On the generalizations of the Choquet integral for application in FRBCs

$\star$

Giancarlo Lucca[1][000−0002−3776−0260], Eduardo N. Borges[1][0000−0003−1595−7676], Rafael A. Berri[1][0000−0002−5125−2756], Leonardo Emmendorfer[1][0000−0002−6950−3947], Graçaliz P. Dimuro[1,2][0000−0001−6986−9888] and Tiago C. Asmus[1,2][0000−0002−7066−7156].

[1] Universidade Federal do Rio Grande, Rio Grande, RS, Brazil
{giancarlo.lucca, eduardoborges, rafaelberri,
leonardoemmendorfer, gracalizdimuro, tiagoasmus}@furg.br
[2] Universidad Publica de Navarra, Pamplona, Navarra, Spain

Abstract. An effective way to cope with classification problems, among others, is by using Fuzzy Rule-Based Classification Systems (FRBCSs). These systems are composed by two main components, the Knowledge Base (KB) and the Fuzzy Reasoning Method (FRM). The FRM is responsible for performing the classification of new examples based on the information stored in the KB. A key point in the FRM is how the information given by the fired fuzzy rules is aggregated. Precisely, the aggregation function is the component that differs from the two most widely used FRMs in the specialized literature. In this paper we provide a revision of the literature discussing the generalizations of the Choquet integral that has been applied in the FRM of a FRBCS. To do so, we consider an analysis of different generalizations, by t-norms, copulas, and by F functions. Also, the main contributions of each generalization are discussed.

Keywords: Choquet integral, generalizations Choquet integral, pre-aggregation function

1 Introduction

A classification problem [1] is a research field in the area of data mining [2], which can be tackled in two different ways. An approach to deal with this problem is known as supervised learning, where a function (classifier) is generated from the available and labeled data (classes). Then, when a new example needs to be classified, the learned classifier is responsible to perform the prediction.

In the literature it is possible to find several methods that aim to cope with these problems using supervised learning, such as Support Vector Machines (SVM) [3], decisions trees [4] and neural networks [5]. Here, the focus is on Fuzzy Rule-Based Classification Systems (FRBCCs) [6], because they provide the user with interpretable models by using linguistic labels [7] in their rules. Another reason is related with their accurate results and versatility, as shown in the many different fields where they have been applied like health [8], security [9], economy [10], food industry [11].

An important role in any FRBCS is played by the Fuzzy Reasoning Method (FRM) [12]. This method is responsible to perform the classification of new examples. For that, it makes usage of the information available in the rule base and the database. Moreover, in order to perform the classification, this mechanism uses an aggregation operator in order to aggregate, by classes, the information provided by the fired fuzzy rules when classifying new examples.

A widely used FRM considers the function Maximum as aggregation operator. By using this aggregation function, for each class, the FRM performs the selection of the best fired rule since it has the highest compatibility with the example [13]. The issue of this inference method is that the information provided by the remainder fired fuzzy rules is ignored. The Maximum is an averaging aggregation operator, since the obtained result is within the range between the minimum and the maximum of the aggregated values (in this case, obviously, the result is always the maximum).

To avoid the problem of ignoring information, it was proposed a FRM that applies the normalized sum [12] to perform the aggregation of the available information given by the fired rules. In this way, for each class, all information is taken into account in the aggregation step. This aggregation operator is considered as non-averaging since the result of this function can leave the range minimum–maximum.

In [14] the authors introduced a FRM considering the usage of the Choquet integral (CI) [15], which is an averaging operator. In this way, this approach mixes the characteristics of the previous FRMs considering an averaging operator that uses the information provided by all the fired rules of the system. Moreover, the CI is defined in terms of a fuzzy measure, which provides it with the nice properties to take into account the interaction among the data to be aggregated [15].

The objective of this paper is to discuss different methodologies that change the aggregation step performed in the FRM, when considering different generalizations of the CI, which are supported by solid theoretical studies [16], varying from the generalization by t-norms (C_T-integrals) [17], by copulas (CC-integrals [18–20]) and functions F(C_F-integrals [21] and C_{F1F2}-integrals [22, 23]). Moreover, for each generalization it is provided provide a discussion of the main obtained results of each study (we highlight that our focus here are related with the main conclusions and not the specific obtained results of each approach).

This paper is organized as follows. Section 2 present the main components of a FRBC, showing an example of how the aggregation function is used in this context. Sections 3 to 6 discuss the theoretical and applied contributions of different generalizations of the CI. Section 8 is the conclusion.

2   The role of aggregation functions in the FRM

Fuzzy Rule-Based Classification Systems (FRBCSs) [6] are extensions of the rule-based system by using fuzzy sets in the antecedents of the rules. The best-known FRBCSs are the ones defined by Takagi-Sugeno-Kang (TSK) [24] and *Mamdani* [25], which is the one that it is adopted. The standard architecture of the Mamdani method is presented in Figure 1.
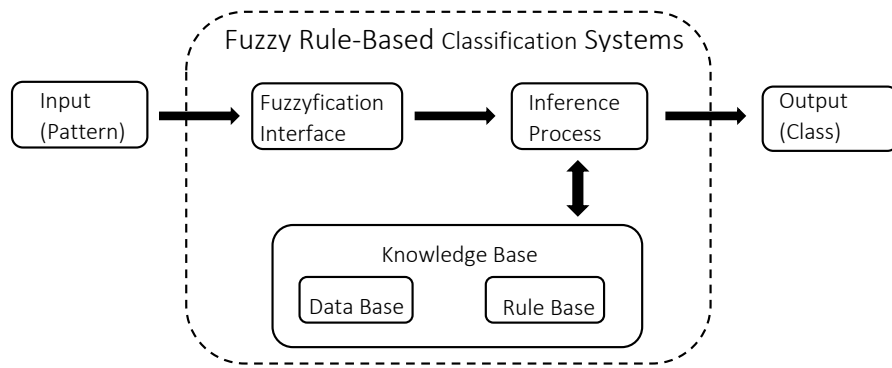

Fig. 1. A structure of FRBCS of the Mamdani type.

Where the Knowledge Base (KB) is composed by:

Data Base (DB) – Stores the membership functions associated with the linguistic labels considered in the fuzzy rules.

Rule Base (RB) – Is composed by a collection of linguistic fuzzy rules that are joined by a connective (operator and). Here we consider that a classification problem ins composed by t training patterns $x_p = (x_{p1}, \ldots, x_{pm}), p = 1, 2, \ldots, t.$ where x_{pi} is the i-th attribute and with the rules having the following structure:

$$\text{Rule } R_j : \text{ If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \tag{1}$$
$$\text{then Class is } C_j \text{ with } RW_j,$$

where R_j is the label of the j-th rule, A_{ji} is a fuzzy set modeling a linguistic term, modeled by a triangular shaped function. C_j is the class label and $RW_j \in [0, 1]$ is the rule weight [26].

The fuzzyfication interface converts the inputs (real values) into fuzzy values. In case of categorical variables, each value is modeled by a singleton and,

consequently, its membership value is either 1 or 0. Once the input is fuzzified, the inference process is the mechanism responsible for the use of the information stored in the KB to determine the class in which the example will be classified. The generalizations discussed in this paper are applied at this point.

Once the knowledge has been learnt and a new example $x_p = x_{p1}, \ldots, x_{pn}$ has to be classified, the FRM [27] is applied to perform this task, where M is the number of classes of the problem and L is the number of rules that compose the RB. The stages of the FRM are:

Matching degree: It represents the importance of the activation of the if-part of the rules for the example to be classified x_p, using a t-norm as conjunction operator:

$$\mu_{A_j}(x) = T(\mu_{A_{j1}}(x_1), \ldots, \mu_{A_{jn}}(x_n)). \tag{2}$$

with $j = 1, \ldots, L$. and $\mu_{A_{j1}}$ as the membership function with relation to a membership function.

Association degree: For each rule, the matching degree is weighted by its rule weight:

$$b_j^k(x) = \mu_{A_j}(x) \cdot RW_j^k, \tag{3}$$

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

Example classification soundness degree for all classes: For each class k, the positive information $b_j^k(x) > 0$, given by the fired fuzzy rules of the previous step, is aggregated by an aggregation function $\mathbb{A}$:

$$S_k(x) = \mathbb{A}_k \left(b_1^k(x), \ldots, b_L^k(x) \right), \tag{4}$$

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

In what follows, three different well-known FRMs are presented. Observe that their main difference is in the use of a different aggregation function to perform the aggregation of the information provided by the rules:

Winning Rule (WR) – For each class, it only considers the rule having the maximum compatibility with the example.

$$S_k(x) = \max_{R_{j_k} \in RB;} b_j(x). \tag{5}$$

Additive combination (AC) – It aggregates all the fired rules, for each class k, by using the normalized sum.

$$S_k(x) = \frac{\sum_{j=1}^{R_{j_k} \in RB} b_j(x)}{f_{1_{max}}}, \tag{6}$$

where $f_{1_{max}} = \max_{k=1,\ldots,M} \sum_{j=1}^{R_{j_k} \in RB} b_j(x)$.

The Choquet integral (CI) – It is the function $\mathfrak{C}_\mathfrak{m} : [0,1]^n \to [0,1]$, defined, for all of $\boldsymbol{x} \in [0,1]^n$, by:

$$\mathfrak{C}_m(\boldsymbol{x}) = \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)} \right) \cdot m \left(A_{(i)} \right), \tag{7}$$

Table 1. Association degrees for each class.

	C_1	C_2	C_3
R_a	0.94	0.15	0.89
R_b	0.10	0.40	0.88
R_c	0.25	0.10	0.85

where $N = \{0, \ldots, n\}$, $\mathfrak{m} : 2^N \to [0, 1]$ is a fuzzy measure[3], $\left(x_{(1)}, \ldots, x_{(n)}\right)$ is an increasing permutation on the input $\boldsymbol{x}$, that is, $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$, with $x_{(0)} = 0$, and $A_{(i)} = \{(i), \ldots, (n)\}$ is the subset of indices corresponding to the $n - i + 1$ largest components of $\boldsymbol{x}$. Then:

$$S_k(x) = \sum_{j=1}^{R_{j_k} \in RB} \mathfrak{C}_{\mathfrak{m}}(b_j(x)). \tag{8}$$

where $\mathfrak{C}$ is the standard CI and $\mathfrak{m}$ the fuzzy measure.

Classification: For the final decision, the class that maximizes all the example classification soundness degrees is considered, using the function $F : [0, 1]^M \to \{1, \ldots, M\}$:

$$F(S_1, \ldots, S_M) = arg \max_{k=1,\ldots,M}(S_k). \tag{9}$$

To exemplify the role of different aggregation operator in the FRM, consider a classification problem composed by 3 classes (C_1, C_2 and C_3). For each one, 3 generic fuzzy rules, R_a, R_b and R_c are fired when classifying a new example (they can be different for each class). We present the information about this problem in Table 1. Notice that the numbers in this table represent the positive association degree (Step 2 of the FRM) obtained for each fired rule. Having into account that three fuzzy rules are fired for each class, by columns, three aggregations have to be computed (one for each class).

Since the CI is defined with respect to a fuzzy measure, in this example the standard cardinality (see [28]) is considered as fuzzy measure. The values computed for each class using these three FRMs are the following ones:

- C_1
    - WR = 0.94
    - AC = $\frac{0.94 + 0.1 + 0.25}{2.62} = 0.49$
    - Choquet = $((0.1 - 0) \frac{3}{3}) + ((0.25 - 0.1) \frac{2}{3}) + ((0.94 - 0.25) \frac{1}{3}) = 0.43$
- C_2
    - WR = 0.4
    - AC = $\frac{0.15 + 0.4 + 0.1}{2.62} = 0.24$
    - Choquet = $((0.1 - 0) \frac{3}{3}) + ((0.15 - 0.1) \frac{2}{3}) + ((0.4 - 0.15) \frac{1}{3}) = 0.21$
- C_3
    - WR = 0.89
    - AC = $\frac{0.89 + 0.88 + 0.85}{2.62} = 1.0$

---

[3] A fuzzy measure $\mathfrak{m}$ is an increasing function on 2^N such that $\mathfrak{m}(\emptyset) = 0$ and $\mathfrak{m}(N) = 1$.

- Choquet = $((0.85 - 0)\frac{3}{3}) + ((0.88 - 0.85)\frac{2}{3}) + ((0.89 - 0.88)\frac{1}{3}) = 0.87$

Once the example classification soundness degree for each class has been computed, the predicted class is the one associated with the largest value (step 4 of the FRM):

- WR = arg max[**0.94**, 0.4, 0.89] = C_1
- AC = arg max[0.49, 0.24, **1.0**] = C_3
- Choquet = arg max[0.43, 0.21, **0.87**] = C_3

It is observable that the usage of the maximum as an aggregation operator predicts class 1, since it only considers the information provided by one fuzzy rule (having the maximum compatibility). However, if we look in detail at the association degrees presented in Table 1, this prediction may not be ideal, since that class 1 has one rule having high compatibility whereas class 3 has three rules having high compatibilities (slightly less than that of class 1). Then, class 3 seems to be the most appropriated option. This fact is taken into account by the CI and the AC, since the information given by all the fuzzy rules and not only by the best one is considered and, consequently, the prediction assigns class 3.

In this example, it is noticeable the non-averaging behavior of AC. Observe that the result of this function for class C_3 is greater than the maximum value. This fact does not occur for averaging functions. In the case of WR, the result is always the maximum, meanwhile for the CI the result is a value between the minimum and the maximum. Another interesting point that raises with this example, is that the usage of different aggregation in the FRM is directly related with the performance of the classifier.

3   The C_T-integral and pre-aggregations

This study was originally based on [14], where the authors modified the FRM of the Chi et al. algorithm [29] by applying the CI to aggregate all available information for each class. Furthermore, they introduced a learning method using a genetic algorithm in which the most suitable fuzzy measure for each class was computed. We highlight that this fuzzy measure is considered in all the applications of the generalizations of the Choquet integral.

For the first proposed generalization, the product operator of the standard CI was replaced by different other t-norms [30]. In this way, the manner how the information was aggregated would be different, consequently leading into different FRMs that could present performances even more accurately. The Choquet integral generalized by t-norms T, known as C_T-integral [17], is defined as:

Definition 1. *[17] Let* $\mathfrak{m} : 2^N \to [0, 1]$ *be a fuzzy measure and* $T : [0, 1]^2 \to [0, 1]$ *be an t-norm. A C_T-integral is the function* $\mathfrak{C}_{\mathfrak{m}}^T : [0, 1]^n \to [0, 1]$*, defined, for all* $\boldsymbol{x} \in [0, 1]^n$*, by*

$$\mathfrak{C}_{\mathfrak{m}}^T(\boldsymbol{x}) = \sum_{i=1}^{n} T\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right), \tag{10}$$

where $x_{(i)}$ and $A_{(i)}$ are defined as in Eq. (7).

Observe that some C_T-integrals are not aggregation function. E.g., take the minimum t-norm $T_M(x, y) = \min(x, y)$ and the cardinality measure (see [17,28]), and consider $\mathbf{x_1} = (0.05, 0.2, 0.7, 0.9)$ and $\mathbf{x_2} = (0.05, 0.1, 0.7, 0.9)$, where $\mathbf{x_1} > \mathbf{x_2}$. However, $\mathfrak{C}_{\mathfrak{m}}^{T_M}(\mathbf{x_1}) = 0.7$ and $\mathfrak{C}_{\mathfrak{m}}^{T_M}(\mathbf{x_2}) = 0.8$. Thus, the primordial condition of increasingness of any aggregation function is not fulfilled by $\mathfrak{C}_{\mathfrak{m}}^{T_M}$.

Yet, it is noticeable that the monotonicity property is not crucial for aggregation functions. Take for example a well-known statistical tool, the mode. It is not considered as an aggregation since the monotonicity of this function is not fulfilled, although it is useful. In [31], Bustince et al. introduced the notion of directional monotonicity, which allows monotonicity to be fulfilled along (some) fixed ray. So, with this in mind, the concept of pre-aggregation functions was introduced in [17]. These functions respect the boundary condition as any aggregation function, however, they are directional increasing:

Definition 2. *[31] 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 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*

$$F(x_1 + cr_1, \ldots, x_n + cr_n) \geq F(x_1, \ldots, x_n). \tag{11}$$

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

Now, as the Chi algorithm it is not a state-of-the-art fuzzy classifier, the C_T-integrals were applied in the FRM of a powerful fuzzy classifier like FARC-HD [32]. The quality of the proposal was analyzed by applying these generalizations to cope with 27 classification problems. The considered datasets are available in KEEL [33] dataset repository. When comparing the different generalizations among themselves, it can be noticed that the one based on Hamacher t-norm was superior to the remaining ones. This fact occurred with four out the five considered fuzzy measures. The best accuracy was obtained when combining the Hamacher product with the power measure. To evaluate the quality of this best generalization, the study has compared it against the classical FRM of WR, since both FRMs apply averaging aggregation functions. In this comparison, it was empirically demonstrated that this generalization is statistically superior to WR and the standard CI.

4   Copulas and CC-integrals

The usage of the generalizations of the CI in a powerful fuzzy classifier has produced satisfactory results to cope with classification problems. However, these generalizations were pre-aggregation functions, that is, the monotonicity is not satisfied. Then, with this in mind, generalizations that are idempotent and averaging aggregation functions were developed. For that, in Eq. (7), firstly the

distributivity property of the product operation is considered with the subtraction and then replaced the two instances of the product by copulas [30], obtaining the CC-integrals [18]:

Definition 3. *Let* $\mathfrak{m} : 2^N \to [0, 1]$ *be a fuzzy measure and* $C : [0, 1]^2 \to [0, 1]$ *be a bivariate copula. The CC-integral is defined as a function* $\mathfrak{C}_\mathfrak{m}^C : [0, 1]^n \to [0, 1]$, *given, for all* $\boldsymbol{x} \in [0, 1]^n$, *by*

$$\mathfrak{C}_\mathfrak{m}^C(\boldsymbol{x}) = \sum_{i=1}^n C\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - C\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right), \qquad (12)$$

where $x_{(i)}$ *and* $A_{(i)}$ *are defined as in Eq. (7).*

To demonstrate the efficiency of the CC-integrals to tackle classification problems, an experimental study considering 30 numerical datasets is considered. This study was conducted in two different ways. The first one was focused on comparisons per family of copulas (t-norms, overlap functions [34,35] and specific copulas), in order to find the function that presented the best generalization. Then, this best generalization is compared with 1) the classical FRM of WR (considering that both functions are averaging); 2) to the standard CI and 3) the best pre-aggregation function achieved in the previous study (C_T-integral), the one based on the Hamacher t-norm. The best CC-integral is the CMin-integral, constructed with the Minimum copula[4]. The obtained results showed that the CMin-Integral is statistically equivalent to the CI and the C_T-integral and superior than the WR..

5 C_F-integrals

The acquired knowledge from the previous studies shows that the function responsible to generalize the CI is very important. At this point only generalizations with averaging characteristics were presented. Having this in mind, the CI was generalized by special functions, in order to produce more competitive generalizations, allowing to produce non-averaging integrals. To achieve it, its used a family of left 0-absorbing aggregation functions F, which satisfy: (LAE) $\forall y \in [0, 1] : F(0, y) = 0$. Moreover, the following two basic properties are also important:

(RNE) Right Neutral Element: $\forall x \in [0, 1] : F(x, 1) = x$;

(LC) Left Conjunctive Property: $\forall x, y \in [0, 1] : F(x, y) \leq x$;

Any bivariate function $F : [0, 1]^2 \to [0, 1]$ satisfying both **(LAE)** and **(RNE)** is called left 0-absorbent **(RNE)**-function.

Then, the so-called C_F-integral [21] is defined as:

Definition 4.   *[21] Let* $F : [0, 1]^2 \to [0, 1]$ *be a bivariate function and* $\mathfrak{m} : 2^N \to [0, 1]$ *be a fuzzy measure. The* C_F-*integral is the function* $\mathfrak{C}_\mathfrak{m}^F : [0, 1]^n \to [0, 1]$,

---

[4] Examples of special CC-integrals were studied in [19, 20].

defined, for all $\boldsymbol{x} \in [0,1]^n$, by

$$\mathfrak{C}_{\mathfrak{m}}^{F}(\boldsymbol{x}) = \min\left\{1, \sum_{i=1}^{n} F\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\}, \tag{13}$$

where $x_{(i)}$ and $A_{(i)}$ are defined as in Eq. (7).

In [21, Theorems 1 and 2], it was proved that the set of conditions that the function F should fulfill for the C_F-integral to be a pre-aggregation function is one of the following ones: Theorem 1 ((LAE) and (RNE)) or Theorem 2 ((LAE), $F(1,1) = 1$ and $(1,0)$-increasingness). Moreover, for the C_F-integral to be averaging F must satisfy (RNE) and (LC). This means that there exist a lot of non-averaging C_F-integrals.

The quality of the C_F-integrals to cope with classification problems was tested considering 33 different datasets. The experimental study was conducted considering C_F-integrals with and without averaging characteristics. Considering the non-averaging functions, six C_F-integrals were studied. In order to support the quality of this approach, a comparison with the best non-averaging C_F-integral with the FRM of AC and a FRM considering the probabilistic sum - PS (since it is an operator with non-averaging characteristics) is provided. The results showed that the non-averaging C_F-integrals-integrals, as expected, offer a performance superior than the averaging ones, and the best C_F-integral, based on the function FNA2 [5] provides results that are statistically superior than all classical FRMs, and also, very competitive with the classical non-averaging FRMs like AC or PS.

6 $C_{F_1 F_2}$-integrals

The previous study demonstrated that the generalization of the standard Choquet integral by functions F resulted in satisfactory results. Then, this study combine the ideas of previous approaches, precisely, it take the same idea of CC-integrals, generalizing the each of the two instances of copulas by a pair of functions F, called F_1 and F_2, as consequence obtaining the $C_{F_1 F_2}$-integrals [22]:

Definition 5. *Let $\mathfrak{m} : 2^N \to [0,1]$ be a symmetric fuzzy measure and $F_1, F_2 : [0,1]^2 \to [0,1]$ be two fusion functions fulfilling:*

(i) *F_1-dominance (or, equivalently, F_2-Subordination): $F_1 \geq F_2$;*
(ii) *F_1 is $(1,0)$-increasing,*

A $C_{F_1 F_2}$-integral is defined as a function $\mathfrak{C}_{\mathfrak{m}}^{(F_1, F_2)} : [0,1]^n \to [0,1]$, given, for all $\boldsymbol{x} \in [0,1]^n$, by

$$\mathfrak{C}_{\mathfrak{m}}^{(F_1, F_2)}(\boldsymbol{x}) = \tag{14}$$

$$\min\left\{1, x_{(1)} + \sum_{i=2}^{n} F_1\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - F_2\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\},$$

---

[5] The function $FNA2 : [0,1]^2$ is defined, for all $x, y \in [0,1]$ by $F(0,y) = 0$, $F(x,y) = \frac{x+y}{2}$ if $0 < x \leq y$ and $F(x,y) = \min(\frac{x}{2}, y)$, otherwise, which satisfies the conditions of [21, Theorems 2].

where $x_{(i)}$ and $A_{(i)}$ are defined as in Eq. (7).

In this paper, twenty-three different functions, F, were considered. As consequence, 201 different pairs of functions that could be used as F_1 and F_2 could be combined, respecting the dominance property. An important question that could appear is related to the choice of the function to be selected as F_1 and the one to act as F_2. Therefore, a methodology to reduce the scope of the study have been proposed by using the concept of Dominance and Subordination Strength degree, DSt and SSt respectively.

Definition 6. *Let $\mathcal{F} = \{F_1, \ldots, F_m\}$ be a set of m fusion functions. The dominance and subordination strength degrees, DSt and SSt, of a fusion function $F_i \in \mathcal{F}$ are defined, respectively, for $j \in \{1, \ldots, m\}$, by as follows:*

$$DSt(F_i) = \frac{1}{m} \sum_{j=1}^{m} \begin{cases} 1 \ if \ F_i \geq F_j, \\ 0 \ otherwise \end{cases} \cdot 100\%$$

$$SSt(F_i) = \frac{1}{m} \sum_{j=1}^{m} \begin{cases} 1 \ if \ F_i < F_j, \\ 0 \ otherwise. \end{cases} \cdot 100\%$$

The generalizations provided in this study are non averaging. Moreover, they satisfy the boundary conditions of any (pre) aggregation function. However, considering the monotonicity, we observed that these functions are neither increasing nor directional increasing. In fact, they are Ordered Directionally (OD) monotone functions [36]. These functions are monotonic along different directions according to the ordinal size of the coordinates of each input.

The $C_{F_1 F_2}$-integrals were used to cope with classification problems in 33 different datasets. When analyzing the results that were obtained by the usage of these generalizations, it is noticeable that the combination of a function having a high dominance as F_1 combined with a function with high subordination as F_2 presented the best results of this study (from the top ten of the best global accuracies from the 81 pairs, eight have this characteristic). We also observed that the opposite, for each function F_2, is also true and that its best results are achieved when using a F_1 with a high dominance.

The performance of this proposal is analyzed by comparing them against distinct state-of-the-art FRBCSs, namely: FARC-HD [32], FURIA [37], IVTURS [38], a classical non-averaging aggregation operator like the probabilistic sum, P^*, and, the best C_F-integral that was selected from the previous study, F_{NA2}. In this comparison, FURIA was the fuzzy classifier that achieved the highest accuracy mean, however, our new approach achieved a close classification rate. Furthermore, the number of specific datasets where the performance of our generalization is the worst among all the methods in the comparison is less than that of FURIA. The function representing the C_F-integrals also achieved good results, meanwhile the remainder cases (IVTURS, P^* and FARC-HD) where inferior and similar among themselves.

The 81 pairs of combinations considered to construct $C_{F_1 F_2}$-integrals were compared against IVTURS, P^*, FARC-HD and F_{NA2}. The results highlighted

the quality of our new method because an equal or greater average result was obtained by 39, 36, 34 and 12 different combinations in these comparisons.

Finally, from the considered pairs, it was observed that five different $C_{F_1 F_2}$-integrals were considered as control variable in the statistical test in which all methods are compared, including FURIA. The last generalization only presented statistical differences with respect to FARC-HD. However, for any remaining pair, it is statistically equivalent when compared to FURIA and to F_{NA2} and superior to IVTURS, P^* and FARC-HD.

7   Detailed Results

In this section the results obtained by the usage of different aggregation operator are shown. We highlight that these results consider the same 33 datasets as in [21, 22] and [28]. Also, the results are related with the power measure, as mentioned previously, take into consideration the 5-fold cross validation technique [2] and are applied in the FRM of the FARC-HD [32] fuzzy classifier[6].

The results are provided in Table-2 where each cell correspond to the mean accuracy among all folds, the rows are related with the different considered datasets and the columns are the results obtained by classical FRMs such as: of the Additive Combination (AC), Probabilistic Sum (PS), Winning Rule (WR), Choquet integral (CI), C_T-integrals (due to lack of space were summarized to int, in all integrals) with is defined by the Hamacher product t-norm, CC-integral that use the copula of the minimum, C_F-integral considering the FNA_2 function and $C_{F_1 F_2}$-integrals using the pair $GM–FBPC$.

From the detailed results, we can noticed that classic FRM of the WR is the one that achieved the lowest global mean, indicating that the usage of all information related with the problem is an interesting alternative. Moreover, it is also observable that all non-averaging generalizations (AC, PS, C_F-integrals and $C_{F_1 F_2}$-integrals) presents superior results when compared against the averaging ones (WR, IC, C_T-integral and CC-integral).

The results also showed that the generalizations of the CI (C_T, CC, C_F, $C_{F_1 F_2}$-integrals) provided a superior performance in comparison to the standard CI. Finally, as mentioned before, the largest performance is obtained when the $C_{F_1 F_2}$-integral is used to cope with classification problems.

8   Conclusions

The application of the Choquet integral (CI) in the Fuzzy Reasoning Method (FRM) of Fuzzy Rule-Based Classification Systems (FRBCSs) modified the way in which the information was used and enhanced the system quality. After that, many generalizations of the CI were proposed and also applied in FRM, obtaining

---

[6] The considered datasets and the fuzzy classifier are available in KEEL repository. Available at https://www.keel.es

Table 2. Detailed results achieved in test by different generalizations of the CI.

Dataset	AC	PS	WR	CI	$C_T - int$	CC-int	C_F-int	$C_{F_1 F_2}$-int
App	83.03	85.84	83.03	80.13	82.99	85.84	85.84	86.80
Bal	85.92	87.20	81.92	82.40	82.72	81.60	88.64	89.12
Ban	85.30	84.85	83.94	86.32	85.96	84.30	84.60	84.79
Bnd	68.28	68.82	69.40	68.56	72.13	71.06	70.48	71.30
Bup	67.25	61.74	62.03	66.96	65.80	61.45	64.64	66.96
Cle	56.21	59.25	56.91	55.58	55.58	54.88	56.55	56.22
Con	53.16	52.21	52.07	51.26	53.09	52.61	53.16	54.72
Eco	82.15	80.95	75.62	76.51	80.07	77.09	80.08	81.86
Gla	65.44	64.04	64.99	64.02	63.10	69.17	66.83	68.25
Hab	73.18	69.26	70.89	72.52	72.21	74.17	71.87	72.53
Hay	77.95	77.95	78.69	79.49	79.49	81.74	79.43	78.66
Ion	88.90	88.32	90.03	90.04	89.18	88.89	89.75	88.33
Iri	94.00	95.33	94.00	91.33	93.33	92.67	94.00	94.00
Led	69.60	69.20	69.40	68.20	68.60	68.40	69.80	70.00
Mag	80.76	80.39	78.60	78.86	79.76	79.81	79.70	80.86
New	94.88	94.42	94.88	94.88	95.35	93.95	96.28	96.74
Pag	95.07	94.52	94.16	94.16	94.34	93.97	94.15	95.25
Pen	92.55	93.27	91.45	90.55	90.82	91.27	92.91	92.91
Pho	81.70	82.51	82.29	82.98	83.83	82.94	81.44	81.42
Pim	74.74	75.91	74.60	74.60	73.44	75.78	74.61	75.38
Rin	90.95	90.00	90.00	90.95	88.78	87.97	89.86	91.89
Sah	68.39	69.69	68.61	69.69	70.77	70.78	70.12	71.43
Sat	79.47	80.40	79.63	79.47	80.40	79.01	80.41	79.47
Seg	93.12	92.94	93.03	93.46	93.33	92.25	92.42	93.29
Shu	95.59	94.85	96.00	97.61	97.20	98.16	97.15	96.83
Son	78.36	82.24	77.42	77.43	79.34	76.95	83.21	85.15
Spe	77.88	77.90	77.90	77.88	76.02	78.99	79.77	79.39
Tit	78.87	78.87	78.87	78.87	78.87	78.87	78.87	78.87
Two	90.95	90.00	86.49	84.46	85.27	85.14	92.57	92.30
Veh	68.56	68.09	66.67	68.44	68.20	69.86	68.08	68.20
Win	96.03	94.92	96.60	93.79	96.63	93.83	96.08	95.48
Wis	96.63	97.22	96.34	97.22	96.78	95.90	96.78	96.78
Yea	58.96	59.03	55.32	55.73	56.53	57.01	57.08	58.56
Mean	80.12	80.07	79.15	79.22	79.69	79.58	80.52	81.02

success as well. In this paper the main contributions, theoretical and applied, of the generalizations are summarized and discussed.

The first generalization was built by the replacement of the product operator of the standard CI by different t-norms. These generalizations were supported by an important theoretical concept known as pre-aggregation functions. Differently from a simple aggregation function, a pre-aggregation function is monotonic only in a determined direction. This first generalization produced averaging functions and its applications to cope with classification problems showed that the generalization by the Hamacher product t-norm was superior than the FRM of the Winning Rule (WR) and the CI.

The second step aimed in generalizations of the CI that produce aggregation functions. To do so, the IC was used in its expanded form and generalized by copula functions, introducing the concept of Choquet-like Copula-Based aggregation functions, the so called CC-integrals. These functions also present averaging characteristics. The results of their applications demonstrated that the classical WR was statistically overcame.

It is observable that up to this point only generalizations with averaging characteristics were presented. On the otter hand, fuzzy classifiers known as

state-of-the-art take into account the usage of non-averaging functions. Thus, to produce more competitive generalizations, a family of fusion functions F were introduced. The generalization of the Choquet integral by F functions introduced the concept of C_F-integrals. This generalization has averaging and non-averaging characteristics, it depends on the considered function. It was observed that the application of any non-averaging function statistically overcome any averaging one. Also, the developed operators outperforms the classical WR and Additive Combination (AC).

The generalization of the expanded CI by two functions F, F_1 and F_2, introduced the concept of $C_{F_1 F_2}$-integrals. These functions present an Ordered Directional increasing functions (OD increasing) and, therefore, represent a different level of aggregation operators. The summit of the performance in the classification problems was reached in this generalization. To do so, a methodology to select different functions as F_1 and F_2 were presented, based on the concept of degrees of dominance and subordination. For the considered $C_{F_1 F_2}$-integrals, in five different cases the generalizations are equivalent, or even superior, in comparison with fuzzy classifiers found in the literature.

Taking as basis the analysis provided by this paper, some interesting research points emerge. For example, the application of these generalizations in the FRM of different fuzzy classifiers. Also, considering that the generalizations are based on the Choquet integral, the usage of a different operator, such as the Sugeno integral can produce even more powerful operators. Finally, the combinations with different fuzzy measures are an alternative with great potential.

References

1. E. Alpaydin, *Introduction to Machine Learning*, 2nd ed.   The MIT Press, 2010.
2. P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining, (First Edition)*.   Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.
3. C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
4. J. Quinlan, *C4.5: Programs for Machine Learning.*   Morgan Kauffman, 1993.
5. N. García-Pedrajas, C. García-Osorio, and C. Fyfe, "Nonlinear boosting projections for ensemble construction," *Journal of Machine Learning Research*, vol. 8, pp. 1–33, 2007.
6. H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and Modeling with Linguistic Information Granules, Advanced Approaches to Linguistic Data Mining*, ser. Advanced Information Processing.   Berlin: Springer, 2005.
7. L. A. Zadeh, "The concept of a linguistic variable and its application to approximate reasoning - I," *Information Sciences*, vol. 8, no. 3, pp. 199–249, 1975.
8. O. Unold, "Diagnosis of cardiac arrhythmia using fuzzy immune approach," in *Adaptive and Natural Computing Algorithms.*   Berlin: Springer, 2011, pp. 265–274.
9. B. Vitoriano, J. T. Rodrígues, G. Tirado, F. J. Martín-Campo, M. T. Ortuño, and J. Montero, "Intelligent decision-making models for disaster management," *Human and Ecological Risk Assessment: An International Journal*, vol. 21, no. 5, pp. 1341–1360, 2015.

10. J. A. Sanz, D. Bernardo, F. Herrera, H. Bustince, and 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 Transactions on Fuzzy Systems*, vol. 23, no. 4, pp. 973–990, Aug 2015.

11. N. Goel and P. Sehgal, "Fuzzy classification of pre-harvest tomatoes for ripeness estimation - an approach based on automatic rule learning using decision tree," *Applied Soft Computing*, vol. 36, pp. 45 – 56, 2015.

12. O. Cordon, M. J. del Jesus, and F. Herrera, "Analyzing the reasoning mechanisms in fuzzy rule based classification systems," *Mathware and Soft Computing*, vol. 5, no. 2-3, pp. 321 – 332, 1998.

13. Z. Chi, H. Yan, and T. Pham, *Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition*.   River Edge, NJ, USA: World Scientific Publishing Co., Inc., 1996.

14. E. Barrenechea, H. Bustince, J. Fernandez, D. Paternain, and J. A. Sanz, "Using the Choquet integral in the fuzzy reasoning method of fuzzy rule-based classification systems," *Axioms*, vol. 2, no. 2, pp. 208–223, 2013.

15. G. Choquet, "Theory of capacities," *Annales de l'Institut Fourier*, vol. 5, pp. 131–295, 1953–1954.

16. G. P. Dimuro, J. Fernández, B. Bedregal, R. Mesiar, J. A. Sanz, G. Lucca, and H. Bustince, "The state-of-art of the generalizations of the choquet integral: From aggregation and pre-aggregation to ordered directionally monotone functions," *Information Fusion*, vol. 57, pp. 27 – 43, 2020.

17. G. Lucca, J. Sanz, G. Pereira Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, and H. Bustince Sola, "Pre-aggregation functions: construction and an application," *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 2, pp. 260–272, April 2016.

18. G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, M. J. Asiain, M. Elkano, and H. Bustince, "CC-integrals: Choquet-like copula-based aggregation functions and its application in fuzzy rule-based classification systems," *Knowledge-Based Systems*, vol. 119, pp. 32 – 43, 2017.

19. G. P. Dimuro, G. Lucca, J. A. Sanz, H. Bustince, and B. Bedregal, *CMin-Integral: A Choquet-Like Aggregation Function Based on the Minimum t-Norm for Applications to Fuzzy Rule-Based Classification Systems*.   Cham: Springer International Publishing, 2018, pp. 83–95.

20. G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, and H. Bustince, "A proposal for tuning the the α parameter in $C_\alpha C$-integrals for application in fuzzy rule-based classification systems," *Natural Computing.*, vol. 19, pp. 533–546, 2020.

21. G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, H. Bustince, and R. Mesiar, "CF-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems," *Information Sciences*, vol. 435, pp. 94 – 110, 2018.

22. G. Lucca, G. P. Dimuro, J. Fernandez, H. Bustince, B. Bedregal, and J. A. Sanz, "Improving the performance of fuzzy rule-based classification systems based on a nonaveraging generalization of CC-integrals named $C_{F_1 F_2}$-integrals," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 1, pp. 124–134, Jan 2019.

23. G. P. Dimuro, G. Lucca, B. Bedregal, R. Mesiar, J. A. Sanz, C.-T. Lin, and H. Bustince, "Generalized cf1f2-integrals: From choquet-like aggregation to ordered directionally monotone functions," *Fuzzy Sets and Systems*, vol. 378, pp. 44 – 67, 2020.

24. T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 1, pp. 116–132, Jan 1985.

25. E. H. Mamdani, "Application of fuzzy algorithms for control of simple dynamic plant," *Electrical Engineers, Proceedings of the Institution of*, vol. 121, no. 12, pp. 1585–1588, December 1974.

26. H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *Fuzzy Systems, IEEE Transactions on*, vol. 9, no. 4, pp. 506–515, 2001.

27. O. Cordón, M. J. del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *International Journal of Approximate Reasoning*, vol. 20, no. 1, pp. 21 – 45, 1999.

28. G. Lucca, J. A. Sanz, G. P. Dimuro, E. N. Borges, H. Santos, and H. Bustince, "Analyzing the performance of different fuzzy measures with generalizations of the choquet integral in classification problems," in *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, June 2019, pp. 1–6.

29. Z. Chi, H. Yan, and T. Pham, *Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition*.   Singapore: World Scientific, 1996.

30. C. Alsina, M. J. Frank, and B. Schweizer, *Associative Functions: Triangular Norms and Copulas*.   Singapore: World Scientific Publishing Company, 2006.

31. H. Bustince, J. Fernandez, A. Kolesárová, and R. Mesiar, "Directional monotonicity of fusion functions," *European Journal of Operational Research*, vol. 244, no. 1, pp. 300–308, 2015.

32. J. Alcala-Fdez, R. Alcala, and F. Herrera, "A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning," *IEEE Transactions on Fuzzy Systems*, vol. 19, no. 5, pp. 857–872, 2011.

33. J. Alcalá-Fdez, L. Sánchez, S. García, M. Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernández, and F. Herrera, "Keel: a software tool to assess evolutionary algorithms for data mining problems," *Soft Computing*, vol. 13, no. 3, pp. 307–318, 2009.

34. H. Bustince, J. Fernandez, R. Mesiar, J. Montero, and R. Orduna, "Overlap functions," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 72, no. 3-4, pp. 1488–1499, 2010.

35. G. P. Dimuro and B. Bedregal, "On the laws of contraposition for residual implications derived from overlap functions," in *Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on*.   Los Alamitos: IEEE, Aug 2015, pp. 1–7.

36. H. Bustince, E. Barrenechea, M. Sesma-Sara, J. Lafuente, G. P. Dimuro, R. Mesiar, and A. Kolesárová, "Ordered directionally monotone functions. justification and application," *IEEE Transactions on Fuzzy Systems*, p. 1, 2017, (In press, corrected proof).

37. J. Hühn and E. Hüllermeier, "Furia: an algorithm for unordered fuzzy rule induction," *Data Mining and Knowledge Discovery*, vol. 19, no. 3, 2009.

38. J. Sanz, A. Fernández, H. Bustince, and 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 Transactions on Fuzzy Systems*, vol. 21, no. 3, pp. 399–411, 2013.