fuzzy association rule extraction for classification: This step is aimed at obtaining the fuzzy rule base. To do so, a search tree (Agrawal and Srikant (1994)) is constructed for each class. The confidence and support degrees are calculated for each class considering each item (the item is a linguistic label). The fuzzy rules are generated by the most frequent itemsets and the number of linguistic terms for each rule is limited to the depth of the tree, which is a parameter of the model.
- Candidate rule prescreening: This step considers a weighting instances scheme (Kavšek et al. (2003)) to preselect the best generated rules.
- Genetic rule selection and lateral tuning: This stage uses an evolutionary algorithm to perform the lateral tuning of the fuzzy sets (Alcalá et al. (2007)) and select the best rules that were generated in the previous steps.

4.1 A method to build fuzzy measures based on fuzzy rules

In this subsection, we describe how the fuzzy measure is built, having in mind the construction method introduced in Section 2.3. Observe that Paternain et al. (2016) adopted the Choquet integral to aggregate the local information given by each fuzzy rule of the system. In this paper, however, we use the $C_\alpha C$-integral, given by Equation (5)).

The basis for this construction method is the generation of the fuzzy sets, which are subsequently processed by overlap functions so that a fuzzy measure can be obtained. In our case, we construct a different fuzzy measure for each class of the problem. Specifically, for each class, the fuzzy measure is constructed using the rule weights of the fuzzy rules that are fired when classifying a new example, that is, when $b_j^p(x_p) > 0$ (see Equation (9) for details). The fuzzy set of the k-th class, E^k, is obtained as:

$$E^k = \{(R_j, RW_j^k)|R_j \text{ such that } Class(R_j) = k \text{ and } \\ b_j^k(x_q) > 0\} \quad (6)$$

Initially the fuzzy set $\widetilde{A}_{E^k}$ is equal to E^k. Then, after aggregating an element with the $C_\alpha C$-integral, the fuzzy set $\widetilde{A}_{E^k}$ is changed so that the rule weight associated to the element that has been aggregated is substituted by 0. Consequently, the monotonicity is guaranteed with this method of constructing the fuzzy sets.

Observe that, in Equation (6), the fuzzy sets used to construct the associated fuzzy measure are the rule weights and, consequently, expressing the interaction among the rules of the different classes. The fuzzy measure for each class is computed using Equation (4), considering the corresponding sets E^k and subsets $\widetilde{A}_{E^k}$, as well as a given overlap index O.

In particular, Paternain et al. (2016) showed that the function GA_{Ov} (Equation (7)) achieved the best results in their study. Then, we also select this function for our proposal, which is defined, for all $A, B \in FS(U)$, as:

$$GA_{Ov}(A,B) = \max_{i=1}^{n}(A(i), B(i))^p. \quad (7)$$

Observe that this function is defined using a parameter p, which model the relation among the rules of different classes. This value may be adapted to each class and, so, it may have different values. Paternain et al. (2016) used the CHC (Eshelman (1991)) evolutionary method to accomplish this task. In this paper, we adopt the same methodology but, additionally, we use this evolutionary algorithm to tune the values of the parameters α, which will be described in details in Subsection 4.3.

4.2 A fuzzy reasoning method using the extended Choquet integral generalized by copulas

We present in this subsection the combination of the $C_\alpha C$-integral and the FRM of FRBCSs. We apply the concepts

presented in Section 2, for example, the product t-norm for the matching degree and the extended Choquet integral to aggregate the information, considering different copulas to define the $C_\alpha C$-integrals.

In the following consider that $x = (x_1, \ldots, x_n)$ is a new example to be classified, L is the number of rules in the rule base and M is the number of classes of the problem. The FRM used in this paper consists of four steps:

1. *Matching degree:* it is the strength of the activation of the if-part of the rules for the example to be classified x, which is computed using a t-norm $T : [0,1]^2 \to [0,1]$, for $j = 1, \ldots, L$:

$$\mu_{A_j}(x) = T(\mu_{A_{j1}}(x_1), \ldots, \mu_{A_{jn}}(x_n)). \tag{8}$$

2. *Association degree:* it is the association degree of the example x with the class of each rule in the rule base:

$$b_j^k(x) = \mu_{A_j}(x) \cdot RW_j^k, \tag{9}$$

with $k = Class(R_j)$ and $j = 1, \ldots, L$.

3. *Example classification soundness degree for all classes:* here the local information given by the fired fuzzy rules is aggregated by classes. In this step, we use the proposed family of $C_\alpha C$-integrals (see Table 3), and the fuzzy measure that is built using the rule weights of the fired fuzzy rules (the ones resulting in positive value for Equation (9), i.e., $b_j^k > 0$). Using these values, we construct the fuzzy sets as introduced in Equation (6), and the fuzzy measures for each class, applying Equation (4). Then, the aggregation of the information is computed as follows:

$$S_k(x) = \mathfrak{C}_{\mathfrak{m}_{GA_{Ov}}}^{C_\alpha} \left(b_1^k(x), \ldots, b_L^k(x) \right), \tag{10}$$

with $k = 1, \ldots, M$.

4. *Classification:* A decision function $F : [0,1]^M \to \{1, \ldots, M\}$ defined over the example classification soundness degrees of all classes and determining the class corresponding to the maximum soundness degree is applied:

$$F(S_1), \ldots, S_M) = arg \max_{k=1,\ldots,M} (S_k). \tag{11}$$

4.3 A Genetic Approach to learn the parameter α for each class

In this paper, we propose the usage of an evolutionary algorithm to learn the value of the parameter α. This section is aimed at explaining the main features of this genetic algorithm.

The main idea is to use the model that Paternain et al. (2016) applied to learn the exponent p and also to apply the same method to learn the parameter α for each class of the problem. The configuration of the evolutionary model is as follows:

1. Coding Scheme: The chromosome is composed of as many genes as classes and its structure is as follows:

$$C = [V_1, \ldots, V_{n \cdot t}, P_1, \ldots, P_M, \alpha_l, \ldots, \alpha_M].$$

It can be observed that the chromosome is divided in three parts, which are devoted to the three following optimization tasks:

   – Lateral tuning (Martínez et al. (2015)): in this part each gene, V_j with $j \in \{1, \ldots, nt\}$, represents the lateral displacement of an specific label. Therefore the total number of genes of this part is the number of variables (n) times the number of labels (t), where each gene is encoded in [-0.5, 0.5].
   – Learning of the exponent of each measure: in this part each gene, P_j with $j \in 1, \ldots, M$, is encoded in the range [0.01, 1.99] although they need to be decoded in the real range [0.01, 100] as described in the original proposal (Barrenechea et al. (2013)). The number of genes equals the number of classes.
   – Tuning of the α parameter: in this part there are as many genes, a_j with $j \in \{1, \ldots, M\}$, as classes and they are encoded in different ranges according to the copula function considered as shown in the column named Range in Table 3. If the value is 0 (these functions are not defined for this value), we assign 0.1 to the parameter α, since it is the best solution achieved in Lucca et al. (2015).

Consequently, the total number of genes is the number of variables times the number of labels plus the number of classes times two. We have to point out that a rule selection is also performed in the evolutionary algorithm as proposed in the FARC-HD algorithm (Alcalá-Fdez et al. (2011)).

2. Chromosome Evaluation: The fitness function considered in this paper is the standard accuracy rate.

3. Initial Gene Pool: Our population is composed by 50 individuals, where we have initialized one chromosome setting the value of all the genes to perform the lateral tuning to 0, those used to learn the exponent of the fuzzy measure to 1.0 and the ones for tuning the parameter α to 0.1.

4. Crossover Operator: We use the Parent Centric BLX (PCBLX) crossover operator (Herrera et al. (2003)). Two parents are crossed if their hamming distance divided by 2 is superior than a predetermined threshold L. We also use the Gray Code to convert each gene to binary code with a fixed number of bits for each gene (BITSGENE).[3]

$$L = (\#Genes \cdot BITSGENE)/4.0$$

where $\#$Genes is the total length of the chromosome.

---

[3] For more information see Barrenechea et al. (2013); Sanz et al. (2010).

5. Restarting Approach: Aiming to increase the convergence of the algorithm, when the threshold value is smaller than zero, we use the best chromosome (elitist scheme) and reset all the population considering the range appropriate for each gene.

5 Experimental Framework and Results

In this section, firstly we present the classification problems and the configuration for the considered approaches (Section 5.1). After that, the study of the obtained results is made in three mains parts:

1. Subsection 5.2 presents the obtained results in test by the four aggregation functions, including for each function the results obtained when the parameter α is fixed to 0.1 as well as when it is genetically learnt.
2. In the Subsection 5.3, we analyze the quality of the approach by comparing the best $C_\alpha C$-integral versus classical FRMs and our previous approaches.
3. The analysis of the different values provided by the evolutionary algorithm for the parameter α is presented in Subsection 5.4.

5.1 Experimental Framework

This study considers 30 datasets selected from the KEEL dataset repository (Alcalá-Fdez et al. (2009)). The properties of the datasets, containing for each dataset, the identifier (Id.), along with the name (Dataset), the number of instances ($\#Inst$), the number of attributes ($\#Att$) and the number of classes ($\#Class$) are summarized in Table 4. The *magic*, *page-blocks*, *penbased*, *ring*, *shuttle*, *satimage* and *twonorm* datasets have been stratified sampled at 10% in order to reduce their size for training. Examples with missing values have been removed, e.g., in the *wisconsin* dataset.

As made in Lucca et al. (2016b); Sanz et al. (2013), we adopt the 5-fold cross-validation model. A dataset is split in five random partitions, where each partition have 20% of the examples, and a combination of four of them is used for training and the remainder one is used for testing. This process is repeated five times by using a different partition to test the created system each time. In order to measure the quality of each partition, the accuracy rate is calculated, that is, we divide the number of correctly classified examples by the total number of examples for each partition. Then, as the final result of the algorithm we consider the average of the achieved accuracy in these five partitions.

In relation to the features of the classifier FARC-HD, we consider the standard configuration, that is:

– Conjunction operator: product t-norm.
– Rule weight: Certainty factor.

Table 4 Summary of the properties of the considered datasets

Id.	Dataset	#Inst	#Att	#Class
App	Appendiciticis	106	7	2
Bal	Balance	625	4	3
Ban	Banana	5300	2	2
Bnd	Bands	365	19	2
Bup	Bupa	345	6	2
Cle	Cleveland	297	13	5
Eco	Ecoli	336	7	8
Gla	Glass	214	9	6
Hab	Haberman	306	3	2
Ion	Ionosphere	351	33	2
Iri	Iris	150	4	3
Mag	Magic	1,902	10	2
New	Newthyroid	215	5	3
Pag	Pageblocks	5,472	10	5
Pen	Penbased	10,992	16	10
Pho	Phoneme	5,404	5	2
Pim	Pima	768	8	2
Rin	Ring	740	20	2
Sat	Satimage	6,435	36	7
Seg	Segment	2,310	19	7
Shu	Shuttle	58,000	9	7
Spe	Spectfheart	267	44	2
Tit	Titanic	2,201	3	2
Two	Twonorm	740	20	2
Veh	Vehicle	846	18	4
Vow	Vowel	990	13	11
Win	Wine	178	13	3
Wqr	Winequality-red	1599	11	11
Wis	Wisconsin	683	11	2
Yea	Yeast	1,484	8	10

– Five linguistic labels per class.
– Minimum support: 0.05.
– Confidence limit: 0.8.
– Depth of the tree: 3.

Regarding the parameters of the genetic algorithm we consider the following ones:

– Population size: 50 individuals
– Number of evaluations: 20.000
– Number of bit for each gene in the gray codification: 30 bits.

5.2 Analysis of the results provided by the genetic learning of the parameter α

This subsection is aimed at analyzing the behavior of the different functions considered in this study (see Table 3), using a fixed or a genetically learnt value for the parameter α in the FRM. We remark that, for the case of the fixed value for α, the system also applies a genetic algorithm to learn both the lateral position of the fuzzy sets and the values of the parameter p of the fuzzy measure. The main objective of this study is to analyze whether our proposal for tuning the parameter α worths being applied or not.

The results achieved in testing by the different $C_\alpha C$-integrals are presented in Table 5 by columns, in which our new proposal for tuning the parameter α is underscored by the tag Gen. In each line, we show the mean of accuracy obtained in the five partitions of the current dataset in testing, where, the best global result for each dataset is highlighted in **boldface**.

From the results shown in Table 5, it is possible to notice that the behavior of the different $C_\alpha C$-integrals related to the four functions with the fixed α value are quite similar among themselves, having a mean accuracy with no great differences, except for the function c that provides better results. However, when using the genetic approach, the results related to the functions a and d are clearly enhanced whereas the ones related to the functions b and c remains similar to those provided when using the fixed values. Specifically, the $C_\alpha C$-integral related to the function a, with α genetically learnt (a_{Gen}), achieves the greatest mean accuracy among all functions and approaches.

In fact, this function achieves the best accuracy for 7 datasets whereas the function a, with the α fixed in 0.1, achieves the best result in only 2 datasets (without considering ties). If we exclusively compare these methods among themselves, the function a_{Gen} achieves the best accuracy in 18 datasets, while the function a obtains the best result in 10 datasets.

Regarding the function b, the fixed approach achieves the best result in 2 datasets, while the genetic approach achieves the best result in 4 datasets. Comparing these methods among themselves, the function b_{Gen} has vantage in 16 datasets, the fixed α (function b), has the best result in 13 datasets.

The function c having the genetic value learnt to the parameter α (c_{Gen}), is the unique case in this study in which the achieved accuracy mean is inferior than the fixed one. However, both functions achieved the best global accuracy mean in 3 dataset. Comparing these functions among themselves, the genetic method achieves the best result in 15 dataset, having the fixed methodology the best result in 13 datasets. We believe that these functions have a similar behavior since they are based on the Lukasiewicz t-norm (see Table 1). Consequently, as we aggregate small values, the result of this function is in general zero, being this function basically represented by α multiplying the minimum (which can be the same number in many cases). For this reason, the influence of α in this function is not very relevant.

Regarding the function d, the genetic approach provides the largest accuracy mean. In fact, the function d_{Gen} has the second greatest mean in the study. More precisely, it achieves the best accuracy mean in 5 datasets. The fixed approach has 2 best global results. Comparing exclusively these two functions, the d_{Gen} achieves a good advantage, since in 20 cases, it obtains a superior accuracy and the function d has superiority in 8 datasets.

Observing that the genetic methodology to learn the most suitable value for the parameter α in different functions achieves the best accuracy in a superior number of datasets is not enough to confirm that this approach has an advantage versus the fixed one. In order to support our conclusions we conduct an statistical study in order to compare each function against the related genetic approach.

Specifically, we have carried out a set of pair-wise statistical comparisons using the well-known Wilcoxon signed-rank test (Wilcoxon (1945)), setting at 0.1 (10%) the level of confidence in all cases. In Table 6, we show the results of these comparisons, where R^+ indicates the ranks achieved by the functions having a fixed value for the parameter α and the R^- indicates the ranks obtained by the functions that consider the genetic approach.

From the results obtained in the statistical test, we can confirm that there are statistical differences between the two approaches when the functions a and d are considered. In the remaining cases, there are not statistical differences between the methodologies. All in all, the obtained ranks are always superior in favor to the genetic method and the obtained accuracy are also superior when the genetic approach is considered. Consequently, we conclude that the genetic approach to learn the parameter α is a good alternative with respect to the fixed one.

5.3 Analyzing the quality of the $C_\alpha C$-integrals

As presented before, to adopt $C_\alpha C$-integrals in the FRM of the FARC-HD could be an interesting alternative, since it achieves satisfactory results. In this subsection, we present a study to compare the best $C_\alpha C$-integral against several related methods, namely the FRM of the Winning Rule (WR) (which uses the maximum as aggregation), the standard Choquet integral (Cho), the best pre-aggregation function obtained in Lucca et al. (2016b), which is based on the Hamacher t-norm, (Ham_{PA}), and the best CC-integral achieved in Lucca et al. (2016c), which is based on the minimum t-norm (CC_{Min}).

In order to test the quality of our approach, we consider the function a_{Gen} as the best representative function of the $C_\alpha C$-integrals. We have taken into consideration that this function achieves the best accuracy mean and it also provides the best result in the largest number of datasets.

We show the results of these approaches by columns in Table 7, highlighting in boldface the best result for each dataset. Moreover, we also present a comparison of the $C_\alpha C$-integral a_{Gen} against each method, where #Wins is the number of datasets in which a_{Gen} achieves a superior accuracy to the compared method, and #Loses represents the opposite case.

From the obtained results it is noticeable that our approach is the one that reaches the best global accuracy (80.14).

Table 5 Accuracy rate achieved in testing for the different $C_\alpha C$-integrals

Dataset	a	a_{Gen}	b	b_{Gen}	c	c_{Gen}	d	d_{Gen}
App	82.16 ± 8.80	84.89 ± 5.25	82.12 ± 5.97	**86.75 ± 4.09**	84.98 ± 5.84	83.03 ± 7.17	83.07 ± 5.11	80.22 ± 9.06
Bal	81.60 ± 3.58	**83.52 ± 1.66**	81.28 ± 3.23	81.76 ± 4.29	82.24 ± 2.15	81.44 ± 2.36	81.76 ± 3.17	82.88 ± 3.47
Ban	85.11 ± 1.59	**85.85 ± 1.15**	85.19 ± 1.01	84.55 ± 1.10	84.83 ± 0.46	85.57 ± 1.75	85.11 ± 1.46	85.34 ± 1.34
Bnd	69.38 ± 8.21	**69.93 ± 6.15**	69.61 ± 5.63	68.03 ± 5.13	67.15 ± 6.47	67.16 ± 5.03	67.76 ± 5.57	67.77 ± 6.73
Bup	61.16 ± 6.43	**64.35 ± 4.30**	63.19 ± 6.03	62.90 ± 3.92	63.48 ± 6.01	62.61 ± 4.96	62.03 ± 8.16	62.32 ± 5.02
Cle	56.56 ± 0.70	56.24 ± 4.47	56.24 ± 3.32	56.21 ± 4.52	56.23 ± 2.85	52.85 ± 3.07	**56.57 ± 2.65**	55.87 ± 5.20
Eco	77.38 ± 6.53	**80.67 ± 4.87**	75.31 ± 6.15	77.38 ± 7.10	77.98 ± 6.52	78.29 ± 6.28	78.59 ± 6.95	77.98 ± 2.39
Gla	62.61 ± 1.69	66.36 ± 5.07	64.47 ± 5.64	**68.19 ± 6.06**	65.40 ± 4.32	67.26 ± 3.97	62.15 ± 3.44	64.97 ± 4.17
Hab	72.86 ± 3.32	72.21 ± 2.72	**74.51 ± 2.53**	71.88 ± 2.83	72.54 ± 2.75	72.22 ± 2.60	72.21 ± 2.17	74.50 ± 3.47
Ion	88.90 ± 5.35	90.03 ± 3.33	88.33 ± 4.60	89.47 ± 4.07	**90.31 ± 4.45**	88.05 ± 5.35	88.61 ± 2.45	89.75 ± 4.88
Iri	93.33 ± 3.33	**94.00 ± 4.35**	93.33 ± 4.71	93.33 ± 5.27	94.00 ± 4.35	**94.00 ± 4.35**	**94.00 ± 4.35**	93.33 ± 4.08
Mag	79.86 ± 1.74	79.39 ± 2.38	80.23 ± 2.64	80.02 ± 2.45	78.91 ± 3.15	**80.39 ± 1.98**	79.44 ± 2.39	79.65 ± 2.42
New	93.49 ± 4.47	93.49 ± 3.45	94.42 ± 2.65	92.56 ± 5.04	94.42 ± 4.82	94.42 ± 5.35	94.88 ± 3.03	**95.35 ± 2.85**
Pag	94.16 ± 1.07	94.34 ± 1.20	93.61 ± 0.67	**94.52 ± 1.13**	94.16 ± 1.23	93.97 ± 1.06	93.79 ± 2.20	93.79 ± 1.52
Pen	90.91 ± 1.29	91.09 ± 2.05	89.27 ± 1.59	**92.09 ± 1.35**	91.09 ± 1.59	89.64 ± 3.54	91.18 ± 3.43	90.91 ± 2.27
Pho	80.90 ± 1.05	81.98 ± 1.77	81.05 ± 1.04	82.09 ± 1.06	81.49 ± 1.68	81.85 ± 1.57	82.85 ± 1.24	**83.14 ± 0.95**
Pim	**75.52 ± 2.02**	73.17 ± 2.78	73.56 ± 3.22	73.82 ± 2.30	73.82 ± 3.56	74.73 ± 3.96	75.00 ± 2.65	74.60 ± 1.68
Rin	89.05 ± 2.84	89.05 ± 2.26	89.19 ± 2.95	88.78 ± 3.66	**90.27 ± 4.42**	89.32 ± 4.26	88.65 ± 4.20	89.32 ± 2.41
Sat	**79.78 ± 1.21**	79.01 ± 1.93	79.16 ± 1.71	78.85 ± 1.43	79.78 ± 1.85	79.31 ± 2.56	78.54 ± 0.49	79.32 ± 2.21
Seg	92.55 ± 0.60	92.34 ± 1.38	91.86 ± 0.64	92.68 ± 1.49	91.65 ± 1.59	92.90 ± 1.36	**93.07 ± 1.34**	91.95 ± 1.54
Shu	97.20 ± 0.44	97.01 ± 0.76	96.87 ± 0.84	97.01 ± 0.67	96.97 ± 0.79	97.20 ± 0.73	**97.29 ± 0.70**	**97.29 ± 1.06**
Spe	76.37 ± 5.56	76.00 ± 5.22	76.38 ± 6.96	77.14 ± 5.42	**77.89 ± 2.58**	76.36 ± 6.76	77.49 ± 6.79	75.62 ± 6.26
Tit	**78.87 ± 1.48**	**78.87 ± 1.48**	**78.87 ± 1.48**	**78.87 ± 1.48**	**78.87 ± 1.48**	**78.87 ± 1.48**	**78.87 ± 1.48**	**78.87 ± 1.48**
Two	83.51 ± 2.97	84.19 ± 2.46	83.24 ± 3.32	84.05 ± 2.51	84.32 ± 1.54	85.54 ± 3.08	84.19 ± 2.81	**85.54 ± 2.73**
Veh	67.50 ± 3.73	67.50 ± 4.33	67.02 ± 1.11	67.37 ± 4.15	68.32 ± 2.17	68.91 ± 3.61	66.55 ± 0.95	**69.98 ± 3.05**
Vow	69.09 ± 3.20	**69.29 ± 1.76**	66.57 ± 1.80	66.87 ± 3.92	66.97 ± 1.62	68.18 ± 2.79	68.28 ± 3.70	68.59 ± 2.12
Win	94.40 ± 4.38	97.19 ± 1.96	96.03 ± 4.77	92.70 ± 4.21	96.57 ± 4.69	96.60 ± 3.70	93.25 ± 5.06	**97.76 ± 3.62**
Wqr	58.60 ± 2.16	58.60 ± 3.42	**59.28 ± 2.84**	58.66 ± 2.34	58.91 ± 2.45	58.35 ± 2.48	58.66 ± 1.79	59.22 ± 2.69
Wis	96.34 ± 1.86	96.49 ± 1.58	96.49 ± 1.08	96.49 ± 0.94	95.90 ± 0.97	**96.93 ± 0.94**	96.49 ± 1.31	96.49 ± 1.30
Yea	55.73 ± 1.96	57.08 ± 2.30	55.79 ± 0.47	55.32 ± 2.73	56.54 ± 1.32	**58.49 ± 1.85**	55.52 ± 1.84	58.09 ± 1.52
Mean	79.50 ± 3.12	**80.14 ± 2.93**	79.42 ± 3.02	79.68 ± 3.22	79.87 ± 2.99	79.82 ± 3.33	79.53 ± 3.10	80.01 ± 3.12

Table 6 Wilcoxon Test to compare the two methods to obtain the parameter α in the copula C_α.

Comparison	R^+	R^-	p-value
a vs. a_{Gen}	127.5	337.5	0.02
b vs. b_{Gen}	188	277	0.33
c vs. c_{Gen}	224.5	240.5	0.93
d vs. d_{Gen}	147.5	317.5	0.09

Regarding the number of datasets where our new proposal provides the best or the worst result, we have to stress its competitive behavior since it provides a larger number of wins than WR and CC_{Min}, whereas the behavior is similar when compared with Choquet and Ham_{PA}.

In order to support our previous results, we have performed once again a set of pair-wise comparisons using the Wilcoxon signed-rank test. The obtained results are available at Table 8, where R^+ indicates the ranks obtained by our approach, a_{Gen}, and R^- represents the ranks achieved by the remainder methods.

The obtained results, as expected, demonstrate that our approach performs similarly to the remainder approaches as the obtained p-values are high, with the exception of the WR method, where we can observe a relatively low p-value show-

ing a positive trend in favor to our proposal. Furthermore, it is possible to observe that the obtained ranks are always superior for our approach, reinforcing the idea that our method is a new possibility of generalization of the Choquet integral, presenting competitive results.

5.4 An analysis of the obtained parameter α

This subsection is aimed at analyzing the resulting values learned for the parameter α in training, considering the 4 distinct functions that are used in the FRM of this study. The results are available in Table 9, where each column represents a different function that uses the genetically learnt parameter α. For each dataset, we show the averaged α value for all the classes and all the partitions along with the standard deviation.

Observe that the obtained mean for the functions b_{Gen} and c_{Gen} are quite similar, whereas this fact does not occur for the functions a_{Gen} and d_{Gen} (these functions are the ones that have the bests results according to the analysis provided in the previous subsection). While the functions b_{Gen} and c_{Gen} have α around 0.5, the functions a_{Gen} and d_{Gen}, have α containing lowers values (close to 0.1 and -0.1, respectively), being this an interesting topic for future studies.

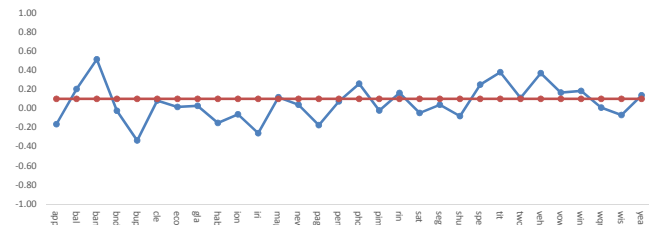**Table 7** Accuracy rate achieved in testing for different FRMs

Dataset	a_{Gen}	WR	Cho	CC_{Min}	Ham_{PA}
App	**84.89 ± 5.25**	83.03 ± 8.61	82.12 ± 5.97	83.98 ± 5.33	**84.89 ± 4.03**
Bal	83.52 ± 1.66	81.92 ± 2.57	**84.16 ± 1.99**	82.72 ± 2.69	81.92 ± 1.84
Ban	85.85 ± 1.15	83.94 ± 1.05	85.19 ± 2.20	**86.09 ± 1.15**	84.77 ± 2.03
Bnd	69.93 ± 6.15	69.40 ± 7.75	70.26 ± 7.22	68.87 ± 6.35	**70.79 ± 6.99**
Bup	**64.35 ± 4.30**	62.03 ± 5.46	63.77 ± 5.89	62.03 ± 5.74	62.32 ± 5.71
Cle	56.24 ± 4.47	**56.91 ± 2.35**	54.55 ± 2.57	54.55 ± 4.78	56.90 ± 2.60
Eco	**80.67 ± 4.87**	75.62 ± 7.37	77.70 ± 5.77	77.69 ± 4.54	78.28 ± 5.71
Gla	66.36 ± 5.07	64.99 ± 5.92	**67.75 ± 4.20**	65.43 ± 4.41	62.64 ± 4.76
Hab	72.21 ± 2.72	70.89 ± 4.70	**72.54 ± 1.96**	72.53 ± 3.68	**72.54 ± 3.20**
Ion	**90.03 ± 3.33**	90.03 ± 2.84	87.48 ± 4.15	89.18 ± 3.85	89.75 ± 3.91
Iri	94.00 ± 4.35	94.00 ± 4.35	94.00 ± 4.35	94.00 ± 4.35	94.00 ± 4.35
Mag	79.39 ± 2.38	78.60 ± 2.71	**80.55 ± 2.02**	79.60 ± 1.81	80.33 ± 2.85
New	93.49 ± 3.45	94.88 ± 4.47	93.49 ± 4.47	**95.35 ± 4.03**	94.42 ± 2.65
Pag	94.34 ± 1.20	94.16 ± 1.06	**94.52 ± 1.84**	94.34 ± 1.65	**94.52 ± 1.32**
Pen	91.09 ± 2.05	91.45 ± 2.21	91.09 ± 2.50	**91.91 ± 2.24**	91.36 ± 2.59
Pho	81.98 ± 1.77	82.29 ± 1.46	**82.48 ± 1.45**	82.40 ± 1.81	81.53 ± 1.66
Pim	73.17 ± 2.78	74.60 ± 3.55	75.26 ± 2.20	**75.39 ± 2.17**	74.60 ± 3.97
Rin	89.05 ± 2.26	**90.00 ± 2.68**	89.46 ± 2.46	**90.00 ± 2.80**	88.65 ± 2.50
Sat	79.01 ± 1.93	**79.63 ± 2.42**	78.85 ± 1.88	78.85 ± 0.94	79.62 ± 2.09
Seg	92.34 ± 1.38	**93.03 ± 0.87**	92.64 ± 1.15	92.12 ± 1.07	92.99 ± 0.86
Shu	97.01 ± 0.76	96.00 ± 1.92	97.15 ± 0.70	96.69 ± 1.32	**97.29 ± 0.95**
Spe	76.00 ± 5.22	**77.90 ± 3.66**	77.49 ± 6.79	77.13 ± 5.32	77.15 ± 6.02
Tit	**78.87 ± 1.48**	**78.87 ± 1.48**	**78.87 ± 1.48**	**78.87 ± 1.48**	**78.87 ± 1.48**
Two	84.19 ± 2.46	**86.49 ± 4.05**	85.14 ± 3.02	84.19 ± 3.04	83.78 ± 3.06
Veh	67.50 ± 4.33	66.67 ± 2.18	68.21 ± 1.89	67.14 ± 1.85	**68.44 ± 3.31**
Vow	**69.29 ± 1.76**	67.98 ± 3.72	67.78 ± 2.21	68.89 ± 3.81	67.47 ± 3.60
Win	97.19 ± 1.96	96.60 ± 3.14	94.37 ± 2.05	**98.30 ± 1.55**	96.60 ± 3.70
Wqr	58.60 ± 3.42	**59.22 ± 3.51**	58.10 ± 1.84	57.66 ± 1.14	58.54 ± 2.08
Wis	96.49 ± 1.58	96.34 ± 1.02	95.61 ± 1.85	**97.07 ± 1.16**	96.05 ± 1.52
Yea	57.08 ± 2.30	55.32 ± 2.07	56.47 ± 2.28	**57.41 ± 1.75**	57.07 ± 1.46
Mean	**80.14 ± 2.93**	79.76 ± 3.37	79.90 ± 3.01	80.01 ± 2.93	79.94 ± 3.09

Table 8 Wilcoxon Test to compare the best $C_\alpha C$-integral versus classical FRMs

Comparison	R^+	R^-	p-value
a_{Gen} vs. WR	290	175	0.22
a_{Gen} vs. Cho	251.5	213.5	0.55
a_{Gen} vs. CC_{Min}	253.5	211.5	0.66
a_{Gen} vs. Ham_{PA}	248.5	216.5	0.72

Fig. 1 Variations of the obtained α in test, considering the function a and a_{Gen}

Specifically, we can observe that the α's values for the function a_{Gen} are close to 0.1, which is the best result according to Lucca et al. (2015). This may seem that this is a good starting value and, by performing small variations (as it can be seen in Table 9), it may result in an enhancement of the performance, as shown in Table 5. The same conclusion can be made for the function d_{Gen} but, in this case, with starting point -0.1 instead of using 0.1, since most of the final values are negative. Finally, the values of the functions b_{Gen} and c_{Gen} end up close to half of the allowed range, since it is (0,1) and the mean value is around 0.5

Figure 1 shows a line graph showing the variation of the parameter α in the function a_{Gen} (since it is the best performing one) against the fixed approach. The line having

the great variation correspond to the method a_{Gen}, while, the constant line is the function a.

6 Conclusion

In this paper, we have introduced a genetic approach to learn the parameter α used in the generalization of the Choquet integral by the family of copulas defined by a parameter α ($C_\alpha C$-integral). Moreover, we have constructed the fuzzy measure by using a novel method in which they are constructed based on the information of the system at execution time. Specifically, we have applied the $C_\alpha C$-integral considering 4 different copulas in the FRM of FRBCSs.

Table 9 Values of the parameter α considered in this study for different functions

Dataset	a_{Gen}	b_{Gen}	c_{Gen}	d_{Gen}
App	-0.16 ±0.05	0.56 ±0.01	0.39 ±0.02	-0.06 ±0.05
Bal	0.20 ±0.03	0.58 ±0.01	0.63 ±0.01	-0.36 ±0.02
Ban	0.51 ±0.00	0.42 ±0.00	0.64 ±0.01	0.38 ±0.01
Bnd	-0.03 ±0.03	0.39 ±0.01	0.57 ±0.01	-0.05 ±0.01
Bup	-0.34 ±0.01	0.29 ±0.01	0.37 ±0.01	-0.27 ±0.01
Cle	0.08 ±0.06	0.36 ±0.04	0.55 ±0.04	-0.10 ±0.05
Eco	0.02 ±0.07	0.53 ±0.04	0.54 ±0.04	-0.06 ±0.09
Gla	0.03 ±0.09	0.49 ±0.04	0.51 ±0.05	0.06 ±0.09
Hab	-0.15 ±0.03	0.45 ±0.01	0.48 ±0.01	0.11 ±0.03
Ion	-0.06 ±0.01	0.51 ±0.01	0.42 ±0.01	0.12 ±0.02
Iri	-0.26 ±0.04	0.41 ±0.03	0.51 ±0.03	-0.08 ±0.07
Mag	0.12 ±0.01	0.72 ±0.00	0.48 ±0.01	-0.28 ±0.01
New	0.04 ±0.04	0.47 ±0.02	0.47 ±0.03	-0.39 ±0.06
Pag	-0.17 ±0.04	0.48 ±0.02	0.53 ±0.02	-0.07 ±0.04
Pen	0.07 ±0.05	0.52 ±0.02	0.51 ±0.02	-0.19 ±0.05
Pho	0.26 ±0.01	0.78 ±0.00	0.68 ±0.00	-0.11 ±0.01
Pim	-0.02 ±0.01	0.49 ±0.01	0.59 ±0.01	-0.12 ±0.01
Rin	0.16 ±0.02	0.65 ±0.01	0.49 ±0.01	-0.28 ±0.01
Sat	-0.05 ±0.05	0.55 ±0.02	0.54 ±0.03	-0.12 ±0.05
Seg	0.04 ±0.03	0.59 ±0.01	0.51 ±0.01	-0.17 ±0.03
Shu	-0.08 ±0.02	0.52 ±0.01	0.49 ±0.01	-0.02 ±0.02
Spe	0.25 ±0.03	0.49 ±0.01	0.50 ±0.02	-0.01 ±0.02
Tit	0.38 ±0.00	0.72 ±0.00	0.68 ±0.00	0.20 ±0.00
Two	0.11 ±0.01	0.61 ±0.00	0.52 ±0.01	-0.81 ±0.01
Veh	0.37 ±0.04	0.68 ±0.01	0.68 ±0.01	-0.24 ±0.03
Vow	0.17 ±0.06	0.62 ±0.03	0.52 ±0.03	-0.07 ±0.04
Win	0.18 ±0.03	0.48 ±0.02	0.47 ±0.03	-0.01 ±0.04
Wqr	0.01 ±0.04	0.57 ±0.02	0.51 ±0.02	-0.02 ±0.04
Wis	-0.07 ±0.02	0.37 ±0.01	0.52 ±0.01	-0.16 ±0.01
Yea	0.14 ±0.04	0.54 ±0.02	0.53 ±0.02	-0.11 ±0.04
Mean	0.06 ±0.03	0.53 ±0.02	0.53 ±0.02	-0.11 ±0.03

We highlight that the adopted genetic approach allows one to conclude that this generalization is statistically superior in half of the functions used, presenting a competitive behavior in all the cases, when compared against the methodology that uses a fixed value for the parameter α. Furthermore, we compare the quality of this approach against classical FRMs and it can be concluded that the methodology proposed in this work is an alternative to the function presented by Lucca et al. (2015), since an improvement in the quality of the new FRM has been found.

Future work in concerned to the study of our generalizations in an interval-valued context, following the approach in Bedregal et al. (2009, 2010); Dimuro (2011).

Acknowledgment

References

Agrawal R, Srikant R (1994) Fast algorithms for mining association rules in large databases. In: Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, VLDB '94, pp 487–499

Alcalá R, Alcalá-Fdez J, Herrera F (2007) A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. IEEE Transactions on Fuzzy Systems 15(4): 616–635,

Alcalá-Fdez J, Sánchez L, García S, Jesus M, Ventura S, Garrell J, Otero J, Romero C, Bacardit J, Rivas V, Fernández J, Herrera F (2009) Keel: a software tool to assess evolutionary algorithms for data mining problems. Soft Computing 13(3): 307–318

Alcalá-Fdez J, Alcalá R, Herrera F (2011) A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. IEEE Transactions on Fuzzy Systems 19(5): 857–872

Alsina C, Frank MJ, Schweizer B (2006) Associative Functions: Triangular Norms and Copulas. World Scientific Publishing Company, Singapore

Barrenechea E, Bustince H, Fernandez J, Paternain D, Sanz JA (2013) Using the Choquet integral in the fuzzy reasoning method of fuzzy rule-based classification systems. Axioms 2(2): 208–223

Bedregal BC, Dimuro GP, Bustince H, Barrenechea E (2013) New results on overlap and grouping functions. Information Sciences 249: 148–170

Bedregal BC, Dimuro GP, Reiser RHS (2009) An approach to interval-valued R-implications and automorphisms. In: Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference, IFSA/EUSFLAT, 2009.

Bedregal BC, Dimuro GP, Santiago RHN, Reiser RHS (2010) On interval fuzzy S-implications. Information Sciences 180 (8): 1373 – 1389.

Beliakov G, Pradera A, Calvo T (2007) Aggregation Functions: A Guide for Practitioners. Springer, Berlin

Bustince H, Fernández J, Mesiar R, Montero J, Orduna R (2009) Overlap index, overlap functions and migrativity. In: Proceedings of IFSA/EUSFLAT Conference, pp 300–305

Bustince H, Fernandez J, Mesiar R, Montero J, Orduna R (2010) Overlap functions. Nonlinear Analysis: Theory, Methods & Applications 72(3-4): 1488–1499

Bustince H, Pagola M, Mesiar R, Hüllermeier E, Herrera F (2012) Grouping, overlaps, and generalized bientropic functions for fuzzy modeling of pairwise comparisons. IEEE Transactions on Fuzzy Systems 20(3): 405–415

Bustince H, Fernandez J, Kolesárová A, Mesiar R (2015) Directional monotonicity of fusion functions. European Journal of Operational Research 244(1): 300–308

Choquet G (1953–1954) Theory of capacities. Annales de línstitut Fourier 5: 131–295

Cordón O, del Jesus MJ, Herrera F (1999) A proposal on reasoning methods in fuzzy rule-based classification systems. International Journal of Approximate Reasoning 20(1): 21–45

Dimuro GP (2011) On interval fuzzy numbers. In: 2011 Workshop-School on Theoretical Computer Science, WEIT 2011, IEEE, Los Alamitos, 2011.

Dimuro GP, Bedregal B (2014) Archimedean overlap functions: The ordinal sum and the cancellation, idempotency and limiting properties. Fuzzy Sets and Systems 252: 39–54

Dimuro GP, Bedregal B (2015) On residual implications derived from overlap functions. Information Sciences 312: 78–88

Dimuro GP, Bedregal B, Bustince H, Asiáin MJ, Mesiar R (2016a) On additive generators of overlap functions. Fuzzy Sets and Systems 287: 76–96

Dimuro GP, Bedregal B, Bustince H, Fernandez J, Lucca G, Mesiar R (2016b) New results on pre-aggregation functions. In: Uncertainty Modelling in Knowledge Engineering and Decision Making, Proceedings of the 12th International FLINS Conference (FLINS 2016), World Scientific Proceedings Series on Computer Engineering and Information Science 10, Singapura, pp 213–219

Dimuro GP, Bedregal B, Bustince H, Jurio A, Baczyński M, Miś K (2017) QL-operations and QL-implication functions constructed from tuples (O, G, N) and the generation of fuzzy subsethood and entropy measures. International Journal of Approximate Reasoning 82: 170–192.

Duda R, Hart P, Stork D (2001) Pattern classification. Wiley

Elkano M, Galar M, Sanz J, Fernández A, Barrenechea E, Herrera F, Bustince H (2015) Enhancing multi-class classification in FARC-HD fuzzy classifier: On the synergy between n-dimensional overlap functions and decomposition strategies. IEEE Transactions on Fuzzy Systems 23(5): 1562–1580

Elkano M, Galar M, Sanz J, Bustince H (2016) Fuzzy rule-based classification systems for multi-class problems using binary decomposition strategies: On the influence of n-dimensional overlap functions in the fuzzy reasoning method. Information Sciences 332: 94–114

Eshelman LJ (1991) The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. In: Rawlings GJE (ed) Foundations of Genetic Algorithms, Morgan Kaufmann, San Francisco, pp 265–283

Garcia-Jimenez S., Bustince H., Hüllermeier E., Mesiar R., Pal N. R., A. Pradera(2015) Overlap indices: Construction of and application to interpolative fuzzy systems. IEEE Transactions on Fuzzy Systems 23(4): 1259–1273

Grabisch M, Labreuche C (2010) A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. Annals of Operations Research 175(1): 247–286

Herrera F, Lozano M, Sánchez AM (2003) A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study. International Journal of Intelligent Systems 18(3): 309–338

Ishibuchi H, Nakashima T (2001) Effect of rule weights in fuzzy rule-based classification systems. Fuzzy Systems, IEEE Transactions on 9(4):506–515

Ishibuchi H, Nakashima T, Nii M (2005) Classification and Modeling with Linguistic Information Granules, Advanced Approaches to Linguistic Data Mining. Advanced Information Processing, Springer, Berlin

Jurio A, Bustince H, Pagola M, Pradera A, Yager R (2013) Some properties of overlap and grouping functions and their application to image thresholding. Fuzzy Sets and Systems 229: 69–90

Kavšek B, Lavrač N, Jovanoski V (2003) Advances in Intelligent Data Analysis V: 5th International Symposium on Intelligent Data Analysis, IDA 2003, Berlin, Germany. Proceedings, Springer Berlin. pp 230–241

Lucca G, Dimuro GP, Mattos V, Bedregal B, Bustince H, Sanz JA (2015) A family of Choquet-based non-associative aggregation functions for application in fuzzy rule-based classification systems. In: 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, Los Alamitos, pp 1–8,

Lucca G, Dimuro GP, Bedregal B, Sanz J, Bustince H (2016a) A proposal for tuning the α parameter in a copula function applied in fuzzy rule-based classification systems. Brazilian Conference on Intelligent Systems

Lucca G, Sanz J, Pereira Dimuro G, Bedregal B, Mesiar R, Kolesárová A, Bustince Sola H (2016b) Pre-aggregation functions: construction and an application. IEEE Transactions on Fuzzy Systems 24(2): 260–272,

Lucca G, Sanz JA, Dimuro GP, Bedregal B, Asiain MJ, Elkano M, Bustince H (2017) CC-integrals: Choquet-like copula-based aggregation functions and its application in fuzzy rule-based classification systems. Knowledge-Based Systems 119: 32–43

Martínez, Luis and Rodríguez, Rosa María and Herrera, Francisco (2015) The 2-tuple Linguistic Model. Computing with Words in Decision Making. Springer International Publishing

Mayor G, Trillas E (1986) On the representation of some aggregation functions. In: Proceedings of IEEE International Symposium on Multiple-Valued Logic, IEEE, Los Alamitos, pp 111–114

Mesiar R, Kolesárová A, Bustince H, Dimuro GP, Bedregal BC (2016) Fusion functions based discrete choquet-like integrals. European Journal of Operational Research 252(2): 601–609

Murofushi T, Sugeno M, Machida M (1994) Non-monotonic fuzzy measures and the Choquet integral. Fuzzy Sets and Systems 64(1): 73–86

Paternain D, Bustince H, Pagola M, Sussner P, Kolesárová A, Mesiar R (2016) Capacities and overlap indexes with an application in fuzzy rule-based classification systems. Fuzzy Sets and Systems 305: 70–94

Samantaray SR, El-Arroudi K, Joos G, Kamwa I (2010) A fuzzy rule-based approach for islanding detection in distributed generation. Power Delivery, IEEE Transactions on 25(3): 1427–1433

Sanz J, Fernández A, Bustince H, Herrera F (2013) IVTURS: A linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection. IEEE Transactions on Fuzzy Systems 21(3): 399–411

Sanz J, Bernardo D, Herrera F, Bustince H, Hagras H (2015) A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data. Fuzzy Systems, IEEE Transactions on 23(4): 973–990,

Sanz JA, Fernández A, Bustince H, Herrera F (2010) Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning. Information Sciences 180(19): 3674 – 3685

Sanz JA, Galar M, Jurio A, Brugos A, Pagola M, Bustince H (2014) Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system. Applied Soft Computing 20: 103 – 111

Sola HB, Tartas EB, Pagola M, Soria F (2006) Weak fuzzy s-subsethood measures: Overlap index. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 14(5): 537–560,

Wilcoxon F (1945) Individual comparisons by ranking methods. Biometrics 1: 80–83

Zadeh LA (1965) Fuzzy sets. Information and Control 8(3): 338–353