Universidad Pública de Navarra

Departamento de Automática y Computación

Aggregation and pre-aggregation functions in fuzzy rule-based classification systems

Giancarlo Lucca

DOCTORAL THESIS

Pamplona, June, 2018

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

Universidad Pública de Navarra

Departamento de Automática y Computación

Aggregation and pre-aggregation functions in fuzzy rule-based classification systems

PhD Student: Giancarlo Lucca

Supervisors: Humberto Bustince

José Antonio Sanz

Pamplona, June, 2018

Autorización

Dr. Humberto Bustince Sola: Catedrático de la Universidad Pública de Navarra en el Área de Conocimiento de Ciencias de la Computación e Inteligencia Artificial

Dr. José Antonio Sanz Delgado: Profesor de la Universidad Pública de Navarra en el Área de Conocimiento de Ciencias de la Computación e Inteligencia Artificial

HACEN CONSTAR que el presente trabajo titulado "*Aggregation and pre-aggregation functions in fuzzy rule-based classification systems*" ha sido realizado bajo su dirección por D. Giancarlo Lucca.

Autorizándole a presentarlo como Memoria para optar al grado de Doctor por la Universidad Pública de Navarra.

El Doctorando

Fdo: Giancarlo Lucca

Los Directores

Fdo: Humberto Bustince Sola

Fdo: José Antonio Sanz Delgado

Pamplona, junio de 2018

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

To Celso, Sônia and Thaís.

This and all my achievements.

Acknowledgments

I start by thanking my parents, Celso and Sônia. You are the reason of all this. I also thank my brothers Marcelo and Márcio, who are an inspiration for me and that I have missed very much during the years I've been doing this work. But, at the end, it is worth it and I know that you are always with me.

A very special thanks to a great woman. My wife. Thank you, Thaís, for all the support, joy and incentives. It is unwritable the quantity of things that we pass throughout to reach this point. Certainly this journey was easier with you at my side. All my love to you.

A special thanks to three giants who carried me in their shoulders. To Humberto, that accepted me as his student and trusted me. You are an exceptional human being, who cares about everybody and that now I can consider a friend of mine. To Josean, who taught me so much during this journey. You made all the difference in my career and in my life. Finally, to Graçaliz, for all the incentive, since the graduation times until nowadays. What a successful partnership.

I also dedicate it to Beatriz, Clea, Madalena, Paulo and all my family. Your support was essential. Also to all my friends from Ijuí and Rio Grande. The distance can not separate us.

A thanks to all my friends at the GIARA group. Thanks for welcoming me and for being so supportive. A special thanks to Cédric, for his friendship, for his advices about the spanish culture and so much more.

Last but not least, I have to mention three important members of our family, Bóris, Luna and Sheldon. Our dogs. The delight of being with them is incomparable.

Mission Accomplished.

Thank you all.

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 the way in which the information given by the fired fuzzy rules is aggregated. Precisely, the aggregation function is the component that differs the two most widely used FRMs in the specialized literature. The first one, known as Winning Rule (WR), applies the maximum as the aggregation function, which has an averaging behavior. This function is limited by the maximum and the minimum of the values to be aggregated and it uses the largest relationship between the new example to be classified and the fuzzy rules. The second one, known as Additive Combination (AC), is used by the most accurate algorithms nowadays and it applies the normalized sum to aggregate the information but, in this case, this aggregation operator has a non-averaging behavior.

In this thesis, we intend to change the way that the information is aggregated in the FRM by applying generalizations of the Choquet integral. To do so, we have developed new theoretical concepts in the field of aggregation operators. These generalizations of the Choquet integral present both averaging and non-averaging behaviors. We use them in the FRM of FARC-HD, which is a state-of-the-art FRBCS. From the obtained results, we show that the new FRM can be used in an efficient way to deal with classification problems, taking into account that the results are statistically comparable, or even superior, to the state-of-the-art fuzzy classifiers.

Resumen

Una manera eficiente de tratar problemas de clasificación, entre otras, es el uso de Sistemas de Clasificación Basados en Reglas Difusas (SCBRDs). Estos sistemas están compuestos por dos componentes principales, la Base de Conocimiento (BC) y el Método de Razonamiento Difuso (MRD). El MRD es el método responsable de clasificar nuevos ejemplos utilizando la información almacenada en la BC. Un punto clave del MRD es la forma en la que se agrega la información proporcionada por las reglas difusas disparadas. Precisamente, la función de agregación es lo que diferencia a los dos MRDs más utilizados de la literatura especializada. El primero, llamado de Regla Ganadora (RG), tiene un comportamiento promedio, es decir, el resultado de la agregación está en el rango delimitado por el mínimo y el máximo de los valores a agregar y utiliza la mayor relación entre el nuevo ejemplo a clasificar y las reglas. El segundo, conocido como Combinación Aditiva (CA), es ampliamente utilizado por los algoritmos difusos más precisos de la actualidad y aplica una suma normalizada para agregar toda la información relacionada con el ejemplo. Sin embargo, este método no presenta un comportamiento promedio.

En este trabajo de tesis, proponemos modificar la manera en la que se agrega la información en el MRD, aplicando generalizaciones de la integral Choquet. Para ello, desarrollamos nuevos conceptos teóricos en el campo de los operadores de agregación. En concreto, definiremos generalizaciones de la Choquet integral con y sin comportamientos promedio. Utilizamos estas generalizaciones en el MRD del clasificador FARC-HD, que es un SCBRD del estado del arte. A partir de los resultados obtenidos, demostramos que el nuevo MRD puede ser utilizado, de manera eficiente, para afrontar problemas de clasificación. Además, mostramos que los resultados son estadísticamente equivalentes, o incluso superiores, a los clasificadores difusos considerados como estado del arte.

Contents

Abstract **ix**

Resumen **xi**

I Introduction **1**

1 Motivation 2

2 Objectives 3

3 Structure 4

II PhD thesis report **5**

1 Background 5

1.1 Aggregation Functions 5

1.1.1 T-norms 7

1.1.2 Overlap functions 7

1.1.3 Copulas 8

1.1.4 Choquet integral 9

1.2 Classification Problems 11

1.2.1 Fuzzy Rule-Based Classification Systems 15

1.2.2 Fuzzy Reasoning Method 19

1.2.2.1 Example of the behavior of different fuzzy reasoning methods 20
1.2.3 Evolutionary Fuzzy Systems 22
1.2.4 The FARC-HD fuzzy classifier 25
2 Discussion 29
2.1 The concept of pre-aggregation functions 29
2.2 The generalization of the extended Choquet integral by copulas . . . 32
2.2.1 A proposal for tuning the alpha parameter in $C_{\alpha}C$-integrals 34
2.3 A generalization of the Choquet integral by left 0-absorbing fusion functions 36
2.4 Generalizations of the Choquet integral by pairs of functions F_1–F_2 under some constraints 38
2.5 The idea of generalized $C_{F_1F_2}$-integrals 41
3 Summaries 43
3.1 Summary of the theoretical developments 43
3.2 Summary of the results of the generalizations of the Choquet integral in the fuzzy reasoning method 45
4 Final Comments 47
4.1 Conclusions 47
4.2 Future research lines 50
5 Introducción y Conclusión (Versión en español) 54
5.1 Introducción 54
5.2 Conclusión 56

III Publications: published, accepted and submitted papers 59

1 Pre-aggregation Functions: Construction and an application 59

2 CC-integrals: Choquet-like Copula-based aggregation functions and its application in fuzzy rule-based classification systems 73

3 A proposal for tuning the alpha parameter in C_αC-integrals for application in fuzzy rule-based classification systems 86

4 C_F-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems 99

5 Improving the performance of fuzzy rule-based classification systems based on a new non averaging generalization of CC-integrals named $C_{F_1F_2}$-integrals . . 117

6 Generalized $C_{F_1F_2}$-integrals: from Choquet-like aggregation to ordered directionally monotone functions 129

References **151**

Chapter I

Introduction

Human beings perform classifications since the primordials. The distinction between edible and non-edible fruits, seeds and roots that were collected is a simple example. These kind of decisions are made based on available and acquired knowledge in the course of time.

Then, we can say that classification is a process in which the data is labeled (classified) based on predetermined characteristics. Observe that there are different complexities in classification problems, from the simplest, as the definition of a breed of a dog based on the characteristics of this dog, to complex, as the classification of a cancer based on blood information. An expert in a certain issue can deal with a determined classification problem, however, he/she can be out of reach and his/her processes can be slow, expensive and even inaccurate.

The usage of an automatic classification system is a good option in the classification process. Notice that the system can not replace the knowledge of an expert, yet, the expert can use this system as an important source of information in the decision making process.

In the literature, classification problems [Alp10, DHS00] are a research field in the area of the so-called, data mining [TSK05]. Classification problems are tackled in two different ways. The first one, which is the approach considered in this thesis, is known as supervised learning. It generates a function (classifier) 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. The second approach deals with unlabeled data (without knowing the classes of the examples used to learn the system) trying to extract the relationships in the data. This method is known as unsupervised learning.

As we have mentioned, this thesis is focused on supervised learning with the objective of tackling classification problems. In the literature we can find several methods that aim to cope with these problems such as Support Vector Machines (SVM) [CV95], decisions trees [Qui93, BFOS84] and neural networks [GPGOF07]. In this thesis we apply Fuzzy Rule-Based Classification Systems (FRBCCs) [INN05], because they provide the user with interpretable models by using linguistic labels (like high, medium or low) [Zad75] in their rules. Another reason is because of their accurate results and versatility, as shown in the many different fields where they have been applied like health [Uno11, SH09], security [GSP+14, VRT+15], economy [SBH+15], food [SFB+16, GS15] and many others.

1 Motivation

An important role in any FRBCS is played by the Fuzzy Reasoning Method (FRM) [CdJH98, CdJH99]. This method is responsible to perform the classification of new examples. To do so, 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 method. By using this aggregation operator, for each class, the FRM performs the selection of the best fired rule since it has the highest compatibility with the example [CYP96a, GP98, INYT94]. The issue of this inference method is that the information provided by the remainder fired fuzzy rules is ignored. This aggregation operator is considered averaging, 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 [CdJH98, CdJH99] 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 [BBF+13] the authors introduced a FRM considering the usage of the Choquet integral [Cho54], 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.

Considering the previous analysis, in this thesis we propose a methodology that changes the aggregation step performed in the FRM. Precisely, we consider the application of different generalizations of the Choquet integral, which are supported by solid theoretical studies. We start by generalizations having averaging characteristics and we end up producing non-averaging ones, in order to produce results that are able to be competitive against state-of-the-art FRBCSs.

2 Objectives

The general objective of this thesis is:

To develop a new methodology to aggregate the information of the fired rules in the fuzzy reasoning method, which leads to an enhancement of the performance of the fuzzy classifier.

To accomplish this main goal, we have some particular objectives:

- To develop new theoretical concepts to aggregate information derived form the standard Choquet integral based on
  1. The replacement of the product operator in the Choquet integral by different t-norms or fusion functions.
  2. The usage of the extended form of the Choquet integral and the usage of well-known functions like copulas or overlaps.
- To construct idempotent functions with averaging or non-averaging characteristics.
- To include the usage of the theoretical developments in the fuzzy reasoning method of FARC-HD, since it is an state-of-the-art fuzzy classifier.
- To introduce a methodology considering the usage of an evolutionary fuzzy system to learn the most appropriate fuzzy measure for each class of the classification problem.
- To compare the results of our methodology versus those of state-of-the-art fuzzy classifiers.

3 Structure

This thesis is dived in two main parts. The first one (Chapter II) contains the body of the text where we start by introducing the background about the aggregation functions that are used to develop this thesis, being them: t-norms, overlaps, copulas and the Choquet integral. Then, we introduce the formal definition of classification problems, the concepts of fuzzy rule-based classification systems, the fuzzy reasoning method, with an example of its behavior, the evolutionary fuzzy systems and the description of the FARC-HD fuzzy classifier. Then, we summarize the theoretical concepts developed in this thesis and their application and finally, we present the conclusions of the thesis as well as some future research lines. Also, we present for each publication associated with this thesis the description of the proposal besides its novelty, the obtained results and the conclusions.

The final part (Chapter III) is composed by the papers that form the core of the thesis. They are the result of the research done throughout this journey. For each publication we present the journal where it was published or submitted as well as its impact factor, the current status and the text of the paper. In the following we present the six publications associated with this thesis, besides the associated publication (if the paper was already published):

- Pre-aggregation Functions: Construction and an Application [LSPD+16].
- CC-integrals: Choquet-like Copula-based aggregation functions and its application in fuzzy rule-based classification systems [LSD+17b].
- A proposal for tuning the alpha parameter in C_αC-integrals for application in fuzzy rule-based classification systems.
- C_F-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems [LSD+18a].
- Improving the performance of fuzzy rule-based classification systems based on a new non averaging generalization of CC-integrals named $C_{F_1F_2}$-integrals.
- Generalized $C_{F_1F_2}$-integrals: from Choquet-like aggregation to ordered directionally monotone functions.

Chapter II

PhD thesis report

In this chapter we present the concepts used in the development of the new methodology created in this thesis. We start by introducing the background necessary for the developments of the thesis, which involves notions about aggregation functions as well as classification problems. Next, we present the core of the thesis where we provide a detailed discussion of each paper that conforms it. Then we present a summary of the developed concepts with their applications and finally, end we draw the conclusions and we mention some open research lines.

1 Background

In this section we introduce the background related to both aggregation operators and classification problems. We start by introducing the concept of aggregation functions, triangular norms, overlap functions, copulas and the Choquet integral.

1.1 Aggregation Functions

We consider that aggregation functions [BPC07, GMMP09] are a special case of functions that combine several values in a determined interval in order to produce a new one, which represents the aggregated information. We must point out that we consider the interval [0, 1] for both the values to be aggregated and the generated one. The mode, the mean and the Ordered Weighted Mean (OWA) [YK97, YKB11] are examples of aggregation functions. Mathematically speaking, we define an aggregation function as:

Definition 1. *A function* $A : [0,1]^n \to [0,1]$ *is an aggregation function if the following conditions hold:*

(A1) *A is increasing*[1] *in each argument:*
for each $i \in \{1,\dots,n\}$, *if* $x_i \le y$, *then* $A(x_1,\dots,x_n) \le A(x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_n)$;

(A2) *A satisfies the boundary conditions:*
(i) $A(0,\dots,0) = 0$ *and*
(ii) $A(1,\dots,1) = 1$.

We say that an aggregation function is averaging [BPC07] if the result of this function is bounded by the minimum and the maximum of the aggregated values.

Definition 2. *An aggregation function* $A : [0,1]^n \to [0,1]$ *is averaging if:*

$$\min(x_1,...,x_n) \le A(x_1,...,x_n) \le \max(x_1,...,x_n)$$

Obviously, we consider that an aggregation function is non-averaging when the result of the aggregation function A is not within the range delimited by the minimum and the maximum of the aggregated values.

Definition 3. *An aggregation function* $A : [0,1]^n \to [0,1]$ *is said to be idempotent if and only if:* $\forall x \in [0,1] : A(x,\dots,x) = x$.

Since aggregation functions are increasing, the idempotent and averaging behaviors are equivalent.

Proof. [BPC07] Take any $\mathbf{x} \in [0,1]$, and denote by $\mathrm{p} = \min(\mathbf{x})$, $\mathrm{q} = \max(\mathbf{x})$. By monotonicity, $\mathrm{p} = f(p,p,\dots,p) \le f(\mathbf{x}) \le f(q,q,\dots,q) = \mathrm{q}$. Hence $\min(\mathbf{x}) \le f(\mathbf{x}) \le \max(\mathbf{x})$. The converse: let $\min(\mathbf{x}) \le f(\mathbf{x}) \le \max(\mathbf{x})$. By taking $\mathbf{x} = (t,t,\dots,t)$, $\min(\mathbf{x}) = \max(\mathbf{x}) = f(\mathbf{x}) = t$, hence idempotency. □

In the remainder of this section we show the different aggregation functions that are considered in this study. Precisely, in subsection 1.1.1 we present t-norms, in 1.1.2 overlaps functions, in 1.1.3 copulas and finally, in the subsection 1.1.4 the Choquet integral.

[1] For an increasing (decreasing) function we do not mean a strictly increasing (decreasing) function.

1.1.1 T-norms

An important class of aggregation functions are the so-called triangular norms (t-norms for short) [Men42, KMP00, SS11].

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: $T(x, T(y,z)) = T(T(x,y), z)$;*

(T4) *Neutral Element: $T(x,1) = x$.*

We present in Table II.1 the t-norms that are considered in this thesis.

Table II.1: T-norms used in this thesis.

Name	Definition	Reference
Minimum	$T_M(x,y) = \min\{x,y\}$	[KMP00]
Algebraic Product	$T_P(x,y) = xy$	[KMP00]
Łukasiewicz	$T_Ł(x,y) = \max\{0, x+y-1\}$	[KMP00]
Drastic Product	$T_{DP}(x,y) = \begin{cases} x & \text{if } y = 1 \\ y & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$	[KMP00]
Nilpotent Minimum	$T_{NM}(x,y) = \begin{cases} \min\{x,y\} & \text{if } x+y > 1 \\ 0 & \text{otherwise} \end{cases}$	[KMP00]
Hamacher Product	$T_{HP}(x,y) = \begin{cases} 0 & \text{if } x = y = 0 \\ \frac{xy}{x+y-xy} & \text{otherwise} \end{cases}$	[KMP00]

1.1.2 Overlap functions

Overlap functions [BFM⁺10, BDBB13, DB15]) are special aggregation functions. They can be used in cases were the associativity property is not strongly required, as in image processing [JBP⁺13] or in decision making based on fuzzy preference relations [BPM⁺12]. Moreover, observe that this kind of function also plays an important role in classification problems [EGSB16, EGS+15].

Definition 5. *An aggregation function, $O : [0,1]^2 \to [0,1]$ is an overlap function if, for all $x, y \in [0,1]$, the following conditions hold:*

(O1) *O is commutative;*

(O2) *$O(x,y) = 0$ if and only if $x = 0$ or $y = 0$;*

(O3) *$O(x,y) = 1$ if and only if $x = y = 1$;*

(O4) *O is continuous.*

The overlap functions considered in this thesis are available in Table II.2.

Table II.2: Overlap functions used in this thesis.

Name	Definition	Reference
O_B	$O_B(x,y) = \min\{x\sqrt{y}, y\sqrt{x}\}$	[BFM+10, Nel99]
O_{mM}	$O_{mM}(x,y) = \min\{x,y\}\max\{x^2,y^2\}$	[DB14, DB13, DBB+16b]
Geometric Mean	$GM(x,y) = \sqrt{xy}$	[EGS+15]
Harmonic Mean	$HM(x,y) = \begin{cases} 0 & \text{if } x = 0 \text{ or } y = 0 \\ \frac{2}{\frac{1}{x}+\frac{1}{y}} & \text{otherwise} \end{cases}$	[EGS+15]
Sine	$S(x,y) = \sin\left(\frac{\pi}{2}(xy)^{\frac{1}{4}}\right)$	[EGS+15]

1.1.3 Copulas

Copulas are aggregation functions that link (two-dimensional) probability distribution functions to their one-dimensional margins, playing an important role in the theory of probabilistic metric spaces and statistics [AFS06].

Definition 6. *A bivariate function $C : [0,1]^2 \to [0,1]$ is said to be a copula if, for all $x, x', y, y' \in [0,1]$ with $x \leq x'$ and $y \leq y'$, the following conditions hold:*

(C1) $C(x,y) + C(x',y') \geq C(x,y') + C(x',y)$;

(C2) $C(x, 0) = C(0, x) = 0$;

(C3) $C(x, 1) = C(1, x) = x$.

We have to point out that t-norms are copulas as stated in the following theorem.

Theorem 1. *[GMMP09, Theorem 9.10] Let $T : [0,1]^2 \to [0,1]$ be a t-norm. Then, the following statements are equivalent:*

(i) *T is copula.*

(ii) *T satisfies the Lipschitz property with constant 1. That is, for each $x_1, x_2, y_1, y_2 \in [0,1]$,*
$|T(x_1, y_1) - T(x_2, y_2)| \leq |x_1 - x_2| + |y_1 - y_2|$

We present in Table II.3 the copulas functions that are used in this thesis.

Table II.3: Copula functions considered in this thesis.

Name	Definition	Reference
C_F	$C_F(x, y) = xy + x^2y(1 - x)(1 - y)$	[KMP00]
C_L	$C_L(x, y) = \max\{\min\{x, \frac{y}{2}\}, x + y - 1\}$	[AFS06]
C_α	$C_\alpha(x, y) = xy(1 + \alpha(1 - x)(1 - y))$, $\alpha \in [-1, 0[\cup]0, 1]$	[AFS06, LDM+15]
C_{Div}	$C_{Div}(x, y) = \frac{xy+\min\{x,y\}}{2}$	[AFS06, LSD+17b]

Observe that the t-norms T_M, T_P, T_L and T_{HP} (See Table II.1) are also copulas. The same happens with some overlaps e.g O_B, O_{mM}. On the other hand, we have some copulas that are also overlaps, as C_{Div} and C_α.

1.1.4 Choquet integral

The Choquet integral is a type of aggregation function which considers the importance of groups of criteria, offering flexibility for modeling aggregations [BPC07]. The Choquet integral, is defined with respect to a fuzzy measure [Cho54, MSM94], providing the relevance of a coalition.

In what follows, denote $N = \{1, \ldots, n\}$, for $n > 0$ and $A \subseteq N$.

Definition 7. *A function* $\mathfrak{m} : 2^N \to [0,1]$ *is said to be a fuzzy measure if, for all* $X, Y \subseteq N$*, the following conditions hold:*

($\mathfrak{m}$1) *Increasingness: if* $X \subseteq Y$*, then* $\mathfrak{m}(X) \leq \mathfrak{m}(Y)$*;*

($\mathfrak{m}$2) *Boundary conditions:* $\mathfrak{m}(\emptyset) = 0$ *and* $\mathfrak{m}(N) = 1$*.*

In the following we provide some examples of fuzzy measures:

1. Cardinality of uniform measure:
$$\mathfrak{m}(A) = \frac{|A|}{n} \tag{II.1}$$

2. Dirac's measure: For a previously fixed $i \in N$,
$$\mathfrak{m}(A) = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A. \end{cases} \tag{II.2}$$

Take an arbitrary vector of weights $(w_1, \ldots, w_n) \in [0,1]^n$ such that $\sum_{i=1}^{n} w_i = 1$.

3. Weighted mean (Wmean): Consider the following values for the fuzzy measure: $m(\{1\}) = w_1, \ldots, m(\{n\}) = w_n$. For $|A| > 1$ the fuzzy measure is,
$$\mathfrak{m}(A) = \sum_{i \in A} m(\{i\}) \tag{II.3}$$

4. Ordered Weighted Averaging (OWA): We assign the following values for the fuzzy measure. $m(\{i\}) = w_j$, with i being the j-th largest component to be aggregated, that is, it consider an OWA operator. For $|A| > 1$ the fuzzy measure is,
$$\mathfrak{m}(A) = \sum_{i \in A} m(\{i\}) \tag{II.4}$$

5. Power Measure:
$$\mathfrak{m}(A) = \left(\frac{|A|}{n}\right)^q, \text{ with } q > 0. \tag{II.5}$$

Then, the discrete Choquet integral is defined as:

Definition 8. *Let* $\mathfrak{m} : 2^N \to [0,1]$ *be a fuzzy measure. The discrete Choquet integral is the function* $\mathfrak{C}_\mathfrak{m} : [0,1]^n \to [0,1]$,

$$\mathfrak{C}_\mathfrak{m}(\vec{x}) = \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)}\right) \cdot \mathfrak{m}\left(A_{(i)}\right), \tag{II.6}$$

where $\left(x_{(1)}, \ldots, x_{(n)}\right)$ *is an increasing permutation on the input* $\vec{x}$*, that is,* $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$*, where* $x_{(0)} = 0$ *and* $A_{(i)} = \{(i), \ldots, (n)\}$ *is the subset of indices corresponding to the* $n - i + 1$ *largest components of* $\vec{x}$*.*

Observe that using the distributivity property of the product, Eq. (II.6) can be also written as:

$$\mathfrak{C}_\mathfrak{m}(\vec{x}) = \sum_{i=1}^{n} \left(x_{(i)} \cdot \mathfrak{m}\left(A_{(i)}\right) - x_{(i-1)} \cdot \mathfrak{m}\left(A_{(i)}\right)\right). \tag{II.7}$$

We call Eq. II.7 as Choquet Integral in its expanded form.

In the following, we define the basic notions about classification problems, fuzzy rule-based classification systems, evolutionary fuzzy systems and the fuzzy classifier used as base for our generalizations, that is, FARC-HD [AFAH11].

1.2 Classification Problems

A classification problem is a situation in which, based on the measured information of a certain object, it is necessary to predict the value of another categorical variable of that object. For example, consider the classification problem of the Iris flowers [FIS36]. In this problem, based on the length and width of the petal and the length and width of the sepal it is necessary to classify the type of the object (type of flower) in one of the tree possible classes namely Setosa, Virginica and Versicolour.

To tackle a classification problem, under a supervised point of view, it is necessary to establish a decision criteria, called model or classifier. To do so, the learning algorithm makes usage of correctly classified examples, known as training set. In this set, each example $e \in E$ is described by the values of N features (also called variables, characteristics or attributes) $X(e) = (e_1, \ldots, e_N)$. The inductive learning process extracts the model from the information of this training set, in order to be able to classify new examples into the known and predefined classes, $C_j \in C = (C_1, \ldots, C_M)$, where M is the number of classes of the problem.

We can say that the objective is to build a model, $D : X(e) \to C$, that should be able to predict the class of the examples having an error rate as small as possible. Once the model is obtained from the training data, it is necessary to measure the quality of the generated classifier. To do so, it is used to classify data that is not considered in the training phase, known as test data. Thus, the prediction of the test data is performed by the learned classifier, which can be also used to classify the training data in order to check if the model presents a good generalization capability. We show the steps of the supervised classification problem in Figure 1.

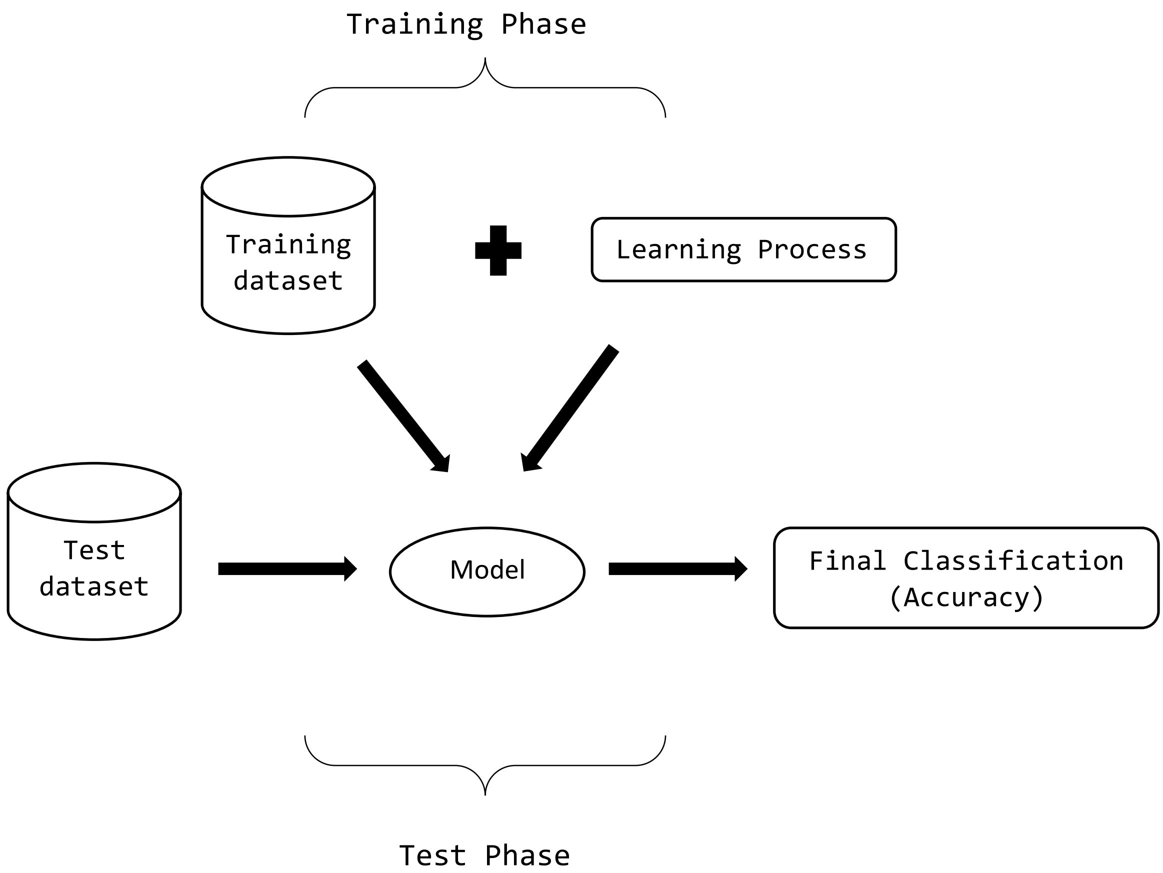

Figure 1: The supervised learning method.

There are many metrics to measure the quality of the generated classifier. They are mainly based on the usage of a confusion matrix. Given a classification problem with M classes, a confusion matrix summarizes the performance of the classification algorithm. We show in Figure 2 a confusion matrix for a binary problem, that is, having two classes. The rows of this matrix correspond with the actual class of the examples to be classified whereas the columns are their predictions. Therefore, the diagonal of the matrix represents the number of correctly classified examples of each class. Furthermore, from a confusion matrix we can obtain some measures of quality, such as:

(TP_R) True Positive Rate: Percentage of correctly classified positive instances

$$TP_R = \frac{TP}{TP + FN}$$

(TN_R) True Negative Rate: Percentage of correctly classified negative instances

$$TN_R = \frac{TN}{TN + FP}$$

(FP_R) False Positive Rate: Percentage of misclassified negative instances

$$FP_R = \frac{FP}{FP + TN}$$

(FN_R) False Negative Rate: Percentage of misclassified positive instances

$$FN_R = \frac{FN}{FN + TP}$$

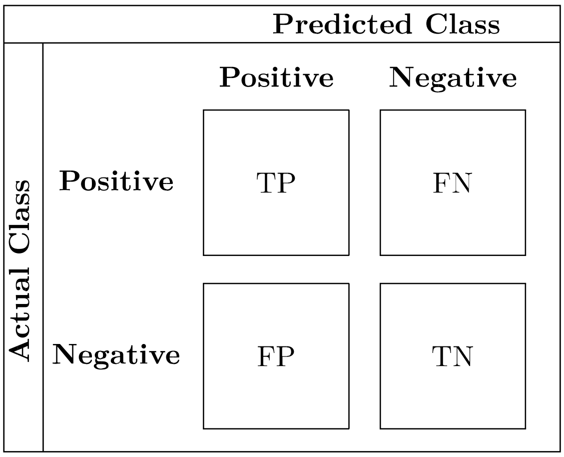

Figure 2: Confusion matrix for binary classification problems.

Using the confusion matrix and the above defined metrics it is possible to define a set of widely used performance metrics like accuracy [DHS00], the Geometric Mean [BSGR03] and Cohen's kappa [Coh60]. In what follows we describe each metric:

- Accuracy:

$$\frac{TP + TN}{TP + TN + FP + FN}$$

- Geometric Mean:

$$\sqrt{TP_R \cdot TN_R}$$

- Cohen's Kappa:

$$\frac{\text{Accuracy} - A_e}{1 - A_e},$$

Where A_e is defined as:

$$A_e = \frac{TP+FN}{TP+TN+FP+FN} \cdot \frac{TP+FP}{TP+TN+FP+FN} + \frac{FP+TN}{TP+TN+FP+FN} \cdot \frac{FN+TN}{TP+TN+FP+FN}$$

Another way to verify the performance of the generated classifier is by using a Receiver Operating Characteristic (ROC) curve. The ROC curve is a graph that plots the false positive rate, FP_R, on the X axis and the true positive rate, TP_R, on the Y axis. We provide, in Figure 3, an example of a ROC curve for a generic method[2].

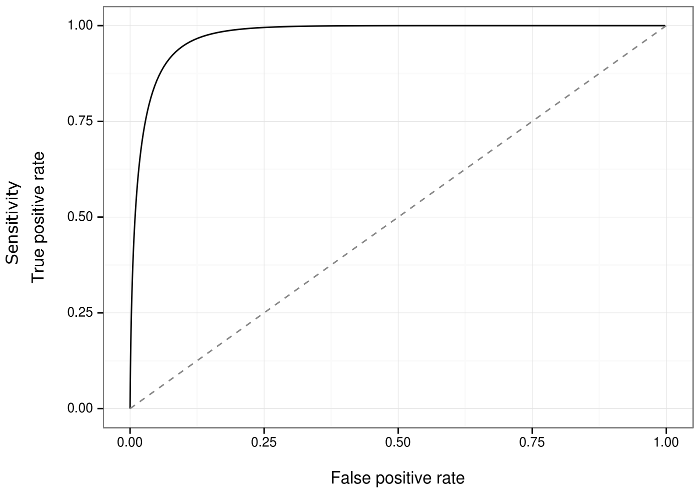

Figure 3: A ROC curve from a determined method.

Observe that the point (0,1), top left, is the perfect classifier since it classifies all positive and negative cases perfectly. The Area Under the ROC Curve (AUC) [HL05] is another measure

[2]The image used as base of this graphic is available at – https://commons.wikimedia.org/wiki/File:Threshold_roc.stack_overflow_answers.svg

of the quality of the classifier, where the larger area (close to the point (0,1)) the better the classifier is.

- ∘ AUC = $\frac{1+TP_R-FP_R}{2}$

In the literature, there can be found several methods to deal with classification problems like Support Vector Machines (SVM) [CV95], decisions trees [Qui93, BFOS84] and neural networks [GPGOF07], among many others. These methods have been applied in different fields of knowledge to help in decision making. For example, to automatically identify chest X-ray reports that support acute bacterial pneumonia [CFCH01], to categorize organic solvents with respect to their dispersibility [Sal15], to detect illegal discharges from ships [Top08] and credit scoring, using models to differentiate good applicants from bad applicants [BVGV+03].

In the next section, we describe the technique applied in this thesis to cope with classification problems.

1.2.1 Fuzzy Rule-Based Classification Systems

A rule-based system [Tun09] is composed by a set of rules in the form IF-THEN for tackling a classification problem. This rules can be expressed in the following form:

$$IF \text{ condition } THEN \text{ decision } .$$

Where the IF part is known as rule antecedent and it consists in one or more attributes (with conditions) that are linked by connectives, also called logical operators. The most used connectives are AND (denoted by $\wedge$) and OR (denoted by $\vee$). The THEN part, is known as rule consequent and it consists in the class label used to perform the prediction.

The antecedents of the rules in a rule-based system are usually categorical or numerical. Having as result only True of False. On other words, using boolean logic. This rigidity in some cases can not be the best option. For example, consider the classification of a person to fit or not fit to receive some benefit, according to its age. The considered rule can be "IF age $\geq$ 70 THEN Benefit granted". Having this benefit granted only for persons with an age $\geq$ 70. It is observed, then, that in this case a person with sixty-nine years is not classified as beneficiary of this grant, which can not be appropriated in some cases. We demonstrate this situation in Figure 4, where it is noticeable the sharp leap that the system produces when the condition is fulfilled.

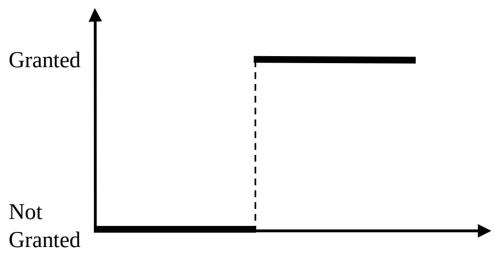

Figure 4: A rule-based system based on boolean logic.

In order to make a smoother system, the fuzzy logic is used. It was introduced by Lofti Asker Zadeh in 1965 [Zad65] in order to address issues related to uncertain, vague or inaccurate information. This theory extended the classical set theory (boolean logic) by introducing the concept of membership degrees. That is, instead of just including (True) or not (False) an element in a set, we can say that this element has a certain membership degree to the considered set.

Let U be an universe of discourse, a fuzzy set F in U is characterized by a prefixed function, called membership function, where each element of the universe U, has a membership degree to the fuzzy set F, $\mu_F : U \to [0, 1]$. The value of $\mu_F(x)$ shows the degree in which the element x, belongs to the fuzzy set F. $\mu_F(x) = 0$ represents that the element do not has pertinence to the fuzzy set, meanwhile, $\mu_F(x) = 1$ means that the element has the total pertinence and the values between 0 and 1 represents different membership degrees. In what follows we present the most common fuzzy sets used in the literature along with their graphical representation.

- A fuzzy set has a **triangular** shape if the membership function is defined in the form:

$$\mu_F(x) = \begin{cases} 0 & \text{if } x \leq a \\ \frac{x-a}{b-a}, & \text{if } a \leq x \leq b \\ \frac{c-x}{c-b}, & \text{if } b \leq x \leq c \\ 0 & \text{if } x \geq c \end{cases}$$

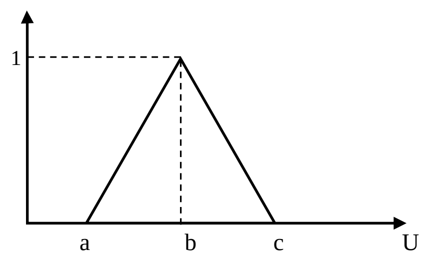

Figure 5: A traingular-shaped membership function.

- A **trapezoidal** membership function has the trapezoidal form if the membership function is defined as:

$$\mu_F(x) = \begin{cases} \frac{x-a}{b-a}, & \text{if } a \leq x \leq b \\ 1 & \text{if } b \leq x \leq c \\ \frac{d-x}{d-c}, & \text{if } c \leq x \leq d \\ 0 & \text{ottherwise} \end{cases}$$

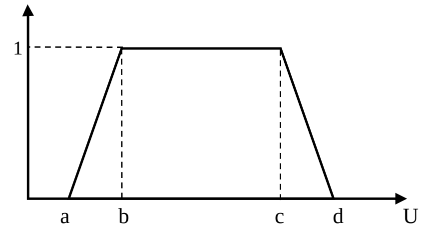

Figure 6: A trapezoidal-shaped membership function.

- ○ A **gaussian** membership function is defined by a central value, μ, and a standard deviation sigma, $\sigma > 0$. Where the σ parameter defines the thickness of the function, being a small σ the responsible for a narrower "bell".

$$\mu_F(x) = e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

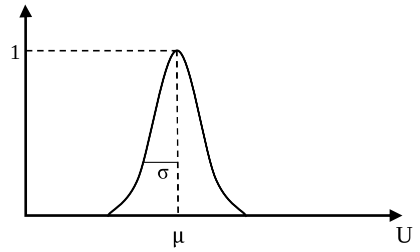

Figure 7: A gaussian-shaped membership function.

An important concept used in the fuzzy logic is the linguistic variable. It is a variable whose values are terms of a language. Take, for example, the age of a person. This variable is linguistic if its values are words, i.e young, adult or elderly. These words are known as linguistic labels and they can be modeled by fuzzy sets, which is a powerful advantage of the fuzzy logic, because the results can be easily interpreted by human beings.

Fuzzy Rule-Based Classification Systems (FRBCSs) [INN05] are an extension of the rule-based system by using fuzzy sets in the antecedents of the rules. Consider the rule, "IF age is elderly THEN Benefit Granted". Observe that the linguistic variable "elderly" can modeled by a fuzzy set. For example, a semi-trapezium, that is, only the left increasing part and the top of the trapezium as shown in Figure 8. As a result the system can also grant persons with ages between sixty and seventy, since they have a positive membership degree to the elderly set. Obviously, as in the boolean system, any person with an age superior to seventy has this benefit granted.

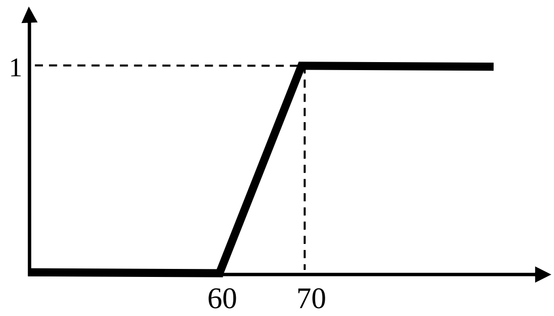

Figure 8: Example of a fuzzy set: variable elderly modeled by a semi-trapezium.

The best-known FRBCSs are the ones defined by Takagi-Sugeno-Kang (TSK) [TS85] and *Mamdani* [Mam74], which is the one used in this thesis. The standard architecture of the Mamdani method is presented in Figure 9.

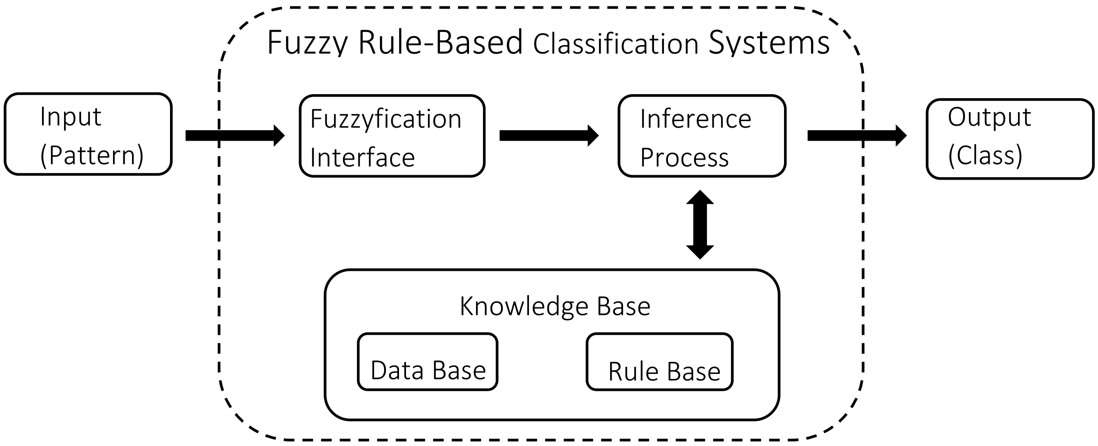

Figure 9: A structure of FRBCS of the type Mamdani.

The Knowledge Base (KB) contains the information in two different parts:

- Data Base (DB) – Stores the definition of 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). In this thesis we consider the usage of rules having the following structure:

$$\text{Rule } R_j: \text{ If } x_{p1} \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_{pn} \text{ is } A_{jn} \text{ then Class is } C_j \text{ with } RW_j, \quad \text{(II.8)}$$

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 [IN01]. Moreover, the rule weight for a class, j, is calculated by the confidence, also known as Certainty Factor [CdJH99], that is:

$$RW_j = CF_j = \frac{\sum_{x_p \in \ Class \ C_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{N} \mu_{A_j}(x_p)}, \tag{II.9}$$

where N is the number of training patterns $x_p = (x_{p1}, \ldots, x_{pm}), p = 1, 2, \ldots, N$.

The fuzzyfication interface converts the inputs (numerical 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 developed in this thesis are applied at this point and, for this reason, in the next subsection we present the steps of this inference process.

1.2.2 Fuzzy Reasoning Method

Once the knowledge has been learnt and a new example, $x_p = x_{p1}, \ldots, x_{pn}$, has to be classified, we apply the FRM [CdJH99] 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:

1. *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.

$$\mu_{A_j}(x) = T(\mu_{A_{j1}}(x_1), \ldots, \mu_{A_{jn}}(x_n)). \tag{II.10}$$

   with $j = 1, \ldots, L$.

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

$$b_j^k(x) = \mu_{A_j}(x) \cdot RW_j^k, \tag{II.11}$$

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

3. *Example classification soundness degree for all classes*: At this point, 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, $\mathbb{A}$.

$$S_k(x) = \mathbb{A}_k\left(b_1^k(x), \ldots, b_L^k(x)\right), \tag{II.12}$$

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

We present in the following three different well-known FRMs, whose difference is in the use of a different aggregation function to perform the aggregation of the information provided by the rules:

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

$$S_k(x) = \underset{R_{j_k} \in RB;}{max} b_j(x) \tag{II.13}$$

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

$$S_k(x) = \frac{\sum_{j=1}^{R_{j_k} \in RB} b_j(x)}{f_{1_{max}}}, \tag{II.14}$$

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

(c) Recently, it was proposed the usage of the Choquet integral (See Eq. II.6) to perform this aggregation.

$$S_k(x) = \sum_{j=1}^{R_{j_k} \in RB} \mathfrak{C}_{\mathfrak{m}}(b_j(x)), \tag{II.15}$$

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

4. *Classification*: The final decision is made in this step. To do so, a function $F : [0,1]^M \rightarrow \{1, \ldots, M\}$ is applied over the results obtained by example classification soundness degrees of all classes:

$$F((S_1, \ldots, S_M) = arg \max_{k=1,\ldots,M} (S_k). \tag{II.16}$$

In order to demonstrate how these FRMs work, in the next subsection we present a short example of how this aggregation process is performed.

1.2.2.1 Example of the behavior of different fuzzy reasoning methods In this subsection we perform the third step of the FRM where the local information, given by the fired rules and obtained after the second step, is aggregated for each class. To do so, we compare the FRM that uses the maximum as aggregation (WR – Eq. II.13), the additive combination (AC – Eq. II.14) and the standard Choquet integral (Eq. II.15).

Example 1. *In this example, a classification problem composed by 3 classes (C_1, C_2 and C_3) is studied. For each class, 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 introduce the information about this problem in Table II.4. 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 (columns of Table 3) three aggregations have to be computed (one for each class).*

Table II.4: Association degrees for each class.

	C_1	C_2	C_3
R_a	0.94	0.15	0.89
R_b	0.1	0.4	0.88
R_c	0.25	0.1	0.85

Since the Choquet integral is defined with respect to a fuzzy measure, in this example we consider as fuzzy measure the standard cardinality (See Equation II.1). 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
  - $WR = 0.89$
  - $AC = \frac{0.89+0.88+0.85}{2.62} = 1.0$
  - $Choquet = ((0.85 - 0) * \frac{3}{3}) + ((0.88 - 0.85) * \frac{2}{3}) + ((0.89 - 0.88) * \frac{1}{3}) = 0.87$

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

We can observe 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 II.4, this prediction may not be ideal, since that class 1 has one rule having a high compatibility whereas class 3 has three rules having high compatibilities (slightly less than that of class 1). Then, class 3 seems to be a most appropriated option. This fact is taken into account by the Choquet integral 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 we can easily notice the reason why AC is non-averaging. Observe that the result of this function for class C_3 is superior than the maximum value considered in the class. This fact does not occur for averaging functions. In the case of WR, the result is always the maximum, meanwhile for the Choquet integral the result is a value between the minimum and the maximum.

1.2.3 Evolutionary Fuzzy Systems

Among the different techniques of the Computational Intelligence, the usage of hybrid techniques has been extended notably in recent years. One of the most common hybridizations is obtained whit the combination of Fuzzy Systems with the Genetic Algorithms (GA), introducing the Evolutionary Fuzzy Sets (EFS) [Her08, CH01, CCJH01]. Basically, an EFS is a fuzzy system improved by a learning process. Having this process based on evolutionary computation e.g. GAs, genetic programming or evolutionary strategies, among others.

The design process of a FRBCS can be seen as an optimization or search problem. Due to this reason, the AGs are a satisfactory mechanism to deal with this issue. Considering that they are a global search technique with the ability to explore large search spaces requiring only a measure of performance. Thus, we can say that AGs are adequate to find almost optimal solutions in complex search spaces. Furthermore, due to its generic coding structure is easy to incorporate prior knowledge. For example, the parameters of the membership functions or the number of rules of the system. We present in Figure 10 an scheme of an evolutionary fuzzy system.

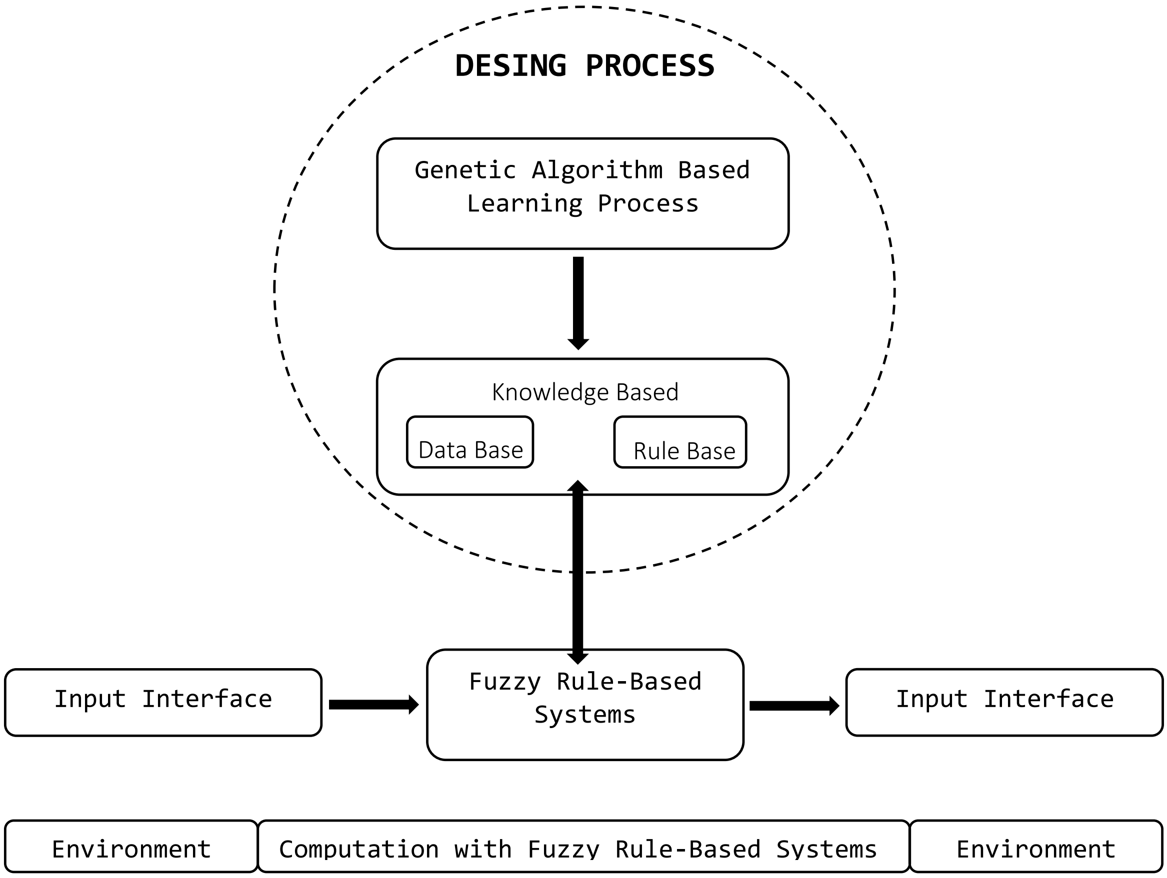

Figure 10: The scheme of an evolutionary fuzzy system [HL09].

In FRBCSs, from a point of view of the optimization, find a good KB is equivalent to codify it as a structure of parameters and find the values of these parameters that give us the optima for a determined measure of performance. The parameters of the KB define the search space and they are adapted according to a genetic representation.

The proposals of EFSs can by divided in two ways: tuning, related to the adjustment of the components of the fuzzy system and, learning, corresponding to the learning of the fuzzy system directly. In the following, we briefly describe them:

1. *Genetic Adjustment* – If exists a Knowledge Base (KB), this method apply a genetic adjustment process to improve the quality of the FRBCSs without modifying the learned RB. There are different groups of techniques within this paradigm:
   - *Genetic adjustment of the parameters of the KB.* This task is performed by a

posteriori adjustment of the parameters of the membership functions. In this way, the RB never changes, that is, it learns a FRBCS with an initial configuration in terms of the number of labels per variable and their shape. Once the fuzzy rules are learned, the parameters that define the membership functions are optimized in order to make the fuzzy rules work better. For more information see [Kar91].

- *Genetic adaptive inference systems.* The main objective of this proposal is to use parametric expressions in the inference system to obtain a better cooperation among the fuzzy rules and more precise fuzzy models, without losing the interpretability inherent in linguistic rules. This method is often called Adaptive Inference Systems. In [AFHHP07, CABFO06, CBM07] we can find proposals in this area which are focused in classification and regression.

2. *Genetic Learning* – In this process we can learn the components of the knowledge base (even including an adaptive FRM). In what follows, we describe the four groups that can be found in the genetic learning:
    - *Genetic learning of the fuzzy rules.* The majority of the approaches that have been proposed to automatically learn the KB, from numerical information, have focused on the rule base (RB) learning, using a predefined data base (DB). The usual way to define the DB demands to pick a number of linguistic terms for each linguistic variable and give it the value of the system parameters by means of an uniform distribution of the linguistic terms considering the universe of discourse of the variables. In [Thr91], it was proposed the first proposal in this area.
    - *Genetic selection of the fuzzy rules.* Once a RB has been learned, this process can be used to select fuzzy rules, in order to avoid including irrelevant, redundant and noisy fuzzy rules. In [AAFH07, CCdJH05], the authors present a methodology for combining the selection of the rules with the genetic adjustment of the parameters.
    - *Genetic learning of the Data Base.* There is another way to generate all the KB, that is, the DB and the RB. The DB generating process allows us to learn the form of membership functions and other components of the DB such as scaling functions or granularity of the diffuse partitions, among others. This process of generating the DB can use a measure to evaluate the quality of the DB, which is called apriori genetic learning of the DB. The second possibility is to consider an embedded genetic learning process where the process of generating the DB is done together with the learning of the RB. In this manner a partitioning of the learning

problem of the KB is required. In [CHV01], we can find a proposal related to the embedded genetic learning of the DB.

- *Simultaneous genetic learning of the components of the knowledge base.* This method intends to learn both components of the KB at the same time. In this way, the obtained KB could be of superior quality. However, the process is slower and hard. See [HM95] for more information.

1.2.4 The FARC-HD fuzzy classifier

In this thesis, we consider the usage of one of the most accurate and interpretable FRBCSs available in the literature, which is the Fuzzy Association Rule-based Classification model for High Dimensional problems (FARC-HD) [AFAH11]. Furthermore, this fuzzy classifier is also an evolutionary fuzzy system. It considers a genetic adjustment process that uses an evolutionary computation to select and to tune fuzzy association rules that have a good classification accuracy in the rule base. In what follows, we describe the learning process of the rules, which is composed by three parts:

1. *Fuzzy association rule extraction for classification*: In this step, an initial fuzzy rule base is obtained. To accomplish it, for each class, a search tree is constructed [AS94] and its depth is limited by a parameter (parameter depth_{max}). It lists all the possible itemsets (set of linguistic labels – items) of a class. This classifier considers five different linguistic labels to model the problem, we show an example in Figure 11.

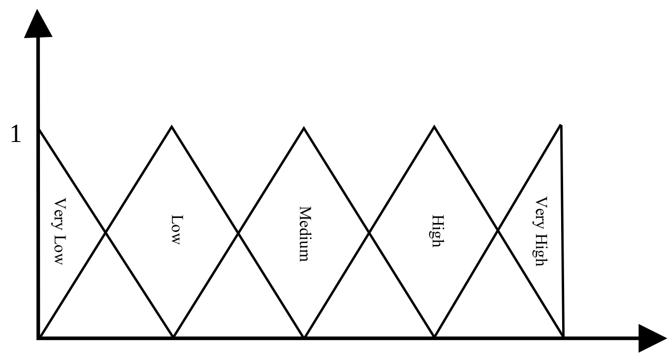

Figure 11: Membership functions used in FARC-HD.

The extraction of the rules is performed by measuring the interest of an association rule. The most common measures are the support and the confidence, being defined for fuzzy association rules as:

$$Support(A \rightarrow B) = \frac{\sum_{x_p \in T} \mu_{AB}(x_p)}{|N|} \tag{II.17}$$

$$Confidence(A \rightarrow B) = \frac{\sum_{x_p \in T} \mu_{AB}(x_p)}{\sum_{x_p \in T} \mu_{A}(x_p)} \tag{II.18}$$

where $|N|$ is the number of transactions in T, $\mu_A(x_p)$ is the matching degree of the transaction x_p with the antecedent part of the rule and $\mu_{AB}(x_p)$ is the matching degree of the transaction x_p with the antecedent and consequent of the rule.

An itemset with a support higher than the minimum support is a frequent itemset. The three is not expanded if the support of a node is inferior to the minimum support (the subsequent nodes will produce supports also smaller than the minimum support). Likewise, if a candidate itemset generates a classification rule with confidence higher than a predefined confidence, this rule has reached the quality level demanded by the user and it is, again, unnecessary to expand the tree further.

To construct the tree, in first place a tree of a single level is constructed. That is, there are as many branches (nodes) as variables times the number of labels used in each variable. The support and the confidence are measured for each node. If they need to be further expanded, the surviving branches are combined. This process is repeated until an stopping criteria is fulfilled.

The number of frequent itemsets generated, and thus the size of the tree, depends on the minimum support (Eq. II.19). The minimum support is generally calculated considering the total number of patterns in the dataset, although the number of patterns for each class (C_j) in a dataset can be different. For this reason, the algorithm determines the minimum support per each class by the distributions of the classes over the dataset.

$$MinimumSupport_{C_j} = minSup \cdot f_{C_j} \tag{II.19}$$

where $minSup$ is the minimum support determined by the expert and f_{C_j}, is the pattern ratio of the class C_j.

2. *Candidate rule prescreening*: The number of candidate fuzzy rules generated in the previous step can be large. In order to reduce the computational cost of the last step, the algorithm considers the usage of subgroup discovery to preselect the most interesting

rules from the RB that were generated in the previous stage. To do so, it applies a pattern weighting scheme [KLJ03] and assigns a counter, i, for each example that counts the number of times each example has been covered (for the selected fuzzy rules).

Weights of positive patterns covered by the rules decrease according to the following formula:

$$w(e_j, i) = \frac{1}{i+1} \tag{II.20}$$

In the first iteration, all patterns are assigned with the same weight, that is, $w(e_j, 0) = 1$. Then, the patterns covered by one or more selected rules decrease their weights in each iteration. In this way, patterns with weights that have not been decreased will have a greater chance of being covered. Covered patterns are eliminated when they have been covered more than k_t times. In each iteration of the process, the rules are sorted from the best to the worst, by a rule evaluation criteria. The best rule is selected, the covered patterns are re-weighted and this process is repeated until all patterns have been covered more than k_t times, or until there are no rules in the RB.

In order to evaluate the quality of the considered rules, the following measure is considered:

$$wWRAcc''(A \to C_j) = \frac{n''(A \cdot C_j)}{n'(C_j)} \cdot (\frac{n''(A \cdot C_j)}{n''(A)} - \frac{n(C_j)}{N}) \tag{II.21}$$

where n'' is the sum of the products of the weights of all patterns covered by their matching degrees with the antecedent part of the rule, $n''(A \cdot C_j)$ is the sum of the products of the weights of all correctly covered patterns by their matching degrees with the antecedent part of the rules, and $n(C_j)$ is the sum of the weights patterns of class C_j.

3. *Genetic rule selection and lateral tuning*: At this point, the previously generated fuzzy rules are optimized so as to enhance as much as possible the system's performance. To do so, it is considered the usage of an evolutionary model (CHC [Esh91]) to select and tune a compact set of fuzzy association rules from the RB obtained in the previous stage. This approach combines the rule selection and the lateral tuning [AAFH07]. In the following, we present the main specific features of this approach that are used by the classifier:

   (a) *Coding Scheme*: In order to combine the rule selection with the lateral tuning of the position of the membership functions, it is considered a double coding scheme:

- ◦ The first one considers a real coding scheme, since it is related to the the tuning of lateral position of the membership functions and it has as many genes as the number of linguistic labels considered in the dataset, having a range in [-0.5,0.5] (for more details see [AAFH07]).
- ◦ The second part of the chromosome uses a binary coding scheme and is related to the rule selection. It has as many genes as rules, and each gene determines if the corresponding rule is used in the FRM or not, by setting it to 1 (selected) or to 0 (not selected).

(b) *Chromosome Evaluation*: to evaluate the quality of the chromosome it is used a fitness function that computes the accuracy rate of the system and it penalizes it by the number of fuzzy rules:

$$Fitness(C) = \frac{\#Hits}{N} - \delta \cdot \frac{NR_{initial}}{NR_{initial} - NR + 1.0}, \tag{II.22}$$

where $\#Hits$ represents the number of examples correctly classified; $NR_{initial}$ is the number of candidate rules; NR is the number of selected rules and δ is a weighting percentage, given by the system expert that determines the trade-off between the accuracy and the complexity.

(c) *Initial Gene Pool*: The population is composed by 50 individuals. The first individual has his first part of the chromosome initialized by setting to 0 the value of all the genes, to perform the lateral tuning. The second part of the chromosome has the genes initialized by setting them to 1, to perform the rule selection. The remainder chromosomes are randomly generated in the corresponding ranges of the genes.

(d) *Crossover Operator*: The crossover operator will depend on the chromosome part considered (lateral tuning **(i)** or rule selection **(ii)**).

(i) The crossover is performed considering the Parent Centric BLX (PCBLX) operator [HLS03]. This operator uses the concept of neighborhood, allowing the offspring genes to be around the genes of one parent or around a zone determined by both parents.

(ii) The half uniform crossover scheme (HUX) [ES93] is applied. This operator crossover interchanges the mid of the alleles that are different in theirs parents (the genes to be crossed are randomly selected among those that are dissimilar in the parents), ensuring the maximum distance of the offspring.

The CHC approach makes usage of an incest prevention mechanism. That is, in order to avoid that similar individuals perform the crossover, two parents are crossed only if their hamming distance divided by 2 is superior to the threshold Th, which is initialized as:

$$Th = \frac{(\#Genes \cdot BITSGENE)}{4.0} \tag{II.23}$$

The Gray code is used to convert each real coded gene to binary coding with a fixed number of bits for each gene (BITSGENE).

(e) *Restarting Approach*: To increase the convergence of the algorithm, the threshold is decreased by BITSGENE if new individuals are not included in the new population. When the threshold is smaller than 0 the best chromosome is picked (elitist scheme) and all the population is reseted with random values.

(f) *Stopping Criteria*: The search process is stopped when:

(i) The maximum number of trials is reached.

(ii) A 100% is obtained as the fitness of the best individual.

2 Discussion

In this section we present a description of the different generalizations of the Choquet integral that were proposed and submitted to journals during the development of this thesis. For each contribution, we present the main acquired knowledge in two ways. The first one is related to the theoretical concepts that were developed, which are used as base for the considered generalization. The second is associated with the application of the generalization to cope with classification problems. Finally, we reinforce that all generalizations were applied in the FRM of the FARC-HD [AFAH11] fuzzy classifier (See subsection 1.2.4).

2.1 The concept of pre-aggregation functions

In this paper, we proposed the first generalization of the Choquet integral. This study was originally based on the paper presented in [BBF+13]. In this paper, the authors modified the FRM of the Chi [CYP96b] et al. algorithm by applying the Choquet integral to aggregate all available information for each class. Furthermore, considering that the Choquet integral is related to a fuzzy measure, they introduced a learning method by means of a genetic algorithm in which the most suitable fuzzy measure for each class was computed.

As the Chi algorithm it is not an state-of-the-art fuzzy classifier, in this first contribution we intended to apply this methodology in the FRM of a powerful fuzzy classifier like FARC-HD. Our goal was to improve the performance of the system by using the methodology proposed in [BBF+13]. But we wanted to go an step further, since the standard Choquet integral (See Eq. II.6) is based on the product, which is a t-norm. So, we also proposed a generalization of the Choquet integral. To do so, we replaced the product operator of the standard Choquet integral by different aggregation functions, precisely, another t-norms. In this way, the manner how the information was aggregated would be different, consequently leading into different FRMs that could present performances even more accurate.

The Choquet integral generalized by an aggregation function, T, is defined as:

Definition 9. *Let* $\mathfrak{m} : 2^N \to [0,1]$ *be a fuzzy measure and* $T, : [0,1]^2 \to [0,1]$ *be an t-norm. Taking as basis the Choquet integral, we define the function* $\mathfrak{C}_{\mathfrak{m}}^{\mathbb{T}} : [0,1]^n \to [0,n]$ *by*

$$\mathfrak{C}_{\mathfrak{m}}^{T}(x) = \sum_{i=1}^{n} T\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right), \tag{II.24}$$

where $N = \{1, \ldots, n\}$, $(x_{(1)}, \ldots, x_{(n)})$ *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 indices of* $n - i + 1$ *largest components of* x.

The first issue arrived when we realized that the simple exchange of one t-norm for another aggregation did not generate an aggregation function. For example, take as aggregation function , the minimum t-norm, $T_M(x,y) = min(x,y)$ and the cardinality measure (Eq. II.1). We want to aggregate the following two set of values: $\mathbf{x_1} = (0.05, 0.2, 0.7, 0.9)$ and $\mathbf{x_2} = (0.05, 0.1, 0.7, 0.9)$. It is easily noticed that $\mathbf{x_1} > \mathbf{x_2}$. However, $\mathfrak{C}_{\mathfrak{m}}^{T_M}(\mathbf{x_1}) = 0.7$ and $\mathfrak{C}_{\mathfrak{m}}^{T_M}(\mathbf{x_2}) = 0.8$. Therefore, the primordial condition of increasingness (monotonicity) of any aggregation function is not fulfilled by $\mathfrak{C}_{\mathfrak{m}}^{T_M}$.

Yet, we noticed 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 [BFKM15], Bustince et al. introduced the notion of directional monotonicity, which allows monotonicity to be fulfilled along (some) fixed ray. So, with this in mind, we have introduced the concept of pre-aggregation functions. These functions respect the boundary condition as any aggregation function, however, they are directional increasing. In what follows, we provide the definition of functions that are directionally increasing and pre-aggregation functions:

Definition 10. *Let* $\vec{r} = (r_1, \dots, r_n)$ *be a real* n*-dimensional vector,* $\vec{r} \neq \vec{0}$*. A function* $F : [0,1]^n \to [0,1]$ *is directionally increasing [BFKM15] with respect to* $\vec{r}$ *(*$\vec{r}$*-increasing, for short) if for all* $(x_1, \dots, x_n) \in [0,1]^n$ *and* $c > 0$ *such that* $(x_1 + cr_1, \dots, x_n + cr_n) \in [0,1]^n$ *it holds that*

$$F(x_1 + cr_1, \dots, x_n + cr_n) \geq F(x_1, \dots, x_n). \tag{II.25}$$

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

Definition 11. *[LSPD+16] Let* $\vec{r} = (r_1, \dots, r_n)$ *be a real* n*-dimensional vector,* $\vec{r} \neq \vec{0}$*. A function* $F : [0,1]^n \to [0,1]$ *is said to be an n-ary* $\vec{r}$*-pre-aggregation function if the following conditions hold:*

(PA1) F *is* $\vec{r}$*-increasing;*

(PA2) F *satisfies the boundary conditions:*
$F(0, \dots, 0) = 0$ *and*
$F(1, \dots, 1) = 1$.

In this contribution we presented three different methods to construct pre-aggregation functions, one of them is by generalizing the standard Choquet integral by t-norms. Furthermore, we have developed an important demonstration that these generalizations, are idempotent and averaging.

Then, to cope with classification problems, we considered the Choquet integral generalized by five different t-norms, namely Minimum (T_M), Łukasiewicz ($T_Ł$), Drastic product (T_{DP}), Nilpotent minimum (T_{NP}) and Hamacher Product (T_{HP}). These functions are available at Table II.1.

Observe that the standard Choquet integral is defined with respect to a fuzzy measure, m. As consequence, the generalizations of the Choquet integral also need a fuzzy measure. Therefore, for each generalization, we considered the same fuzzy measures as [BBF+13]. That is, Cardinality (Uniform), Dirac, Ordered Weighted Average (OWA), Weighted mean (Wmean) and the Power measure. This last measure was the one that achieved the best performance in [BBF+13]. It is the cardinality measure but raised to the power q. In [BBF+13] this exponent q is adapted for each class of the problem by using a genetic algorithm. Thus, in order to also consider this measure in the study, the CHC evolutionary model used by FARC-HD was adapted. To do so, for this parameter, one gene per class in the coding scheme was added. Additionally, the genes related to this parameter were set to 1.0, in the initial gene pool. In this way, at least the standard cardinality is obtained. The power measure, for class k, in this paper was defined as:

$$\mathfrak{m}_k(X) = \left(\frac{|A|}{N}\right)^{q_k}, \text{ with } q_k > 0, \tag{II.26}$$

where $|A|$ is the number of elements to be aggregated, N the total number of elements and q_k is the exponent genetically learnt to obtain the most suitable value for each class k. Consequently, this measure uses a different value for q, for each class.

The quality of the proposal was analyzed by applying these generalizations to cope with 27 classification problems. The considered datasets are available in KEEL [AFSG+09] dataset repository[3]. When comparing the different generalizations among themselves, we 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 Dirac measure achieved a bigger result with the product t-norm). Finally, the best accuracy was obtained when combining the Hamacher product with the power measure, which is similar to the results obtained in [BBF+13].

In order to evaluate the quality of this best generalization, we 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 Choquet integral.

The associated publication is available in Chapter III.1, and it is the following:

- 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 24 (2) (2016) 260 – 272.

2.2 The generalization of the extended Choquet integral by copulas

The usage of the generalizations of the Choquet integral 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, in this contribution we aimed to develop generalizations that were aggregation func-

[3] http://keel.es/

tions (satisfying the monotonicity). To do so, we considered the same methodology used in the previous study, however, in this paper we considered the Choquet integral in its expanded form, which is defined as:

$$\mathfrak{C}_{\mathfrak{m}}(\vec{x}) = \sum_{i=1}^{n} \left(x_{(i)} \cdot \mathfrak{m}\left(A_{(i)}\right) - x_{(i-1)} \cdot \mathfrak{m}\left(A_{(i)}\right) \right).$$

Then, in order to produce generalizations that are aggregation functions, we replaced the product operator by copulas. This generalization introduced the concept of Choquet-like Copula-based aggregation functions (CC-integral for short).

Definition 12. *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 Choquet-like copula-based integral with respect to* $\mathfrak{m}$ *is defined as a function* $\mathfrak{C}^C_{\mathfrak{m}} : [0,1]^n \to [0,1]$*, given, for all* $x \in [0,1]^n$*, by*

$$\mathfrak{C}^C_{\mathfrak{m}}(\vec{x}) = \sum_{i=1}^{n} C\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - C\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right), \tag{II.27}$$

where $(x_{(1)}, \ldots, x_{(n)})$ *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 indices of* $n - i + 1$ *largest components of* $\vec{x}$*.*

We demonstrated that CC-integrals are increasing functions and that they respect the boundary conditions. These are the conditions necessary for a function to be an aggregation function. Furthermore, we shown that this generalization is idempotent and averaging.

We conducted a study considering nine different functions: three t-norms (T_M, $T_Ł$ and T_{HP}), three Overlap functions (O_b, O_mM and O_α) and three copulas that are neither t-norms nor overlap functions (C_F, C_L and C_{Div}). All these generalizations were applied considering the power measure (Eq. II.26), since it was the measure that achieved the best results in [BBF$^+$13] and in our previous study.

To demonstrate the efficiency of the CC-integrals to tackle classification problems, we developed an experimental study considering 30 numerical datasets. This study was conducted in two different ways. The first one was focused on comparisons per family of copulas (t-norms, overlaps and specific copulas), in order to find the function that presented the best generalization. Then, we compared this best generalization with 1) the classical FRM of WR (considering that both functions are averaging); 2) to the standard Choquet integral and

3) the best pre-aggregation function achieved in the previous study, the one based on the Hamacher t-norm.

Analyzing the families of copulas we noticed that:

1. The minimum, T_M, is the generalization that best represents the family of t-norms.
2. The function O_α was the one that best represented the group of overlap functions.
3. For the specific copulas, the function C_F was chosen to represent this family.

Comparing these three functions (T_M, O_α and C_F) among themselves, we have noticed that there were no statistical difference among them. However, we selected the generalization by the minimum t-norm as the best CC-integral. This was due to the fact that this generalization was considered as control variable in the statistical test and also, it is the function that achieved the the biggest accuracy mean in the study.

The final comparison that aimed at showing the quality of the generalization, demonstrated that the best CC-integral was the function that achieved the biggest accuracy mean, closely followed by the best pre-aggregation and that both are superior than the standard Choquet integral and the WR. The statistical comparisons between these methods showed that there are no differences when comparing best CC-integral versus the Hamacher or the product (Choquet integral) t-norms. However, this CC-integral is statistically superior when compared to the WR method.

The associated publication is available in Chapter III.2 and it is the following one:

- 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 119 (2017) 32 – 43.

2.2.1 A proposal for tuning the alpha parameter in C_αC-integrals

As can be observed in our previous paper, we considered a generalization of the Choquet by a copula function, C_α (Eq. II.28). This function was also considered the best overlap-based generalization (O_α). Observe that it takes into consideration a parameter α in its definition. So, we presented in [LDM+15] a study considering different fixed values for this function, showing that the chosen value affects the performance of the system.

$$C_\alpha(x, y) = xy(1 + \alpha(1 - x)(1 - y)), \alpha \in [1, 0[\cup]0, 1]. \tag{II.28}$$

Then, after that, we published a contribution in the Brazilian Conference on Intelligent Systems (BRACIS) [LDB+16] in which we presented a method to assign a value to this α parameter using a genetic algorithm in order to find the value that best fits for each class. Specifically, we adapted the genetic part of the FARC-HD fuzzy classifier to also optimize the α variable used in the functions for each class of the problem. The tuning of the α parameter has as many genes as classes. These genes are encoded in different ranges according to the copula considered (the range of the parameter α can vary in each function). If the value is 0 (these functions are not defined for this value), we assign 0.1 to the parameter, since it is the best solution achieved in [LDM+15].

Meanwhile we were developing the method, it was published in the literature a fuzzy measure that was based on fuzzy sets derived from FRBCSs [PBP+16]. This method generates the fuzzy sets using the rule weights and then it applies overlap functions over them to build the fuzzy measure. Precisely, it considers, for each class the fuzzy measure constructed using the rule weights of the fuzzy rules of each class that are fired when classifying a new example. The authors of [PBP+16] demonstrated empirically that the overlap function GA_{OV} was the best one. So, we also picked this function in our study. This fuzzy measure was applied in four different CC-integrals, that are defined by functions that make usage of the α parameter.

The experimental study was performed considering 30 different numerical datasets. For each CC-integral based on a different C_α we considered a fixed (setting 0.1 to the α variable) and a genetic approach (adapting the α for each class). Then, we showed that, in general, the genetic approach presents superior accuracy mean.

To test the quality of this generalization, we picked the function that achieved the highest accuracy results and we compared it to the classical averaging FRM of WR, to the best CC-integral, to the best pre-aggregation function and to the standard Choquet integral. In this comparison we showed that this generalization presented the best performance. However, no statistical differences were found comparing these approaches among themselves.

Finally, we presented a study of the variation of this α parameter for each class, considering the function that achieved the best results. We noticed that the values are close to 0.1, which reinforces that this was a good starting value as we concluded in [LDM+15].

The associated publication is available in Chapter III.3 and it is the following one:

- G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, and H. Bustince, "A proposal for tuning the alpha parameter in $C_\alpha C$-integrals for application in fuzzy rule-based classification systems", Natural Computing (special issue BRACIS – 2016) (Accepted).

2.3 A generalization of the Choquet integral by left 0-absorbing fusion functions

The acquired knowledge from our previous studies led us to realize that the function responsible to generalize the Choquet integral is very important. At this point we have only produced generalizations with averaging characteristics. Having this in mind, we intended to generalize the standard Choquet integral by special aggregation functions, in order to produce more competitive generalizations of the Choquet integral without being limited by the maximum (non-averaging).

To achieve it, we introduced the family of left 0-absorbing aggregation functions F, which was defined as:

Definition 13. *A bivariate function* $F : [0,1]^2 \to [0,1]$ *with 0 as left annihilator element, that is, satisfying:*

(LAE) $\forall y \in [0,1] : F(0,y) = 0$*,*

is said to be left 0-absorbent.

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.

So, in this paper we replaced the product operator of the standard Choquet integral in the same way as a pre-aggregation function. However, instead of applying as aggregation, a t-norm, we applied a function, F. As consequence, we introduced the so-called C_F-integral, which is defined as:

Definition 14. *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 Choquet-like integral based on* F *with respect to* $\mathfrak{m}$*, called* C_F*-integral, is the function* $\mathfrak{C}_{\mathfrak{m}}^F : [0,1]^n \to [0,1]$*, defined, for all* $x \in [0,1]^n$*, by*

$$\mathfrak{C}_{\mathfrak{m}}^F(\vec{x}) = \min\left\{1, \sum_{i=1}^{n} F\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\}, \quad \text{(II.29)}$$

where $(x_{(1)}, \ldots, x_{(n)})$ *is an increasing permutation on the input* $\vec{x}$*, 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 indices of* $n - i + 1$ *largest components of* $\vec{x}$*.*

Observe that a C_F-integral is limited by the minimum between 1 and the aggregated value. This, is due to the fact that the function F can produce values superior to the maximum (it is non-averaging) and even to 1. Then, in order to respect the maximal boundary of the unit interval, we bound this function if the result is superior to 1. We demonstrated the conditions that the F function must fulfill for the C_F-integral to be a pre-aggregation function and to present averaging or non-averaging characteristics.

The quality of the C_F-integrals to cope with classification problems was tested considering 33 different datasets. Then, the experimental study was conducted considering generalizations with and without averaging characteristics. Furthermore, the fuzzy measure used in this study was the power measure.

Taking into account the averaging generalizations, we studied nine different C_F-integrals, based on distinct functions, namely O_α, O_B, O_{mM}, O_{Div}, C_F, C_L, F_{FBPC}, F_{BD1} and F_{NA}. This last function produced the best generalization, having the highest accuracy among the averaging ones and, also, it was adopted as control variable by the statistical method. Yet, the statistical test did not present differences between this approach and the remainder methods. But, we used this generalization, F_{NA}, as representative of the averaging C_F-integrals. After that, we compared it against FRMs with averaging operators as the WR, the standard Choquet integral, the best CC-integral and the best pre-aggregation function. The results showed that the pre-aggregation was the function that achieved the best global accuracy mean, followed closely by our C_F-integral and by the CC-min. On the other side the standard Choquet integral and the WR achieved the smallest accuracy mean.

Considering the non-averaging functions, we conducted a study using six C_F-integrals. Precisely, we applied the following six F functions: GM, HM, Sin, OR, F_{GL} and F_{NA2}. The first observation was that any non-averaging C_F-integral had an accuracy mean superior to all averaging ones. However, we picked F_{NA2} as the function representative for the non-averaging C_F-integrals. This was done because this function achieved the biggest accuracy mean in the non-averaging study and, additionally, it was considered as control variable in the statistical test. In order to support the quality of this approach, we compared the best non-averaging C_F-integral with both the FRM of AC and a FRM considering the probabilistic sum, P^* (since it is another known operator with non-averaging characteristics). The results showed that the function F_{NA2} presented a superior accuracy mean when compared against these two methods, although without statistical differences. We finished the study comparing the best non-averaging C_F-integral against all the averaging ones. We showed that the former statistically overcame all the averaging methods. Therefore, it reinforced the idea that not being limited by the maximum is a good approach, remembering that the state-of-the-art fuzzy classifier apply non averaging aggregation operators.

The associated publication is available in Chapter III.4 and it is the following:

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

2.4 Generalizations of the Choquet integral by pairs of functions F_1–F_2 under some constraints

In the previous work we have seen that not being limited by the maximum is a good approach to deal with classification problems. In addition, we have shown that the generalization of the standard Choquet integral by F functions resulted in satisfactory results. So, we wanted to improve the results obtained by that methodology. To do so, we used the concept of expanded integral Choquet, where we generalized it by two F functions (F_1 and F_2). We name this generalization $C_{F_1F_2}$-integrals and its definition is:

Definition 15. *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*

(ii) F_1 *is* $(1,0)$*-increasing,* *A $C_{F_1F_2}$-integral is defined as a function* $\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)} : [0,1]^n \to [0,1]$*, given, for all* $x \in [0,1]^n$*, by*

$$\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x}) = \min\left\{1, x_{(1)} + \sum_{i=2}^{n} F_1\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - F_2\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\}, \quad \text{(II.30)}$$

where $(x_{(1)}, \ldots, x_{(n)})$ *is an increasing permutation on the input* $\vec{x}$*, 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 indices of* $n - i + 1$ *largest components of* $\vec{x}$*.*

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. To answer it, we used the concept of dominance and subordination, that is:

(DM) F_1-Dominance (or, equivalently, F_2-Subordination): $F_1 \geq F_2$, that is: ∀ x,y ∈ [0,1]: F_1 (x,y) $\geq F_2$(x,y)

In this paper, twenty-three different functions, F, were considered. As consequence, we could combine 201 different pairs of functions that could be used as F_1 and F_2, respecting the dominance property. Therefore, we proposed a methodology to reduce the scope of the study by using the concept of Dominance and Subordination Strength degree, DSt and SSt respectively.

Definition 16. *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:*

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

To reduce the number of combinations, we categorized each function according to their associated DSt and SSt, as Low, Medium and High. We picked tree functions per considered category, as shown in table II.5. In this way, we reduced the scope of the functions to 81 different combinations.

We demonstrated that the selected functions considered as $C_{F_1F_2}$-integrals 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

Table II.5: The selected functions according to dominance/subordination strength degrees.

Strength degree	Dominance (F_1)	Subordination (F_2)
Low	T_{DP}	S
	F_{NA}	GM
	O_B	F_α
Medium	T_{HP}	T_M
	T_M	F_{NA}
	F_{IM}	T_P
High	GM	$T_Ł$
	F_{GL}	F_{BPC}
	S	T_{DP}

nor directional increasing. In fact, they are Ordered Directionally (OD) monotone functions [BBSS+17]. These functions are monotonic along different directions according to the ordinal size of the coordinates of each input.

We used the $C_{F_1F_2}$-integrals to cope with classification problems in 33 different datasets. Furthermore, we must point out that we applied these generalizations with the power measure. When analyzing the results that were obtained by the usage of these generalizations, we noticed 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.

We analyzed the performance of this proposal by comparing them against distinct state-of-the-art FRBCSs, namely: FARC-HD (See subsection 1.2.4), FURIA [HH09], IVTURS [SFBH13], 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 biggest 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.

We compared the 81 pairs of combinations considered to construct $C_{F_1F_2}$-integrals 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 we also observed that five different $C_{F_1F_2}$-integrals were considered as control variable in the statistical test in which we compared all methods, including FURIA. These functions are based on the pairs: GM–F_{BPC}, GM–$T_{Ł}$, F_{GL}–T_M, F_{GL}–F_{BPC} and GM–F_α. We must point out that 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.

The associated publication is available in Chapter III.5 and it is the following one:

- G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, H. Bustince and R. Mesiar, "Improving the performance of fuzzy rule-based classification systems based on a new non averaging generalization of CC-integrals named $C_{F_1F_2}$-integrals", IEEE Transactions on Fuzzy Systems (submitted).

2.5 The idea of generalized $C_{F_1F_2}$-integrals

In the previous study we have introduced the concept of $C_{F_1F_2}$-integrals, where we generalized the Choquet integral by two different F functions. Then, in this paper we generalized this concept by pseudo pre-aggregation pairs, introducing the so-called $gC_{F_1F_2}$-integrals. These functions are based on a solid theoretical framework. To understand it we first introduced the concept of pseudo pre-aggregation pairs (F_1, F_2).

Definition 17. *Consider two bivariate functions* $F_1, F_2 : [0,1]^2 \to [0,1]$. *The pair* (F_1, F_2) *is said to be a pseudo pre-aggregation function pair whenever the following conditions hold, for all* $y \in [0,1]$*:*

(DI) *Directional Increasingness:* F_1 *is* $(1,0)$*-increasing;*

(BC0) *Boundary Conditions for 0:*

(i) $F_1(0,y) = F_2(0,y)$ *and*

(ii) $F_1(0,1) = 0$*;*

(BC1) *Boundary Condition for 1:* $F_1(1,1) = 1$;

(DM) F_1*-Dominance (or, equivalently,* F_2*-Subordination):* $F_1 \geq F_2$.

Then, we need the concept of dimension reduction, since the $gC_{F_1F_2}$-integrals are functions well defined to aggregate elements without repetition.

Definition 18. *A* $(n \mapsto k)$*-dimension reduction function is a function* $R_{n\mapsto k} : [0,1]^n \to [0,1]^k$*, with* $k \leq n$*, defined, for all* $(x_1, \ldots, x_n) \in [0,1]^n$*, by:*

$$R_{n\mapsto k}(x_1, \ldots, x_n) = (y_1, \ldots, y_k), \tag{II.31}$$

such that:

(R1) $\{x_1, \ldots, x_n\} = \{y_1, \ldots, y_k\}$ *and*

(R2) $y_1 < \ldots < y_k$.

Note that the function $R_{n\mapsto k}$ is well defined and, in case some components of the input $\vec{x}$ are repeated, they collapse into one single value. With this definition at hand, we denote, for each $j \in K$:

$$B_j^R(\vec{x}) = \{i \in N \mid x_i = y_j\}. \tag{II.32}$$

Thus, the $gC_{F_1F_2}$-integrals considering the pseudo pre-aggregation pair is defined as:

Definition 19. *Let* $F_1, F_2 : [0,1]^2 \to [0,1]$ *be a pair of functions such that* $F_1 \geq F_2$ *(i.e.,* F_1 *dominates* F_2*) and* F_1 *is* $(1,0)$*-increasing, and consider a fuzzy measure* $\mathfrak{m} : 2^N \to [0,1]$*. Let* $R_{n\mapsto k} : [0,1]^n \to [0,1]^k$ *be a* $(n \mapsto k)$*-dimension reduction function given in Definition 18. The generalized* $C_{F_1F_2}$*-integral based on* (F_1, F_2) *with respect to* $\mathfrak{m}$ *is defined as a function* $g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)} : [0,1]^n \to [0,1]$*, given, for all* $\vec{x} \in [0,1]^n$*, by*

$$g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x}) = \min\left\{1, \sum_{j=1}^{k} F_1\left(y_j, \mathfrak{m}\left(\cup_{p=j}^{k} B_p^R(\vec{x})\right)\right) - F_2\left(y_{j-1}, \mathfrak{m}\left(\cup_{p=j}^{k} B_p^R(\vec{x})\right)\right)\right\}, \tag{II.33}$$

with the convention that $y_0 = 0$ *and* B_j^R *is as defined in Equation (II.32).*

Another concept used in this paper, is related to the pairwise increasingness, which is defined as:

Definition 20. *A pseudo pre-aggregation function pair* (F_1, F_2) *is pairwise increasing if, for all* $x, y_1, y_2 \in [0,1]$ *and* $h > 0$ *such that* $x + h \in [0,1]$*, the following condition holds:*

(PI) *If* $y_2 \leq y_1$ *then* $F_1(x, y_1) - F_2(x, y_2) \leq F_1(x+h, y_1) - F_2(x+h, y_2)$.

Using the previous concepts, we demonstrated that for any fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$ and pseudo pre-aggregation function pair (F, F) satisfying **(PI)**, $g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}$ is an aggregation function. Moreover, it is an averaging aggregation function if and only if $F(x, 1) = x$, for all $x \in [0,1]$, which is a generalization of the CC-integrals.

The associated publication is available in Chapter III.6 and it is the following one:

- G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, H. Bustince and R. Mesiar, "Generalized $C_{F_1F_2}$-integrals: from Choquet-like aggregation to ordered directionally monotone functions", Fuzzy Sets and Systems (submitted).

3 Summaries

In this section we summarize the concepts that were developed and used in this thesis. We start by introducing the summary of the theoretical developments. After that, we present a summary of the application of the generalizations to cope with classification problems.

3.1 Summary of the theoretical developments

All contributions presented in this thesis are based in solid theoretical studies. Therefore, in this subsection we summarize them in order to ease their comprehension.

To do so, we start presenting in Figure 12, a scheme of these concepts. Then, based on this scheme, we describe in the following all developed generalizations. This generalizations are divided in two groups. The first group contains Pre-Aggregation Functions (PAF), which also include Aggregation Functions (AF). The second group contains the ordered directional-monotone functions (OD-MF). Observe that OD-MF functions in a certain direction (k, ..., k) are also pre-aggregation functions, consequently, both groups have an intersection. Furthermore, it is also observable that all pre-aggregation functions are aggregation functions, however, the opposite is not true.

We started our generalizations by replacing the product t-norm of the standard Choquet integral by t-norms. Introducing the C_T-integrals [LSPD+16, BSL+16, DBB+16a, LSD+15].

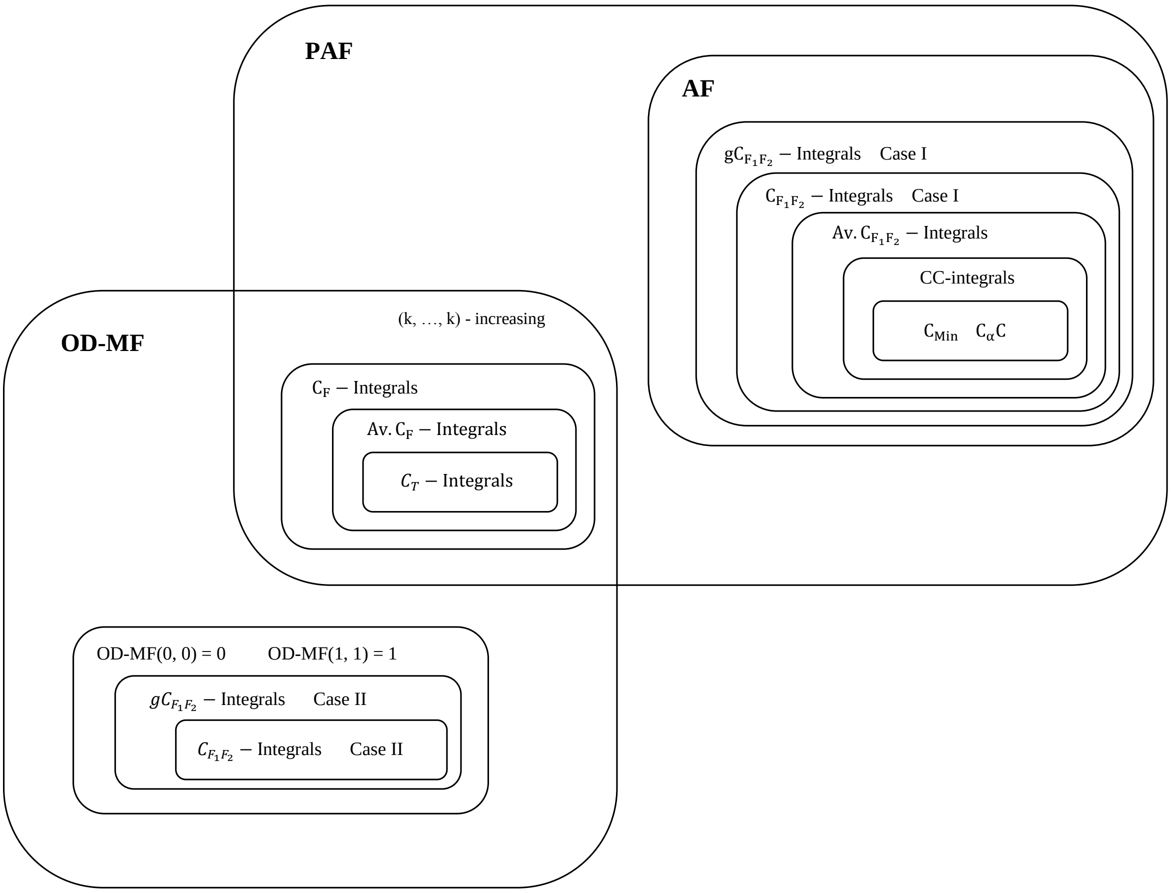

Figure 12: The scheme of the theoretical concepts provided by this thesis.

We realized that these functions are directional increasing, that is, along a specific direction, resulting in the concept of pre-aggregation functions. These functions are located in the intersection between PAF and OD-MF.

C_T-integrals were generalized by the C_F-integrals, which are based on the usage of left 0-absorbing functions F. These functions are located in the same "place" than CT integrals but are more general since they can be averaging (Av. C_F-integrals) o nor averaging (simply C_F-integrals).

Another set of functions were created when using the expanded Choquet integral, which considers the distributivity property of the product in the Choquet integral. When this expanded function was generalized by copulas, we obtained the so-called CC-integrals [LSD$^+$17b], which are aggregation functions with averaging characteristics [LSD$^+$17a]. We have constructed CC-integrals using different copulas, as: minimum (C_{Min}-integral) [DLS$^+$18] or by functions considering α parameters ($C_\alpha C$-integrals) [LDM+15, LDB+16]. We must point out that $C_\alpha C$-integrals are a specific case of CC-integrals.

Our two last new theoretical concepts are based on the expanded form of the Choquet integral and the usage of two fusion functions F_1 and F_2. The so-called $C_{F_1F_2}$-integrals, are Ordered Directional increasing functions (OD increasing) [BBSS+17], that is, increasing along different directions according to the ordinal size of the coordinates of each input. When we generalize $C_{F_1F_2}$-integrals by a pseudo pre-aggregation function pair, we obtained the $gC_{F_1F_2}$-integrals that, according to the properties of F_1 and F_2, may be either aggregation functions (Case I) with averaging properties or OD-MF functions (Case II) with non-averaging characteristics.

Observe that $C_{F_1F_2}$-integrals are a subfamily of $gC_{F_1F_2}$-integrals, in two senses. First, when F_1 and F_2 are both equal to a copula C, we have the subfamily of CC-integrals, which are aggregation functions. On the other hand, when F_1 and F_2 are different functions, the generated $C_{F_1F_2}$-integrals are a subfamily of $gC_{F_1F_2}$-integrals that are OD-MF.

3.2 Summary of the results of the generalizations of the Choquet integral in the fuzzy reasoning method

In this section, we present a summary of the results obtained in this thesis by the application of the generalizations of the Choquet integral to deal with classification problems. We must recall that the developed generalizations of the Choquet integral were applied in the FRM of the fuzzy classifier FARC-HD to cope with classification problems.

We present, in Figure 13, a summary of the obtained accuracy mean in testing by our different generalizations, according to each publication considered in this thesis in the x axis. The numbers in the x axis represent a different paper as follows:

1. C_T-integrals
2. CC-integrals
3. $C_\alpha C$-integrals
4. C_F-integrals
5. $C_{F_1F_2}$-integrals

For each paper, we present the methods used to support the quality of the proposal made in the paper. Moreover, to facilitate the visualization of the quality of each generalization, we put a horizontal line in the results obtained by the classical FRMs, where the orange line is related to the FRM of the WR and the green line is related to the FRM of the AC (labeled as FARC-HD, since this method uses it). Finally, it can be observed that the results from a same aggregation can vary from study to study. This is due to the fact that we have added more datasets in the course of this thesis, in order to perform a more complete study.

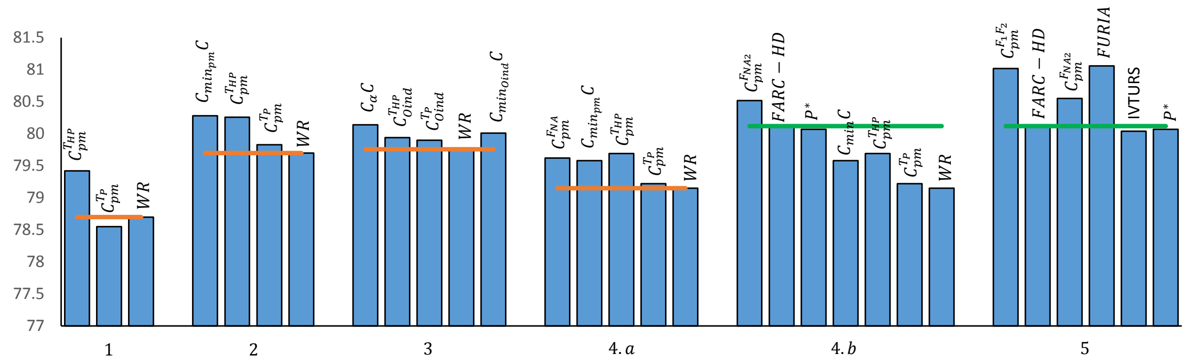

Figure 13: Results obtained in testing by different generalizations provided by this thesis.

Considering the results presented in Figure 13, in the following we summarize the results for each paper:

1) In this paper we presented the concept of C_T-integrals [LSPD+16]. The largest accuracy mean was obtained by the Hamacher t-norm, considering the power measure (See Eq. II.26) – $\mathfrak{C}_{\mathrm{pm}}^{T_{HP}}$. The quality of this approach was observed by comparing it versus two other averaging operators: 1) to the standard Choquet integral, $\mathfrak{C}_{\mathrm{pm}}^{T_P}$, and 2) the WR (See 1.2.2). In the statistical study we showed that our proposal was better than WR.

2) In this paper we introduced the concept of CC-integrals [LSD+17b]. The largest accuracy mean was obtained by the CC-integral using the minimum t-norm (which is also a copula). This CC-integral, when compared versus the classical FRM of the WR is statistically superior. However, no statistical differences were found with respect to the standard Choquet integral ($\mathfrak{C}_{\mathrm{pm}}^{T_P}$) and the best pre-aggregation, $\mathfrak{C}_{\mathrm{pm}}^{T_{HP}}$.

3) In this paper we presented the $C_\alpha C$-integrals and it uses a fuzzy measure adapted to the data to be aggregated [PBP+16]. From the obtained results we have observed that the genetic tuning of the α parameter has a beneficial effect and when we compare the best $C_\alpha C$-integral against the remaining methods, despite of the achieved superior classification rate, no statistical evidence was found.

4) In this paper we presented the concept of C_F-integrals. These functions can be averaging or non averaging.

4.a) We discovered that the best averaging generalization was achieved by the function F_{NA}. From the obtained results, we can observe that it is statistically superior than WR but there are not statistical differences versus the best pre-aggregation, $\mathfrak{C}_{\mathsf{pm}}^{T_{HP}}$.

4.b) The best non-averaging generalization was obtained by the function F_{NA2}. We showed that this generalization, $\mathfrak{C}_{\mathsf{pm}}^{F_{NA2}}$, achieved the largest accuracy mean in the study. We showed that there are differences when comparing this method against any averaging approach. This differences can be easily observed in Fig. 13. However, when comparing it against the FARC-HD fuzzy classifier or against the non-averaging operator P^*, there are no statistical difference among the methods.

5) In this paper we introduced the concept of $C_{F_1F_2}$-integrals. This generalization achieved the highest classification rate in this thesis. In order to support the quality of the method, we compared it against the state-of-the-art fuzzy classifiers. Our generalization is statistically better than IVTURS, FARC-HD and P^*. On the other hand it is equivalent to FURIA and the C_F-integrals. Furthermore, we have to stress that this methodology present five combinations statistically equivalent to FURIA.

4 Final Comments

In this section we discuss the main contributions that we have reached after each study of this thesis, the general conclusions and, finally, some open research lines.

4.1 Conclusions

The research carried out in this thesis aims to improve the quality of FRBCSs. To do so, we have developed both a theoretical and a practical part.

The generalization of the FRM of the Chi et al. algorithm by the Choquet integral led to an increase of the system's performance. In this thesis, we intended to apply this methodology in the FRM of the FARC-HD fuzzy classifier. This classifier is considered as one of the most interpretable and accurate fuzzy classifiers nowadays and we tried to enhance its performance.

We have followed an incremental research methodology. Which means that we started from generalizations with averaging characteristics (delimited by the maximum of the elements to be aggregated) and went to generalizations with non-averaging characteristics (not limited by their maximum). The mentioned incremental line, as will be seen in the next paragraphs.

To start, a first generalization was constructed by the replacement of the product of the standard Choquet integral by different t-norms. This generalization was supported by an important theoretical concept that we introduced: the pre-aggregation functions. Differently of a standard aggregation function, pre-aggregation functions are monotone along some direction, being an important contribution in the field of aggregation operators. We noticed that this first generalization, produced averaging functions and, when used to cope with classification problems, enhanced the performance of the classifier. Thus, we continued this line of work.

At this point, we were aiming at producing a generalization of the Choquet integral that resulted in aggregation functions. To do so, we used the distributivity property of the product used in the Choquet integral, which was called, the Choquet integral in its expanded form. Then, we generalized this expanded form by copulas, introducing the concept of Choquet-like Copula-Based aggregation functions (CC-integral for short). These CC-integrals are averaging generalizations of the expanded Choquet integral. However, we demonstrated that they could produce results that could compete with the pre-aggregations and could be even more accurate than the standard Choquet integral or than the classical FRM of the Winning Rule.

In the previous step, we introduced the CC-integrals. One of them was based on a copula that uses an α parameter, $C_\alpha C$-integral. Then, at this point, we also introduced a methodology to tune this parameter by adapting the evolutionary part of the FARC-HD algorithm. The $C_\alpha C$-integral is a CC-integral, consequently, is also an averaging operator. We showed in the experimental, that this approach is also able to increase the performance of the classifier.

Up to this point, we have only presented generalizations of the Choquet integral with averaging characteristics. However, the state-of-the-art fuzzy classifiers use a non-averaging approach. Thus, in order to produce even more competitive generalizations, we introduced the family of left 0-absorbing fusion functions F. Additionally, the generalization of the standard Choquet integral by a function F introduced the concept of C_F-integrals. These functions are averaging or non-averaging, it depends on the considered function F that generalizes the Choquet integral. We demonstrated that the averaging C_F-integrals present good results when compared with another averaging operators. Furthermore, the non-averaging C_F-integrals were comparable with classical non-averaging operators. Finally, we showed that the non-averaging C_F-integrals statistically overcame the averaging ones, reinforcing the idea that not being limited by the maximum is a good option to tackle classification problems.

The summit of our generalizations was reached when we generalized the extended Choquet integral by two functions F_1 and F_2. The result of this generalization was named $C_{F_1F_2}$-integrals. These functions are Ordered Directional increasing functions (OD increasing) and, therefore, represent a different level of aggregation operators. We showed a methodology to select different functions as F_1 and F_2, based on the concept of dominance and strength degrees. Then, for the considered $C_{F_1F_2}$-integrals we demonstrated that in five different combinations of F_1 and F_2 we produced generalizations that are equivalent, or even superior, than classical fuzzy classifiers like FARC-HD, IVTURS and FURIA.

Finally, we draw the main conclusions of the thesis related to our initial objectives:

- We have applied the developed generalizations of the Choquet integral in the FRM of the FARC-HD fuzzy classifier, which is one of the most accurate and interpretable fuzzy classifier nowadays, and we enhanced its performance. We must point out, that this was the main objective of this thesis.
- The generalizations were constructed by replacing the product operator of the original Choquet integral and its extended form by different aggregation functions. This allowed us to define important concepts in the field of aggregation operators, like:

  C_T-integrals: Generalizations of the original Choquet integral by t-norms.

  CC-integrals: Generalizations of the extended Choquet integral by copulas.

  C_F-integrals: Generalizations of the standard Choquet integral by functions F. Where F, is a family of fusion functions that we have introduced.

  $C_{F_1F_2}$-integrals: Generalizations of the extended Choquet integral by two functions, F_1 and F_2.

- We have introduced a methodology to adapt the evolutionary part of the FARC-HD to learn a fuzzy measure which is adapted for each class of the problem. In this way we increased the performance of the system.

- The developed generalizations presented both averaging and non-averaging behavior. In this way, we compared them against classical FRMs presented in the literature and state-of-the-art fuzzy classifiers.
  - When comparing our averaging methods against the classical averaging FRM of WR we have that most of the developed generalizations are superior in terms of performance.
  - Comparing the generalizations having non-averaging characteristics versus state-of-the-art fuzzy classifiers, we have that the generalizations are equivalent, or even superior, than the considered methods.

4.2 Future research lines

In this subsection we describe some open research lines based on the methodologies that have proposed in this thesis.

Generalizations of the Choquet integral using different fuzzy measures

The usage of the Choquet integral to deal with classification problems was firstly introduced in [BBF+13]. Furthermore, in this study, the authors have introduced a fuzzy measure that adapts itself for each class of the problem. In our first study [LSPD+16] we considered generalizations that were combined with the same measures used in [BBF+13]. The achieved results confirmed that the proposed fuzzy measure was superior to the remainder, for this reason in this thesis we basically consider the same fuzzy measure in all papers.

We have used a different fuzzy measure only in the paper related to the tunning of the α parameter. In that study we have considered as a measure a method that is built based on an overlap index, which achieved satisfactory results. However, when the different generalizations of the Choquet integral proposed in this thesis were combined with this method, poorer results were obtained when compared against the generalizations of the Choquet integral using the power measure.

Both fuzzy measures are calculated in execution time, then, it is possible to explore even more the potency of the fired fuzzy rules if we had a fuzzy measure that is able to represent in a better way the relationship among all the rules of each class. Therefore, having a different fuzzy measure per class as in this thesis. The open problem is how to learn these fuzzy measures in order to use them when classifying new examples. A possible solution could be the usage of the Sugeno fuzzy measure [Sug74], which is a fuzzy measure that only needs to define the values for the sets having an unique element and the remainder are automatically obtained. Therefore, we could use an evolutionary algorithm to learn these values in order to obtain the best possible fuzzy measures for the faced problem.

Analyzing the behavior of the generalizations of the Choquet integral

As can be noticed in the published papers, we have applied different generalizations of the Choquet integral in the FRM of the FARC-HD fuzzy classifier. This generalizations were used to tackle different classification problems by aggregating the information provided by all fired rules in the FRM.

From the obtained results, it is noticeable that some generalizations achieved a better performance than others in determined datasets. Thus, an important open research line is related to a pre-definition of the characteristics of the dataset that allows one to determine in advance the most appropriate generalization to face it. We have made preliminary studies in this field in [LSD+17a, LSD+18b] by analyzing the behavior of a CC-integral in the FRM and by using data complexity measures, respectively. However, despite the interesting results in both cases, we must dig deeper in this subject.

Applying the generalizations with different classifiers

The generalizations of the Choquet integral developed in this thesis were always applied in the FRM of the FARC-HD [AFAH11] fuzzy classifier. However, observe that the first proposal of the usage of the standard Choquet integral in a FRM [BBF+13] was applied in the fuzzy classifier of Chi et al [CYP96b], and, also enhanced the performance of the algorithm.

Then, we could select the different approaches developed in this thesis and apply them in different fuzzy classifiers. In this way, we would research if these generalizations can also improve their performance, which could support the hypothesis that the Choquet integrals and their generalizations are a good option to improve the quality of all the FRBCSs.

The construction of a model based on ensemble methods

There are different techniques to improve the performance of the classification systems. For example, Evolutionary Fuzzy Systems (See subsection 1.2.3) and Ensembles [OM99]. Ensembles of classifiers usually enhance the performance of single classifiers by inducing several classifiers and combining them so that they outperform all the models conforming it [GFB+12].

Therefore, the basic idea of an ensemble method is to construct several classifiers from the original data and, then, aggregate their predictions when unknown instances are presented. It is observable that different generalizations of the Choquet integral achieved different performances in the same dataset. Thus, another open research line is related to the construction of an ensemble that considers different FRMs, that is, different generalizations of the Choquet integral. Then we could made the final decision based on the outputs of the different generalizations. The open problem is to know if this methodology offers diversity enough as it is a key factor when constructing ensembles of classifiers [Kun05]. If this combination is successful we could improve even more the behavior of FRBCSs to deal with classification problems.

The problem of imbalanced datasets

In a classification problem, whenever the classes are not represented equally, we have the so-called, imbalanced data problem [FGH11]. That is, when the number of instances which represent one class is smaller than the ones from the remainder class.

To ease the explanation of why the usage of generalizations of the Choquet integral could be an interesting research line to cope with this kind of problem, consider the following situation: let the elements to be aggregated for class 1, $C_1 = [0.75, 0.8]$ and the elements of class 2, $C_2 = [0.3, 0.4, 0.5, 0.6, 0.7]$. We highlight that class 1 has less examples (positive class) than class 2 (negative class). As more rules of the negative class are built in runtime it is easier to have many rules fired for class 2 which can alter the classification of the positive class.

For instance, if we aggregated these values using AC the result would be $C_1 = \frac{0.75+0.8}{2.5} = 0.62$, and, $C_2 = \frac{0.3+0.4+0.5+0.6+0.7}{2.5} = 1$. Although the values for C_1 are better, when performing the aggregation, C_2 becomes the chosen one, for the simple reason that it has more fired rules. However, considering the averaging generalizations of the Choquet integral presented in this thesis. The result for C_1 would be a value between $[0.75, 0.8]$, and for C_2 a value between $[0.3, 0.7]$. It could result in a fairness representation for the problem of imbalanced data. For this reason, we consider it an interesting future research line as it may enhance the results of FRBCSs in imbalanced classification problems.

Extension to interval-valued fuzzy sets

Interval-Valued Fuzzy Sets (IVFSs) [Sam75] have proved to be a suitable tool to represent the system uncertainties and the ignorance in the definition of the linguistic fuzzy terms. An IVFS provides an interval, instead of a single number, as the membership degree of each element to this set. To ease this comprehension, we present in Figure 14, an example of a triangular shaped interval-valued fuzzy set[4]. Where $A(u_i) = [\overline{A}(u_i), \overline{A}(u_i)]$, is the membership degree of the element $u_i \in U$ to the set inferior $(\overline{A}(u_i))$ and superior $(\overline{A}(u_i))$.

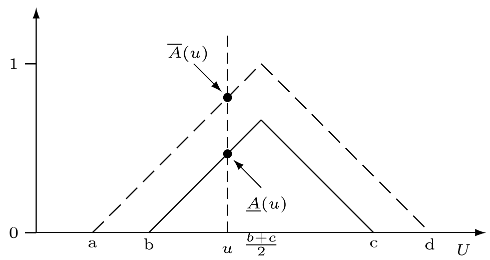

Figure 14: An example of an interval-valued fuzzy set.

The Interval-Valued fuzzy reasoning method (IV-FRM) with TUning and Rule Selection (IVTURS) [SFBH13] is an state-of-the-art FRBCS that works with IVFSs. It achieves good results dealing with classification problems. Therefore, a natural step of this thesis, and an open research line, is the adaption of the presented generalizations to the context of IVFSs, to take into account aggregations on intervals. In this way, we will be able to analyze if the improvement produced when working with fuzzy sets is also materialized in the interval-valued context.

[4] Image available at [SD11].

5 Introducción y Conclusión (Versión en español)

5.1 Introducción

Los seres humanos afrontan problemas de clasificación desde el principio. La distinción sobre si las frutas, semillas y raíces son comestibles o no, es un ejemplo simple. Este tipo de decisión se realiza en base al conocimiento adquirido en el transcurso del tiempo.

Se puede decir que un problema de clasificación es un proceso en que un dato se etiqueta (clasifica) en base a sus características. Existen distintos niveles de complejidad en los problemas de clasificación, desde los más sencillos, como la definición de la raza de un perro basada en las características del perro. Hasta complejas, como la clasificación de si un paciente tiene cáncer utilizando la información de los test sanguíneos. Una persona especialista en el problema podría abordar determinados problemas de clasificación, sin embargo, esta persona puede ser difícil de encontrar y su proceso puede ser lento, caro y hasta impreciso.

La utilización de un sistema de clasificación automático puede ser una buena opción en el proceso de clasificación. Nótese que el sistema no puede reemplazar el conocimiento de un experto, pero el experto puede usar este sistema como una importante fuente de información en el proceso de la toma de decisión.

En la literatura, los problemas de clasificación [Alp10, DHS00] son un campo de investigación de la minería de datos [TSK05]. Los problemas de clasificación son abordados de dos maneras distintas. La primera se conoce como aprendizaje supervisado. Este paradigma de aprendizaje genera una función (clasificador) a partir de los datos etiquetados (clases) que están disponibles y son conocidos. Por tanto, cuando se necesita clasificar un nuevo ejemplo, esta función es la responsable de hacer la predicción. El segundo método utiliza datos sin etiquetar (sin clases conocidas), intentando extraer las relaciones de dichos datos. Este paradigma de aprendizaje es conocido como aprendizaje no supervisado.

En esta tesis nos centramos en aprendizaje supervisado para tratar problemas de clasificación. En la literatura se pueden encontrar diferentes maneras de tratar con estos problemas como las máquinas de soporte vectorial [CV95], los árboles de decisión [Qui93, BFOS84] y las redes neuronales [GPGOF07], entre otras muchas. En esta tesis vamos a trabajar con Sistemas de Clasificación Basados en Reglas Difusas (SCBRDs) [INN05]. Éstos sistemas proporcionan al usuario modelos interpretables mediante el uso de etiquetas lingüísticas (como alto, medio o bajo) [Zad75] en sus reglas. Además obtienen resultados precisos y por ello los SCBRDs han sido utilizado en numerosos problemas como salud [Uno11, SH09], seguridad [GSP+14, VRT+15], economía [SBH+15] o alimentación [SFB+16, GS15], entre otros.

Motivación

Un papel importante en cualquier SCBRDs es el que juega el Método de Razonamiento Difuso (MRD) [CdJH98, CdJH99]. Este método es el responsable de clasificar nuevos ejemplos. Para ello, utiliza la información disponible en la base de reglas y en la base de datos. Para determinar la clase del ejemplo a clasificar, el MRD usa una función para agregar la información de cada clase dada por las reglas difusas compatibles con el ejemplo.

Un MRD muy utilizado considera como método de agregación la función máximo. Utilizando este operador de agregación el MRD selecciona la mejor regla difusa disparada para cada clase, ya que esa regla tiene la mayor compatibilidad con el ejemplo [CYP96a, GP98, INYT94]. El problema de este método de razonamiento es que la información proporcionada por el resto de las reglas es ignorada. Además, el operador de agregación es promedio, es decir, el resultado obtenido esta en el rango delimitado por el mínimo y el máximo de los valores a agregar.

Para evitar este problema, se utiliza el MRD que aplica la suma normalizada [CdJH98, CdJH99] para agregar de la información dada por las reglas disparadas. De esta manera, se utiliza toda la información disponible para cada clase. Sin embargo, esta función se sale del rango mínimo–máximo, y por lo tanto, es considerada como no promedio.

En [BBF+13] los autores introdujeron un MRD que considera la utilización de la integral Choquet [Cho54], que es un operador promedio. Este método combina las características buenas de los dos MRDs anteriores porque considera un operador promedio sin ignorar la información ofrecida por todas las reglas disparadas del sistema.

Teniendo en cuenta las consideraciones anteriores, en este trabajo de tesis proponemos una metodología que cambia la fase de agregación del MRD. Específicamente, consideramos la aplicación de diferentes generalizaciones de la integral Choquet, soportadas por un estudio teórico sólido. Comenzamos por generalizaciones con características promedio e iremos hasta generalizaciones sin esa característica para producir funciones que sean capaces de ser competitivas contra SCBRDs del estado del arte.

5.2 Conclusión

La investigación realizada en esta memoria tiene como objetivo aumentar la calidad de los SCDBRs. Para conseguirlo, hemos desarrollado una parte teórica y una aplicada.

La adaptación del MRD del algoritmo de Chi et al. en la que se utiliza la integral Choquet permitió mejorar el rendimiento del sistema. En esta tesis, intentamos aplicar esta metodología en el MRD del clasificador difuso FARC-HD. Este clasificador es considerado como uno de los más interpretables y precisos de la actualidad y hemos intentando mejorar su calidad.

Hemos seguido una metodología de investigación incremental. Lo que significa que empezamos con generalizaciones con características promedio (delimitadas por el máximo de los elementos a agregar) y acabamos con generalizaciones sin esa característica.

La primera generalización fue construida reemplazando el producto de la integral Choquet original por diferentes t-normas. Estas generalizaciones fueron soportadas por un importante concepto teórico que definimos: las funciones de pre-agregación. A diferencia de una agregación normal, las pre-agregaciones son monótonas en una determinada dirección, siendo una contribución importante en el campo de los operadores de agregación. Esta primera generalización genera funciones promedio y, cuando son usadas para tratar problemas de clasificación, aumentan el rendimiento del clasificador.

En este punto, queríamos obtener una generalización de la Choquet integral que fuera una agregación. Para ello, utilizamos la propriedad de la distributividad del producto, usado por la integral Choquet (integral Choquet en su forma expandida). Reemplazamos el producto de la integral Choquet en su forma expandida por cópulas, introduciendo el concepto de CC-integrales. Las CC-integrales son generalizaciones promedio de la integral Choquet extendida. Mostramos que las CC-integrales pueden producir resultados competitivos con respecto a las pre-agregaciones e incluso más precisos que la integral Choquet original y el MRD clásico de la regla ganadora.

En el paso anterior introdujimos las CC-integrales. Una de ellas estaba basada en una cópula que hacia uso de un parámetro α, $C_\alpha C$-integral. Para intentar mejorar el rendimiento de este nuevo concepto creamos una metodología para ajustar el valor de dicho parámetro. Para ello adaptamos el algoritmo evolutivo de FARC-HD. Debemos destacar que una $C_\alpha C$-integral es una CC-integral y, como consecuencia, también es un operador promedio. En el estudio experimental mostramos que este método también permite mejorar el rendimiento del

SCBRDs.

Hasta este punto, solo habíamos presentado generalizaciones de la integral Choquet con características promedio. Sin embargo, los clasificadores difusos del estado del arte usan agregaciones sin características promedio. Por lo tanto, para producir generalizaciones aún mas competitivas, definimos la familia de funciones de fusión 0-absorventes por la izquierda, F. Además, la generalización de la integral Choquet por una función F permite definir el concepto de C_F-integrales. Estas funciones son promedio o no, en función de la función F que generaliza la integral Choquet. Demostramos que las C_F-integrales promedio presentan buenos resultados cuando las comparamos contra otros operadores promedio. Además, las C_F-integrales no promedio son comparables con operadores de agregación clásicos no promedio. Finalmente, mostramos que las C_F-integrales no promedio superan estadisticamente todas las promedio, reforzando la idea de que no obtener un resultado limitado por el máximo es una buena opción para abordar problemas de clasificación.

El culmen de nuestras generalizaciones lo alcanzamos cuando generalizamos la Choquet integral extendida por dos funciones, F_1 y F_2. El resultado de esta generalización lo hemos llamado $C_{F_1F_2}$-integrales. Estas funciones tienen un crecimiento ordenado en una dirección (OD-crecimiento) y, por lo tanto, representan un nivel diferente de operadores de agregación. Hemos creado una metodología para seleccionar la función que actúe como F_1 y la que lo haga como F_2, utilizando el concepto de grados de fuerza y de dominancia. Para las $C_{F_1F_2}$-integrales consideradas en el estudio experimental, hemos mostrado que cinco combinaciones diferentes de F_1 y F_2 producen generalizaciones que son equivalentes, o incluso mejores, que clasificadores difusos del estado del arte como FARC-HD, IVTURS o FURIA.

Finalmente vamos a realizar las conclusiones generales de la tesis en relación a los objetivos que nos habíamos marcado al comienzo de la misma:

- Hemos aplicado el desarrollo de las generalizaciones de la integral Choquet en el MRD del clasificador difuso FARC-HD, que es uno de los clasificadores difusos más precisos e interpretables de la actualidad, y hemos mejorado su calidad. Destacamos que éste era el objetivo principal de la tesis.
- Hemos construido las generalizaciones cambiando el operador producto en la Choquet integral original, y en su forma expandida, por diferentes funciones de agregación. Esto nos permitió definir importantes conceptos en el campo de los operadores de agregación como:

C_T-integrales: Generalizaciones de la integral Choquet por t-normas.

CC-integrales: Generalizaciones de la integral Choquet expandida por cópulas.

C_F-integrales: Generalizaciones de la integral Choquet estándar por funciones F. Donde F, es una familia de funciones de fusión que hemos creado en la tesis.

$C_{F_1F_2}$-integrales: Generalizaciones de la integral Choquet expandida por dos funciones, F_1 and F_2.

- Hemos introducido una metodología para adaptar el modelo evolutivo de FARC-HD para aprender una medida difusa, de forma que se adapta a cada clase del problema. De esta manera, hemos aumentado el rendimiento del sistema.
- Las generalizaciones desarrolladas presentan comportamientos promedio y no promedio. De esta forma, las hemos comparado contra MRDs clásicos publicados en la literatura y clasificadores difusos estados del arte.
  - Al comparar nuestros métodos promedio contra el MRD clásico de la regla ganadora, en la mayoría de las ocasiones nuestras propuestas dan un rendimiento superior
  - Comparando las generalizaciones con características no promedio contra clasificadores difusos estado del arte, tenemos que nuestros métodos son equivalentes, o incluso mejores, a los métodos considerados.

Chapter III

Publications: published, accepted and submitted papers

1 Pre-aggregation Functions: Construction and an application

Related publication:

- 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 24 (2) (2016) 260 – 272.
  - Journal: IEEE Transactions on fuzzy Systems
  - Status: Published
  - Impact Factor (JCR 2016): 7.671
  - Knowledge Area:
    - Artificial Intelligence: Ranking 4/133 (Q1)
    - Computer Science: Ranking 4/133 (Q1)
    - Engineering, Electrical & Electronic: Ranking 9/260 (Q1)

Artículo eliminado por restricciones de derechos de autor

Publicado en:
G. Lucca et al., "Preaggregation Functions: Construction and an Application," in *IEEE Transactions on Fuzzy Systems*, vol. 24, no. 2, pp. 260-272, April 2016.
doi: 10.1109/TFUZZ.2015.2453020

2 CC-integrals: Choquet-like Copula-based aggregation functions and its application in fuzzy rule-based classification systems

Related publication:

- 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 119 (2017) 32 – 43.
  - Journal: Knowledge-Based Systems
  - Status: Published
  - Impact Factor (JCR 2016): 4,627
  - Knowledge Area:
    - Computer Science: Ranking 16/133 (Q1)
    - Artificial Intelligence: Ranking 16/133 (Q1)

Knowledge-Based Systems 119 (2017) 32–43

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

CC-integrals: Choquet-like Copula-based aggregation functions and its application in fuzzy rule-based classification systems

Giancarlo Lucca [a,d], José Antonio Sanz [a,c], Graçaliz Pereira Dimuro [a,d,*], Benjamín Bedregal [e], Maria José Asiain [b], Mikel Elkano [a,c], Humberto Bustince [a,c]

[a] Departamento de Automática y Computación, Universidad Publica de Navarra, Campus Arrosadía s/n, P.O. Box 31006, Pamplona, Spain
[b] Departamento de Matemáticas, Universidad Publica de Navarra, Campus Arrosadía s/n, P.O. Box 31006, Pamplona, Spain
[c] Institute of Smart Cities, Universidad Publica de Navarra, Campus Arrosadía, Navarra, 31006, Spain
[d] Centro de Ciências Computacionais, Universidade Federal do Rio Grande, Av. Itália km 08, Campus Carreiros, Rio Grande, 96201-900, Brazil
[e] Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte, Campus Universitário, Natal, 59072-970, Brazil

ARTICLE INFO

Article history:
Received 3 May 2016
Revised 21 October 2016
Accepted 1 December 2016
Available online 2 December 2016

Keywords:
Aggregation functions
Choquet integral
Copula
t-norm
Overlap function
Fuzzy rule-based classification systems
Fuzzy reasoning method

ABSTRACT

This paper introduces the concept of Choquet-like Copula-based aggregation function (CC-integral) and its application in fuzzy rule-based classification systems. The standard Choquet integral is expanded by distributing the product operation. Then, the product operation is generalized by a copula. Unlike the generalization of the Choquet integral by t-norms using its standard form (i.e., without distributing the product operator), which results in a pre-aggregation function, the CC-integral satisfies all the conditions required for an aggregation function. We build some examples of CC-integrals considering different examples of copulas, including t-norms, overlap functions and copulas that are neither t-norms nor overlap functions. We show that the CC-integral based on the minimum t-norm, when applied in fuzzy rule-based classification systems, obtains a performance that is, with a high level of confidence, better than that which adopts the winning rule (maximum). We concluded that the behavior of CC-integral is similar to the best Choquet-like pre-aggregation function. Consequently, the CC-integrals introduced in this paper can enlarge the scope of the applications by offering new possibilities for defining fuzzy reasoning methods with a similar gain in performance.

1. Introduction

Whenever we face the task of organizing instances in various predefined categories according to their different features, we have the so-called classification problem [1]. In general, one applies a learning algorithm, which, using the available information about the instances, is able to learn a decision function, called classifier, which is used to perform the classification for future instances.

Classification problems are present in many different real-world problems. For example, Patidar et al. [2] developed a method for diagnosis of coronary artery disease classifying heart rate signals, Galar et al. [3,4] presented a survey in fingerprint classification, Huang and Lin [5] developed a system to classify multiple harmonic sources (which distort the original frequency) in a power quality problem and finally, a study using different classifiers in earthquake predictions is done in [6].

Fuzzy Rule-Based Classification Systems (FRBCSs) [7] are widely used to deal with classification problems, since they usually present a good performance besides providing an interpretable model. The usage of linguistic terms for modeling the problem domain allows these systems to be easily applied and understood by the final users of real world applications [8,9]. Moreover, FRBCSs can combine information coming from different sources, that is, expert knowledge, mathematical models, data bases, or empirical measures [9].

An important issue in any FRBCS is the considered Fuzzy Reasoning Method (FRM) [10], which consists in an inference procedure that uses the information stored in the knowledge base to determine the class in which new instances will be classified. To do so, in first place, one computes the compatibility between the example to be classified and each rule, which provides local information. Then, this local information is aggregated, obtaining the global information associated with each known class. Finally, the class determined for the new example is the one presenting the maximum global information.

* Corresponding author.
E-mail addresses: lucca.112793@e.unavarra.es (G. Lucca), joseantonio.sanz@unavarra.es (J.A. Sanz), gracalizdimuro@furg.br, gracaliz@gmail.com (G.P. Dimuro), bedregal@dimap.ufrn.br (B. Bedregal), asiain@unavarra.es (M.J. Asiain), elkano.mikel@gmail.com (M. Elkano), bustince@unavarra.es (H. Bustince).

http://dx.doi.org/10.1016/j.knosys.2016.12.004
0950-7051/

The FRM of the Winning Rule (WR), which uses the maximum as aggregation function to obtain the global information, is widely used in the literature [11]. However, this method considers, for each class, just the information given by the fuzzy rule with the highest compatibility with the example and, thus, the available information provided by all the other fired fuzzy rules is ignored. In order to avoid this shortcoming using averaging operators, Barrenechea et al. [12] introduced a FRM that aggregates the information given by all the fired fuzzy rules using the Choquet integral [13]. Specifically, they achieved the best results when using the Choquet integral considering the power measure and applying the CHC evolutionary algorithm [14] to learn a different fuzzy measure for each class of the problem.

Aimed at improving the results in [12], Lucca et al. [15] introduced the concept of pre-aggregation function. These functions satisfy the same boundary conditions of an aggregation function, but are just directionally increasing [16], that is, they are increasing along some specific ray (direction), but not for all directions. Furthermore, they presented the Choquet-like construction method of pre-aggregation functions, which are built by replacing the product operation of the Choquet integral by other t-norms. Finally, they applied these Choquet-like pre-aggregation functions in the FRM of the FRBCSs, showing that the use of Hamacher Product t-norm along with the power measure, outperformed the performance of the method by Barrenechea et al. [12] as well as that of the WR.

From the methods introduced in [12,15], the questions that arise are: “Is there a way to build a Choquet-like integral that is based on a more general aggregation than the product operator, but that yields to an aggregation function instead of a pre-aggregation function? If so, is the FRBCS using this new method able to achieve a better or a similar performance than that of the most accurate Choquet-like pre-aggregation function?”

To answer the questions above, the objectives of this paper are:

1. To introduce the notion of Choquet-like Copula-based aggregation functions, called CC-integrals;
2. To define different types of CC-integrals, using t-norms [17], overlap functions [18–21], and copulas [22] that are neither t-norms nor overlap functions;
3. To analyze the behavior of CC-integrals when applied in FRBCSs.

To test the quality of the FRBCS in which the FRM uses CC-integrals we consider 30 real-world problems that are publicly available in the KEEL database repository [23]. The performance of the classifiers is measured using the standard accuracy rate and the results are supported by appropriate statistical tests [24–26]. Specifically, we analyze the behavior of the different CC-integrals considered in this paper and we test if the best performing one is competitive versus the FRM of the WR, the classical Choquet integral and the best Choquet-like based pre-aggregation function, that is, the one using the Hamacher t-norm (with the power measure[1]), which is denoted by Ham_{PA} in this paper.

The paper is organized as follows. Section 2 presents some basic concepts that are necessary to develop the paper. Section 3 introduces the concept of CC-integral. The FRBCS using CC-integrals is presented in Section 4. In Section 5, we explain the experimental framework, including the description of the 30 datasets considered in this paper and statistical tests, used for comparing the achieved results. The analysis of the application and run-time of the CC-Integral in classification problems are presented in Section 6 and Section 7 is the Conclusion.

In order to ease the readability of the paper, we have created three appendixes: the first one (Appendix A) containing a summary of the abbreviations used in this paper, the second (Appendix B) in which is present an example of an overlap function that is a copula and finally in the last one (Appendix C), where is shown the proof of the CC-integral is an increasing function.

[1] For more explanation of the power measure see, [12,15].

2. Preliminary concepts

This section aims at introducing the background necessary to develop the paper. First, we present some basic concepts, and then some new results that are important in the context of the paper.

2.1. Basic concepts

The key concept in this paper is one of aggregation functions [11,27]:

Definition 1. A function A: $[0,1]^n \to [0,1]$ is said to be an aggregation function whenever the following conditions are satisfied:

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

Definition 2. 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 ***(T3)*** (and also $T(1,x)=x$ only), then it is called a semi-copula.

Definition 3. A function O: $[0,1]^2 \to [0,1]$ is said to be an overlap function if it satisfies the following conditions:

(O1) O is commutative;
(O2) $O(x,y)=0$ if and only if $xy=0$;
(O3) $O(x,y)=1$ if and only if $xy=1$;
(O4) O is increasing;
(O5) O is continuous.

Definition 4. 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 $x \le x'$ and $y \le y'$:

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

Copulas are functions that link (two-dimensional) probability distribution functions to their one-dimensional margins, playing an important role in the theory of probabilistic metric spaces and statistics [22].

Proposition 1 [17, Proposition 9.8] [28, Lemma 6.1.8, Lemma 6.3.1]**.** *For each copula C: $[0,1]^2 \to [0,1]$, the Lukasiewicz and Minimum T-norms (T_L, T_M) T: $[0,1]^2 \to [0,1]$ it holds that:*

(i) $T_L \le C \le T_M$;
(ii) *C is increasing;*
(iii) *C satisfies the Lipschitz property with constant 1, that is, for all $x_1, x_2, y_1, y_2 \in [0,1]$, one has that:*

$$|\,C(x_1,y_1)-C(x_2,y_2)\,| \le |\,x_1-x_2\,|+|\,y_1-y_2\,|\,.$$

[2] For an increasing (decreasing) function we do not mean a strictly increasing (decreasing) function.

Table 1
Examples of Copulas.

(I) T-norms		
Definition	Name	Observations
$T_M(x,y)=\min\{x,y\}$	Minimum	Overlap function
$T_P(x,y)=xy$	Algebraic Product	Overlap function
$T_L(x,y)=\max\{0,x+y-1\}$	Lukasiewicz	
$T_{HP}(x,y)=\begin{cases}0 & \text{if } x=y=0\\ \frac{xy}{x+y-xy} & \text{otherwise}\end{cases}$	Hamacher Product	Overlap function
(II) Non-associative overlap functions		
Definition	Reference	Observations
$O_B(x,y)=\min\{x\sqrt{y},y\sqrt{x}\}$	[18, Theorem 8]	Cuadras-Augé family of copulas [32]
$O_{mM}(x,y)=\min\{x,y\}\max\{x^2,y^2\}$	[33, Example 3.1.(i)], [34, Example 4] [35, Example 3.1]	
$O_\alpha(x,y)=xy(1+\alpha(1-x)(1-y)),\ \alpha\in[-1,0[\cup]0,1]$	[22, Apendix A (A.2.1)], [36]	
(III) Other non-associative copulas		
Definition	Reference	Observations
$C_F(x,y)=xy+x^2y(1-x)(1-y)$	[17, Example 9.5 (v)]	Non-commutative
$C_L(x,y)=\max\{\min\{x,\frac{y}{2}\},x+y-1\}$	[22, Apendix A (A.5.3a)]	Non-commutative
$C_{Div}(x,y)=\frac{xy+\min\{x,y\}}{2}$	[22, Apendix A (A.8.7)]	

An immediate consequence of Proposition 1 is that any copula is continuous. Then, each associative copula is a continuous t-norm [17, Corollary 9.9].

Theorem 1 [17, Theorem 9.10]. *Let* T: $[0, 1]^2 \to [0, 1]$ *be a t-norm. Then, the following statements are equivalent:*

(i) *T is copula.*
(ii) *T satisfies the Lipschitz property with constant 1.*

The definitions of overlap, t-norms and copulas can be easily extended to n-ary functions. [17,29–31].

Table 1 presents the copulas used in the rest of the paper, in particular, some t-norms T: $[0, 1]^2 \to [0, 1]$ (Table 1 (I)), overlap functions O: $[0, 1]^2 \to [0, 1]$ (Table 1 (II)) and copulas C: $[0, 1]^2 \to [0, 1]$ that are neither t-norms nor overlap functions (Table 1 (III)). Observe that the overlap functions and copulas shown in Table 1 are all non-associative.

Now, we present the concept of fuzzy measure [13,37], which is a central tool for defining the Choquet integral. In what follows, denote $N=\{1,\ldots,n\}$, for an arbitrary $n>0$.

Definition 5. A function $\mathfrak{m}: 2^N \to [0,1]$ is said to be a fuzzy measure if, for all $X, Y \subseteq N$, it satisfies the following properties:

($\mathfrak{m}$1) Increasing: if $X \subseteq Y$, then $\mathfrak{m}(X) \leq \mathfrak{m}(Y)$;
($\mathfrak{m}$2) Boundary conditions: $\mathfrak{m}(\emptyset)=0$ and $\mathfrak{m}(N)=1$.

Regarding aggregation functions, we use fuzzy measures to analyze the relationship among the elements that we are aggregating, obtaining the relevance of a coalition. In this paper, we adopt the power measure $\mathfrak{m}_{PM}: 2^N \to [0,1]$, which is defined, for all $X \subseteq N$, by

$$\mathfrak{m}_{PM}(X)=\left(\frac{|X|}{n}\right)^q, \text{ with } q>0. \quad (1)$$

The choice for this fuzzy measure was based on the results obtained by Barrenechea et al. [12], who introduced an evolutionary algorithm to define the most suitable q to be used in the definition of the measure for each class. See also the results shown in [15,36,38], which also make use of such approach.

The Choquet integral generalizes the Lebesgue integral, defined considering additive measures. The Choquet integral, however, considers fuzzy measures. The discrete Choquet integral [13] is defined on finite spaces:

Definition 6 [11, Definition 1.74]**.** Let $\mathfrak{m}: 2^N \to [0,1]$ be a fuzzy measure. The discrete Choquet integral is the function $\mathfrak{C}_{\mathfrak{m}}: [0,1]^n \to [0,1]$, defined, for all of $\vec{x}=(x_1,\ldots,x_n)\in[0,1]^n$, by:

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

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

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

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

which we call the Choquet integral in its expanded form.

When using the Choquet integral in order to aggregate the inputs, it is possible to consider the relevance of the different coalitions (groups of inputs).

2.2. Some new results on copulas and overlap functions

Theorem 2. *A copula* C: $[0, 1]^2 \to [0, 1]$ *is an overlap function if it is commutative and positive.*

Proof. Is immediate that any positive and commutative copula C satisfies **(O1)** and **(O2)**. It follows that:

(O3): From **(C3)**, we have that $C(1,y)=1$ if and only if $y=1$. From the monotonicity of C, this implies that $C(x,y)=1$ if and only if $x=y=1$. Thus, one concludes that C satisfies **(O3)**.
(O4): From Proposition 1 **(ii)**, it follows that C is increasing.
(O5): It follows from Proposition 1 **(iii)** that, since C satisfies the Lipschitz condition for a constant 1, then C is continuous. □

Notice that any overlap function which is also a copula, by **(C3)**, necessarily has 1 as neutral element and, by Proposition 1, it satisfies the Lipschitz property with constant 1.

Theorem 3. *Let* O: $[0, 1]^2 \to [0, 1]$ *be an overlap function with neutral element and satisfying the Lipschitz condition for the constant 1.*

O is a copula if, for all x, y, x', $y' \in [0, 1]$ with $x \leq x'$ and $y \leq y'$, there exists $z \leq z'$ such that

$$O(O(x, y'), z') = O(x, y) \quad (4)$$

and

$$O(O(x', y'), z) = O(x', y). \quad (5)$$

Proof. Trivially, **(C2)** and **(C3)** follows from **(O2)** and the well-known fact that whenever an overlap function has a neutral element then this element is 1. To prove **(C1)**, consider x, x', y, y', z, $z' \in [0, 1]$ such that $x \leq x'$, $y \leq y'$ and the Eqs. (4) and (5) hold. It follows that:

$$\begin{aligned} O\big(x', y'\big) - O\big(x, y'\big) &\geq O\big(O\big(x', y'\big), z'\big) \\ &\quad - O\big(O\big(x, y'\big), z'\big) \text{ Lipschitz cond. for 1} \\ &= O\big(O\big(x', y'\big), z'\big) - O(x, y) \text{ by Eq. (4)} \\ &\geq O\big(O\big(x', y'\big), z\big) - O(x, y) \\ &= O\big(x', y\big) - O(x, y) \text{ by Eq. (5)} \end{aligned}$$

and then $O(x, y) + O(x', y') \geq O(x, y') + O(x', y)$. □

Observe that when O has a neutral element, by its continuity and isotonicity, there always exist z, $z' \in [0, 1]$ satisfying the Eqs. (4) and (5). An example of an overlap function that is also a copula function, is introduced in Appendix B.

3. Constructing Choquet-like Copula-based aggregation functions

In this section, we introduce a method for constructing a family of aggregation functions obtained by combining the discrete Choquet Integral in its expanded form (Eq. (3)) with copulas, just substituting the product operation in Eq. (3) by a copula. Such functions are called CC-integrals.

In the following, consider $N = \{1, \ldots, n\}$.

Definition 7. 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 Choquet-like Copula-based integral with respect to $\mathfrak{m}$ is defined as a function $\mathfrak{C}_{\mathfrak{m}}^{C} : [0, 1]^n \to [0, 1]$, given, for all $x \in [0, 1]^n$, by

$$\mathfrak{C}_{\mathfrak{m}}^{C}(\vec{x}) = \sum_{i=1}^{n} C\big(x_{(i)}, \mathfrak{m}\big(A_{(i)}\big)\big) - C\big(x_{(i-1)}, \mathfrak{m}\big(A_{(i)}\big)\big), \quad (6)$$

where $(x_{(1)}, \ldots, x_{(n)})$ 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 indices of $n - i + 1$ largest components of $\vec{x}$.

Proposition 2. *Under the conditions of Definition 7, $\mathfrak{C}_{\mathfrak{m}}^{C}$ is well defined, for any copula C: $[0, 1]^2 \to [0, 1]$.*

Proof. Since $x_{(i)} \geq x_{(i-1)}$ and, by Proposition 1, C is increasing, then it is immediate that $\mathfrak{C}_{\mathfrak{m}}^{C}(\vec{x}) \geq 0$, for any $\vec{x} \in [0, 1]$. On the other hand, by Proposition 1, C satisfies the Lipschitz property with constant 1, that is, for all x_1, x_2, y_1, $y_2 \in [0, 1]$, one has that:

$$\mid C(x_1, y_1) - C(x_2, y_2) \mid \leq \mid x_1 - x_2 \mid + \mid y_1 - y_2 \mid .$$

Then, since $x_{(i)} - x_{(i-1)} \geq 0$ and C is increasing, one has that:

$$\begin{aligned} C(x_{(i)}, \mathfrak{m}(A_{(i)})) - C(x_{(i-1)}, \mathfrak{m}(A_{(i)})) &\leq x_{(i)} - x_{(i-1)} + \mathfrak{m}(A_{(i)}) - \mathfrak{m}(A_{(i)}) \\ &= x_{(i)} - x_{(i-1)}. \end{aligned}$$

Thus, for any $\vec{x} \in [0, 1]$, it follows that,

$$\begin{aligned} \mathfrak{C}_{\mathfrak{m}}^{C}(\vec{x}) &= \sum_{i=1}^{n} C\big(x_{(i)}, \mathfrak{m}\big(A_{(i)}\big)\big) - C\big(x_{(i-1)}, \mathfrak{m}\big(A_{(i)}\big)\big) \\ &\leq \sum_{i=1}^{n} x_{(i)} - x_{(i-1)} \\ &= x_{(n)} \\ &\leq 1. \end{aligned}$$

□

Consider a fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$ and $\vec{x} \in [0, 1]^n$. The CC-integrals with respect to $\mathfrak{m}$, for each copula of Table 1, assume a form included in Table 2.

Proposition 3. *For any copula C: $[0, 1]^2 \to [0, 1]$ and fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$, $\mathfrak{C}_{\mathfrak{m}}^{C}$ is idempotent.*

Proof. Considering $\vec{x} = (x, \ldots, x) \in [0, 1]^n$, one has that:

$$\begin{aligned} \mathfrak{C}_{\mathfrak{m}}^{C}(\vec{x}) &= C(x, \mathfrak{m}(A_{(1)})) - C(0, \mathfrak{m}(A_{(1)})) \\ &\quad + \sum_{i=2}^{n} C(x, \mathfrak{m}(A_{(i)})) - C(x, \mathfrak{m}(A_{(i)})) \\ &= C(x, 1) - C(0, 1) + 0 \\ &= x. \end{aligned}$$

□

Proposition 4. *For any copula C: $[0, 1]^2 \to [0, 1]$ and fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$, $\mathfrak{C}_{\mathfrak{m}}^{C}$ satisfies the boundary conditions* **(A2)**.

Proof. Considering $\vec{0} = (0, \ldots, 0) \in [0, 1]^n$ and $\vec{1} = (1, \ldots, 1) \in [0, 1]^n$, by Proposition 3, one has that $\mathfrak{C}_{\mathfrak{m}}^{C}(\vec{0}) = 0$ and $\mathfrak{C}_{\mathfrak{m}}^{C}(\vec{1}) = 1$. □

Proposition 5. *For any copula C: $[0, 1]^2 \to [0, 1]$ and fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$, $\mathfrak{C}_{\mathfrak{m}}^{C}$ is increasing* **(A1)**.

In order to ease the readability of the paper, the proof to the Proposition 5 was moved to Appendix C and from there, we can imply that the following results are immediate.

Corollary 1. *For any bivariate copula C: $[0, 1]^2 \to [0, 1]$ and fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$, it holds that* $\min \leq \mathfrak{C}_{\mathfrak{m}}^{C} \leq \max$.

Theorem 4. *For any bivariate copula C: $[0, 1]^2 \to [0, 1]$ and fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$, $\mathfrak{C}_{\mathfrak{m}}^{C}$ is an average aggregation function.*

Proof. It follows from Propositions 4 and 5, and Corollary 1. □

4. Applying CC-integrals as aggregation functions in a fuzzy reasoning algorithm

In this section, we recall firstly the main concepts of FRBCSs and, then, the algorithm of the proposed FRM using CC-integrals is presented.

4.1. Fuzzy rule-based classification systems

A classification problem is composed by m training examples characterized by n input attributes, denoted by

$$\vec{x}_p = (x_{p1}, \ldots, x_{pn}, y_p),$$

where

- $p = 1, \ldots, m$;
- x_{pi}, with $i = 1, \ldots, n$, is the value of the ith attribute;

Table 2
Choquet-like Copula-based integral (CC-integral).

Copula	CC-integral
TM	$\mathfrak{C}_{\mathfrak{m}}^{C_M}(\vec{x}) = \sum_{i=1}^{n}(\min\{x_{(i)}, \mathfrak{m}(A_{(i)})\} - \min\{x_{(i-1)}, \mathfrak{m}(A_{(i)})\})$
TL	$\mathfrak{C}_{\mathfrak{m}}^{C_L}(\vec{x}) = \sum_{i=1}^{n}(\max\{0, x_{(i)} + \mathfrak{m}(A_{(i)}) - 1\} - \max\{0, x_{(i-1)} + \mathfrak{m}(A_{(i)}) - 1\})$
TH	$\mathfrak{C}_{\mathfrak{m}}^{C_{HP}}(\vec{x}) = \begin{cases} 0, & \text{if } x = y = 0 \\ \frac{x_{(i)}\mathfrak{m}(A_{(i)})}{x_{(i)}+\mathfrak{m}(A_{(i)})-x_{(i)}\mathfrak{m}(A_{(i)})} - \frac{x_{(i-1)}\mathfrak{m}(A_{(i)})}{x_{(i-1)}+\mathfrak{m}(A_{(i)})-x_{(i-1)}\mathfrak{m}(A_{(i)})}, & \text{otherwise} \end{cases}$
OB	$\mathfrak{C}_{\mathfrak{m}}^{C_{O_B}}(\vec{x}) = \sum_{i=1}^{n}(\min\{x_{(i)}\sqrt{\mathfrak{m}(A_{(i)})}, \mathfrak{m}(A_{(i)})\sqrt{x_{(i)}}\} - \min\{x_{(i-1)}\sqrt{\mathfrak{m}(A_{(i)})}, \mathfrak{m}(A_{(i)})\sqrt{x_{(i-1)}}\})$
Oα	$\mathfrak{C}_{\mathfrak{m}}^{C_{O_\alpha}}(\vec{x}) = \sum_{i=1}^{n}(x_{(i)}\mathfrak{m}(A_{(i)})(1+\alpha(1-x_{(i)})(1-\mathfrak{m}(A_{(i)}))) - x_{(i-1)}\mathfrak{m}(A_{(i)})(1+\alpha(1-x_{(i-1)})(1-\mathfrak{m}(A_{(i)}))))$
OmM	$\mathfrak{C}_{\mathfrak{m}}^{C_{O_{mM}}}(\vec{x}) = \sum_{i=1}^{n}(\min\{x_{(i)}, \mathfrak{m}(A_{(i)})\}\max\{{x_{(i)}}^2, \mathfrak{m}(A_{(i)})^2\} - \min\{x_{(i-1)}, \mathfrak{m}(A_{(i)})\}\max\{{x_{(i-1)}}^2, \mathfrak{m}(A_{(i)})^2\})$
CF	$\mathfrak{C}_{\mathfrak{m}}^{C_F}(\vec{x}) = \sum_{i=1}^{n}(x_{(i)}\mathfrak{m}(A_{(i)}) + {x_{(i)}}^2\mathfrak{m}(A_{(i)})(1-x_{(i)})(1-\mathfrak{m}(A_{(i)})) - x_{(i-1)}\mathfrak{m}(A_{(i)}) + {x_{(i-1)}}^2\mathfrak{m}(A_{(i)})(1-x_{(i-1)})(1-\mathfrak{m}(A_{(i)})))$
CL	$\mathfrak{C}_{\mathfrak{m}}^{C_{C_L}}(\vec{x}) = \sum_{i=1}^{n}(\max\{\min\{x_{(i)}, \frac{\mathfrak{m}(A_{(i)})}{2}\}, x_{(i)} + \mathfrak{m}(A_{(i)}) - 1\} - \max\{\min\{x_{(i-1)}, \frac{\mathfrak{m}(A_{(i)})}{2}\}, x_{(i-1)} + \mathfrak{m}(A_{(i)}) - 1\})$
CDiv	$\mathfrak{C}_{\mathfrak{m}}^{C_{C_{Div}}}(\vec{x}) = \sum_{i=1}^{n}\left(\frac{x_{(i)}\mathfrak{m}(A_{(i)})+\min\{x_{(i)},\mathfrak{m}(A_{(i)})\}}{2} - \frac{x_{(i-1)}\mathfrak{m}(A_{(i)})+\min\{x_{(i-1)},\mathfrak{m}(A_{(i)})\}}{2}\right)$

- $y_p \in \mathbb{C} = \{C_1, C_2, \ldots, C_M\}$ is the label of the class of the pth training example.

Although one may find in the literature several techniques to deal with classification problems, FRBCSs [39] are indeed the one most frequently adopted. Observe that FRBCSs allow us to have all the available information in the system modeling, obtaining an interpretable model and producing quite accurate results. A FRBCSs is composed by:

(i) The Knowledge Base: it contains the Rule and the Data Bases, with, respectively, the fuzzy inference rules and the membership functions;
(ii) The Fuzzy Reasoning Algorithm: it is the inference procedure used to classify examples considering all the information stored in the Knowledge Base.

We adopt the following structure for the fuzzy rules:

$$\text{Rule } R_j: \text{ If } x_{p1} \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_{pn} \text{ is } A_{jn} \text{ then Class is } C_j \text{ with } RW_j, \tag{7}$$

where:

- $x_p = (x_{p1}, \ldots, x_{pn})$ is the n-dimensional vector of attribute values corresponding to an example x_p of the pth example;
- R_j is the label of the jth rule;
- A_{ji} is an antecedent fuzzy set which models a linguistic term;
- C_j is the class of the jth rule;
- $RW_j \in [0, 1]$ is the rule weight [40], which, in this case, is computed using the certainty factor.

We adopt the FARC-HD [38] fuzzy classifier (Fuzzy Association Rule-based Classification model for High Dimensional problems), to accomplish the fuzzy rule learning process. The learning process of FARC-HD is composed of the following three stages:

1) *Fuzzy association rule extraction for classification:* This step is aimed at generating fuzzy association rules from frequent itemsets. To do so, a search tree is constructed for each class computing the confidence and support of each item or itemset (in this algorithm an item is a linguistic label). We have to point out that the number of antecedents of the different fuzzy rules can be different and their maximum length is determined by the maximum depth allowed for the search tree.
2) *Candidate rule prescreening:* This phase uses an instance weighting scheme to preselect the most interesting rules and consequently, to decrease the computational cost of the evolutionary process.
3) *Genetic rule selection and lateral tuning:* The final stage of the method consists in the application of an evolutionary algorithm to perform both the lateral tuning of the fuzzy sets [41] and the selection of the best rules generated in the previous step.

4.2. The new fuzzy reasoning method

Once the fuzzy rules composing the system have been created, it is necessary a mechanism for classifying new examples. Specifically, let $x_p = (x_{p1}, \ldots, x_{pn})$ be a new example to be classified, L being the number of rules in the rule base and M being the number of classes of the problem. The new algorithm where our CC-functions are applied in the FRM is composed of the following steps:

Step 1 To compute the *matching degree*, that is, the strength of the activation of the if-part of the rules for the example x_p, which is computed using a t-norm T': $[0, 1]^n \to [0, 1]$:

$$\mu_{A_j}(x_p) = T'(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})), \text{ with } j = 1, \ldots, L. \tag{8}$$

Step 2 The *Association degree* is computed, that is, for the class of each rule, the matching degree is weighted with the corresponding rule weight, given by:

$$b_j^k(x_p) = \mu_{A_j}(x_p) \cdot RW_j^k, \text{ with } k = Class(R_j), j = 1, \ldots, L. \tag{9}$$

Step 3 The *example classification soundness degree for all classes* is calculated, applying the CC-functions (Eq. (6)) to combine the association degrees obtained in the previous step, as follows:

$$Y_k(x_p) = \mathfrak{C}_{\mathfrak{m}}^{C}\left(b_1^k(x_p), \ldots, b_L^k(x_p)\right), \text{ with } k = 1, \ldots, M, \tag{10}$$

where $\mathfrak{C}_{\mathfrak{m}}^{C}$ is the obtained CC-integral, for the copula C: $[0, 1]^2 \to [0, 1]$ and fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$ (Table 2). Since, whenever $b_j^k(x_p) = 0$, it holds that:

$$\mathfrak{C}_{\mathfrak{m}}^{C}(b_1^k(x_p), \ldots, b_L^k(x_p)) = \mathfrak{C}_{\mathfrak{m}}^{C}(b_1^k(x_p), \ldots, b_{j-1}^k(x_p), b_{j+1}^k(x_p), \ldots, b_L^k(x_p)),$$

then, for practical reasons, only those $b_j^k > 0$ are considered in Eq. (10).

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

$$F(Y_1, \ldots, Y_M) = \min_{k=1\ldots M} k \text{ s.t. } Y_k = \max_{w=1,\ldots,M}(Y_w). \tag{11}$$

In practical applications, it is sufficient to consider

$$F(Y_1, \ldots, Y_M) = \arg\max_{k=1,\ldots,M}(Y_k). \tag{12}$$

Table 3
Association degrees for each class.

	C_1	C_2	C_3
R_x	0.9	0.89	0.1
R_y	0.2	0.88	0.3
R_z	0.1	0.86	0.25

As it can be observed, in the third step of the FRM we propose to use CC-integrals, which are associated with a fuzzy measure. According to the results obtained in several works (see, e.g., in [12,15,36]), we have selected the power measure where the exponent q is genetically learned. Therefore, for each class of the problem we create a different fuzzy measure by learning the most appropriate exponent to model the interaction among the rules of that class. To do so, the CHC evolutionary algorithm [14] is considered since it is widely used in this domain [24,42]. The main specific features of our evolutionary model are the following ones.[3]

- **Coding Scheme:** A chromosome is formed of as many genes as classes (G_k, $k = 1, \ldots, M$), using a real codification. That is:

  $C_{CHOQUET} = \{G_1, \ldots, G_M\}$

  where $G_k \in [0.01, 1.99]$ with $k = 1, \ldots, M$. However, the suggested range of the real values the exponents, q_k, is [0.01, 100]. For this reason, it is necessary to adapt the value of the genes to the real range, as following:

$$q_k = \begin{cases} G_k & \text{if } 0 < G_k \leq 1 \\ \frac{1}{2-G_k} & \text{if } 1 < G_k < 2 \end{cases} \quad (13)$$

- **Initial Gene Pool:** An individual containing all genes with value 1 is included in the population. In this manner, the cardinality measure is obtained.
- **Chromosome Evaluation:** The most common metric for classification problems is used, i.e. the accuracy rate. This metric is the percentage of correctly classified examples.
- **Crossover Operator:** The parent centrix BLX crossover operator [43] is selected as it is usually applied for real codifications.
- **Restarting Approach** In order to avoid the local optima, the algorithm uses a restarting approach since it does not apply mutation during the recombination phase. To do so, a threshold is firstly initialized and it is decreased when no new individuals are included in the population after the crossover stage. When the threshold value is less than zero, all chromosomes (except the best one, as in the elitist scheme) are regenerated randomly with the purpose of introducing new diversity to the search. The evolutionary process also finishes when three restarts are sequentially done without improving the best solution.

In order to clarify how the new FRM works, we introduce a short example considering different aggregation functions to perform the third step of the FRM where the local information given by several rules (obtained after step 2 of the FRM) is aggregated according to the classes of the rules. More specifically, we compare the FRM that uses the maximum as aggregation function and our new methodology considering the CC-integral based on the minimum t-norm, *TM*.

Example 1. A classification problem composed of 3 classes is studied. In Table 3 we introduce the association degrees obtained by the fuzzy rules fired when classifying a new example. As it can be observed there are three fired rules for each class (columns of Table 3) and consequently, three aggregations have to be computed (one for each class) in the third step of the FRM.

[3] For more details about the evolutionary method see [12].

In order to show the behavior of our new method we consider the usage of the CC-integral based on the minimum t-norm (T_M) and the cardinality measure, that is, the power measure setting the exponent q to 1 (in order to ease the understanding of the calculations). We compare it versus the classical FRM of the winning rule, which uses the maximum as the aggregation function. The values computed for each class using the two approaches are the following ones:

- C_1
  - Sorted association degrees: [0.1, 0.2, 0.9]
    - Maximum = 0.9
    - TM $= (\min(0.1, \frac{3}{3}) - \min(0, \frac{3}{3})) + (\min(0.2, \frac{2}{3}) - \min(0.1, \frac{2}{3})) + (\min(0.9, \frac{1}{3}) - \min(0.2, \frac{1}{3})) = 0.33$
- C_2
  - Sorted association degrees: [0.86, 0.88, 0.89]
    - Maximum = 0.89
    - TM $= (\min(0.86, \frac{3}{3}) - \min(0, \frac{3}{3})) + (\min(0.88, \frac{2}{3}) - \min(0.86, \frac{2}{3})) + \min((0.89, \frac{1}{3}) - \min(0.88, \frac{1}{3})) = 0.86$
- C_3
  - Sorted association degrees: [0.1, 0.25, 0.3]
    - Maximum = 0.3
    - TM $= (\min(0.1, \frac{3}{3}) - \min(0, \frac{3}{3})) + (\min(0.25, \frac{2}{3}) - \min(0.1, \frac{2}{3})) + (\min(0.3, \frac{1}{3}) - \min(0.25, \frac{1}{3})) = 0.3$

Once the association degree for each class has been computed, the predicted class is the one associated with the largest value (last step of the FRM):

- Maximum = Max[**0.9**, 0.89, 0.3] = C_1
- TM = Max[0.33, **0.86**, 0.3] = C_2

It can be noticed that the usage of the maximum as an aggregation operator predicts class 1, since it only considers the information provided by one rule per class. However, if we look in detail at the association degrees presented in Table 3, this prediction may not be ideal, since that the class 1 has one rule having a high compatibility whereas the class 2 has three rules having high compatibilities (slightly less than that of class 1). Therefore, class 2 seems to be a most appropriate option. This fact is taken into account by our new approach since the information given by all the fuzzy rules and not only by the best one is considered and consequently, class 2 is predicted.

Furthermore, in the example it can be observed that the behavior of the CC-integrals can be different depending on the values to be aggregated. For example, for class 2 the result is the minimum of the values to be aggregated, for class 3 is the maximum whereas for class 1 is an intermediate value. Therefore, the FRM is more flexible, which may imply a enhancement of its behavior.

5. Experiment specifications

In this section, firstly we present the real world classification problems besides the configuration for the considered approaches. After that, we introduce the statistical tests that are necessary to compare the achieved results.

5.1. Datasets

We have selected 30 real world datasets from the KEEL dataset repository [23]. Table 4 presents the characteristics of the these datasets: the identifier (Id.), the name (Dataset), the number of instances (*#Inst*), the number of attributes (*#Att*) and the number of classes (*#Class*).

Some datasets, namely: *magic, page-blocks, penbased, ring, satimage* and *twonorm*, were stratified sampled at 10% in order to reduce their size for training. Some examples containing missing information were removed, e.g., in the *wisconsin* dataset.

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
Hay	Hayes-Roth	160	4	3
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
Sah	Saheart	462	9	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
Wis	Wisconsin	683	11	2
Yea	Yeast	1,484	8	10

The model used in this paper, which was proposed in [12,36,42], consists of a 5-fold cross-validation model, where a dataset is split in five partitions randomly. Each partition presents 20% of the examples. Four partitions are used for training, and the other is used for testing. This process is repeated five times, using a different partition to test the formed system each time. The performance of the approaches are measured considering each partition, based on the accuracy rate, which is defined by the number of correctly classified examples divided by the total number of examples for each partition. After, is calculate the average result of the five testing partitions, which is output of the algorithm.

5.2. Configuration of the different classifiers used in the study

The parameter set-up for the FARC-HD algorithm is the one suggested by the authors:

- The conjunction operator T' is the product t-norm;
- The rule weight RW_j is the certainty factor;
- 5 linguistic labels per variable;
- The minimum support is 0.05;
- The threshold for the confidence is 0.8;
- The depth of the search trees is limited to 3;
- k_t, the parameter that determines the number of fuzzy rules that cover each example, is equal to 2.

The features considered for the evolutionary process are the following ones:

- The populations are composed of 50 individuals;
- 30 bits per gen are considered for the Gray codification;
- The maximum number of evaluations is 20.000.

Finally, for the copula O_α, the value adopted for the α parameter is 0.1 ($\alpha = 0.1$), since this value was tested in [36] previously, although in a different method, but providing better performance than the other values.

5.3. Statistical tests for performance comparison

In order to give statistical support for the analysis of the results, we consider the usage of hypothesis validation techniques [24,25]. Specifically, we use non-parametric tests, since the initial conditions that guarantee the reliability of the parametric tests cannot be performed [26].

In fact, we use the aligned Friedman rank test [44] to detect statistical differences among a group of results and to show how good a method is with respect to its partners. In this method, the algorithm achieving the lowest average ranking is the best one. Additionally, we have graphically shown the obtained ranks to easily observe which is the best method.

Furthermore, we apply the post-hoc Holm's test [45] to study whether the best ranking method rejects the equality hypothesis with respect to its partners. The post-hoc procedure allows us to know if a hypothesis of comparison could be rejected at a specified level of significance α. Specifically, we compute the adjusted p-value (APV) to take into account that multiple tests are conducted. As a result, we can directly compare the APV with the level of significance α so as to be able to reject the null hypothesis.

Finally, we also consider the usage of the Wilcoxon test [46] in order to perform pair-wise comparisons.

6. Analysis of the application of the CC-integral in classification problems

This section is aimed at providing an analysis of the application of CC-integral in real-world classification problems. As we have mentioned in Section 4, we use the FARC-HD fuzzy classifier whose FRM is adapted to use CC-integrals (the ones introduced in Table 2). Specifically, the goal of the analysis is to study whether the usage of CC-integrals in the FRM allows us to obtain results as accurate as those achieved by the best performing pre-aggregation functions introduced in [15], which is named in this paper Ham_{PA}.

To do so, the study is divided in three main parts:

- In first place, we want to determine the best CC-Integral. For that, we conduct a study comparing the different copulas considered in this work (the ones shown in Table 1), which produce all the CC-integrals presented in Table 2. For this task, we divide the study according to the type of copula: t-norm, overlap or copula (which is neither a t-norm nor an overlap function), and finally, we compare the best of each group to determine the best CC-integral (Section 6.1).
- Then, we compare if the best CC-integral is able to improve the results provided by the classical FRM of the WR as well as the standard Choquet integral and those provided by the best pre-aggregation function introduced [15], that is, the Ham_{PA} (Section 6.2).
- Finally, we analyze the execution time taken by the different approaches (Section 6.3).

6.1. Analyzing the behavior of the different CC-integrals

This subsection is aimed at analyzing the behavior of the proposed generalizations in the FRM. In order to determine the best CC-integral in this study we compare them by groups of copulas according to their types (copulas that are t-norms, overlaps and those that are neither t-norms nor overlap functions, the last denote by n-copulas). From each group, we select the best one and we compare them to determine the best CC-integral.

The results achieved in testing by the different CC-integrals are presented in Table 5 by columns, and in each line, containing the mean of accuracy obtained in the five partitions of the dataset where the best global result for each dataset is highlighted in **boldface**.

Table 5
Accuracy rate achieved in testing for the different CC-integrals.

Dataset	TM	TL	TH	OB	Oα	OmM	CF	CL	Cdiv
App	**85.84**	82.99	84.94	81.13	**85.84**	81.13	83.98	83.03	83.03
Bal	81.60	80.64	80.96	81.28	79.52	80.32	81.28	**82.24**	78.88
Ban	84.30	84.43	85.91	86.17	86.11	**87.00**	85.85	86.79	85.38
Bnd	**71.06**	68.24	69.38	66.01	67.72	67.41	70.47	68.54	67.70
Bup	61.45	65.51	64.06	63.19	**67.25**	66.38	65.51	62.90	64.93
Cle	54.88	**57.57**	56.91	56.92	55.23	56.55	55.92	54.54	54.89
Eco	77.09	73.53	76.20	74.41	**78.86**	76.50	77.40	78.58	77.38
Gla	**69.17**	61.71	64.97	66.39	63.55	60.29	60.76	63.12	66.82
Hab	74.17	73.21	70.90	71.22	73.51	69.93	**74.48**	72.86	72.86
Hay	**81.74**	79.46	79.49	78.75	79.49	78.72	78.75	79.49	78.75
Iri	92.67	93.33	93.33	93.33	92.67	**94.67**	94.00	92.67	93.33
Mag	79.81	**80.34**	79.86	79.34	80.07	79.02	80.02	79.97	79.07
New	93.95	93.02	93.95	94.42	94.42	**96.28**	94.42	93.95	94.88
Pag	93.97	93.98	93.61	93.97	**94.88**	94.16	94.16	93.79	94.34
Pen	91.27	90.82	**91.91**	91.45	90.09	89.82	91.36	90.91	90.73
Pho	82.94	81.83	83.40	**83.55**	83.46	82.62	82.73	83.53	83.10
Pim	75.78	74.87	74.87	75.38	75.78	73.57	74.86	76.04	**76.95**
Rin	87.97	89.46	89.19	89.73	**90.41**	88.38	89.59	88.65	89.59
Sah	**70.78**	69.03	69.48	67.53	66.89	67.95	68.83	68.82	69.69
Sat	79.01	78.54	**79.94**	79.62	78.85	78.69	79.63	79.16	79.00
Seg	92.25	91.90	93.12	92.68	92.86	92.77	92.55	**93.20**	92.94
Shu	**98.16**	97.66	97.29	97.52	97.47	97.24	96.41	97.79	98.07
Spe	78.99	78.27	79.73	77.13	**79.74**	77.51	79.00	77.86	77.88
Tit	**78.87**	**78.87**	**78.87**	**78.87**	**78.87**	**78.87**	**78.87**	**78.87**	**78.87**
Two	85.14	**85.68**	84.59	84.73	84.59	85.14	84.32	85.14	83.92
Veh	**69.86**	66.67	67.97	67.85	67.73	68.32	68.56	67.85	68.56
Vow	**68.89**	66.26	67.88	66.57	67.47	67.37	66.97	65.76	68.38
Win	93.83	95.48	96.08	94.35	94.38	96.62	96.62	**98.32**	95.51
Wis	95.90	96.34	96.63	96.34	**97.22**	96.34	97.22	96.34	96.63
Yea	57.01	57.75	57.61	57.28	56.40	57.34	57.55	57.82	**57.88**
Mean	**80.28**	79.58	80.10	79.57	80.04	79.56	80.07	79.95	80.00

Table 6
Aligned Friedman and Holm test to compare the different CC-integrals derived from different families of copulas.

t-Norm	TM	TL	TH
	39.61	56.31 (0.02)	40.56 (0.88)
Overlap	OB	O_α	OmM
	47.63 (0.23)	**39.08**	47.78 (0.23)
n-Copula	CL	CF	Cdiv
	39.76 (0.33)	**49.08**	47.65 (0.33)

From the results shown in the previous table, it is possible notice that the behavior of the different CC-integrals derived from the same family of copulas are quite similar among themselves with the exception of Lukasiewicz (for t-norms) and OmM (for overlaps), which provided the worst results. Furthermore, it seems that the CC-integrals associated with the minimum t-norm, TM, achieved the best result in 9 datasets and the best average result among all CC-integrals under study. Therefore, a priori, TM would be the best CC-integral. Anyway, these conclusions cannot be meaningful without conducting the appropriate study.

Specifically, we have carried out three multiple comparison statistical tests (one for each family of aggregation functions) by applying the aligned Friedman test and the post-hoc Holm's test, whose results are shown in Table 6. These results are grouped by rows according to the family of aggregation functions. In each row, we show the results of the three functions considered in this paper, for each family. The value of each cell correspond to the rank obtained by the aligned Friedman test, when comparing the different functions belonging to the same family. The value shown in brackets represent the APV obtained by the Holm post-hoc test, using as the control method the approach having the smallest rank, which is shown in **boldface**. The APV is underlined when there are statistical differences (considering a significance level of $\alpha = 0.1$).

As is shown in the first row of Table 6, when considering t-norms, the results provided by TM are statistically better than those by TL and similar to those by TH. However, the fact that TM obtained a better overall mean providing the best result in 12 datasets leads us to select it as the best t-norm.

Regarding CC-integrals derived from overlap functions, we can see in the second row of Table 6 that the best ranked one is O_α, but there are not statistical differences among them. From this group, we select O_α as the representative, since it obtained the best overall mean and achieved the highest accuracy rate in 14 data-sets and tie only once.

Finally, the third row of Table 6 allows us to state that CF is the best option among the group of n-copulas. Although there are not statistical differences among the functions in this family, this copula obtained the best ranking, which is based on obtaining the best result in 11 out of the 30 datasets, having a tie in two datasets.

Once the best aggregation is determined for each family of copulas, namely, TM, Oα and CF, we conduct another statistical study to compare them among themselves. We have carried out again the aligned Friedman test to detect whether there are statistical differences among them or not. The obtained rankings are shown in Fig. 1. Then we run the Holm's statistical test using TM as control method, since it provides the best ranking in the previous statistical test. The obtained results by the Holm's test are shown in Table 7. From these results we can conclude that TH does not statistically outperform neither O_α nor CF. However, we select TM as the best option due to the fact that it obtained the best overall result, as shown in Table 5.

6.2. Comparison of the best CC-integral versus other averaging operators

As we have shown in the previous section, the best CC-integral among the ones considered in this is study is the one related to the minimum t-norm, TM. In this section our aim is to compare this generalization against the classical FRM of the WR (it uses the maximum as aggregation function), the standard Choquet In-

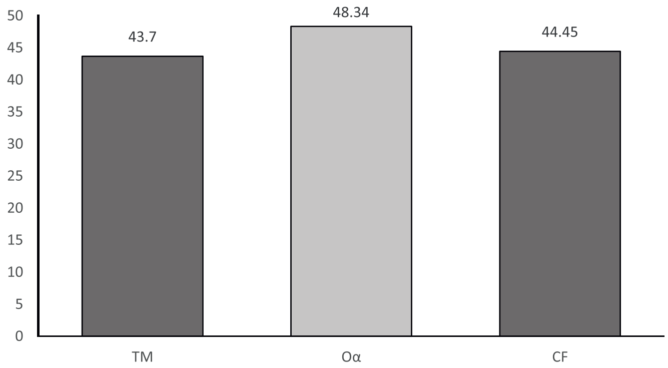

Fig. 1. Rankings obtained by the best CC-integrals for each family of copulas.

Table 7
Holm test to compare the best CC-integrals for each family of copulas.

Generalization	APV
Oα	0.98
CF	0.98

Table 8
Accuracy results achieved in test provided by the best CC-integral and other averaging operators.

Dataset	TM	Choquet	Ham_{PA}	WR
App	**85.84**	80.13	82.99	83.03
Bal	81.60	82.40	**82.72**	81.92
Ban	84.30	**86.32**	85.96	83.94
Bnd	71.06	68.56	**72.13**	69.40
Bup	61.45	**66.96**	65.80	62.03
Cle	54.88	55.58	55.58	**56.91**
Eco	77.09	76.51	**80.07**	75.62
Gla	**69.17**	64.02	63.10	64.99
Hab	**74.17**	72.52	72.21	70.89
Hay	**81.74**	79.49	79.49	78.69
Iri	92.67	91.33	93.33	**94.00**
Mag	**79.81**	78.86	79.76	78.60
New	93.95	94.88	**95.35**	94.88
Pag	93.97	94.16	**94.34**	94.16
Pen	91.27	90.55	90.82	**91.45**
Pho	82.94	82.98	**83.83**	82.29
Pim	**75.78**	74.60	73.44	74.60
Rin	87.97	**90.95**	88.78	90.00
Sah	**70.78**	69.69	70.77	68.61
Sat	79.01	79.47	**80.40**	79.63
Seg	92.25	**93.46**	93.33	93.03
Shu	**98.16**	97.61	97.20	96.00
Spe	**78.99**	77.88	76.02	77.90
Tit	**78.87**	**78.87**	**78.87**	**78.87**
Two	85.14	84.46	85.27	**86.49**
Veh	**69.86**	68.44	68.20	66.67
Vow	**68.89**	67.58	68.18	67.98
Win	93.83	93.79	**96.63**	96.60
Wis	95.90	**97.22**	96.78	96.34
Yea	**57.01**	55.73	56.53	55.32
Mean	**80.28**	79.83	80.26	79.70

tegral (the one that uses the product t-norm) and the best pre-aggregation function from [15] (Ham_{PA}). Table 8 shows the results achieved in testing by these four approaches, where the best global result for each dataset is highlighted in **boldface**.

Looking at the results of these four approaches, it is noticeable that TM obtained the best average global result, which is similar to that of Ham_{PA} and it is superior to those of WR and the standard Choquet integral. In a closer look we can observe that TM obtains the best result in 12 datasets, whereas Ham_{PA}, Choquet and WR provide the best results in 8, 5 and 4 datasets, respectively.[4] In order to support our previous results, we have carried out a set of pair-wise statistical comparisons using the well-known Wilcoxon signed-rank test [46]. Specifically, we have compared the best CC-integral, TM, versus WR, Ham_{PA} and Choquet integral. Table 9 shows the results of these comparisons, where R^+ indicates the ranks obtained by TM and R^- represents the ranks achieved by the method used in each comparison.

Table 9
Wilcoxon test to compare the best CC-integral versus the Ham_{PA}, the Winning Rule and the standard Choquet integral.

Comparison	R^+	R^-	p-value
TM vs. Ham_{PA}	215.0	250.0	0.72
TM vs. WR	311.5	153.5	0.09
TM vs. Choquet	303.5	161.5	0.14

According to the obtained statistical results presented in Table 9, we can affirm, with a high level of confidence, that the CC-integral defined by the minimum is better than the WR. Regarding the standard Choquet integral, we can observe that, although there are not statistical differences, the obtained p-value is low. Furthermore, TM improves the results of the Choquet integral in 18 out of the 30 datasets considered in this study. These two facts, show that TM is enhancing the results provided by the standard Choquet integral. Finally, we point out that, when comparing TM and Ham_{PA}, the obtained p-value is high, which implies that the behavior of these two approaches is similar.

6.3. Analyzing the execution time of the different CC-integrals

In this subsection we present the analysis of the results in terms of the execution time. We must point out that this execution time refers to all process, including the reading of the files containing the datasets, the execution of the learning algorithm as well as the classification of the training and testing examples and the writing of the output files.

The experimnets have been carried out in a 8 nodes cluster connected via 1GB/s Ethernet LAN network. Half of these nodes are composed by 2 Intel Xeon E5-2620 v3 processors at 2.4 GHz (3.2 GHz with Turbo Boost) with 12 virtual cores in each one (where 6 of them are physical). Three of the remaining node are equipped with 2 Intel Xeon E5-2620 v2 processors at 2.1 GHz with the same number of cores than the previous ones. The last node is the master node, composed of an Intel Xeon E5-2609 processors with 4 physical cores at 2.4 GHz. All slaves nodes are equipped with 32GB of RAM memory, while the master works with 8GB of RAM memory. With the respect to storage specifications, all nodes uses Hard Disk Drivers featuring read/write performance of 128 MB/s.

We present in Table 10 the mean execution time of all different approaches considered in the study and we highlight in **boldface** the aggregation that achieves the lowest execution time. From the execution times achieved, it is noticeable that all CC-integrals have a similar run-time (more or less one minute), being TL the quickest. In relation to the standard Choquet integral and the pre-aggregation function that considers the Hamacher product (Ham_{PA}) it is possible to observe that they have a similar execution time, even similar to the CC-integrals. This behavior is normal since all of them are based on the Choquet integral and the number of computations is the same. The small differences are due to the different aggregation function used.

Regarding the FRM of the WR, as expected, it is the approach having the less execution time. This occurs because in this FRM it

[4] We have not considered in the count the Titanic dataset which have the same classification rate for all methods.

Table 10
Average execution time considering all aggregations.

Dataset	TM	TL	TH	OB	Oα	OmM	CF	CL	CDiv	Choquet	Ham_{PA}	WR
App	0:05	0:05	0:05	0:04	0:06	0:06	0:06	0:06	0:05	0:05	0:05	**0:01**
Bal	0:57	0:57	0:59	0:59	0:57	0:48	0:47	0:42	0:41	0:54	1:09	**0:23**
Ban	2:37	2:16	2:36	2:32	2:15	3:40	3:13	2:11	2:50	2:41	2:19	**0:25**
Bnd	0:28	0:27	0:33	0:34	0:32	0:32	0:29	0:25	0:29	0:31	0:29	**0:14**
Bup	0:19	0:15	0:18	0:19	0:19	0:20	0:17	0:16	0:19	0:19	0:18	**0:03**
Cle	0:29	0:25	0:30	0:32	0:33	0:32	0:34	0:30	0:28	0:35	0:31	**0:19**
Eco	0:21	0:17	0:20	0:22	0:22	0:20	0:22	0:22	0:23	0:22	0:24	**0:09**
Gla	0:11	0:10	0:13	0:14	0:14	0:13	0:14	0:11	0:12	0:12	0:13	**0:05**
Hab	0:12	0:11	0:12	0:12	0:14	0:12	0:12	0:12	0:13	0:12	0:11	**0:02**
Hay	0:05	0:05	0:05	0:05	0:06	0:05	0:05	0:04	0:04	0:05	0:05	**0:03**
Iri	0:02	0:02	0:02	0:02	0:03	0:02	0:02	0:03	0:02	0:02	0:03	**0:00**
Mag	2:22	2:41	2:50	2:39	2:41	4:29	3:41	2:23	2:27	2:47	2:54	**0:55**
New	0:07	0:06	0:08	0:07	0:09	0:07	0:07	0:07	0:07	0:08	0:09	**0:02**
Pag	0:23	0:20	0:23	0:22	0:24	0:26	0:23	0:25	0:23	0:23	0:22	**0:09**
Pen	2:07	1:28	1:44	1:41	1:40	1:29	1:37	1:26	1:37	2:10	1:48	**1:14**
Pho	3:35	3:36	4:04	3:58	5:27	3:46	3:26	2:57	3:07	5:04	4:57	**1:11**
Pim	1:27	1:23	1:20	1:25	1:32	2:03	1:39	1:23	1:18	1:22	1:09	**0:27**
Rin	0:41	0:48	0:40	0:42	0:32	0:40	0:37	0:31	0:29	0:46	0:33	**0:23**
Sah	0:44	0:38	0:43	0:40	0:36	0:36	0:37	0:33	0:33	0:48	0:35	**0:18**
Sat	1:17	1:03	1:01	0:58	1:00	0:52	0:52	**0:48**	0:50	1:14	1:04	0:57
Seg	3:15	**1:56**	2:46	2:50	3:50	2:36	2:34	2:36	2:38	3:56	3:52	1:57
Shu	1:04	0:48	1:00	0:50	1:04	0:48	0:44	0:58	0:57	1:15	0:54	**0:18**
Spe	0:36	0:33	0:34	0:31	0:35	0:28	0:28	**0:26**	0:27	0:37	0:29	0:27
Tit	0:51	0:50	0:53	0:51	1:02	0:44	0:42	0:39	0:39	0:52	1:02	**0:10**
Two	0:45	0:46	0:43	0:45	0:47	0:59	0:54	0:42	0:44	0:44	0:44	**0:29**
Veh	2:07	1:32	1:50	1:40	1:32	1:30	1:36	1:26	1:25	1:52	1:42	**1:07**
Vow	1:55	1:16	1:56	1:55	1:31	1:56	2:07	1:47	1:51	1:53	1:39	**0:49**
Win	0:02	0:02	0:02	0:02	0:02	**0:01**	0:02	0:02	0:03	0:02	0:02	**0:01**
Wis	0:30	0:29	0:28	0:32	0:21	0:36	0:33	0:28	0:24	0:33	0:27	**0:10**
Yea	1:49	1:38	1:57	1:57	2:22	2:15	2:30	1:41	1:47	2:04	2:17	**0:46**
Mean	01:03	0:54	1:02	1:01	1:05	1:06	1:03	0:53	0:55	1:09	01:05	**0:27**

is not necessary to perform as many calculations as with the CC-integrals based FRM (see Example 1).

7. Conclusion

In this paper, we introduce the notion of Choquet-like Copula-based aggregation function (CC-integral). We applied the CC-integral based on several kinds of copulas in FRBCSs, showing that the one based on the minimum t-norm presented the best results among the CC-integrals considered in this work. Furthermore, we highlight that this CC-integral allows to enhance the results of the classical FRM of the WR as well as those of the standard Choquet integral, and provides results that are competitive with those obtained by the Ham_{PA}, offering new possibilities in defining FRMs with similar gain in performance.

In future works, we intend to study the properties satisfied by the CC-integrals. We will also consider CC-integral in a fuzzy interval approach [47–51], as, e.g., in [52,53].

Acknowledgment

This work is supported by Brazilian National Counsel of Technological and Scientific Development CNPq (Proc. 232827/2014-1, 233950/2014-1, 481283/2013-7, 306970/2013-9, 307681/2012-2) and by the Spanish Ministry of Science and Technology under project TIN2016-77356-P

Appendix A. Abbreviations used in the paper

In order to make a paper easy to read and clearer, we present in this appendix a table containing the abbreviations used throughout the paper. In Table A.11 is presented in each line the acronym followed by its meaning.

Table A.11
Abbreviations used in the paper.

Abbreviation	Meaning
CC-integral	Choquet-Like Copula-Based integral
t-norm	Triangular norm
WR	Winning Rule
FRBCSs	Fuzzy Rule-Based Classification Systems
FRM	Fuzzy Reasoning Method
CHC	Crossover using generational elitist selection with Heterogeneous recombination and Cataclysm mutation
FARC-HD	Fuzzy Association Rule-based Classification model for High Dimensional problems

Appendix B. Example of an overlap function that is a non-associative copula

We show that the overlap function O_B is a copula. Consider $x, y, x', y' \in [0, 1]$ such that $x \leq x'$ and $y \leq y'$. First observe that an equivalent definition for O_B is the following one: $O_B(x, y) = \min\{x, y\}\sqrt{\max\{x, y\}}$. Since O_B is commutative, then it is sufficient to consider the following three cases:

Case 1: $x \leq x' \leq y \leq y'$. In this case, one has that

$$O_B(x, y) = x\sqrt{y},$$
$$O_B(x, y') = x\sqrt{y'},$$
$$O_B(x', y) = x'\sqrt{y},$$
$$O_B(x', y') = x'\sqrt{y'}.$$

So, it holds that

$$O_B(x', y') - O_B(x', y) = x'\sqrt{y'} - x'\sqrt{y} = x'(\sqrt{y'} - \sqrt{y}),$$
$$O_B(x, y') - O_B(x, y) = x\sqrt{y'} - x\sqrt{y} = x(\sqrt{y'} - \sqrt{y}).$$

Since $x \leq x'$, then one concludes that

$$O_B(x', y') - O_B(x', y) \geq O_B(x, y') - O_B(x, y).$$

Case 2: $x \le y \le x' \le y'$. In this case, one has that

$$\begin{aligned} O_B(x,y) &= x\sqrt{y},\\ O_B(x,y') &= x\sqrt{y'},\\ O_B(x',y) &= y\sqrt{x'},\\ O_B(x',y') &= x'\sqrt{y'}. \end{aligned}$$

So, it holds that

$$O_B(x',y') - O_B(x,y') = x'\sqrt{y'} - x\sqrt{y'} = \sqrt{y'}(x'-x),$$

and

$$\begin{aligned} O_B(x',y) - O_B(x,y) &= y\sqrt{x'} - x\sqrt{y}\\ &= \sqrt{y}(\sqrt{y}\sqrt{x'} - x)\\ &= \sqrt{y}(\sqrt{yx'} - x). \end{aligned}$$

Since $\sqrt{y} \le \sqrt{y'}$ and $\sqrt{yx'} \le \sqrt{x'x'} = x'$, then it follows that

$$O_B(x',y') - O_B(x,y') \ge O_B(x',y) - O_B(x,y).$$

Case 3: $x \le y \le y' \le x'$. In this case, one has that

$$\begin{aligned} O_B(x,y) &= x\sqrt{y},\\ O_B(x,y') &= x\sqrt{y'},\\ O_B(x',y) &= y\sqrt{x'},\\ O_B(x',y') &= y'\sqrt{x'}. \end{aligned}$$

So, it holds that

$$\begin{aligned} O_B(x',y') - O_B(x',y) &= y'\sqrt{x'} - y\sqrt{x'} = \sqrt{x'}(y'-y),\\ O_B(x,y') - O_B(x,y) &= x\sqrt{y'} - x\sqrt{y} = x(\sqrt{y'} - \sqrt{y}). \end{aligned}$$

Since $x \le x'$ and $(\sqrt{y'} - \sqrt{y}) \le y' - y$, then one concludes that

$$O_B(x',y') - O_B(x',y) \ge O_B(x,y') - O_B(x,y).$$

Therefore, in all cases, $O_B(x',y') + O_B(x,y) \ge O_B(x,y') + O_B(x',y)$, that is, O_B satisfy **(C1)**. Since **(C2)** and **(C3)** are trivially satisfied by O_B then O_B is non-associative copula.

Appendix C. Proofs

This appendix is aimed to present the proof of Proposition 5, which is recalled below.

Proposition 5. *For any copula* C: $[0,1]^2 \to [0,1]$ *and fuzzy measure* $\mathfrak{m} : 2^N \to [0,1]$, $\mathfrak{C}_{\mathfrak{m}}^{C}$ *is increasing* **(A1)**.

Proof. Since $\mathfrak{C}_{\mathfrak{m}}^{C}$ is trivially commutative, then it is sufficient to consider the case when the input $\vec{x}$ is ordered, that is, $x_i = x_{(i)}$, for each $i = 1, \ldots, n$. Also, by the transitivity, is it is sufficient to consider the following cases:

(i) Consider $x_{(j)} \le y \le x_{(j+1)}$, for some $j = 1, \ldots, n-1$. Since, C is a copula, then it holds that:

$$\begin{aligned} &C(x_{(j)}, \mathfrak{m}(A_{(j+1)})) + C(y, \mathfrak{m}(A_{(j)}))\\ &\quad \ge C(x_{(j)}, \mathfrak{m}(A_{(j)})) + C(y, \mathfrak{m}(A_{(j+1)})). \end{aligned}$$

Thus, one has that

$$\begin{aligned} &C(x_{(j)}, \mathfrak{m}(A_{(j)})) - C(x_{(j)}, \mathfrak{m}(A_{(j+1)}))\\ &\quad \le C(y, \mathfrak{m}(A_{(j)})) - C(y, \mathfrak{m}(A_{(j+1)})) \end{aligned}$$

and, therefore, it holds that

$$\begin{aligned} &C(x_{(j)}, \mathfrak{m}(A_{(j)})) - C(x_{(j-1)}, \mathfrak{m}(A_{(j)})) + C(x_{(j+1)}, \mathfrak{m}(A_{(j+1)}))\\ &\quad - C(x_{(j)}, \mathfrak{m}(A_{(j+1)}))\\ &\quad \le C(y, \mathfrak{m}(A_{(j)})) - C(x_{(j-1)}, \mathfrak{m}(A_{(j)})) + C(x_{(j+1)}, \mathfrak{m}(A_{(j+1)}))\\ &\quad - C(y, \mathfrak{m}(A_{(j+1)})). \end{aligned}$$

It follows that

$$\begin{aligned} &\mathfrak{C}_{\mathfrak{m}}^{C}(x_1, \ldots, x_j, \ldots, x_n)\\ &= \sum_{i=1}^{n} C(x_{(i)}, \mathfrak{m}(A_{(i)})) - C(x_{(i-1)}, \mathfrak{m}(A_{(i)}))\\ &\le \left(\sum_{i=1}^{j-1} C(x_{(i)}, \mathfrak{m}(A_{(i)})) - C(x_{(i-1)}, \mathfrak{m}(A_{(i)}))\right)\\ &\quad + \left(C(y, \mathfrak{m}(A_{(j)})) - C(x_{(j-1)}, \mathfrak{m}(A_{(j)}))\right)\\ &\quad + \left(C(x_{(j+1)}, \mathfrak{m}(A_{(j+1)})) - C(y, \mathfrak{m}(A_{(j+1)}))\right)\\ &\quad + \left(\sum_{i=j+2}^{n} C(x_{(i)}, \mathfrak{m}(A_{(i)})) - C(x_{(i-1)}, \mathfrak{m}(A_{(i)}))\right)\\ &= \mathfrak{C}_{\mathfrak{m}}^{C}(x_1, \ldots, x_{j-1}, y, x_{j+1}, \ldots, x_n). \end{aligned}$$

(ii) Consider $x_{(n)} \le y$. Since C is increasing, it holds that

$$C(x_{(n)}, \mathfrak{m}(A_{(n)})) \le C(y, \mathfrak{m}(A_{(n)})),$$

and, therefore, one has that

$$\begin{aligned} &C(x_{(n)}, \mathfrak{m}(A_{(n)})) - C(x_i, A_{(n-1)}, \mathfrak{m}(A_{(n)}))\\ &\quad \le C(y, \mathfrak{m}(A_{(n)})) - C(x_i, A_{(n-1)}, \mathfrak{m}(A_{(n)})). \end{aligned}$$

It follows that:

$$\begin{aligned} \mathfrak{C}_{\mathfrak{m}}^{C}(x_1, \ldots, x_n) &= \sum_{i=1}^{n} C(x_{(i)}, \mathfrak{m}(A_{(i)})) - C(x_{(i-1)}, \mathfrak{m}(A_{(i)}))\\ &\le \left(\sum_{i=1}^{n-1} C(x_{(i)}, \mathfrak{m}(A_{(i)})) - C(x_{(i-1)}, \mathfrak{m}(A_{(i)}))\right)\\ &\quad + \left(C(y, \mathfrak{m}(A_{(n)})) - C(x_{(n-1)}, \mathfrak{m}(A_{(n)}))\right)\\ &= \mathfrak{C}_{\mathfrak{m}}^{C}(x_1, \ldots, x_{n-1}, y). \end{aligned}$$

□

References

[1] E. Alpaydin, Introduction to Machine Learning, 2nd ed., The MIT Press, 2010.

[2] S. Patidar, R.B. Pachori, U.R. Acharya, Automated diagnosis of coronary artery disease using tunable-q wavelet transform applied on heart rate signals, Knowl. Based Syst. 82 (2015) 1–10.

[3] M. Galar, J. Derrac, D. Peralta, I. Triguero, D. Paternain, C. Lopez-Molina, S. García, J.M. Benítez, M. Pagola, E. Barrenechea, H. Bustince, F. Herrera, A survey of fingerprint classification part i: taxonomies on feature extraction methods and learning models, Knowl. Based Syst. 81 (2015a) 76–97.

[4] M. Galar, J. Derrac, D. Peralta, I. Triguero, D. Paternain, C. Lopez-Molina, S. García, J.M. Benítez, M. Pagola, E. Barrenechea, H. Bustince, F. Herrera, A survey of fingerprint classification part II: experimental analysis and ensemble proposal, Knowl. Based Syst. 81 (2015b) 98–116.

[5] C.-H. Huang, C.-H. Lin, Multiple harmonic-source classification using a self-organization feature map network with voltage-current wavelet transformation patterns, Appl. Math. Model. 39 (19) (2015) 5849–5861.

[6] G. Asencio-Cortés, F. Martínez-Álvarez, A. Morales-Esteban, J. Reyes, A sensitivity study of seismicity indicators in supervised learning to improve earthquake prediction, Knowl. Based Syst. 101 (2016) 15–30.

[7] H. Ishibuchi, Y. Nojima, Pattern classification with linguistic rules, in: H. Bustince, F. Herrera, J. Montero (Eds.), Fuzzy Sets and Their Extensions: Representation, Aggregation and Models, Vol. 220 of Studies in Fuzziness and Soft Computing, Springer, Berlin, 2008, pp. 377–395.

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

[9] J.A. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, H. Bustince, Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system, Appl. Soft Comput. 20 (2014) 103–111.

[10] O. Cordón, M.J. del Jesus, F. Herrera, A proposal on reasoning methods in fuzzy rule-based classification systems, Int. J. Approximate Reasoning 20 (1) (1999) 21–45.
[11] G. Beliakov, A. Pradera, T. Calvo, Aggregation Functions: A Guide for Practitioners, Springer, Berlin, 2007.
[12] E. Barrenechea, H. Bustince, J. Fernandez, D. Paternain, J.A. Sanz, Using the Choquet integral in the fuzzy reasoning method of fuzzy rule-based classification systems, Axioms 2 (2) (2013) 208–223.
[13] G. Choquet, Theory of capacities, Annales de l'Institut Fourier 5(1953-1954) 131–295.
[14] L.J. Eshelman, The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination, in: G.J.E. Rawlings (Ed.), Foundations of Genetic Algorithms, Morgan Kaufmann, San Francisco, 1991, pp. 265–283.
[15] G. Lucca, J. Sanz, G.P. Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, H.B. Sola, Pre-aggregation functions: construction and an application, IEEE Trans. Fuzzy Syst. 24 (2) (2016) 260–272.
[16] H. Bustince, J. Fernandez, A. Kolesárová, R. Mesiar, Directional monotonicity of fusion functions, Eur. J. Oper. Res. 244 (1) (2015) 300–308.
[17] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms, Kluwer Academic Publisher, Dordrecht, 2000.
[18] H. Bustince, J. Fernandez, R. Mesiar, J. Montero, R. Orduna, Overlap functions, Nonlinear Anal. 72 (3-4) (2010) 1488–1499.
[19] B.C. Bedregal, G.P. Dimuro, H. Bustince, E. Barrenechea, New results on overlap and grouping functions, Inf. Sci. 249 (2013) 148–170.
[20] G.P. Dimuro, B. Bedregal, On residual implications derived from overlap functions, Inf. Sci. 312 (2015a) 78–88.
[21] G.P. Dimuro, B. Bedregal, On the laws of contraposition for residual implications derived from overlap functions, in: IEEE International Conference on Fuzzy Systems, Los Alamitos, 2015b, pp. 1–7.
[22] C. Alsina, M.J. Frank, B. Schweizer, Associative Functions: Triangular Norms and Copulas, World Scientific Publishing Company, Singapore, 2006.
[23] J. Alcalá-Fdez, L. Sánchez, S. García, M. Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernández, F. Herrera, Keel: a software tool to assess evolutionary algorithms for data mining problems, Soft comput. 13 (3) (2009) 307–318.
[24] S. García, A. Fernández, J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics–based machine learning: accuracy and interpretability, Soft comput. 13 (10) (2009) 959–977.
[25] D. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, 2nd ed., Chapman & Hall/CRC, 2006.
[26] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[27] G. Mayor, E. Trillas, On the representation of some aggregation functions, in: Proceedings of IEEE International Symposium on Multiple-Valued Logic, IEEE, Los Alamitos, 1986, pp. 111–114.
[28] B. Schweizer, A. Sklar, Probabilistic Metric Spaces, North-Holland, New York, 1983.
[29] D. Gómez, J.T. Rodríguez, J. Montero, H. Bustince, E. Barrenechea, n-dimensional overlap functions, Fuzzy Sets Syst. 287 (2016) 57–75. Theme: Aggregation Operations.
[30] M. Elkano, M. Galar, J. Sanz, H. Bustince, 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, Inf. Sci. 332 (2016) 94–114.
[31] M. Elkano, M. Galar, J. Sanz, A. Fernández, E. Barrenechea, F. Herrera, H. Bustince, Enhancing multi-class classification in FARC-HD fuzzy classifier: on the synergy between n-dimensional overlap functions and decomposition strategies, IEEE Trans. Fuzzy Syst. 23 (5) (2015) 1562–1580.
[32] R.B. Nelsen, An Introduction to Copulas, Lecture Notes in Statistics, 139, Springer, New York, 1999.
[33] G.P. Dimuro, B. Bedregal, Archimedean overlap functions: the ordinal sum and the cancellation, idempotency and limiting properties, Fuzzy Sets Syst. 252 (2014) 39–54.
[34] G.P. Dimuro, B. Bedregal, Additive generators of overlap functions, in: H. Bustince, J. Fernandez, R. Mesiar, T. Calvo (Eds.), Aggregation Functions in Theory and in Practice, Vol. 228 of Advances in Intelligent Systems and Computing, Springer, Berlin, 2013, pp. 167–178.
[35] G.P. Dimuro, B. Bedregal, H. Bustince, M.J. Asiáin, R. Mesiar, On additive generators of overlap functions, Fuzzy Sets Syst. 287 (2016) 76–96. Theme: Aggregation Operations
[36] G. Lucca, G.P. Dimuro, V. Mattos, B. Bedregal, H. Bustince, J.A. Sanz, 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, 2015, pp. 1–8.
[37] T. Murofushi, M. Sugeno, M. Machida, Non-monotonic fuzzy measures and the Choquet integral, Fuzzy Sets Syst. 64 (1) (1994) 73–86.
[38] J. Alcala-Fdez, R. Alcala, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE Trans. Fuzzy Syst. 19 (5) (2011) 857–872.
[39] H. Ishibuchi, T. Nakashima, M. Nii, Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining, Springer-Verlag, Berlin, 2004.
[40] H. Ishibuchi, T. Nakashima, Effect of rule weights in fuzzy rule-based classification systems, IEEE Trans. Fuzzy Syst. 9 (4) (2001) 506–515.
[41] R. Alcala, J. Alcala-Fdez, F. Herrera, A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection, IEEE Trans. Fuzzy Syst. 15 (4) (2007) 616–635.
[42] J. Sanz, A. Fernández, H. Bustince, F. Herrera, IVTURS: a linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection, IEEE Trans. Fuzzy Syst. 21 (3) (2013) 399–411.
[43] F. Herrera, M. Lozano, A.M. Sánchez, A taxonomy for the crossover operator for real-coded genetic algorithms: an experimental study, Int. J. Intell. Syst. 18 (3) (2003) 309–338.
[44] J.L. Hodges, E.L. Lehmann, Ranks methods for combination of independent experiments in analysis of variance, Ann. Math. Stat. 33 (1962) 482–497.
[45] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (1979) 65–70.
[46] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics 1 (1945) 80–83.
[47] B.C. Bedregal, G.P. Dimuro, R.H.S. Reiser, An approach to interval-valued R-implications and automorphisms, in: J.P. Carvalho, D. Dubois, U. Kaymak, J.M.d. C. Sousa (Eds.), Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference, IFSA/EUSFLAT, 2009, pp. 1–6.
[48] B.C. Bedregal, G.P. Dimuro, R.H.N. Santiago, R.H.S. Reiser, On interval fuzzy S-implications, Inf. Sci. 180 (8) (2010) 1373–1389.
[49] G.P. Dimuro, On interval fuzzy numbers, in: 2011 Workshop-School on Theoretical Computer Science, WEIT 2011, IEEE, Los Alamitos, 2011, pp. 3–8.
[50] G.P. Dimuro, B.C. Bedregal, R.H.N. Santiago, R.H.S. Reiser, Interval additive generators of interval t-norms and interval t-conorms, Inf. Sci. 181 (18) (2011) 3898–3916.
[51] R.H.S. Reiser, G.P. Dimuro, B.C. Bedregal, R.H.N. Santiago, Interval valued QL-implications, in: D. Leivant, R. Queiroz (Eds.), Logic, Language, Information and Computation, no. 4576 of Lecture Notes in Computer Science, Springer, Berlin, 2007, pp. 307–321.
[52] H. Bustince, M. Galar, B. Bedregal, A. Kolesárová, R. Mesiar, A new approach to interval-valued Choquet integrals and the problem of ordering in interval-valued fuzzy set applications, IEEE Trans. Fuzzy Syst. 21 (6) (2013) 1150–1162.
[53] J.A. Sanz, A. Fernández, H. Bustince, F. Herrera, Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning, Inf. Sci. 180 (19) (2010) 3674–3685.

3 A proposal for tuning the alpha parameter in C_αC-integrals for application in fuzzy rule-based classification systems

Related publication:

- G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, and H. Bustince, "A proposal for tuning the alpha parameter in C_αC-integrals for application in fuzzy rule-based classification systems", Natural Computing (special issue BRACIS – 2016) (Accepted).
  - Journal: Natural Computing – Special Issue BRACIS
  - Status: Published[1]
  - Impact Factor (JCR 2016): 0,778
  - Knowledge Area:
    - Computer Science: Ranking 111/133 (Q4)
    - Artificial Intelligence: Ranking 111/133 (Q4)
    - Theory & Methods: Ranking 80/104 (Q4)

[1] Paper accepted, however, have not been published yet.

Artículo eliminado por restricciones de derechos de autor

Publicado en:

Lucca, G., Sanz, J. A., Dimuro, G. P., Bedregal, B., & Bustince, H. (2018). A proposal for tuning the α parameter in CαC -integrals for application in fuzzy rule-based classification systems. *Natural Computing*, 1-14. http://doi.org/10.1007/s11047-018-9678-x

4 C_F-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems

- G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, H. Bustince and R. Mesiar, "C_f-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems", Information Sciences 435 (2018) 94 – 110.
  - Journal: Information Sciences
  - Status: Published
  - Impact Factor (JCR 2016): 4,832
  - Knowledge Area:
    * Computer Science: Ranking 7/146 (Q1)
    * Information Systems: Ranking 7/146 (Q1)

Information Sciences 435 (2018) 94–110

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

C_F-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems

Giancarlo Lucca[a,*], José Antonio Sanz[a,b], Graçaliz Pereira Dimuro[b,c], Benjamín Bedregal[d], Humberto Bustince[a,b], Radko Mesiar[e,f]

[a] Depto. of Automática y Computación, Universidad Publica de Navarra, Campus Arrosadía s/n, P.O. Box 31006, Pamplona, Spain
[b] Institute of Smart Cities, Universidad Publica de Navarra, Campus Arrosadía, Navarra, 31006, Spain
[c] Centro de Ciências Computacionais, Universidade Federal do Rio Grande, Av. Itália km 08, Campus Carreiros, Rio Grande, 96201-900, Brazil
[d] Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte, Campus Universitário, Natal, 59072-970, Brazil
[e] Department of Mathematics, Slovak University of Technology, Radlinského 11, Bratislava 81005, Slovakia
[f] Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague 18208, Czech Republic

A R T I C L E I N F O

Article history:
Received 28 April 2017
Revised 29 September 2017
Accepted 23 December 2017
Available online 24 December 2017

Keywords:
C_F-integral
Choquet integral
Pre-aggregation function
Classification problems
Fuzzy reasoning method
Fuzzy rule-based classification systems

A B S T R A C T

This paper introduces the family of C_F-integrals, which are pre-aggregations functions that generalizes the Choquet integral considering a bivariate function F that is left 0-absorbent. We show that C_F-integrals are $\vec{1}$-pre-aggregation functions, studying in which conditions they are idempotent and/or averaging functions. This characterization is an important issue of our approach, since we apply these functions in the Fuzzy Reasoning Method (FRM) of a fuzzy rule-based classification system and, in the literature, it is possible to observe that non-averaging aggregation functions usually provide better results. We carry out a study with several subfamilies of C_F-integrals having averaging or non-averaging characteristics. As expected, the proposed non-averaging C_F-integrals obtain more accurate results than the averaging ones, thus, offering new possibilities for aggregating accurately the information in the FRM. Furthermore, it allows us to enhance the results of classical FRMs like the winning rule and the additive combination.

1. Introduction

An effective approach to handle classification problems [25] is through the application of the Fuzzy Rule-Based Classification Systems (FRBCSs) [35], since they provide the user with interpretable models by using linguist labels in their rules and, moreover, achieving accurate results. FRBCSs have been applied in several problems, including real-time vehicle classification [57], health [52] or economy [49], among many others.

A key component in any FRBCS is the Fuzzy Reasoning Method (FRM) [14], which determines how the information learned in form of fuzzy rules will be used to classify new examples. A crucial point in any FRM is the way to obtain the

* Corresponding author.
E-mail addresses: lucca.112793@e.unavarra.es (G. Lucca), joseantonio.sanz@unavarra.es (J. Antonio Sanz), gracalizdimuro@furg.br, gracaliz.pereira@unavarra.es (G.P. Dimuro), bedregal@dimap.ufrn.br (B. Bedregal), bustince@unavarra.es (H. Bustince), mesiar@math.sk (R. Mesiar).

https://doi.org/10.1016/j.ins.2017.12.029

information associated with each class of the problem. This is done by applying an aggregation [8,30,44] or, more recently, a pre-aggregation [11,12,22,40,42] function over the local information given by each fired rule of the FRBCS.

In the literature, it is possible to find classical FRMs that consider, as the aggregation operator, the maximum (Winning Rule – WR) or the normalized sum (Additive Combination – AC). The first method takes into consideration only one rule (the one having the maximum compatibility with the example to be classified) and, obviously, has an averaging and idempotent behavior. On the other hand, the second method aggregates the information of all triggered rules and it has neither an averaging nor an idempotent behavior. Usually, the FRM of AC provides better performance than that the FRM of the WR, as it can be observed widely in the literature, since the most accurate FRBCSs currently (FURIA [33], IVTURS [51] and FARC-HD [2]) make use of the AC.

Recently, several works were proposed to apply aggregation and pre-aggregation functions (with averaging and idempotent characteristics) to aggregate the local information associated with each rule. The initial idea was proposed by Barrenechea et al. [4], where the Choquet integral [13] was used to perform this aggregation in a way that also took into account the correlation between the rules. After that, this method was improved by Lucca et al. [40], introducing the concept of pre-aggregation function, which is a generalization of the Choquet integral where the product operator of this function is replaced by a t-norm [36]. In [41], the Choquet integral in its expanded form was generalized using copula functions [3], instead of the product operator,obtaining aggregation functions called CC-integrals.

In this paper, the product operator of the Choquet integral is replaced by a more general function F: $[0, 1]^2 \to [0, 1]$. We study which are the minimal requirements that this function F must satisfy so that the obtained generalization of the Choquet integral is a pre-aggregation function. Specifically, we have found that the key property to achieve this is the presence of 0 as a left annihilator element, in which case the function F is called left 0-absorbent.

The general aim is to apply such pre-aggregation functions in the FRM of a FRBCS, searching for more flexible ways of aggregating information. In this manner, it is possible to make an in-depth analysis of the their performances according to their averaging or non-averaging behavior. Observe that the non-averaging behavior is a novel approach, since we have not considered it in our previous works.

Then, the first objective of this paper is the definition of the concept of C_F-integral, which is a generalization of the Choquet integral based on a left 0-absorbent function F satisfying a minimal set of properties that guarantees that any C_F-integral is a pre-aggregation function. Secondly, we analyze under which conditions such C_F-integrals are idempotent and/or averaging pre-aggregation functions. In the sequence, we study subfamilies of C_F-integrals, considering left 0-absorbent functions F that are (I) t-norms [36], (II) overlap functions [5,10,19,20,23,24], (III) copulas [3] that are neither t-norms nor overlap functions, (IV) other kinds of aggregation functions and (V) pre-aggregation functions.

As done in [4,40,41], we apply this generalization in the FRM of FRBCSs and we conduct an experimental study composed of two steps. The first one is based on C_F-integrals having averaging characteristics, where we compare them among themselves in order to choose the representative for this family. After that, we compare this representative against the classical FRM of WR, the standard Choquet integral, the best pre-aggregation achieved in [40] and the best CC-integral obtained in [41].

The second part of this analysis is concerned with C_F-integrals having non-averaging characteristics. As done in the first part of the experimental study, firstly we determine the best function of this family and compare it against the classical non-averaging FRMs of AC and probabilistic sum.

The experimental study was performed considering 33 datasets that are available in the KEEL database repository [1]. The standard accuracy rate is used to measure the performance of the classifiers and the results are supported by appropriate statistical tests [15,28,53].

The paper is organized in the following way. Section 2 is aimed at introducing the basic concepts that are necessary to understand the paper. The concept of C_F-integral is introduced in Section 3, where we analyze several properties, such as idempotency and averaging behaviors. The Section 4 presents the methodology to build a generalized FRM of FRBCSs using different C_F integrals, configurations of the classifier used in this paper and the experimental framework. In Section 5 we show the experimental study, showing the results achieved in test considering this new approach, and the appropriate analysis. The conclusions are drawn in Section 6.

2. Basic concepts

This section presents the preliminary concepts that are used in the development of this work. In our approach, the basic property that is considered for any bivariate function defined on [0, 1], is the following.

Definition 1. A bivariate function F: $[0, 1]^2 \to [0, 1]$ with 0 as left annihilator element, that is, satisfying:

(**LAE**) $\forall y \in [0,1] : F(0,y) = 0$,

is said to be left 0-absorbent.

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.

Now, we recall the concepts of aggregation and pre-aggregation functions, and specific types of aggregation functions, such as t-norms, overlap and copulas.

Definition 2 [8,30,44]**.** A function A: $[0, 1]^n \to [0, 1]$ is an aggregation function if the following conditions hold:

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

Definition 3 [36]**.** An aggregation function T: $[0, 1]^2 \to [0, 1]$ is said to be a t-norm if, for all $x, y, z \in [0, 1]$, the following conditions hold:

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

Definition 4 [10,19,21]**.** A function O: $[0, 1]^2 \to [0, 1]$ is an overlap function if, for all $x, y, z \in [0, 1]$, the following conditions hold:

(**O1**) O is commutative;
(**O2**) $O(x, y) = 0$ if and only if $x = 0$ or $y = 0$;
(**O3**) $O(x, y) = 1$ if and only if $x = y = 1$;
(**O4**) O is increasing;
(**O5**) O is continuous.

Definition 5 [3]**.** A bivariate function C: $[0, 1]^2 \to [0, 1]$ is said to be a copula if, for all $x, x', y, y' \in [0, 1]$ with $x \le x'$ and $y \le y'$, the following conditions hold:

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

Observe that overlap functions, t-norms and copulas can be extended to n-ary functions (see, e.g., [26,27,29,36]).

Definition 6 [9]**.** Let $\vec{r} = (r_1, \ldots, r_n)$ be a real n-dimensional vector, $\vec{r} \neq \vec{0}$. A function F: $[0, 1]^n \to [0, 1]$ is $\vec{r}$-increasing if, for all vectors $(x_1, \ldots, x_n) \in [0, 1]^n$ and for all $c > 0$ such that $(x_1 + cr_1, \ldots, x_n + cr_n) \in [0, 1]^n$, it holds

$$F(x_1 + cr_1, \ldots, x_n + cr_n) \ge F(x_1, \ldots, x_n). \tag{1}$$

Similarly, ones defines an $\vec{r}$-decreasing function.

Definition 7 [22,40]**.** Let $\vec{r} = (r_1, \ldots, r_n)$ be a real n-dimensional vector, $\vec{r} \neq \vec{0}$. A function PA: $[0, 1]^n \to [0, 1]$ is said to be an n-ary pre-aggregation function if it satisfies **(A2)** and it is $\vec{r}$-increasing. We say that PA is an $\vec{r}$-pre-aggregation function.

Example 1. In this example, we analyze the basic properties (LAE), (RNE) and (LC) for some pre-aggregation functions.

1. The function F_{NA}: $[0, 1]^2 \to [0, 1]$, defined by:

$$F_{NA}(x, y) = \begin{cases} x & \text{if } x \le y \\ \min\{\frac{x}{2}, y\} & \text{otherwise} \end{cases}$$

is a left 0-absorbent pre-aggregation function. In fact, it is immediate that F_{NA} satisfies **(A2)**. Moreover, consider $x, y \in [0, 1]$ and $c > 0$ such that $y + c \in [0, 1]$. To show that F_{NA} is (0, 1)-increasing, consider the following cases:

$x \le y$: In this case, it holds that $x \le y + c$. It follows that:

$$F_{NA}(x, y + c) = x = F_{NA}(x, y).$$

$x > y$: If $x > y + c$, then one has that:

$$F_{NA}(x, y + c) = \min\left\{\frac{x}{2}, y + c\right\} \ge \min\left\{\frac{x}{2}, y\right\} = F_{NA}(x, y).$$

Now suppose that $x \le y + c$. Then, it follows that:

$$F_{NA}(x, y + c) = x > \min\left\{\frac{x}{2}, y\right\} = F_{NA}(x, y).$$

[1] For an increasing (decreasing) function we do not mean a strictly increasing (decreasing) function.

Thus, F_{NA} is a (0, 1)-pre-aggregation function. In fact, F_{NA} is $\vec{r}$-increasing whenever the non-zero vector $\vec{r} = (r_1, r_2)$ satisfies $r_2 \geq r_1 \geq 0$. Hence, F_{NA} is also (1, 1)-increasing. Finally, observe that F_{NA} is left 0-absorbing **(LAE)**, since $F_{NA}(0, y) = 0$, for all $y \in [0, 1]$. Additionally, F_{NA} satisfies **(RNE)** and **(LC)**.

2. Consider now the function $F_{NA1}\colon [0, 1]^2 \to [0, 1]$, defined by

$$F_{NA1}(x, y) = \begin{cases} \frac{x+y}{2} & \text{if } x \leq y \\ \min\{\frac{x}{2}, y\} & \text{otherwise.} \end{cases}$$

Similarly, one can show that F_{NA1} is a (0, 1)-pre-aggregation function. However, it is not left 0-absorbent, since, for example $F_{NA1}(0, 0.2) = 0.1 \neq 0.2$. Moreover, F_{NA1} satisfies neither **RNE** nor **LC**.

3. Consider a slight modification in the definition of the function F_{NA1}, obtaining the function $F_{NA2}\colon [0, 1]^2 \to [0, 1]$, defined by

$$F_{NA2}(x, y) = \begin{cases} 0 & \text{if } x = 0 \\ \frac{x+y}{2} & \text{if } 0 < x \leq y \\ \min\{\frac{x}{2}, y\} & \text{otherwise.} \end{cases}$$

Again, analogously, it is possible to show that F_{NA2} is a (0, 1)-pre-aggregation function and it is immediate that F_{NA2} is left 0-absorbent **(LAE)**. However, F_{NA2} does not satisfy neither **RNE** nor **LC**.

4. Similarly, one can modify F_{NA2} into $F_{NA3}\colon [0, 1]^2 \to [0, 1]$, defined by

$$F_{NA3}(x, y) = \begin{cases} x & \text{if } y = 1 \\ F_{NA2}(x, y) & \text{otherwise,} \end{cases}$$

and then F_{NA3} satisfies all three properties **(LAE), (RNE)** and **(LC)**. However, although F_{NA3} satisfies **(A2)**, it is not (0, 1) increasing, since, for example, for $x = 0.4,\ y = 0.8$ and $c = 0.2$, one has that $F_{NA3}(0.4, 0.8 + 0.2) = 0.4 <= 0.6 = \frac{0.4+0.8}{2} = F_{NA3}(0.4, 0.8)$.

Finally, it is worth mentioning that F_{NA}, F_{NA1} and F_{NA2} are all (1, 0)-increasing, but F_{NA3} is not.

Definition 8 [40, Thorem 4.1]**.** Let $\vec{r} = (r_1, \ldots, r_n)$ be a real n-dimensional vector, $\vec{r} \neq \vec{0}$. An $\vec{r}$-pre-aggregation function PA: $[0, 1]^n \to [0, 1]$ is averaging if

$$\min \leq PA \leq \max .$$

Observe that there exist pre-aggregation functions that are averaging but are not aggregation functions, for example, the mode.

Remark 1. Observe that all $\vec{r}$-pre-aggregation functions PA that are averaging are also idempotent. However the converse does not hold. For example, consider the (0, 1)-pre-aggregation function F_{NA} of Example 1, which is obviously idempotent. F_{NA} is not averaging, since, for example:

$$F_{NA}(0.5, 0.4) = \min\{0.25, 0.4\} = 0.25 < \min\{0.5, 0.4\}.$$

Fuzzy integrals are well known aggregation operators. However, their use is not easy as their interpretation is not straightforward. In [54], Torra and Narukawa study the interpretation of fuzzy integrals, focusing on Sugeno ones, showing their application in fuzzy inference systems when the rules are not independent, for control problems.

The Choquet integral is a type of aggregation function which considers the relationship among the elements that are being aggregated, providing the relevance of a coalition by fuzzy measures.

In what follows, denote $N = \{1, \ldots, n\}$, for $n > 0$.

Definition 9 [13,47]**.** A function $\mathfrak{m} : 2^N \to [0, 1]$ is said to be a fuzzy measure if, for all $X, Y \subseteq N$, the following conditions hold:

($\mathfrak{m}$1) Increasingness: if $X \subseteq Y$, then $\mathfrak{m}(X) \leq \mathfrak{m}(Y)$;
($\mathfrak{m}$2) Boundary conditions: $\mathfrak{m}(\emptyset) = 0$ and $\mathfrak{m}(N) = 1$.

In this paper, we have selected the power measure according to the results in [4,11,39,40,43]. The power measure is defined as $\mathfrak{m}_{PM} : 2^N \to [0, 1]$, which is given, for all $X \subseteq N$, by

$$\mathfrak{m}_{PM}(X) = \left(\frac{|X|}{n}\right)^q, \text{ with } q > 0. \tag{2}$$

Definition 10 [8, Definition 1.74]**.** Let $\mathfrak{m} : 2^N \to [0, 1]$ be a fuzzy measure. The discrete Choquet integral is the function $\mathfrak{C}_{\mathfrak{m}} : [0, 1]^n \to [0, 1]$, defined, for all of $\vec{x} = (x_1, \ldots, x_n) \in [0, 1]^n$, by:

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

where $\left(x_{(1)}, \ldots, x_{(n)}\right)$ is an increasing permutation on the input $\vec{x}$, that is, $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$, where $x_{(0)} = 0$ and $A_{(i)} = \{(i), \ldots, (n)\}$ is the subset of indices corresponding to the $n - i + 1$ largest components of $\vec{x}$.

3. Construction of pre-agregation functions using Choquet integrals and left 0-absorbent functions

In [40], we introduced the concept of a pre-aggregation function, presenting a construction method of idempotent and averaging pre-aggregation functions by means of the Choquet integral. To do it, we replace the product operation in Eq. (3) by functions F that are (1, 0)-pre-aggregation functions satisfying **(LAE), (RNE)** and **(LC)** (see [40, Theorem 4.1]). The application shown in that paper considered just the case when F is a t-norm, which obviously satisfies those three properties.

In this paper, we intend to propose a more general way for this construction, since we do not require F to be an (1, 0)-pre-aggregation function. That is, just the conditions **(LAE)** and **(RNE)** are necessary to have idempotent pre-aggregation functions. In case we want to obtain also averaging pre-aggregation functions the functions F also have to fulfill the **(LC)** property.

In the following, we present the method for constructing a family of pre-aggregation functions defined by generalizing the discrete Choquet Integral using left 0-absorbent functions F: $[0, 1]^2 \to [0, 1]$, obtaining the so-called C_F-integrals.

Definition 11. 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 Choquet-like integral based on F with respect to $\mathfrak{m}$, called C_F-integral, is the function $\mathfrak{C}_{\mathfrak{m}}^F : [0, 1]^n \to [0, 1]$, defined, for all $x \in [0, 1]^n$, by

$$\mathfrak{C}_{\mathfrak{m}}^F(\vec{x}) = \min\left\{1, \sum_{i=1}^{n} F\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\}, \tag{4}$$

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

Proposition 1. *$\mathfrak{C}_{\mathfrak{m}}^F$ is well defined, for any function F: $[0, 1]^2 \to [0, 1]$ and fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$.*

Proof. It is immediate. □

Remark 2. There are some other approaches presenting Choquet-like integrals or generalizations of the Choquet integrals, mostly not restricted to discrete domains. In [45], Mesiar introduced some Choquet-like integrals defined in terms of pseudo-addition and pseudo-multiplication, presenting similar properties than those of the standard Choquet Integral. Murofushi and Sugeno [46] defined the fuzzy t-conorm integral, which is a generalization of Sugeno integral and Choquet integral based on a t-system composed by continuous t-conorms and a continuous t-norm which all are either idempotent (then the Sugeno integral is obtained), or all are Archimedean (then a transform of the Choquet integral is obtained). Differently, our Choquet-like integrals, introduced in Definition 11, are obtained in the context of the discrete Choquet integral. They are based on the standard summation + (i.e., in this item less general than the two above mentioned types of integrals) and on a rather general function F (much more general than the pseudo-multiplications considered in the two above integrals).

Remark 3. In the literature, there exist also other kinds of integrals not defined in terms of the Choquet/Sugeno integral but related to them. For example, Wang et al. [56] introduced a nonlinear integral with respect to set functions vanishing at the empty set which need not be monotone. Observe that if a fuzzy measure $\mathfrak{m}$ is considered, then this integral coincides with the concave integral introduced by Lehrer [37] (see also [38]). This integral is just the Choquet integral whenever the considered fuzzy measure $\mathfrak{m}$ is superadditive, i.e., if $\mathfrak{m}(E_1 \cup E_2) + \mathfrak{m}(E_1 \cap E_2) \ge \mathfrak{m}(E_1) + m(E_2)$, for any sets E_1, $E_2 \subseteq N$. In particular, these equalities hold when $\mathfrak{m}$ is a belief measure [55]. We point out that Wang et al.'s integral coincides with our C_F-integral only if F is the standard product and $\mathfrak{m}$ is a supermodular fuzzy measure (and then they are just the standard Choquet integral, as the authors showed in [56, Corollary 2]).

Proposition 2. *For any fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$ and left 0-absorbent **(RNE)**-function F: $[0, 1]^2 \to [0, 1]$, $\mathfrak{C}_{\mathfrak{m}}^F$ is idempotent.*

Proof. Considering $\vec{x} = (x, \ldots, x) \in [0, 1]^n$, one has that:

$$\begin{aligned}\mathfrak{C}_{\mathfrak{m}}^{(F)}(\vec{x}) &= \min\left\{1, F(x - 0, 1) + \sum_{i=2}^{n} F\left(x - x, \mathfrak{m}(A_{(i)})\right)\right\} \text{ by Eq. (4)} \\ &= \min\{1, x\} \quad \text{by } (\mathbf{RNE}) \text{ and } (\mathbf{LAE}) \\ &= x.\end{aligned}$$

□

Proposition 3. *For any fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$ and F: $[0, 1]^2 \to [0, 1]$ satisfying **(RNE)**, it holds that $\mathfrak{C}_{\mathfrak{m}}^F \ge \min$.*

Proof. Let $(x_{(1)}, \ldots, x_{(i-1)}, x_{(i)}, \ldots, x_{(n)})$ be an increasing permutation of $\vec{x} \in [0,1]^n$. It follows that:

$$\begin{aligned}
\mathfrak{C}_{\mathfrak{m}}^{F}(\vec{x}) &= \min\left\{1, F(x_{(1)} - 0, 1) + \sum_{i=2}^{n} F(x_{(i)} - x_{(i-1)}, \mathfrak{m}(A_{(i)}))\right\} \text{ by Eq. (4)} \\
&= \min\left\{1, x_{(1)} + \sum_{i=2}^{n} F(x_{(i)} - x_{(i-1)}, \mathfrak{m}(A_{(i)}))\right\} \text{ by}(\textbf{RNE}) \\
&\geq x_{(1)} \\
&= \min \vec{x}.
\end{aligned}$$

□

Proposition 4. *For any fuzzy measure* $\mathfrak{m}: 2^N \to [0,1]$ *and* $F: [0,1]^2 \to [0,1]$ *satisfying **(LC)**, it holds that* $\mathfrak{C}_{\mathfrak{m}}^{F} \leq \max$.

Proof. Let $(x_{(1)}, \ldots, x_{(i-1)}, x_{(i)}, \ldots, x_{(n)})$ be an increasing permutation of $\vec{x} \in [0,1]^n$. It follows that:

$$\begin{aligned}
\mathfrak{C}_{\mathfrak{m}}^{F}(\vec{x}) &= \min\left\{1, \sum_{i=1}^{n} F(x_{(i)} - x_{(i-1)}, \mathfrak{m}(A_{(i)}))\right\} \text{ by Eq. (4)} \\
&\leq \min\left\{1, \sum_{i=1}^{n} \left(x_{(i)} - x_{(i-1)}\right)\right\} \text{ by}(\textbf{LC}) \\
&= \min\{1, x_{(n)}\} \\
&= x_{(n)} \\
&= \max \vec{x}.
\end{aligned}$$

□

Proposition 5. *For any fuzzy measure* $\mathfrak{m}: 2^N \to [0,1]$ *and left 0-absorbent function* $F: [0,1]^2 \to [0,1]$, *if F satisfies **(A2, ii)**, then* $\mathfrak{C}_{\mathfrak{m}}^{F}$ *satisfies the boundary conditions **(A2)**.*

Proof. Consider $\vec{0} = (0, \ldots, 0) \in [0,1]^n$ and $\vec{1} = (1, \ldots, 1) \in [0,1]^n$. It follows that:

$$\begin{aligned}
\mathfrak{C}_{\mathfrak{m}}^{F}(\vec{0}) &= \min\left\{1, \sum_{i=1}^{n} F(0 - 0, \mathfrak{m}(A_{(i)}))\right\} \text{ by Eq. (4)} \\
&= 0 \quad \text{by}(\textbf{LAE})
\end{aligned}$$

and

$$\begin{aligned}
\mathfrak{C}_{\mathfrak{m}}^{F}(\vec{1}) &= \min\{1, F(1 - 0, \mathfrak{m}(A_{(1)})) + \sum_{i=2}^{n} F(1 - 1, \mathfrak{m}(A_{(i)}))\} \text{ by Eq. (4)} \\
&= \min\{1, F(1,1) + \sum_{i=2}^{n} F(0, \mathfrak{m}(A_{(i)}))\} \\
&= 1. \quad \text{by } (\textbf{A2})(\textbf{ii}) \text{ and } \textbf{LAE}
\end{aligned}$$

□

Proposition 6. *For any fuzzy measure* $\mathfrak{m}: 2^N \to [0,1]$, *if the function* $F: [0,1]^2 \to [0,1]$ *satisfies one of the following conditions:*

(i) *F is* (1, 0)*-increasing*
(ii) *F satisfies **(RNE)***

then $\mathfrak{C}_{\mathfrak{m}}^{F}$ *is* $\vec{1}$*-increasing.*

Proof. Let $(x_{(1)}, \ldots, x_{(i-1)}, x_{(i)}, \ldots, x_{(n)})$ be an increasing permutation of $\vec{x} \in [0, 1]^n$. Suppose that **(i)** holds and consider $c > 0$ such that $\vec{x} + c \in [0, 1]^n$. Then, it follows that:

$$\begin{aligned} \mathfrak{C}_{\mathfrak{m}}^{F}(x_1 + c, \ldots, x_n + c) &= \min\left\{1, F(x_{(1)} + c - 0, \mathfrak{m}(A_{(1)})) + \sum_{i=2}^{n} F(x_{(i)} + c - (x_{(i-1)} + c), \mathfrak{m}(A_{(i)}))\right\} \\ &\quad \text{by Eq. (4)} \\ &\geq \min\left\{1, F(x_{(1)}, \mathfrak{m}(A_{(1)})) + \sum_{i=2}^{n} F(x_{(i)} - x_{(i-1)}, \mathfrak{m}(A_{(i)}))\right\} \text{ by}(\mathbf{i}) \\ &= \mathfrak{C}_{\mathfrak{m}}^{F}(x_1, \ldots, x_n). \text{ by Eq. (4)} \end{aligned}$$

Now consider that **(ii)** holds. It follows that:

$$\begin{aligned} \mathfrak{C}_{\mathfrak{m}}^{F}(x_1 + c, \ldots, x_n + c) &= \min\left\{1, F(x_{(1)} + c - 0, 1) + \sum_{i=2}^{n} F(x_{(i)} + c - (x_{(i-1)} + c), \mathfrak{m}(A_{(i)}))\right\} \\ &\quad \text{by Eq. (4)} \\ &= \min\left\{1, x_{(1)} + c + \sum_{i=2}^{n} F(x_{(i)} - x_{(i-1)}, \mathfrak{m}(A_{(i)}))\right\} \text{ by}(\mathbf{ii}) \\ &\geq \min\left\{1, x_{(1)} + \sum_{i=2}^{n} F(x_{(i)} - x_{(i-1)}, \mathfrak{m}(A_{(i)}))\right\} \\ &= \mathfrak{C}_{\mathfrak{m}}^{F}(x_1, \ldots, x_n). \quad \text{by Eq. (4)} \end{aligned}$$

□

Theorem 1. *For any fuzzy measure* $\mathfrak{m} : 2^N \to [0, 1]$ *and left 0-absorbent **(RNE)**-function* F: $[0, 1]^2 \to [0, 1]$, $\mathfrak{C}_{\mathfrak{m}}^{F}$ *is a* $\vec{1}$*-pre-aggregation function.*

Proof. It follows from Propositions 5 and 6, observing that the property **(RNE)** implies **(A2)**(ii). □

Corollary 1. *For any fuzzy measure* $\mathfrak{m} : 2^N \to [0, 1]$ *and left 0-absorbent **(RNE)**-function* F: $[0, 1]^2 \to [0, 1]$ *satisfying **(LC)**,* $\mathfrak{C}_{\mathfrak{m}}^{F}$ *is an idempotent averaging* $\vec{1}$*-pre-aggregation function.*

Proof. It follows from Propositions 3 and 4, and Theorem 1. □

Remark 4. Observe that, even when a left 0-absorbent function F: $[0, 1]^2 \to [0, 1]$ is not an averaging function, we may obtain an averaging C_F-integral. For example, consider the left 0-absorbent function F_{NA}: $[0, 1]^2 \to [0, 1]$ of Example 1. By Remark 1, we know that that F_{NA} is idempotent but not averaging. However, it is immediate that F_{NA} satisfies (RNE) and (LC), and, therefore, by Corollary 1, the C_F-integral $\mathfrak{C}_{\mathfrak{m}}^{F_{NA}}$, for a fuzzy measure $\mathfrak{m}$, is an averaging idempotent $\vec{1}$-pre-aggregation function.

Theorem 2. *For any fuzzy measure* $\mathfrak{m} : 2^N \to [0, 1]$ *and left 0-absorbent* $(1, 0)$*-pre-aggregation function* F: $[0, 1]^2 \to [0, 1]$, $\mathfrak{C}_{\mathfrak{m}}^{F}$ *is a* $\vec{1}$*-pre-aggregation function.*

Proof. It follows from Propositions 5 and 6, observing that any left 0-absorbent pre-aggregation function satisfies **(LAE)**. □

In Table 1 we show a set of bivariate functions F: $[0, 1]^2 \to [0, 1]$ that belong to different families like (I) t-norms, (II) overlap functions, (III) copulas that are neither t-norms nor overlap functions, (IV) aggregation functions not included in (I)-(III) and (V) left-0 absorbent (0, 1)-pre-aggregation functions. For each function F, we show its definition and reference (if they are new this field is left empty), as well as whether or not they satisfy **(LAE), (RNE), (LC), (A2)** and (1, 0)-increasingness. Then, in the last but two column, according to properties analyzed in the previous columns, we indicate whether or not the obtained C_F-integral (constructed using Eq. (4)) is a $\vec{1}$-pre-aggregation function (PA). Observe that the set of conditions that 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)**).
- Theorem 2 (**(LAE), (A2)**, (1, 0)-increasingness).

Finally, the last but one column shows if the obtained C_F-integral is averaging **(AV)** (that is, if it satisfies Propositions 3 and 4), and the last column if it is idempotent **(ID)** (that is, if it satisfies Proposition 2).

Table 1
Analysis of the conditions of Theorems 1 and 2, Propositions 2–4, for families of left 0-absorbent functions F.

Definition	Name/Reference	(LAE)	(RNE)	LC	(A2)	(1, 0)-inc.	PA	AV	ID
(I) T-norms [36]									
$T_M(x,y)=\min\{x,y\}$	Minimum	✓	✓	✓	✓	✓	✓	✓	✓
$T_P(x,y)=xy$	Algebraic Product	✓	✓	✓	✓	✓	✓	✓	✓
$T_L(x,y)=\max\{0,x+y-1\}$	Łukasiewicz	✓	✓	✓	✓	✓	✓	✓	✓
$T_{HP}(x,y)=\begin{cases}0 & \text{if } x=y=0\\ \frac{xy}{x+y-xy} & \text{otherwise}\end{cases}$	Hamacher Product	✓	✓	✓	✓	✓	✓	✓	✓
(II) Overlap functions [5,10,19,21]									
$O_B(x,y)=\min\{x\sqrt{y},y\sqrt{x}\}$	[10, Theorem 8]	✓		✓	✓	✓	✓		
$O_{mM}(x,y)=\min\{x,y\}\max\{x^2,y^2\}$	Cuadras-Augé copula [48] [19, Example 3.1.(i)], [18, Example 4] [21, Example 3.1]	✓	✓	✓	✓	✓	✓	✓	✓
$O_\alpha(x,y)=xy(1+\alpha(1-x)(1-y))$, $\alpha\in[-1,0[\,\cup\,]0,1]$	[3, Apendix A (A.2.1)],[39] Farlie-Gumbel-Morgenstern copula family*	✓	✓	✓	✓	✓	✓	✓	✓
$O_{Div}(x,y)=\frac{xy+\min\{x,y\}}{2}$	[3, Apendix A (A.8.7)], [41, Table 1]	✓	✓	✓	✓	✓	✓	✓	✓
$GM(x,y)=\sqrt{xy}$	Geometric Mean [27, Example 1]	✓			✓	✓	✓		
$HM(x,y)=\begin{cases}0 & \text{if } x=0 \text{ or } y=0\\ \frac{2}{\frac{1}{x}+\frac{1}{y}} & \text{otherwise}\end{cases}$	Harmonic Mean [27, Example 1]	✓			✓	✓	✓		
$S(x,y)=\sin\left(\frac{\pi}{2}(xy)^{\frac{1}{4}}\right)$	Sine [27, Example 1]	✓			✓	✓	✓		
$O_{RS}(x,y)=\min\left\{\frac{(x+1)\sqrt{y}}{2},y\sqrt{x}\right\}$		✓			✓	✓	✓		
(III) Copulas that are neither t-norms nor overlap functions [3]									
$C_F(x,y)=xy+x^2y(1-x)(1-y)$	[36, Example 9.5 (v)], [41, Table 1]	✓	✓	✓	✓	✓	✓	✓	✓
$C_L(x,y)=\max\{\min\{x,\frac{y}{2}\},x+y-1\}$	[3, Apendix A (A.5.3a)], [41, Table 1]	✓	✓	✓	✓	✓	✓	✓	✓
(IV) Aggregation functions other than (I)–(III)									
$F_{GL}(x,y)=\sqrt{\frac{x(y+1)}{2}}$		✓			✓	✓	✓		
$F_{BPC}(x,y)=xy^2$	[8, Example 1.80]	✓	✓	✓	✓	✓	✓	✓	✓
(V) Left 0-absorbent (0, 1)-pre-aggregation functions									
$F_{BD1}(x,y)=\min\{x,1-x+\min\{x,y^q\}\},\ 0<q\leq 1$		✓	✓	✓	✓	✓	✓	✓	✓
$F_{NA}(x,y)=\begin{cases}x & \text{if } x\leq y\\ \min\{\frac{x}{2},y\} & \text{otherwise}\end{cases}$		✓	✓	✓	✓	✓	✓	✓	✓
$F_{NA2}(x,y)=\begin{cases}0 & \text{if } x=0\\ \frac{x+y}{2} & \text{if } 0<x\leq y\\ \min\{\frac{x}{2},y\} & \text{otherwise}\end{cases}$		✓			✓	✓	✓		✓

* When $\alpha=0$, we have that $O_\alpha=T_P$, the product t-norm, which was considered in the first part of table.

4. Applying C_F-integrals in fuzzy rule-based classification systems

In this section, we firstly present the application of C_F-integrals in classification problems [25], adopting it to aggregate the information given by the fuzzy rules. To do so, consider that a classification problem consists of m training examples, $\mathbf{x}_p = (x_{p1}, \ldots, x_{pn}, y_p)$, with $p = 1, \ldots, m$, where x_{pi}, with $i = 1, \ldots, n$, is the value of the ith attribute and $y_p \in \mathbb{C} = \{C_1, \ldots, C_M\}$ is the label of the class of the pth training example, where M is the number of classes.

In this work, we use FRBCSs to tackle this kind of problems. Specifically, we have selected FARC-HD [2] to accomplish the learning of the fuzzy rules, since it is one of the most precise fuzzy classifiers nowadays. The form of the fuzzy rules used by this algorithm is:

$$\text{Rule } R_j: \text{ If } x_{p1} \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_{pn} \text{ is } A_{jn} \text{ then } x_p \text{ is } C_j \text{ with } RW_j,$$

where $x_p = (x_{p1}, \ldots, x_{pn})$ is the n-dimensional vector of attribute values corresponding to an example $\mathbf{x}_p$, R_j is the label of the jth rule, A_{ji} is an antecedent fuzzy set modeling a linguistic term, C_j is the label of the class of the rule R_j, with $C_j \in \{1, \ldots, M\}$ and $RW_j \in [0, 1]$ is the rule weight [34], which, in this case, is computed using the certainty factor.

We have used the set up suggested by the authors of FARC-HD, which is as follows: the product t-norm as the conjunction operator, five linguistic labels per variable, modeled by triangular shaped membership functions, the minimum support is set to 0.05, the threshold for the confidence is 0.8 and the maximum depth of the search tree is limited to 3.

In this paper, we propose a new FRM, where C_F-integrals are used to obtain the information associated with each class of the problem, that is, to aggregate the local information given by the fired rules of the system when classifying a new example, x_p. Specifically, the predicted class for a new example x_p is computed by:

$$class = \arg \max_{k=\{1,\ldots,M\}} (\mathfrak{C}^F_{m_k}(\mu_{A_j}(x_p) * RW_j \mid Class(R_j) = k)) \text{ with, } j = 1, \ldots, L. \quad (5)$$

where $\mathfrak{C}^F_{m_k}$ is the C_F-integral (associated with the fuzzy measure m_k) considered to aggregate information given by the fired rules for the class k, μ_{A_j} is the matching degree of the example x_p with the antecedent of the jth fuzzy rule, RW_j is its rule weight and L is the number of fuzzy rules in the system.

From Eq. (5) it can be observed that we consider a different C_F-integral for each class of the problem. This is due to the fact that we construct a different fuzzy measure for each class of the problem. Specifically, we use the power measure (see Eq. (2)) in which a different q exponent is learnt for each class of the problem using a genetic algorithm, as we have done in our previous papers in the topic (see [4,40,41] for details of the evolutionary algorithm). Regarding the parameters of this genetic algorithm, we consider a population composed of 50 individuals, 20.000 evaluations and 30 bits for each gene in the gray codification.

In the remainder of this section we present the experimental framework used to test the quality of our new FRM. In first place we present the considered datasets (Ssection 4.1) followed by the statistical tests that are used in this paper for performing comparisons (Section 4.2).

4.1. Datasets

In this paper, to analyze the performance of our proposal, we consider 33 different datasets selected from the KEEL[2] dataset repository [1]. The properties of the selected datasets are summarized in Table 2, showing for each dataset the identification of this dataset (ID), followed by the name of the dataset (Dataset), the number of examples (#Ex.), the number of attributes (#Atts.) and the number of classes (#Class).

Some datasets, namely: *magic, page-blocks, penbased, ring, satimage* and *twonorm*, were stratified sampled at 10% in order to reduce their size for training. Some examples containing missing information were removed, e.g., in the *wisconsin* dataset.

We have applied a 5-fold cross-validation technique, that is, split the dataset into five partitions randomly. Each partition has 20% of the examples. We use four partitions for training, and the other is used for testing. This process is repeated five times, using a different partition for testing each time. In each iteration we measure the quality of the classifier using the accuracy rate, which is defined as the number of correctly classified examples divided by the total number of examples for each partition. We then compute the average result of the five testing partitions, which is the output of the algorithm.

4.2. Statistical tests for performance comparisons

For the statistical analysis of the results, we use hypothesis validation techniques [28,53], namely, non-parametric tests, since the initial conditions that guarantee the reliability of the parametric tests cannot be guaranteed [15].

We apply the aligned Friedman rank test [31] to detect statistical differences among a group of results and to verify the quality of a method in comparison to its partners. The algorithm achieving the lowest average ranking is the best one.

Additionally, to analyze if the best ranking method rejects the equality hypothesis with respect to its partners we use the post-hoc Holm's test [32]. This method allows us to see whenever a hypothesis of comparison could be rejected at a specified level of significance α. We compute the adjusted p-value (APV) to take into account that multiple tests are

[2] http://www.keel.es.

Table 2
Summary of the properties of the datasets considered in this study.

Id.	Dataset	#Ex.	#Atts.	#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
Con	Contraceptive	1473	9	3
Eco	Ecoli	336	7	8
Gla	Glass	214	9	6
Hab	Haberman	306	3	2
Hay	Hayes-Roth	160	4	3
Ion	Ionosphere	351	33	2
Iri	Iris	150	4	3
Led	led7digit	500	7	10
Mag	Magic	1902	10	2
New	Newthyroid	215	5	3
Pag	Pageblocks	5472	10	5
Pen	Penbased	10,992	16	10
Pho	Phoneme	5404	5	2
Pim	Pima	768	8	2
Rin	Ring	740	20	2
Sah	Saheart	462	9	2
Sat	Satimage	6435	36	7
Seg	Segment	2310	19	7
Shu	Shuttle	58,000	9	7
Son	Sonar	208	60	2
Spe	Spectfheart	267	44	2
Tit	Titanic	2201	3	2
Two	Twonorm	740	20	2
Veh	Vehicle	846	18	4
Win	Wine	178	13	3
Wis	Wisconsin	683	11	2
Yea	Yeast	1484	8	10

performed. Then, we can directly compare the APV with the level of significance α, and, thus, we are able to reject the null hypothesis.

5. Experimental study and results

In this section, we present the results achieved in testing when using the FRM generalized by C_F-integrals. To do so, the results of our proposal are analyzed considering three main steps:

1. Firstly, we present the results and analyze the performance of each averaging C_F-integral. After that, we compare them among themselves in order to discover which generalization is the one that best represent the family of averaging C_F-integrals.
2. The second part of the study is related to C_F-integrals that are not averaging. In order to find the best representative method of this family, we firstly analyze the achieved results and after that, compare them among themselves.
3. Once we have found the two best C_F-integrals obtained in the previous steps, in order to test the quality of our approach, we perform the following comparisons:
   (a) The best averaging C_F-integral versus classical averaging functions (FRM of the WR) and our previous averaging pre-aggregation functions.
   (b) The best non-averaging C_F-integral versus the classical non averaging functions, like the FRM of AC or the usage of the probabilistic sum.
   (c) Finally, we test whether the application of the best non averaging function enhances the results of the averaging operators or not.

5.1. Analysis of the performance of averaging C_F-integrals

This subsection is aimed at analyzing the performance of the averaging C_F-integrals considered in this study (see Table 1) in the FRM. The results achieved in test by the 9 averaging functions are available in Table 3, by columns. In each row we introduce the results of each dataset, highlighting the best global result in **boldface**. We also include in this table the number of datasets where each function achieves the best (#Wins) and the worst (#Loses) result, respectively.

Table 3
Accuracy achieved in test by different averaging C_F-integrals.

Dataset	O_α	O_B	O_{mM}	O_{Div}	CF	CL	F_{BPC}	F_{BD1}	F_{NA}
App	83.03	83.03	84.94	83.94	82.99	**85.84**	83.07	83.94	82.99
Bal	80.32	82.08	82.56	81.60	82.24	84.00	80.32	**84.80**	82.56
Ban	86.09	86.81	86.85	85.79	85.70	85.04	**87.02**	83.09	86.09
Bnd	68.26	**71.83**	67.70	69.68	69.13	69.66	66.62	71.09	69.40
Bup	63.77	65.51	66.09	65.51	65.22	62.32	66.67	64.06	**67.83**
Cle	55.23	56.24	54.55	55.90	52.18	55.54	54.88	57.92	**57.92**
Con	52.00	52.89	52.75	**53.83**	52.54	50.78	52.95	52.41	52.27
Eco	76.49	76.20	77.08	75.61	77.09	78.87	77.69	**79.17**	78.88
Gla	62.14	**66.82**	60.75	63.10	63.58	63.09	63.54	65.44	64.51
Hab	72.85	72.86	72.21	72.52	69.92	72.86	73.17	72.21	**73.51**
Hay	78.75	78.72	78.01	**80.26**	78.75	79.46	78.01	77.95	78.72
Ion	90.04	88.32	87.47	89.47	89.46	89.75	89.47	89.46	**90.60**
Iri	92.00	94.00	94.00	93.33	**94.67**	93.33	92.00	93.33	93.33
Led	68.00	68.40	67.60	68.80	**69.00**	68.00	68.80	68.00	68.60
Mag	**80.18**	79.86	79.97	79.91	78.81	79.76	78.86	79.55	80.02
New	**95.35**	94.88	94.42	**95.35**	94.88	94.42	94.88	93.49	93.49
Pag	93.43	94.52	93.98	93.97	**94.89**	93.61	94.34	94.34	93.97
Pen	90.09	91.09	89.45	90.82	90.55	90.27	90.09	**92.55**	91.45
Pho	83.05	82.92	83.29	82.81	83.23	**83.88**	82.70	81.96	82.86
Pim	75.13	75.38	**76.17**	74.48	75.78	75.52	73.82	73.56	75.13
Rin	89.19	89.32	90.00	89.86	89.73	89.46	88.38	88.78	**90.27**
Sah	**71.85**	69.48	70.78	68.18	69.48	68.39	71.21	69.70	68.61
Sat	79.16	78.54	79.00	78.23	**80.72**	79.16	78.38	78.70	78.54
Seg	92.73	92.51	**93.33**	92.77	92.94	93.20	92.42	93.16	92.55
Shu	97.01	97.29	97.01	**97.84**	97.10	97.01	97.10	97.15	96.78
Son	**80.78**	75.49	77.93	77.42	77.89	79.83	74.05	80.29	78.85
Spe	77.48	77.88	76.75	76.39	77.51	76.77	**79.76**	74.92	78.26
Tit	78.87	78.87	78.87	78.87	78.87	78.87	78.87	**79.06**	78.87
Two	84.73	83.78	84.46	85.14	**85.68**	85.41	84.19	85.14	83.92
Veh	68.21	67.73	66.55	67.02	**70.33**	67.38	68.20	69.26	67.97
Win	95.48	94.97	**97.21**	93.24	94.38	92.13	96.62	96.62	96.03
Wis	96.49	96.63	96.34	96.34	96.05	96.19	96.34	**96.64**	96.34
Yea	56.33	57.35	57.08	**57.48**	57.34	56.87	**57.48**	55.12	56.40
Mean	79.23	79.46	79.25	79.26	79.35	79.29	79.15	79.48	**79.62**
#Wins	4	2	3	5	6	2	3	5	5
#Loses	4	2	6	4	5	4	7	6	4

Table 4
Statistical analysis of the methods based on averaging C_F-integrals.

Algorithm	Ranking	APV
$\mathbf{F}_{NA}$	**129.31**	–
O_B	138.93	1.0
F_{BD1}	143.21	1.0
CF	145.60	1.0
CL	149.71	1.0
O_{Div}	152.74	1.0
O_{mM}	154.01	1.0
F_{BPC}	162.77	0.79
O_α	164.68	0.75

From the results shown in Table 3, it is possible to notice that all the averaging C_F-integrals, except F_{NA}, present a mean performance between 79.15 and 79.48. We have to highlight the function F_{NA} since it achieves the best global mean and also the best accuracy rate in 5 out of the 33 datasets considered in this study. Moreover, observe that the functions O_α, O_{Div}, *CF*, F_{BD1} and F_{NA} achieves similar results in terms of number of datasets with the best and worst performance respectively, while the remainder functions achieves worse results.

In order to select objectively the best function among this group, we have carried out a statistical study according to the recommendations made in the specialized literature [16,28,53].

Specifically, we have performed the aligned Friedman rank test to compare the 9 approaches, whose obtained rankings are presented in the second column of Table 4. In this table we sort the values from the lowest to highest obtained ranking, where the best one is highlighted in **boldface**. Then, we apply the Holm's post-hoc test, to check whether the control approach (the one associated with the best ranking) is statistically better than the remainder approaches, showing the obtained APV in the last column of this table.

Table 5
Accuracy achieved in test by different non-averaging C_F-integrals.

Dataset	GM	HM	Sin	O_{RS}	F_{GL}	F_{NA2}
App	82.08	83.98	85.80	83.98	82.08	**85.84**
Bal	88.48	86.40	**89.44**	88.00	89.12	88.64
Ban	85.28	**86.19**	82.79	85.58	83.42	84.60
Bnd	71.30	70.51	**72.69**	69.19	71.01	70.48
Bup	61.16	**66.96**	63.48	66.09	62.03	64.64
Cle	57.23	57.24	**57.55**	55.88	57.25	56.55
Con	53.77	52.21	**54.31**	53.84	54.24	53.16
Eco	81.26	79.47	**82.45**	80.07	81.55	80.08
Gla	65.44	**69.17**	66.83	65.89	66.33	66.83
Hab	71.54	71.88	71.87	**72.87**	70.24	71.87
Hay	79.49	79.43	77.98	**81.77**	78.69	79.43
Ion	**90.89**	88.91	87.46	88.32	90.04	89.75
Iri	**94.67**	94.00	94.00	94.00	94.00	94.00
Led	68.00	68.40	69.60	69.20	**68.80**	**69.80**
Mag	80.02	80.23	79.34	**80.23**	79.70	79.70
New	**97.67**	95.81	95.35	95.81	**97.67**	96.28
Pag	94.34	93.97	94.34	**94.52**	94.34	94.15
Pen	92.18	92.09	91.45	92.00	92.73	**92.91**
Pho	82.07	**83.73**	80.96	82.72	81.27	81.44
Pim	74.87	74.87	75.13	75.00	**76.82**	74.61
Rin	90.95	90.00	88.51	90.27	**91.35**	89.86
Sah	69.04	68.84	71.20	**71.86**	70.33	70.12
Sat	79.01	78.69	77.45	**80.72**	78.53	80.41
Seg	**93.46**	92.73	92.47	92.77	93.07	92.42
Shu	97.06	96.92	96.69	96.37	96.69	**97.15**
Son	82.73	81.28	81.74	83.19	**83.69**	83.21
Spe	77.51	79.76	78.65	78.27	79.76	**79.77**
Tit	**78.87**	**78.87**	**78.87**	**78.87**	**78.87**	**78.87**
Two	89.19	86.89	91.49	89.05	90.00	**92.57**
Veh	67.97	68.79	64.77	67.38	**69.03**	68.08
Win	96.08	96.05	**97.17**	97.16	95.49	96.08
Wis	96.93	**97.07**	96.34	97.06	95.76	96.78
Yea	56.94	**58.15**	56.47	57.28	57.68	57.08
Mean	80.23	80.29	80.14	80.46	80.35	**80.52**
#Wins	5	8	7	7	7	7
#Loses	4	7	11	3	5	2

From these results, it is noticeable that there are no statistical differences among the averaging functions. However, for the sake of selecting a representative for this family, we choose the function F_{NA} as it obtains the best global mean and it is selected as the control method in the Holm's test.

5.2. Studying the quality of non-averaging C_F-integrals

In this subsection, we present the achieved results in testing when one considers C_F-integrals that do not have averaging characteristics. We present the results of the 6 functions of this type in Table 5, by columns. In each row we introduce the results of each dataset where the best result is highlighted in **boldface**. Like in Table 3, #Wins and #Loses represent the number of datasets where the function obtains the best and worst result, respectively.

From the results presented in Table 5, we can directly conclude that the non-averaging functions have a superior mean in relation to the averaging functions (Table 3), since the smallest obtained mean (80.14 by Sin) is superior than the best averaging C_F-integral (79.62 by F_{NA}). Additionally, we have to highlight the leap in performance provided by the usage of F_{NA2} and O_{RS}. The first function has the best accuracy in 7 datasets and the worst accuracy in only 2 dataset. The function O_{RS} also achieves a good mean, with best accuracy in 7 datasets and the worst one in 3 datasets. Furthermore, we should stress that although the number of dataset in which the remainder functions provide the best results is similar the number of datasets where they provide the worst result is larger, which implies a decrease on the overall performance as shown in Table 5.

According to the obtained results, it is necessary to conduct a statistical analysis to select the best function among this group. In order to do it, we have performed the same statistical study as in the previous section. The results of the aligned Friedman and Holm's tests are presented in Table 6. As expected, according to Table 5, all methods present a similar behavior, therefore, we select F_{NA2} as representative of this family since it is considered as control variable and it also achieves the best global mean.

Table 6
Average rankings of the non-averaging C_F-integrals (aligned Friedman).

Algorithm	Ranking	APV
$\mathbf{F_{NA2}}$	**91.78**	-
F_{GL}	95.12	1.0
O_{RS}	95.39	1.0
GM	99.22	1.0
HM	105.13	1.0
Sin	110.33	0.94

Table 7
Results achieved in test by the averaging FRMs.

Dataset	F_{NA}	Choquet	Ham_{PA}	CP_{Min}	WR
App	82.99	80.13	82.99	**85.84**	83.03
Bal	82.56	82.40	**82.72**	81.60	81.92
Ban	86.09	**86.32**	85.96	84.30	83.94
Bnd	69.40	68.56	**72.13**	71.06	69.40
Bup	**67.83**	66.96	65.80	61.45	62.03
Cle	**57.92**	55.58	55.58	54.88	56.91
Con	52.27	51.26	**53.09**	52.61	52.07
Eco	78.88	76.51	**80.07**	77.09	75.62
Gla	64.51	64.02	63.10	**69.17**	64.99
Hab	73.51	72.52	72.21	**74.17**	70.89
Hay	78.72	79.49	79.49	**81.74**	78.69
Ion	**90.60**	90.04	89.18	88.89	90.03
Iri	93.33	91.33	93.33	92.67	**94.00**
Led	68.60	68.20	68.60	68.40	**69.40**
Mag	**80.02**	78.86	79.76	79.81	78.60
New	93.49	94.88	**95.35**	93.95	94.88
Pag	93.97	94.16	**94.34**	93.97	94.16
Pen	**91.45**	90.55	90.82	91.27	**91.45**
Pho	82.86	82.98	**83.83**	82.94	82.29
Pim	75.13	74.60	73.44	**75.78**	74.60
Rin	90.27	**90.95**	88.78	87.97	90.00
Sah	68.61	69.69	70.77	**70.78**	68.61
Sat	78.54	79.47	**80.40**	79.01	79.63
Seg	92.55	**93.46**	93.33	92.25	93.03
Shu	96.78	97.61	97.20	**98.16**	96.00
Son	78.85	77.43	**79.34**	76.95	77.42
Spe	78.26	77.88	76.02	**78.99**	77.90
Tit	**78.87**	**78.87**	**78.87**	**78.87**	**78.87**
Two	83.92	84.46	85.27	85.14	**86.49**
Veh	67.97	68.44	68.20	**69.86**	66.67
Win	96.03	93.79	**96.63**	93.83	96.60
Wis	96.34	**97.22**	96.78	95.90	96.34
Yea	56.40	55.73	56.53	**57.01**	55.32
Mean	79.62	79.22	**79.69**	79.58	79.15
#Wins	6	5	11	11	5
#Loses	4	7	3	8	10

5.3. Comparisons of the best C_F-integrals against classical FRMs

Once we have selected the functions that represent the family of C_F-integrals with averaging or non-averaging characteristics (F_{NA} and F_{NA2}), we compare them against classical averaging (Section 5.3.1) and non-averaging functions (Section 5.3.2), respectively.

5.3.1. Analyzing the behavior of the representative averaging C_F-integral

In first place, we compare the best averaging function against FRMs where averaging aggregations are applied. Namely, the FRM of the Winning Rule (WR) [8], the standard Choquet integral (Choquet) [4], the best pre-aggregation function presented in [40] (which is named Ham_{PA} since it is based on the Hamacher t-norm) and the best Choquet-Like Copula-based [41] (which is named CP_{Min} as it is based on the Minimum t-norm). We have to point out that the pre-aggregation function based on the Hamacher t-norm (Ham_{PA}) is also a C_F-integral (where F is the Hamacher t-norm).

The results achieved in test by this averaging FRMs are available in Table 7, using the same structure as the tables presented before.

Table 8
Statistical analysis of the FRMs based on averaging operators.

Algorithm	Ranking	APV
$\mathbf{Ham}_{PA}$	**68.96**	
F_{NA}	76.25	0.62
CP_{Min}	80.87	0.62
Choquet	92.98	0.12
WR	95.90	<u>0.08</u>

Table 9
Results achieved in test by classical non-averaging operators.

Dataset	F_{NA2}	AC	ProbSum
App	**85.84**	83.03	85.84
Bal	**88.64**	85.92	87.20
Ban	84.60	**85.30**	84.85
Bnd	**70.48**	68.28	68.82
Bup	64.64	**67.25**	61.74
Cle	56.55	56.21	**59.25**
Con	**53.16**	53.16	52.21
Eco	80.08	**82.15**	80.95
Gla	**66.83**	65.44	64.04
Hab	71.87	**73.18**	69.26
Hay	**79.43**	77.95	77.95
Ion	**89.75**	88.90	88.32
Iri	94.00	94.00	**95.33**
Led	**69.80**	69.60	69.20
Mag	79.70	**80.76**	80.39
New	**96.28**	94.88	94.42
Pag	94.15	**95.07**	94.52
Pen	92.91	92.55	**93.27**
Pho	81.44	81.70	**82.51**
Pim	74.61	74.74	**75.91**
Rin	89.86	**90.95**	90.00
Sah	**70.12**	68.39	69.69
Sat	**80.41**	79.47	80.40
Seg	92.42	**93.12**	92.94
Shu	**97.15**	95.59	94.85
Son	**83.21**	78.36	82.24
Spe	**79.77**	77.88	77.90
Tit	**78.87**	**78.87**	**78.87**
Two	**92.57**	90.95	90.00
Veh	68.08	**68.56**	68.09
Win	**96.08**	96.03	94.92
Wis	96.78	96.63	**97.22**
Yea	57.08	58.96	**59.03**
Mean	**80.52**	80.12	80.07
#Wins	17	10	8
#Loses	11	12	11

From these results, it is possible to observe that WR and Choquet have a low mean while Ham_{PA} is the one obtaining the best global mean, followed by our new averaging C_F-integral, F_{NA} and CP_{Min}. Moreover, observe that Ham_{PA} is also the function that has the biggest number of good results (along with CP_{Min}) and the lowest number of cases having bad results.

We have conducted the same statistical study than in the previous sections, where the achieved results are presented in Table 8. These results are sorted according to the ranking and highlighting in **boldface** the control ranking. Whenever there is an statistical difference in favor to the control method the APV is underlined.

Observing these results, we can see that the pre-aggregation function based on the Hamacher t-norm is considered as control variable, and it also presents differences against the FRM of the WR and a positive trend versus the standard Choquet integral. On the other hand, it is not possible to affirm that there are differences against the remainder functions. Therefore, we consider Ham_{PA} as the best averaging C_F-integral since it achieves the best mean and the largest number of datasets having the best accuracy.

5.3.2. Analyzing the behavior of the best non-averaging C_F-integral

Next, we study the behavior of the non-averaging operators. Specifically, we compare our representative for the family of non-averaging C_F-integral, F_{NA2}, against the classical FRMs of the additive combination (AC) and the probabilistic sum (PS). The results are available in Table 9, having the same structure of our results presented before.

Table 10
Statistical analysis of the methods based on non-averaging operators.

Algorithm	Ranking	APV
$\mathbf{F}_{NA2}$	**41.80**	
AC	53.65	0.14
PS	54.54	0.14

Table 11
Statistical analysis of the best non-averaging C_F-integral against the averaging operators.

Algorithm	Ranking	APV
$\mathbf{F}_{NA2}$	**64.93**	
Ham_{PA}	90.77	<u>0.06</u>
F_{NA}	99.40	<u>0.02</u>
CP_{Min}	102.95	<u>0.02</u>
Cho	117.43	<u>7.91E-4</u>
WR	121.48	<u>3.05E-4</u>

We can see in the obtained results that the C_F-integral based on the function F_{NA2} presents the highest global mean. If we look closer, this function achieves the best accuracy in almost half of the datasets considered in this study. On the other hand, the classical aggregation functions applied in the FRM provide the best result in a lower number of datasets (10 and 8, respectively).

Again, to statistically compare these methods among themselves, we perform the Friedman rank test and the Holm's post hoc test. The obtained results are available in Table 10. These results show that our new function F_{NA2} has the best rank and, consequently, it is considered as control variable as it was expected according to the previous results. When we observe the obtained APVs we can see that their APVs are low, which shows a trend pointing out that our new non-averaging function is very competitive versus these two classical aggregation functions. Therefore, the quality of our proposal is proved since it is enhancing the results of the classical FRM of the AC and PS.

To finish our study, for the sake of certifying the quality of our new function (F_{NA2}), we also compare it versus the five averaging FRMs studied in the previous Section 5.3.1. To accomplish this comparison, we have performed again the Friedman rank test and Holm's post hoc test among these approaches. The results of the statistical test and the obtained APVs are shown in Table 11, where the ranking related to the function considered as control method is highlighted in **boldface**. Furthermore, the APV is underlined when there are statistical differences favorably to the control approach versus the opponent method.

The obtained statistical results clearly show the superiority of the C_F-integral based on the function F_{NA2}, since it achieves statistical differences versus all the remainder methods. All in all, the non-averaging C_F-integral constructed using the function F_{NA2}, has proven to be the best choice among all developed functions, since it offers the best performance, it statistically outperforms classical averaging functions applied in the FRM and it is competitive with respect classical non averaging FRMs.

6. Conclusion

In this paper we have proposed a generalization of the Choquet integral by replacing its product operator by a function F with some weak properties. As a result, we have defined the C_F-integrals, a new family of pre-aggregation functions with some particular characteristics, which allows us to enlarge the scope of the methodology that we proposed in [40]. The main advantages of this approach in relation to our previous work concerning generalizations of the Choquet integral are:

- The function F used for the generalization may satisfy a less number of properties, and we still have a pre-aggregation function.
- The resulting pre-aggregation function does not need to be neither an averaging nor idempotent function.

We have applied these averaging and non-averaging C_F-integrals in FRBCSs to tackle classification problems. Precisely, in this work we performed a study considering 33 different public datasets, and the conclusions we draw are the following ones:

1. The considered averaging C_F-integrals present a similar performance than that of our previous generalizations.
2. The best averaging C_F-integral is Ham_{PA}, which was previously introduced in another paper. However, it is also a C_F-integral (based on the Hamacher t-norm).
3. The non-averaging C_F-integrals, as expected, offer a performance superior than the averaging ones.
4. The best C_F-integral, F_{NA2}, provides results that are statistically superior than all classical FRMs, and also, very competitive with the classical non-averaging FRMs like AC or PS.

Consequently, we have created a new family of pre-aggregation functions, which provides accurate results when they have non-averaging features.

Future work is concerned with two lines of research. In one hand, we will search for a generalization of our CC-integrals [41], by means of two arbitrary functions F_1 and F_2, to put in the place of each copula, satisfying a minimal set of properties that guarantee that the generalized CC-integral is, at least, a pre-aggregation function. On the other hand, we will study our generalizations in an interval-valued context, following the approach in [6,7,17], as in [49–51].

Acknowledgments

This work is supported by Brazilian National Counsel of Technological and Scientific Development CNPq (Proc. 233950/2014-1, 306970/2013-9, 307781/2016-0), by grant APVV-14-0013, by the Spanish Ministry of Science and Technology (under project TIN2016-77356-P (AEI/FEDER, UE)), and by Caixa and Fundación Caja Navarra of Spain.

References

[1] J. Alcalá-Fdez, L. Sánchez, S. García, M. Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernández, F. Herrera, Keel: a software tool to assess evolutionary algorithms for data mining problems, Soft Comput. 13 (3) (2009) 307–318.
[2] J. Alcalá-Fdez, R. Alcalá, F. Herrera, A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning, IEEE Trans. Fuzzy Syst. 19 (5) (2011) 857–872.
[3] C. Alsina, M.J. Frank, B. Schweizer, Associative Functions: Triangular Norms and Copulas, World Scientific Publishing Company, Singapore, 2006.
[4] E. Barrenechea, H. Bustince, J. Fernandez, D. Paternain, J.A. Sanz, Using the Choquet integral in the fuzzy reasoning method of fuzzy rule-based classification systems, Axioms 2 (2) (2013) 208–223.
[5] B.C. Bedregal, G.P. Dimuro, H. Bustince, E. Barrenechea, New results on overlap and grouping functions, Inf. Sci. 249 (2013) 148–170.
[6] B.C. Bedregal, G.P. Dimuro, R.H.S. Reiser, An approach to interval-valued R-implications and automorphisms, in: J.P. Carvalho, D. Dubois, U. Kaymak, J.M.d. C. Sousa (Eds.), Proceedings of the Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference, IFSA/EUSFLAT, 2009.
[7] B.C. Bedregal, G.P. Dimuro, R.H.N. Santiago, R.H.S. Reiser, On interval fuzzy S-implications, Inf. Sci. 180 (8) (2010) 1373–1389.
[8] G. Beliakov, A. Pradera, T. Calvo, Aggregation Functions: A Guide for Practitioners, Springer, Berlin, 2007.
[9] H. Bustince, J. Fernandez, A. Kolesárová, R. Mesiar, Directional monotonicity of fusion functions, Eur. J. Oper. Res. 244 (1) (2015) 300–308.
[10] H. Bustince, J. Fernandez, R. Mesiar, J. Montero, R. Orduna, Overlap functions, nonlinear analysis: theory, Methods Appl. 72 (3–4) (2010) 1488–1499.
[11] H. Bustince, J.A. Sanz, G. Lucca, G.P. Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, G. Ochoa, Pre-aggregation functions: definition, properties and construction methods, in: Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on, IEEE, 2016.
[12] H. Bustince, J.A. Sanz, G. Lucca, G.P. Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, G. Ochoa, Pre-aggregation functions: definition, properties and construction methods, in: 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, Los Alamitos, 2016.
[13] G. Choquet, Theory of capacities, Annales de l'Institut Fourier 5 (1953-1954) 131–295.
[14] O. Cordón, M.J. del Jesus, F. Herrera, A proposal on reasoning methods in fuzzy rule-based classification systems, Int. J. Approx. Reason. 20 (1) (1999) 21–45.
[15] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[16] J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm Evol. Comput. 1 (1) (2011) 3–18.
[17] G.P. Dimuro, On interval fuzzy numbers, 2011 Workshop-School on Theoretical Computer Science, WEIT 2011, IEEE, Los Alamitos, 2011.
[18] G.P. Dimuro, B. Bedregal, J. Fernandez, Additive generators of overlap functions, in: H. Bustince, R. Mesiar, T. Calvo (Eds.), Aggregation Functions in Theory and in Practice, vol. 228 of Advances in Intelligent Systems and Computing, Springer, Berlin, 2013, pp. 167–178.
[19] G.P. Dimuro, B. Bedregal, Archimedean overlap functions: the ordinal sum and the cancellation, idempotency and limiting properties, Fuzzy Sets Syst. 252 (2014) 39–54.
[20] G.P. Dimuro, B. Bedregal, On residual implications derived from overlap functions, Inf. Sci. 312 (2015) 78–88.
[21] G.P. Dimuro, B. Bedregal, H. Bustince, M.J. Asiáin, R. Mesiar, On additive generators of overlap functions, Fuzzy Sets Syst. 287 (2016) 76–96. Theme: Aggregation Operations
[22] G.P. Dimuro, B. Bedregal, H. Bustince, J. Fernandez, G. Lucca, R. Mesiar, New results on pre-aggregation functions, in: Uncertainty Modelling in Knowledge Engineering and Decision Making, Proceedings of the 12th International FLINS Conference (FLINS 2016), vol. 10 of World Scientific Proceedings Series on Computer Engineering and Information Science, World Scientific, Singapura, 2016, pp. 213–219.
[23] G.P. Dimuro, B. Bedregal, H. Bustince, A. Jurio, M. Baczyński, K. Miś, *QL*-operations and *QL*-implication functions constructed from tuples (*O, G, N*) and the generation of fuzzy subsethood and entropy measures, Int. J. Approx. Reason. 82 (2017) 170–192, doi:10.1016/j.ijar.2016.12.013. ISSN 0888-613X.
[24] G.P. Dimuro, B. Bedregal, R.H.N. Santiago, On (*G, N*)-implications derived from grouping functions, Inf. Sci. 279 (2014) 1–17.
[25] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd, Wiley-Interscience, 2000.
[26] M. Elkano, M. Galar, J. Sanz, H. Bustince, 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, Inf. Sci. 332 (2016) 94–114.
[27] M. Elkano, M. Galar, J. Sanz, A. Fernández, E. Barrenechea, F. Herrera, H. Bustince, Enhancing multi-class classification in FARC-HD fuzzy classifier: on the synergy between n-dimensional overlap functions and decomposition strategies, IEEE Trans. Fuzzy Syst. 23 (5) (2015) 1562–1580.
[28] S. García, A. Fernández, J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics–based machine learning: accuracy and interpretability, Soft Comput. 13 (10) (2009) 959–977.
[29] D. Gómez, J.T. Rodríguez, J. Montero, H. Bustince, E. Barrenechea, N-dimensional overlap functions, Fuzzy Sets Syst. 287 (2016) 57–75. Theme: Aggregation Operations
[30] M. Grabisch, J. Marichal, R. Mesiar, E. Pap, Aggregation Functions, Cambridge University Press, Cambridge, 2009.
[31] J.L. Hodges, E.L. Lehmann, Ranks methods for combination of independent experiments in analysis of variance, Ann. Math. Stat. 33 (1962) 482–497.
[32] S. Holm, A simple sequentially rejective multiple test procedure, Scand. J. Stat. 6 (1979) 65–70.
[33] J. Hühn, E. Hüllermeier, FURIA: an algorithm for unordered fuzzy rule induction, Data Min. Knowl. Discov. 19 (3) (2009) 293–319.
[34] H. Ishibuchi, T. Nakashima, Effect of rule weights in fuzzy rule-based classification systems, Fuzzy Syst. IEEE Trans. 9 (4) (2001) 506–515.
[35] H. Ishibuchi, T. Nakashima, M. Nii, Classification and Modeling with Linguistic Information Granules, Advanced Approaches to Linguistic Data Mining, Advanced Information Processing, Springer, Berlin, 2005.
[36] E.P. Klement, R. Mesiar, E. Pap, Triangular Norms, Kluwer Academic Publisher, Dordrecht, 2000.
[37] E. Lehrer, A new integral for capacities, Econ. Theory 39 (1) (2009) 157–176.
[38] E. Lehrer, R. Teper, The concave integral over large spaces, Fuzzy Sets Syst. 159 (16) (2008) 2130–2144.
[39] G. Lucca, G.P. Dimuro, V. Mattos, B. Bedregal, H. Bustince, J.A. Sanz, 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, 2015.
[40] G. Lucca, J. Sanz, G.P. Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, H.B. Sola, Pre-aggregation functions: construction and an application, IEEE Trans. Fuzzy Syst. 24 (2) (2016) 260–272.
[41] G. Lucca, J.A. Sanz, G.P. Dimuro, B. Bedregal, M.J. Asiain, M. Elkano, H. Bustince, CC-integrals: Choquet-like copula-based aggregation functions and its application in fuzzy rule-based classification systems, Knowl. Based Syst. 119 (2017) 32–43.

[42] G. Lucca, J.A. Sanz, G.P. Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, H. Bustince, The notion of pre-aggregation function, in: V. Torra, T. Narukawa (Eds.), Modeling Decisions for Artificial Intelligence: 12th International Conference, MDAI 2015, Skövde, Sweden, September 21–23, 2015, Proceedings, Springer International Publishing, Cham, 2015, pp. 33–41.
[43] G. Lucca, R. Vargas, G.P. Dimuro, J. Sanz, H. Bustince, B. Bedregal, Analysing some t-norm-based generalizations of the Choquet integral for different fuzzy measures with an application to fuzzy rule-based classification systems, in: P. Santos, R. Prudencio (Eds.), ENIAC 2014 - Encontro Nacional de Inteligência Artificial e Computacional, SBC, São Carlos, 2014.
[44] G. Mayor, E. Trillas, On the representation of some aggregation functions, in: Proceedings of IEEE International Symposium on Multiple-Valued Logic, IEEE, Los Alamitos, 1986.
[45] R. Mesiar, Choquet-like integrals, J. Math. Anal. Appl. 194 (2) (1995) 477–488.
[46] T. Murofushi, M. Sugeno, Fuzzy t-conorm integral with respect to fuzzy measures: generalization of Sugeno integral and Choquet integral, Fuzzy Sets Syst. 42 (1) (1991) 57–71.
[47] T. Murofushi, M. Sugeno, M. Machida, Non-monotonic fuzzy measures and the Choquet integral, Fuzzy Sets Syst. 64 (1) (1994) 73–86.
[48] R.B. Nelsen, An introduction to copulas, vol. 139 of Lecture Notes in Statistics, Springer, New York, 1999.
[49] J. Sanz, D. Bernardo, F. Herrera, H. Bustince, H. Hagras, A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data, IEEE Trans. Fuzzy Syst. 23 (4) (2015) 973–990.
[50] J. Sanz, H. Bustince, A. Fernández, F. Herrera, IIVFDT: ignorance functions based interval-valued fuzzy decision tree with genetic tuning, Int. J. Uncertain. Fuzziness Knowl. Based Syst. 20 (supp02) (2012) 1–30.
[51] J. Sanz, A. Fernández, H. Bustince, F. Herrera, IVTURS: a linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection, IEEE Trans. Fuzzy Syst. 21 (3) (2013) 399–411.
[52] J.A. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, H. Bustince, Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system, Appl. Soft Comput. 20 (2014) 103–111.
[53] D. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, 2nd, Chapman & Hall/CRC, 2006.
[54] V. Torra, Y. Narukawa, The interpretation of fuzzy integrals and their application to fuzzy systems, Int. J. Approx. Reason. 41 (2006) 43–58.
[55] Z. Wang, G.J. Klir, Generalized Measure Theory, Springer, Boston, 2009.
[56] Z. Wang, K.-S. Leung, M.-L. Wong, J. Fang, A new type of nonlinear integrals and the computational algorithm, Fuzzy Sets Syst. 112 (2) (2000) 223–231.
[57] X. Wen, L. Shao, Y. Xue, W. Fang, A rapid learning algorithm for vehicle classification, Inf. Sci. 295 (2015) 395–406.

5 Improving the performance of fuzzy rule-based classification systems based on a new non averaging generalization of CC-integrals named $C_{F_1F_2}$-integrals

- G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, H. Bustince and R. Mesiar, "Improving the performance of fuzzy rule-based classification systems based on a new non averaging generalization of CC-integrals named $C_{F_1F_2}$-integrals", IEEE Transactions on Fuzzy Systems (submitted).
  - Journal: IEEE Transactions on fuzzy Systems
  - Status: Submitted
  - Impact Factor (JCR 2016): 7.671
  - Knowledge Area:
    - Artificial Intelligence: Ranking 4/133 (Q1)
    - Computer Science: Ranking 4/133 (Q1)
    - Engineering, Electrical & Electronic: Ranking 9/260 (Q1)

IEEE Computational Intelligence Society Transactions on Fuzzy Systems

Improving the performance of fuzzy rule-based classification systems based on a new non-averaging generalization of CC-integrals named C_{F1F2}-integrals

Journal:	*Transactions on Fuzzy Systems*
Manuscript ID	TFS-2018-0151
Manuscript Type:	Full Paper
Date Submitted by the Author:	23-Feb-2018
Complete List of Authors:	Lucca, Giancarlo; Universidad Publica de Navarra, Automática y computación Pereira Dimuro, Graçaliz; Universidade Federal do Rio Grande, Centro de Ciências Computacionais Fernandez, Javier; Universidad Publica de Navarra, Automatica y Computacion Bustince, Humberto; Universidad Publica de Navarra, Automatica y Computacion; Bedregal, Benjamin; Federal University of Rio Grande do Norte, Department of Informatics and Applied Mathematics Sanz, José; Universidad Publica de Navarra, Automatica y Computacion
Keywords:	Fuzzy rule-based classification systems, Choquet Integral, CC-integrals, OD monotone functions

Improving the performance of fuzzy rule-based classification systems based on a new non-averaging generalization of CC-integrals named $C_{F_1F_2}$-integrals

Giancarlo Lucca, Graçaliz Pereira Dimuro *IEEE Member*, Javier Fernández, Humberto Bustince, *IEEE Senior Member*, Benjamín Bedregal *IEEE Member*, José Antonio Sanz

***Abstract*—A key component of Fuzzy Rule-Based Classification Systems (FRBCSs) is the Fuzzy Reasoning Method (FRM), since it infers the class predicted for new examples. A crucial stage in any FRM is the way in which the information given by the fired rules during the inference process is aggregated. A widely used FRM is the winning rule, which applies the maximum to accomplish this aggregation. The maximum is an averaging operator, which means that its result is within the range delimited by the minimum and the maximum of the aggregated values. Over the last years, new averaging operators based on generalizations of the Choquet integral were also proposed to perform this aggregation process. However, the most accurate FRBCSs use the FRM known as additive combination, that considers the normalized sum as the aggregation operator, which is non-averaging. For this reason, this paper is aimed at introducing a new non averaging operator named $C_{F_1F_2}$-integral, which is a generalization of the Choquet-like Copula-based integral (CC-integral). $C_{F_1F_2}$-integrals present the desired properties of an aggregation-like operator, since they satisfy appropriate boundary conditions and have some kind of increasingness property. We show that $C_{F_1F_2}$-integrals, when used to cope with classification problems, enhance the results of the previous averaging generalizations of the Choquet integral and they provide competitive results (even better) when compared with state-of-the-art FRBCSs.**

***Index Terms*—Fuzzy rule-based classification systems, Choquet Integral, $C_{F_1F_2}$-integrals, CC-integrals, OD monotone functions.**

I. Introduction

In a supervised classification problem [1] it is necessary to determine the class of an example based on the information given by labeled examples. Among others, an accurate way to tackle classification problems is by using Fuzzy Rule-Based Classification Systems (FRBCSs) [2]. This technique achieves accurate results taking into consideration linguistic labels in the rules, which leads to obtaining an interpretable model that can be easily used in the decision making process.

The two main components of FRBCSs are the knowledge base, which is composed of the rule base and the data base, and the Fuzzy Reasoning Method (FRM) [3]. The latter is a mechanism that uses the information available in the knowledge base to assign a class to new examples that have to be classified. The FRM of the winning rule is a classical inference process found in the literature that assigns the class of the fuzzy rule whose compatibility with the example to be classified is maximum. To do so, it applies the maximum as the aggregation operator, which has an averaging characteristic, i.e., its result is delimited by the minimum and the maximum of the values to be aggregated. Consequently, to classify an example it only uses the information given by one fuzzy rule and it disregards the remainder information.

Barrenechea et. al proposed in [4] a FRM that takes into account the information provided by all the fired rules using the Choquet integral [5]. After that, the Choquet integral was generalized by replacing the standard product operator by different t-norms, which led to the concept of pre-aggregation functions [6]. Next, aiming at producing an aggregation function, in [7] the authors presented the Choquet-like Copula-based integral (CC-integrals for short). They swapped the product operator of the extended form of the Choquet integral by two identical copulas C. These three approaches have averaging characteristics [8] and they provide competitive results in classification problems.

However, the state of the art FRBCSs algorithms, like IVTURS [9], FARC-HD [10] or FURIA [11] apply the FRM known as additive combination that is based on the usage of the normalized sum as aggregation function [3], which has a non-averaging behavior. Taking this fact into account, in [12] the standard Choquet integral was generalized by replacing the product operation by different functions F, introducing the concept of C_F-integrals. These integrals are pre-aggregations that may have either averaging or non-averaging characteristics according to the considered function F. The authors showed that the non-averaging C_F-integrals statistically overcome the averaging ones, reinforcing the quality of the usage of non-averaging functions in this domain.

For this reason, in this paper we define a new generalization of the CC-integral, named $C_{F_1F_2}$-integral, substituting the copula C by two fusion functions F_1 and F_2 satisfying some special conditions. The new $C_{F_1F_2}$-integrals are non-averaging

G. Lucca is with Departamento of Automática y Computación, Universidad Publica de Navarra, Navarra, 31006 Spain e-mail: lucca.112793@e.unavarra.es

H. Bustince, J. Fernández and J. Sanz are with the Departamento of Automática y Computación and with the Institute of Smart Cities, Universidad Publica de Navarra, Navarra, 31006 Spain e-mails: {joseantonio.sanz,bustince}@unavarra.es.

G. Pereira Dimuro is with Institute of Smart Cities, Universidad Publica de Navarra, Pamplona, Spain, and Centro de Ciências Computacionais, Universidade Federal do Rio Grande, Rio Grande, Brazil, e-mail: gracalizdimuro@furg.br.

B. Bedregal is with Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte, Natal, Brazil, e-mail: bedregal@dimap.ufrn.br.

Ordered Directionally (OD) increasing functions satisfying the required boundary conditions for any "aggregation-like operator". Moreover, we present a methodology to select the best pairs of fusion functions F_1 and F_2 to define the $C_{F_1F_2}$-integrals to tackle classification problems. Finally, we also introduce a new FRM based on this new concept.

In the experimental study, we consider 33 different datasets available in Keel dataset repository [13]. We analyze the quality of our method by selecting the best $C_{F_1F_2}$-integrals and comparing them against the state-of-the-art fuzzy classifiers, the best C_F-integral presented in [12] and the classical probabilistic sum [14] applied in the FRM of a FRBCS as it is a known non-averaging function. The quality of the results is supported by a proper statistical study as suggested in the specialized literature [15], [16], [17].

The paper is organized as follows. In Section II, we introduce the background necessary to understand the paper. In Section III, we introduce the concept of $C_{F_1F_2}$-integrals, showing that these functions are OD increasing functions satisfying appropriate conditions. We describe in Section IV the use of $C_{F_1F_2}$-integrals in the FRM and the evolutionary learning of the fuzzy measure. Section V introduces the experimental framework, describing the datasets along with the setup configuration for the different methodologies and the statistical tests used for performance comparison. In Section VI we present the experimental results achieved in testing by $C_{F_1F_2}$-integrals and we draw the main conclusions in Section VII.

II. Preliminaries

In this section, we present some basic theoretical concepts that are necessary to develop the paper. Any n-ary function $F : [0,1]^n \to [0,1]$ is named fusion function as it receives n values and it fuses them returning a single one.

Definition 1. *[18], [19] A fusion function $A : [0,1]^n \to [0,1]$ is an aggregation function whenever the following conditions hold:*

(A1) *A is increasing[1] in each argument: for each $i \in \{1,\dots,n\}$, if $x_i \leq y$, then $A(x_1,\dots,x_n) \leq A(x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_n)$;*

(A2) *A satisfies the boundary conditions: (i) $A(0,\dots,0) = 0$ and (ii) $A(1,\dots,1) = 1$.*

An aggregation function $A : [0,1]^n \to [0,1]$ is said to be averaging if and only if: **(AV)** $\forall(x_1,\dots,x_n) \in [0,1]^n$: $\min\{x_1,\dots,x_n\} \leq A(x_1,\dots,x_n) \leq \max\{x_1,\dots,x_n\}$.

Definition 2. *[20] Let $\vec{r} = (r_1,\dots,r_n)$ be a real n-dimensional vector, $\vec{r} \neq \vec{0}$. A function $F : [0,1]^n \to [0,1]$ is said to be $\vec{r}$-increasing if for all $\vec{x} = (x_1,\dots,x_n) \in [0,1]^n$ and for all $c > 0$ such that $\vec{x}+c\vec{r} = (x_1+cr_1,\dots,x_n+cr_n) \in [0,1]^n$ it holds*

$$F(\vec{x} + c\vec{r}) \geq F(x_1,\dots,x_n). \qquad (1)$$

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

[1]For an increasing (decreasing) function we do not mean a strictly increasing (decreasing) function.

Definition 3. *[6] A function $PA : [0,1]^n \to [0,1]$ is said to be an n-ary pre-aggregation function if the following conditions hold:*

(PA1) *Directional increasingness: there exists $\vec{r} = (r_1,\dots,r_n) \in [0,1]^n$, $\vec{r} \neq \vec{0}$, such that PA is $\vec{r}$-increasing;*

(PA2) *Boundary conditions: (i) $PA(0,\dots,0) = 0$ and (ii) $PA(1,\dots,1) = 1$.*

If F is a pre-aggregation function with respect to a vector $\vec{r}$ we just say that F is an $\vec{r}$-pre-aggregation function.

In what follows, denote $N = \{1,\dots,n\}$ for an arbitrary $n > 0$.

Definition 4. *[5], [21][22, Definition 1.77] A function $\mathfrak{m} : 2^N \to [0,1]$ is said to be a fuzzy measure if, for all $X, Y \subseteq N$, the following conditions hold:*

($\mathfrak{m}$1) *Increasingness: if $X \subseteq Y$, then $\mathfrak{m}(X) \leq \mathfrak{m}(Y)$;*

($\mathfrak{m}$2) *Boundary conditions: $\mathfrak{m}(\emptyset) = 0$ and $\mathfrak{m}(N) = 1$.*

A fuzzy measure $\mathfrak{m}$ is symmetric whenever $\mathfrak{m}(X) = \mathfrak{m}(Y)$ for all $X, Y \subseteq N$ such that $\mid X \mid = \mid Y \mid$.

Definition 5. *[5] Let $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure. The discrete Choquet integral is the function $\mathfrak{C}_\mathfrak{m} : [0,1]^n \to [0,1]$, defined, for all of $\vec{x} = (x_1,\dots,x_n) \in [0,1]^n$, by:*

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

where $\left(x_{(1)},\dots,x_{(n)}\right)$ is an increasing permutation on the input $\vec{x}$, that is, $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$, where $x_{(0)} = 0$ and $A_{(i)} = \{(i),\dots,(n)\}$ is the subset of indices corresponding to the $n-i+1$ largest components of $\vec{x}$.

The CC-integral [7] is a generalization of the Choquet integral using copulas [14].

Definition 6. *[7, Definition 7] Let $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure and $C : [0,1]^2 \to [0,1]$ a copula. The CC-integral with respect to $\mathfrak{m}$ is the function $\mathfrak{C}^C_\mathfrak{m} : [0,1]^n \to [0,1]$, defined, for all of $\vec{x} = (x_1,\dots,x_n) \in [0,1]^n$, by:*

$$\mathfrak{C}^C_\mathfrak{m}(\vec{x}) = \sum_{i=1}^{n} \left(C\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - C\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right). \qquad (3)$$

III. A generalization of CC-integrals using two fusion functions F_1 and F_2

In this section, we introduce a method for constructing a generalization of CC-integrals, named $C_{F_1F_2}$-integral, using two fusion functions F_1 and F_2 satisfying some specific properties instead of the same copula C (Section III.A). We also present a mechanism for choosing the functions F_1 and F_2 that should work fine when applied in the FRM of FRBCSs (Section III.B). Finally, we prove that $C_{F_1F_2}$-integrals built with some specific pairs of fusion functions F_1 and F_2 are non-averaging OD increasing functions satisfying proper boundary conditions to be applied in the FRM (Section III.C).

A. Defining the $C_{F_1F_2}$-integrals

In this subsection, we aim at introducing the definition of $C_{F_1F_2}$-integrals and analyzing some properties for specific pairs of fusion functions F_1 and F_2.

An important concept used in this paper is the dominance (or, conversely, subordination) property:

(DM) F_1-Dominance (or, equivalently, F_2-Subordination): $F_1 \geq F_2$, that is: $\forall$ x,y $\in$ [0,1]: F_1 (x,y) $\geq$ F_2(x,y)

Definition 7. *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*
(ii) *F_1 is $(1,0)$-increasing,*

A $C_{F_1F_2}$-integral is defined as a function $\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)} : [0,1]^n \to [0,1]$, given, for all $x \in [0,1]^n$, by

$$\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x}) = \min\left\{1, x_{(1)} + \sum_{i=2}^{n} F_1\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - F_2\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\}, \quad (4)$$

where $(x_{(1)}, \ldots, x_{(n)})$ is an increasing permutation on the input $\vec{x}$, 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 indices of $n - i + 1$ largest components of $\vec{x}$.

Observe that it is immediate that $\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}$ is well defined, for any pair $F_1, F_2 : [0,1]^2 \to [0,1]$ and a symmetric fuzzy measure $\mathfrak{m}$.

Remark 1. *Observe that the first element of the summation in the definition of $\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}$ is just $x_{(1)}$ instead of*

$$F_1\left(x_{(1)}, \mathfrak{m}\left(A_{(1)}\right)\right) - F_2\left(x_{(0)}, \mathfrak{m}\left(A_{(1)}\right)\right).$$

This is considered to avoid the initial discrepant behavior of non-averaging functions in the initial phase of the aggregation process. For example, consider an unitary vector $\vec{x} = 0.9 \in [0,1]$ and $F_1 = F_2 = AVG$, where $AVG(x,y) = \frac{x+y}{2}$ is the arithmetic mean. Then, if we included the first element in the summation of the integral the result would be:

$$\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x}) = \min\left\{1, \sum_{i=1}^{n} F_1\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - F_2\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\}, \quad (5)$$

$$= \min\left\{1, \frac{0.9+1}{2} - \frac{0+1}{2}\right\} = 0.45 \neq 0.9.$$

Observe here the large discrepancy of the result, since one expects that the aggregated value would be 0.9. Using our definition of $C_{F_1F_2}$-integral (Equation (4)) this unexpected behavior is avoided and the result is 0.9.

Table I shows the definitions of fusion functions $F : [0,1]^2 \to [0,1]$ that are $(1,0)$-increasing (condition **(ii)** of Definition 7) and, thus, candidates to be used as F_1 and/or F_2 in the definition of $C_{F_1F_2}$-integrals. The expression of the function is introduced in the first column whereas in the second and in the third columns we show the family (or families) of the function and the source where they were published, respectively.

As we have mentioned, all these functions fulfill condition **(ii)** of Definition 7. Therefore, we need to study whether they fulfill condition **(i)**. Consequently, we conduct the study about the dominance property in the next subsection.

B. Analyzing the Dominance (Subordination) property

In this paper, $C_{F_1F_2}$-integrals are applied in the FRM of FRBCSs (see Section IV). In this environment, the discrimination power is usually driven by fuzzy rules having the best compatibility with the examples. Therefore, we need to accentuate the differences among large values. According to Equation (5), F_1 deals with the larger value in each iteration of the summation. For this reason, it is desirable to have a high domination of F_1 over F_2, so that the difference among the two values becomes larger, which can lead to improve the system's performance.

Consequently, the dominance property is an important concept used in this work, playing a central role in the construction of $C_{F_1F_2}$-integrals applied in the FRM of FRBCSs discussed in this paper. For this reason, we analyze such property in order to determine which fusion functions, among those presented in Table I, are more suitable to be F_1 and F_2 in the construction the $C_{F_1F_2}$-integrals.

To do so, we define the concepts of dominance and subordination strength degrees. 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 for $j \in \{1, \ldots, m\}$ as follows:

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

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

That is, the DSt and SSt degrees of a fusion function F take into account the number of functions in which F dominates, or is subordinated to, respectively.

Table II presents the analysis of the dominance property for the functions presented in Table I. In this table a cell is marked with the ✓ symbol when the function introduced in the row dominates the one shown in the column. Furthermore, we also show in this table the DSt of the function in the row (it conforms the last column) and the SSt of the function in the column (it conforms the last row), which imply a total number of possible combinations of 50.

Since the number of possible combinations of fusion functions, marked with ✓ in Table II, for F_1 and F_2 is too high (201 different combinations), we propose a methodology to reduce the scope of this study. We consider the DSt and SSt degrees to be Low, Medium and High when they are less than 33%, between 34% and 66% and larger than 66%, respectively. Then, we have selected three functions of each category (Low, Medium, High) for both DSt and SSt to play the role of functions F_1 and/or F_2 respectively. The selected functions

TABLE II: Analysis of the dominance property of the fusion functions introduced in Table I

	T_P	T_M	$T_Ł$	T_{DP}	T_{HP}	O_B	O_{mM}	O_α	O_{Div}	GM	HM	S	F_{RS}	C_F	C_L	F_{GL}	F_{BPC}	F_{NA}	F_α	F_{NA2}	AVG	F_{IM}	F_{IP}	DSt (%)
T_P	✓		✓	✓			✓										✓							21.74
T_M	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓	✓		✓	✓						56.52
$T_Ł$			✓	✓																				8.70
T_{DP}				✓																				4.35
T_{HP}	✓		✓	✓	✓		✓	✓						✓			✓							34.78
O_B	✓		✓	✓		✓	✓	✓									✓							30.43
O_{mM}				✓			✓										✓							13.04
O_α	✓		✓	✓			✓	✓									✓							26.09
O_{Div}	✓		✓	✓		✓	✓	✓	✓								✓							34.78
GM	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓						69.57
HM	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓			✓	✓		✓	✓						60.87
S	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	100
F_{RS}	✓		✓	✓		✓	✓	✓					✓				✓							34.78
C_F	✓		✓	✓			✓							✓			✓							26.09
C_L			✓	✓											✓		✓							17.39
F_{GL}	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓					78.26
F_{BPC}				✓											✓		✓							13.04
F_{NA}				✓											✓			✓						13.04
F_α				✓											✓				✓					13.04
F_{NA2}				✓											✓			✓		✓				17.39
AVG	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓			82.61
F_{IM}	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓	✓		✓	✓	✓			✓		65.22
F_{IP}	✓	✓	✓	✓	✓	✓	✓	✓	✓					✓	✓		✓	✓	✓			✓	✓	69.57
SSt (%)	65.21	34.78	73.91	100	39.13	47.82	69.56	56.52	39.13	17.39	21.73	4.34	21.73	43.47	56.52	8.69	78.26	43.47	26.08	13.04	8.69	13.04	8.69	

TABLE I: $(1, 0)$-increasing fusion functions initially considered in this study

Definition	Family	Reference
$T_M(x,y) = \min\{x,y\}$	t-norm, copula	[23]
$T_P(x,y) = xy$	t-norm, copula	[23]
$T_Ł(x,y) = \max\{0, x+y-1\}$	t-norm, copula	[23]
$T_{HP}(x,y) = \begin{cases} 0 & \text{if } x=y=0 \\ \frac{xy}{x+y-xy} & \text{otherwise} \end{cases}$	t-norm, copula	[23]
$T_{DP}(x,y) = \begin{cases} x & \text{if } y=1 \\ y & \text{if } x=1 \\ 0 & \text{otherwise} \end{cases}$	t-norm	[23]
$O_B(x,y) = \min\{x\sqrt{y}, y\sqrt{x}\}$	overlap function, copula	[24], [25]
$O_{mM}(x,y) = \min\{x,y\}\max\{x^2,y^2\}$	overlap function, copula	[26], [27], [28], [29]
$O_\alpha(x,y) = xy(1+\alpha(1-x)(1-y))$, $\alpha \in [-1,0[\cup]0,1]$	overlap function, copula	[12], [14]
$O_{Div}(x,y) = \frac{xy+\min\{x,y\}}{2}$	overlap function, copula	[14]
$GM(x,y) = \sqrt{xy}$	overlap function	[30]
$HM(x,y) = \begin{cases} 0 & \text{if } x=0 \text{ or } y=0 \\ \frac{2}{\frac{1}{x}+\frac{1}{y}} & \text{otherwise} \end{cases}$	overlap function	[30]
$S(x,y) = \sin\left(\frac{\pi}{2}(xy)^{\frac{1}{4}}\right)$	overlap function	[30]
$C_F(x,y) = xy + x^2y(1-x)(1-y)$	copula	[23], [7]
$C_L(x,y) = \max\{\min\{x, \frac{y}{2}\}, x+y-1\}$	copula	[14], [7]
$AVG(x,y) = \frac{x+y}{2}$	overlap	
$F_{RS}(x,y) = \min\left\{\frac{(x+1)\sqrt{y}}{2}, y\sqrt{x}\right\}$	aggregation function	[12]
$F_{GL}(x,y) = \sqrt{\frac{x(y+1)}{2}}$	aggregation function	[12]
$F_{BPC}(x,y) = xy^2$	aggregation function	[18]
$F_\alpha(x,y) = \begin{cases} \alpha x & \text{if } x<y \\ \max\{\alpha x, y\} & \text{otherwise} \end{cases}, 0<\alpha<1$	aggregation function	[12]
$F_{NA}(x,y) = \begin{cases} x & \text{if } x \le y \\ \min\{\frac{x}{2}, y\} & \text{otherwise} \end{cases}$	Pre-aggregation function	[12]
$F_{NA2}(x,y) = \begin{cases} 0 & \text{if } x=0 \\ \frac{x+y}{2} & \text{if } 0<x\le y \\ \min\{\frac{x}{2}, y\} & \text{otherwise} \end{cases}$	Pre-aggregation function	[12]
$F_{IM}(x,y) = \max\{1-y, x\}$	Non Pre-Aggregation function	[12]
$F_{IP}(x,y) = 1-y+xy$	Non Pre-Aggregation function	[12]

* When $\alpha = 0$, we have that $O_\alpha = T_P$, the product t-norm.

TABLE III: Summary of the adopted functions according to dominance/subordination strength degrees

Strength degree	Dominance (F_1)	Subordination (F_2)
Low	T_{DP}	S
	F_{NA}	GM
	O_B	F_α
Medium	T_{HP}	T_M
	T_M	F_{NA}
	F_{IM}	T_P
High	GM	$T_Ł$
	F_{GL}	F_{BPC}
	S	T_{DP}

according to this methodology are presented in Table III, which imply a total number of possible combinations of 81.

C. *The selected $C_{F_1F_2}$-integrals as non-averaging OD monotone functions*

For the aggregation process in the FRM to be well defined it is necessary an operator that has two characteristics. First, some kind of increasingness property is required in order to guarantee that the more information is provided the higher is the aggregated value (condition **(A1)** of Def. 1 and Def. 3). Second, the aggregation operator must satisfy boundary conditions related to the domain $[0, 1]$ (condition **(A2)** of Def. 1 and Def. 3).

Our selected $C_{F_1F_2}$-integrals (Table III) satisfy the boundary conditions **(A2)** of an (pre) aggregation function. However, our selected $C_{F_1F_2}$-integrals are neither increasing nor directionally increasing. However, we have noticed that they do present some kind of increasingness property. In fact, they are Ordered Directionally (OD) monotone functions [31]. Such functions are monotonic along different directions according to the ordinal size of the coordinates of each input.

In this section, we prove such properties for our best $C_{F_1F_2}$-integral, according to the results shown in Section VI (GM–F_{BPC}), since the proofs for the other pairs of fusion functions F_1 and F_2, considered in this paper, are analogous. We also show that they are non-averaging functions.

Definition 8. *[31] Consider a function $F : [0,1]^n \to [0;1]$ and let $\vec{r} = (r_1, \ldots, r_n)$ be a real n-dimensional vector, $\vec{r} \neq \vec{0}$. F is said to be ordered directionally (OD) $\vec{r}$-increasing if, for each $\vec{x} \in [0,1]^n$, any permutation $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ with $x_{\sigma(1)} \geq \ldots \geq x_{\sigma(n)}$, and $c > 0$ such that $1 \geq x_{\sigma(1)} + cr_1 \geq \ldots \geq x_{\sigma(n)} + cr_n$, it holds that $F(\vec{x} + c\vec{r}_{\sigma^{-1}}) \geq F(\vec{x})$, where $\vec{r}_{\sigma^{-1}} = (r_{\sigma^{-1}(1)}, \ldots, r_{\sigma^{-1}(n)})$. Similarly, one defines an ordered directionally (OD) $\vec{r}$-decreasing function.*

Theorem 1. *For any symmetric fuzzy measure $\mathfrak{m} : 2^N \to [0, 1]$*

and $k > 0$, $\mathfrak{C}_{\mathfrak{m}}^{(GM,F_{BPC})}$, where GM and F_{BPC} are defined in Table I, is an (OD) $(k, 0, \ldots, 0)$-increasing function.

Proof. For all $\vec{x} \in [0,1]^n$ and permutation $\sigma : \{1,\ldots,n\} \to \{1,\ldots,n\}$ with $x_{\sigma(1)} \geq \ldots \geq x_{\sigma(n)}$ and $c > 0$ such that $x_{\sigma(i)} + cr_i \in [0,1]$, for $i \in \{1,\ldots,n\}$, and $1 \geq x_{\sigma(1)} + cr_1 \geq \ldots \geq x_{\sigma(n)} + cr_n$, for $\vec{r}_{\sigma^{-1}} = (r_{\sigma^{-1}(1)}, \ldots, r_{\sigma^{-1}(n)})$, one has that:

$$\begin{aligned}
&\mathfrak{C}_{\mathfrak{m}}^{(GM,F_{BPC})}(\vec{x} + c\vec{r}_{\sigma^{-1}}) \\
&= \min\Big\{1, (x_{(1)} + c \cdot r_{\sigma^{-1}(1)}) \\
&\quad + \sum_{i=2}^{n-1} \Big(\sqrt{(x_{(i)} + c \cdot r_{\sigma^{-1}(i)}) \cdot \mathfrak{m}(A_{(i)})} \\
&\qquad - (x_{(i-1)} + c \cdot r_{\sigma^{-1}(i-1)}) \cdot \mathfrak{m}(A_{(i)})^2 \Big) \\
&\quad + \sqrt{(x_{(n)} + c \cdot r_{\sigma^{-1}(n)}) \cdot \mathfrak{m}(A_{(n)})} \\
&\quad - (x_{(n-1)} + c \cdot r_{\sigma^{-1}(n-1)}) \cdot \mathfrak{m}(A_{(n)})^2 \Big\} \\
&= \min\Big\{1, (x_{(1)} + c \cdot 0) \\
&\quad + \sum_{i=2}^{n-1} \Big(\sqrt{(x_{(i)} + c \cdot 0) \cdot \mathfrak{m}(A_{(i)})} \\
&\qquad - (x_{(i-1)} + c \cdot 0) \cdot \mathfrak{m}(A_{(i)})^2 \Big) \\
&\quad + \sqrt{(x_{(n)} + c \cdot k) \cdot \mathfrak{m}(A_{(n)})} \\
&\quad - (x_{(n-1)} + c \cdot 0) \cdot \mathfrak{m}(A_{(n)})^2 \Big\} \\
&\geq \min\Big\{1, x_{(1)} + \\
&\quad \sum_{i=2}^{n-1} \Big(\sqrt{x_{(i)} \cdot \mathfrak{m}(A_{(i)})} - x_{(i-1)} \cdot \mathfrak{m}(A_{(i)})^2 \Big) \\
&\quad + \sqrt{x_{(n)} \cdot \mathfrak{m}(A_{(n)})} - x_{(n-1)} \cdot \mathfrak{m}(A_{(n)})^2 \Big\} \\
&= \mathfrak{C}_{\mathfrak{m}}^{(GM,F_{BPC})}(\vec{x}),
\end{aligned}$$

since GM dominates F_{BPC} and GM is $(1,0)$-increasing. Thus, $\mathfrak{C}_{\mathfrak{m}}^{(GM,F_{BPC})}$ is OD ($k, 0, \ldots, 0)$-increasing, for $k > 0$. □

Theorem 2. *For any symmetric fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$, $\mathfrak{C}_{\mathfrak{m}}^{(GM,F_{BPC})}$ satisfies the boundary conditions* **(A2)**.

Proof. Consider $\vec{0} = (0,\ldots,0) \in [0,1]^n$ and $\vec{1} = (1,\ldots,1) \in [0,1]^n$. It follows that:

$$\begin{aligned}
&\mathfrak{C}_{\mathfrak{m}}^{(GM,F_{BPC})}(\vec{0}) \\
&= \min\left\{1, 0 + \sum_{i=2}^{n} \sqrt{0 \cdot \mathfrak{m}(A_{(i)})} - 0 \cdot \mathfrak{m}(A_{(i)})^2 \right\} = 0
\end{aligned}$$

$$\begin{aligned}
&\mathfrak{C}_{\mathfrak{m}}^{(GM,F_{BPC})}(\vec{1}) \\
&= \min\{1, 1 + \sum_{i=2}^{n} \sqrt{1 \cdot \mathfrak{m}(A_{(i)})} - 1 \cdot \mathfrak{m}(A_{(i)})^2\} = 1
\end{aligned}$$

□

Proposition 1. *For any symmetric fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$, $\mathfrak{C}_{\mathfrak{m}}^{(GM,F_{BPC})}$ is non-averaging.*

Proof. Suppose that $\mathfrak{C}_{\mathfrak{m}}^{(GM,F_{BPC})}$ is averaging. Now take $\vec{x} = (0.2, 0.5, 0.7, 0.9)$ and the power measure (Equation (10)), with $q = 1$. It follows that

$$\begin{aligned}
&\mathfrak{C}_{\mathfrak{m}}^{(GM,F_{BPC})}(\vec{x}) \\
&= \min\Big\{1, 0.2 + \sum_{i=1}^{3} (GM(x_{(i)}, \mathfrak{m}(A_{(i)})) \\
&\qquad - F_{BPC}(x_{(i-1)}, \mathfrak{m}(A_{(i)})))\Big\} \\
&= \min\left\{1, 0.2 + \sum_{i=1}^{3} \left(\sqrt{x_{(i)} \cdot \mathfrak{m}(A_{(i)})} - x_{(i-1)} \cdot \mathfrak{m}(A_{(i)})^2 \right) \right\} \\
&= \min\Big\{1, 0.2 + \sqrt{0.5 \cdot 0.75} - 0.2 \cdot 0.75^2 \\
&\quad + \sqrt{0.7 \cdot 0.5} - 0.5 \cdot 0.5^2 + \sqrt{0.9 \cdot 0.25} - 0.7 \cdot 0.25^2 \Big\} \\
&= \min\{1, 1.59\} > 0.9 = \max\{0.2, 0.5, 0.7, 0.9\}.
\end{aligned}$$

This a contradiction with the averaging property. □

IV. Using $C_{F_1F_2}$-integrals in fuzzy rule-based classification systems

In this section, our goal is to describe the main components of FRBCSs and the used fuzzy classifier. Furthermore, we present the considered FRM containing the main modification with respect to the original, which consist of the inclusion of the $C_{F_1F_2}$-integrals in the aggregation stage.

A classification problem consists of t training examples $\mathrm{x}_p = (x_{p1}, \ldots, x_{pn}, y_p)$, with $p = 1, \ldots, t$, where x_{pi}, with $i = 1, \ldots, n$, is the value of the i-th variable and $y_p \in \mathbb{C} = \{C_1, \ldots, C_M\}$ is the label of the class of the p-th training example, and M is the number of classes.

In this paper, we focus on FRBCSs. Specifically, we use the Fuzzy Association Rule-based Classification model for High Dimensional Problems (FARC-HD [10]). The structure of the fuzzy rules generated by this classifier has the following form:

$$\begin{aligned}
\text{Rule } R_j :\ & \text{If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \\
& \text{then Class is } C_j \text{ with } RW_j,
\end{aligned} \tag{6}$$

where R_j is the label of the j-th rule, A_{ji} is a fuzzy set representing a linguistic term modeled by a triangular shaped membership function. C_j is the class label and $RW_j \in [0,1]$ is the rule weight [32], which in this case is computed as the confidence of the fuzzy rule.

In order to generate the set of fuzzy rules, FARC-HD applies the following three stages:

- *Fuzzy association rule extraction for classification*: In this step, an initial fuzzy rule base is obtained. To accomplish it, for each class, a search tree [33] is constructed, whose maximum depth is limited (parameter depth$_{max}$) . For each linguistic label (item), the support and confidence are calculated in order to obtain the frequent itemsets (set of items). Then, a fuzzy rule is generated for each frequent itemset.
- *Candidate rule prescreening*: This stage considers a weighting pattern scheme [34] to select the best generated fuzzy rules.
- *Genetic rule selection and lateral tuning*: At this point, the previously generated fuzzy rules are optimized so as

to enhance as much as possible the system's performance. To do so, the CHC evolutionary algorithm [35] is applied to carry out a rule selection process and the lateral tuning of the fuzzy sets [36].

A. Application of $C_{F_1F_2}$ in the fuzzy reasoning method

Once the knowledge base has been learnt and a new example has to be classified, the FRM is responsible to perform this task. As we have mentioned, we modify the classical FRM of FARC-HD [10] to include the usage of $C_{F_1F_2}$-integrals in its third stage. The steps of the new FRM are the following ones:

1) *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 $T : [0,1]^2 \to [0,1]$.

$$\mu_{A_j}(x) = T(\mu_{A_{j1}}(x_1), \ldots, \mu_{A_{jn}}(x_n)), \quad j = 1, \ldots, L, \tag{7}$$

where L is the number of rules.

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

$$b_j^k(x) = \mu_{A_j}(x) \cdot RW_j^k, \tag{8}$$

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

3) *Example classification soundness degree for all classes*: This is the stage in which the $C_{F_1F_2}$-integrals are applied. At this point, for each class, all information given by the fired fuzzy rules is aggregated. To do so, the positive information provided by the previous step is aggregated by Equation (9).

$$S_k(x) = \mathfrak{C}_{\mathfrak{m}_k}^{C_{F_1F_2}}\left(b_1^k(x), \ldots, b_L^k(x)\right), \tag{9}$$

with k = Class(R_j) and $b_j^k > 0$.
where $C_{F_1F_2}$ is the $C_{F_1F_2}$-integral considered to perform the aggregation. We remind that we use as F_1 and F_2 the functions presented Table III (Section III-B). $C_{F_1F_2}$-integrals are functions that generalize the Choquet integral (Equation (2)) and consequently, they use a fuzzy measure. In this work, we use the symmetric fuzzy measure applied in our previous papers, that is, the power measure:

$$\mathfrak{m}_k(X) = \left(\frac{|X|}{n}\right)^{q_k}, \text{ with } q_k > 0, \tag{10}$$

where the exponent q_k is genetically learnt (see section IV-B) by an evolutionary algorithm, to obtain the most suitable value, q_k, for each class k. Consequently, we use a different measure for each class.

4) *Classification*: The final decision is made in this step. To do so, a function $F : [0,1]^M \to \{1, \ldots, M\}$ is applied over the results obtained by example classification soundness degrees of all classes:

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

B. Evolutionary learning of the fuzzy measure for each class

The original FARC-HD algorithm makes usage of the CHC evolutionary algorithm [35] to perform the lateral tuning of the fuzzy sets [36] and select the best set of fuzzy rules. In this paper we also learn a fuzzy measure for each class [4], k, by learning the q_k parameter as shown in Equation (10). The specific features of our evolutionary model are:

1) *Coding Scheme*: The chromosome is divided into three parts.
   (i) The first one considers the genes related to the tuning of lateral position of the membership functions and it has as many genes as the number of linguistic labels, where the range of each gene is [-0.5, 0.5] (for more details see [36]).
   (ii) The second part has one gene per class, k, and it is used to learn the exponent q_k. It is encoded in the range [0.01, 1.99]. However, as the real range is [0.01, 100] as showed in [4] the values of the genes have to be decoded in this range (See [4], [7] for details).
   (iii) The last part of the chromosome is related to the rule selection and it has as many genes as rules. Each gene determines if the corresponding rule is used in the FRM or not, by setting it to 1 (selected) or to 0 (not selected).
2) *Chromosome Evaluation*: We use as fitness function the standard accuracy rate.
3) *Initial Gene Pool*: The population is composed by 50 individuals. Having one chromosome initialized by setting to 0 the value of all the genes to perform the lateral tuning, those used to learn the exponent of the fuzzy measure are set to 1.0 to obtain the classical cardinality fuzzy measure and the genes to perform the rule selection process are set to 1. The remainder chromosomes are randomly generated in the corresponding ranges of the genes.
4) *Crossover Operator*: We use the Parent Centric BLX (PCBLX) crossover operator [37] for the real coding part and the HUX [38] for the binary coded part. Two parents are crossed if their hamming distance divided by 2 is superior than the threshold Th, which is initialized as:

$$Th = \frac{(\#Genes \cdot BITSGENE)}{4.0} \tag{12}$$

We use the Gray code to convert each real coded gene to binary coding with a fixed number of bits for each gene (BITSGENE).
5) *Restarting Approach*: To increase the convergence of the algorithm, if new individuals are not included in the new population, we decrease the threshold by BITSGENE. When the threshold is smaller than 0 we pick the best chromosome (elitist scheme) and reset all the population with random values.
6) *Stopping Criteria*: The search process is stopped when:
   (i) The maximum number of trials is reached.
   (ii) A 100% is obtained as the fitness of the best individual.

V. Experimental Framework

In this section we present the experimental framework used in this paper. We start by describing the datasets along with the configuration of the classifiers considered in this paper. After that, the statistical methods that are used for performance comparison.

A. Datasets and classifiers' set-up

In this study, to assess the performance of our approach, we consider 33 numeric datasets selected from the KEEL[2] dataset repository [39]. The features of the datasets are summarized in Table IV, showing for each one its identification (ID), followed by the name of the dataset (Dataset), the number of samples (#Samp.), the number of features (#Feat.) and the number of classes (#Class).

Examples containing missing information were removed, e.g., in the *wisconsin* dataset. Also, the datasets *magic*, *page-blocks*, *penbased*, *ring*, *satimage* and *twonorm* were stratified sampled at 10% in order to reduce their size for training.

For each dataset, we have considered a 5-fold cross-validation technique, that is, the dataset is splitted into five random partitions, with 20% of the examples and maintaining the class distribution. Then, we use four partitions for training, and the remainder is used for testing. This process is repeated five times, considering a different partition for testing each time. We measure the quality of the classifier in each iteration using the accuracy rate, which is defined as the number of correctly classified examples divided by the total number of examples for each partition. At the end, we compute the average result of the five testing partitions, which is the result we show for each algorithm.

In order to show the quality of our method, we compare it versus three state-ofthe-art FRBCSs, namely, FURIA [11], IVTURS [9] and the original FARC-HD [10]. We show the configuration of these algorithms in Table V. In this table, we have to stress that our new proposal and IVTURS share the same fuzzy rule learning algorithm than that of FARC-HD and consequently, we use the same values for their parameters to perform a fair comparison.

B. Statistical tests for comparing performances

To give statistical support to the analysis of the results, we consider some hypothesis validation techniques [16], [40], that is, non-parametric tests, taking into account that the conditions that guarantee the reliance of the parametric tests cannot be warranted [17].

Specifically, we use the aligned Friedman rank test [41] to discover statistical differences among a group of results and to verify the quality of a method in comparison to otters approaches. Observe that the algorithm achieving the lowest average ranking is the best one.

Moreover, we also use the Holm post-hoc test [42] to find the method that reject the equivalence hypothesis with respect to the best approach found with the aligned Friedman rank test. We compute the adjusted p-value (APV) considering that multiple tests are performed. Then, it is possible to directly compare the APV with the level of significance α, and, thus, we are capable of reject the null hypothesis.

[2] http://www.keel.es

TABLE IV: Properties of the datasets considered in this study

Id.	Dataset	#Samp.	#Feat.	#Class
App	Appendicitis	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
Con	Contraceptive	1473	9	3
Eco	Ecoli	336	7	8
Gla	Glass	214	9	6
Hab	Haberman	306	3	2
Hay	Hayes-Roth	160	4	3
Ion	Ionosphere	351	33	2
Iri	Iris	150	4	3
Led	led7digit	500	7	10
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
Sah	Saheart	462	9	2
Sat	Satimage	6,435	36	7
Seg	Segment	2,310	19	7
Shu	Shuttle	58,000	9	7
Son	Sonar	208	60	2
Spe	Spectfheart	267	44	2
Tit	Titanic	2,201	3	2
Two	Twonorm	740	20	2
Veh	Vehicle	846	18	4
Win	Wine	178	13	3
Wis	Wisconsin	683	11	2
Yea	Yeast	1,484	8	10

TABLE V: Parameter setup of the considered algorithms

Algorithm	Configuration
FURIA	Number of optimizations: 2 Number of folds: 3
FARC-HD, IVTURS and $C_{F_1F_2}$-integrals	Linguistic labels per variable: 5 Conjunction operator: Product t-norm Rule weight: Confidence Minimum support: 0.05 Minimum confidence: 0.8 Depth of the search tree: 3 Number of fuzzy rules that cover each example: 2 Population size: 50 Gray codification: 30 bits per gene Number of evaluations: 20.000

VI. Experimental Results

This section is aimed at analyzing the performance of our new approach. To do so, we have separated the study in two parts. In the first one, we present the results obtained by the $C_{F_1F_2}$-integrals constructed using the pairs selected in section III-B (Table III). In the second one, in order to show the quality of our method, we perform comparisons against different state-of-the-art FRBCSs.

A. Analysis of the results of different $C_{F_1F_2}$-integrals

The results achieved in test by all the constructed $C_{F_1F_2}$-integrals are presented in Table VI. The rows represent the functions used as F_1, which are dominant in relation to the functions F_2, which are shown by columns and they have subordination characteristics. The result of each cell is the average testing result among the 33 datasets considered in the study.

We have to point out that we only show the averaged results due to space limitations. The complete results can be accessed in – https://github.com/Giancarlo-Lucca/TablesCF1F2. In Table VI we highlight in **boldface** the maximum accuracy per row and we underline the best accuracy for each column. We have to point out that blank spaces are related to combinations that could not be performed, since the dominance property is not satisfied for the specific pair of functions.

In a general looking, it is possible to observe that the largest accuracy is obtained by picking the function GM as F_1 and F_{BPC} as function F_2. This pair is a combination of a function having a high dominance as F_1 combined with a function with a high subordination as F_2. We can observe that for the functions to be F_1 the results are better when they are paired with a function F_2 with a high subordination degree (results highlighted in **boldface**). The opposite is also observed, since for each F_2 function, the best results are obtained when considering a F_1 whit high dominance (underlined results). In this manner, we can conclude that it is a good choice to select pairs of functions whose F_1 has a large dominance strength degree and its F_2 function is a highly dominated one.

Analyzing the results by categories (high, medium and low) according to the functions F_1, we have that:

- Using a function with high dominance characteristics as F_1 provides good results, since eight of the top ten best classifications, are pairs with this characteristic. Observe that, if we pick the functions GM and F_{GL} as F_1, the results tend to present a stability since the accuracies could be considered as similar. Regarding the sine function, S, its unsatisfactory behavior could occur since the differences between the pair of functions are too wide, which may imply a decrease on the performance of the classifier. Observe that in [30] this function also presents a similar behavior.
- The usage of functions having medium dominance characteristic as F_1 (T_{HP}, T_M and F_{IM}) tends to maintain competitive results. From the 21 possible combinations of these functions, only in four cases the obtained accuracy are less than 80%.
- Applying functions with low dominance as F_1, in general, does not fulfill the dominance property and, for this reason, less pairs can be used to construct $C_{F_1F_2}$-integrals. However, from the seven pairs constructed in this study, three of them provide poor results (less than 80%) and the remainder ones obtain satisfactory results.

B. Comparisons against other non-averaging aggregation functions and state-of-the-art fuzzy classifiers

As mentioned before, in general the obtained results tend to be stable an satisfactory. Thus, in order to demonstrate the quality of our approach, we compare the performance of the $C_{F_1F_2}$-integral that achieved the highest accuracy (GM–F_{BPC}) against the best non-averaging function of our previous paper [12] (F_{NA2}), a classical non-averanging aggregation operator like the probabilistic sum (P^*) and three state-of-the-art fuzzy classifiers, namely, FURIA [11], IVTURS [9] and FARC-HD [10].

TABLE VI: Accuracy mean achieved in testing by different $C_{F_1F_2}$-integrals

			F_2								
			Low			Medium			High		
			S	GM	F_α	T_M	F_{NA}	T_P	$T_{Ł}$	F_{BPC}	T_{DP}
		T_{DP}									**77.17**
	Low	F_{NA}					77.58				**80.42**
		O_B						79.86	80.44	80.30	<u>**80.57**</u>
		T_{HP}						80.22	80.25	**80.47**	80.46
F_1	Medium	T_M				79.54	79.02	80.42	**80.68**	80.52	80.41
		F_{IM}		80.05	80.20	79.92	80.04	80.26	**80.45**	79.88	80.08
		GM		79.76	<u>80.78</u>	80.28	<u>80.70</u>	80.52	<u>80.97</u>	<u>**81.02**</u>	80.56
	High	F_{GL}		<u>80.55</u>	80.56	<u>**80.92**</u>	80.68	<u>80.61</u>	80.27	80.80	80.50
		S	<u>77.19</u>	79.78	79.70	79.61	79.74	79.88	**80.17**	79.66	80.11

The results achieved in testing by the different methods are detailed in Table VII by columns. In each row of this table we present the accuracy obtained per each dataset. Furthermore, we highlight in **boldface** the best achieved result for each one and, in the two last rows, we present the number of datasets in which the classifier achieves the best (#Wins) and the worst result (#Loses).

From the obtained results, performing just a simple numerical comparison, it is possible to observe that FURIA is the method achieving the best global mean and the largest number of best classification results. The $C_{F_1F_2}$-integral achieves the second position in both criterias. However, we have to stress that whilst FURIA provides the worst result in seven datasets, our approach achieves the worst results in a single case. Therefore, we can observe that our method provides a good performance in a regular way. This affirmation can also be made to the CF-integral, F_{NA2}, but in this case it provides the best results in a less number of datasets. For the remainder methods, the results are worse than those of FURIA and our new approach, since the number of datasets having the best results are less and the number of loses cases are larger.

In order to highlight the behavior or our new method, if we look at the results in Table VI and we compare them against the ones of Table VII, we can find a large number of combinations leading to a global mean equal or larger than that of the compared methods (except that of FURIA). Specifically, the number of combinations having an equal or greater average result is 39, 36, 34 and 12 when compared against IVTURS, P^*, FARC-HD and F_{NA2}, respectively.

To support these findings we have conducted a set of statistical studies (as many as combinations in Table VII) using the aligned Friedman rank test to compare each $C_{F_1F_2}$-integral with the remainder methods considered in this section, whose obtained results are available in Table VIII. Specifically, in this table we only show the results of those $C_{F_1F_2}$-integrals (in columns) that obtain the best rank and consequently, they are used as control method in the post-hoc Holm's test, whose obtained APV is shown in brackets. If there are statistical differences between two methods, considering 0.05 (5%) as the level of confidence, we underline the APV.

From the obtained results, it is noticeable that five $C_{F_1F_2}$-integrals are considered as control method, presenting statistical differences against IVTURS, FARC-HD and P^* in almost all cases. Regarding FURIA and F_{NA2}, we can infer that there are no statistical differences between both methods, since the obtained APVs are high. Therefore, we can conclude that our new method obtains competitive results as it statistically

TABLE VII: Results achieved in testing by different FRMs

Dataset	FURIA	IVTURS	FARC-HD	P*	F_{NA2}	(GM, F_{BPC})
App	**87.71**	84.94	83.03	85.84	85.84	86.80
Bal	83.68	85.76	85.92	87.20	88.64	**89.12**
Ban	**88.57**	81.70	85.30	84.85	84.60	84.79
Bnd	69.40	67.70	68.28	68.82	70.48	**71.30**
Bup	**70.14**	67.54	67.25	61.74	64.64	66.96
Cle	56.57	**59.60**	56.21	59.25	56.55	56.22
Con	54.17	53.36	53.16	52.21	53.16	**54.72**
Eco	80.06	78.58	**82.15**	80.95	80.08	81.86
Gla	**72.91**	67.31	65.44	64.04	66.83	68.25
Hab	72.55	72.85	**73.18**	69.26	71.87	72.53
Hay	**81.00**	80.23	77.95	77.95	79.43	78.66
Ion	89.75	**92.89**	88.90	88.32	89.75	88.33
Iri	94.00	**96.00**	94.00	95.33	94.00	94.00
Led	**71.80**	70.40	69.60	69.20	69.80	70.00
Mag	80.65	79.76	80.76	80.39	79.70	**80.86**
New	94.88	95.35	94.88	94.42	96.28	**96.74**
Pag	**95.25**	95.07	95.07	94.52	94.15	**95.25**
Pen	92.45	92.18	92.55	**93.27**	92.91	92.91
Pho	**85.90**	80.00	81.70	82.51	81.44	81.42
Pim	**76.17**	74.73	74.74	75.91	75.52	75.38
Rin	85.54	87.57	90.95	90.00	89.86	**91.89**
Sah	70.33	70.99	68.39	69.69	70.12	**71.43**
Sat	**82.27**	75.90	79.47	80.40	80.41	79.47
Seg	**97.32**	90.30	93.12	92.94	92.42	93.29
Shu	**99.68**	91.82	95.59	94.85	97.15	96.83
Son	78.90	80.33	78.36	82.24	83.21	**85.15**
Spe	77.88	**80.52**	77.88	77.90	79.77	79.39
Tit	78.51	**78.87**	**78.87**	**78.87**	**78.87**	**78.87**
Two	88.11	92.30	90.95	90.00	**92.57**	92.30
Veh	**70.21**	67.38	68.56	68.09	68.08	68.20
Win	93.78	**97.19**	96.03	94.92	96.08	95.48
Wis	96.63	96.49	96.63	**97.22**	96.78	96.78
Yea	58.22	55.86	58.96	**59.03**	57.08	58.56
Mean	**81.06**	80.04	80.12	80.07	80.55	81.02
#Wins	13	6	3	4	2	10
#Loses	7	12	7	8	3	1

TABLE VIII: Statistical results using the align Friedman rank test and Holm's post hoc test

	GM–F_{BPC}	GM–$T_Ł$	F_{GL}–T_M	F_{GL}–F_{BPC}	GM–F_α
$C_{F_1F_2}$	74.87 (–)	75.45 (–)	76.28 (–)	74.87 (–)	83.86 (–)
FURIA	83.81 (0.52)	85.10 (0.49)	84.12 (0.57)	83.81 (0.52)	80.27 (0.79)
F_{NA2}	97.09 (0.23)	96.81 (0.25)	97.19 (0.27)	97.09 (0.23)	97.04 (0.46)
IVTURS	112.63 (0.02)	110.07 (0.04)	109.93 (0.05)	112.63 (0.02)	108.68 (0.13)
P*	112.56 (0.02)	112.81 (0.03)	113.22 (0.03)	112.56 (0.02)	111.01 (0.11)
FARC	116.01 (0.01)	116.72 (0.01)	116.22 (0.02)	116.01 (0.01)	116.12 (0.05)

improves other state-of-the-art fuzzy classifiers and it provides similar results than those of FURIA.

VII. Conclusion

In this paper we have defined the concept of $C_{F_1F_2}$-integrals, which are a generalization of the CC-integral introduced in [7]. Specifically, these integrals use two different fusion functions, F_1 and F_2, in order to try to enhance the behavior of FRBCSs. The constructed $C_{F_1F_2}$-integrals are non-averaging as most of the aggregation operator used by state-of-the-art fuzzy classifiers are. Furthermore, these integrals are OD increasing functions satisfying proper boundary conditions. We have presented a method to select the best combination of functions to be F_1 and F_2, which is based on the concept of dominance and subordination.

From the obtained results, we can conclude that the results of this approach could be considered as satisfactory and stable, since the results are quite similar in many cases. Furthermore, we showed that 5 different $C_{F_1F_2}$-integrals provide competitive results when compared against FURIA and the best C_F-integral published in our previous work on the topic. Moreover, we have to highlight that we achieve a statistical superior performance versus two state-of-the-art fuzzy classifiers like IVTURS and FARC-HD. All these facts support that this approach is an efficient option and it expands the scope of the generalizations of the Choquet integral.

Acknowledgment

Supported by CNPq (Proc. 233950/2014-1, 306970/2013-9, 307781/2016-0), the Spanish Ministry of Science and Technology (project TIN2016-77356-P (AEI/FEDER, UE)), and by Caixa and Fundación Caja Navarra of Spain.

References

[1] R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification (2Nd Edition), Wiley-Interscience, 2000.

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

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

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

[5] G. Choquet, Theory of capacities, Annales de l'Institut Fourier 5 (1953–1954) 131–295.

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

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

[8] G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, J. Fernández, H. Bustince, Analyzing the behavior of a CC-integral in a fuzzy rule-based classification system, in: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2017, pp. 1–6.

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

[10] J. Alcalá-Fdez, R. Alcalá, 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 19 (5) (2011) 857–872.

[11] J. Hühn, E. Hüllermeier, Furia: an algorithm for unordered fuzzy rule induction, Data Mining and Knowledge Discovery 19 (3).

[12] G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, H. Bustince, R. Mesiar, CF-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems, Information Sciences 435 (2018) 94 – 110.

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

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

[15] J. Derrac, S. García, D. Molina, F. Herrera, A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms, Swarm and Evolutionary Computation 1 (1) (2011) 3 – 18.

[16] S. García, A. Fernández, J. Luengo, F. Herrera, A study of statistical techniques and performance measures for genetics–based machine learning: Accuracy and interpretability, Soft Computing 13 (10) (2009) 959–977.

[17] J. Demšar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research 7 (2006) 1–30.

[18] G. Beliakov, A. Pradera, T. Calvo, Aggregation Functions: A Guide for Practitioners, Springer, Berlin, 2007.

[19] G. Mayor, E. Trillas, On the representation of some aggregation functions, in: Proceedings of IEEE International Symposium on Multiple-Valued Logic, IEEE, Los Alamitos, 1986, pp. 111–114.
[20] H. Bustince, J. Fernandez, A. Kolesárová, R. Mesiar, Directional monotonicity of fusion functions, European Journal of Operational Research 244 (1) (2015) 300–308.
[21] T. Murofushi, M. Sugeno, M. Machida, Non-monotonic fuzzy measures and the Choquet integral, Fuzzy Sets and Systems 64 (1) (1994) 73 – 86.
[22] G. Beliakov, H. Bustince, T. Calvo, A Practical Guide to Averaging Functions, Springer, Berlin, New York, 2016.
[23] E. P. Klement, R. Mesiar, E. Pap, Triangular Norms, Kluwer Academic Publisher, Dordrecht, 2000.
[24] H. Bustince, J. Fernandez, R. Mesiar, J. Montero, R. Orduna, Overlap functions, Nonlinear Analysis: Theory, Methods & Applications 72 (3-4) (2010) 1488–1499.
[25] R. B. Nelsen, An introduction to copulas, Vol. 139 of Lecture Notes in Statistics, Springer, New York, 1999.
[26] G. P. Dimuro, B. Bedregal, Archimedean overlap functions: The ordinal sum and the cancellation, idempotency and limiting properties, Fuzzy Sets and Systems 252 (2014) 39 – 54.
[27] G. P. Dimuro, B. Bedregal, Additive generators of overlap functions, in: H. Bustince, J. Fernandez, R. Mesiar, T. Calvo (Eds.), Aggregation Functions in Theory and in Practice, Vol. 228 of Advances in Intelligent Systems and Computing, Springer, Berlin, 2013, pp. 167–178.
[28] G. P. Dimuro, B. Bedregal, H. Bustince, M. J. Asiáin, R. Mesiar, On additive generators of overlap functions, Fuzzy Sets and Systems 287 (2016) 76 – 96, theme: Aggregation Operations.
[29] G. P. Dimuro, B. Bedregal, On residual implications derived from overlap functions, Information Sciences 312 (2015) 78 – 88.
[30] M. Elkano, M. Galar, J. Sanz, A. Fernández, E. Barrenechea, F. Herrera, H. Bustince, 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) (2015) 1562–1580.
[31] H. Bustince, E. Barrenechea, M. Sesma-Sara, J. Lafuente, G. P. Dimuro, R. Mesiar, A. Kolesárová, Ordered directionally monotone functions. justification and application, IEEE Transactions on Fuzzy Systems (2017) 1(In press, corrected proof). doi:10.1109/TFUZZ.2017.2769486.
[32] H. Ishibuchi, T. Nakashima, Effect of rule weights in fuzzy rule-based classification systems, Fuzzy Systems, IEEE Transactions on 9 (4) (2001) 506–515.
[33] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: Proceedings of the 20th International Conference on Very Large Data Bases, VLDB '94, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1994, pp. 487–499.
[34] B. Kavšek, N. Lavrač, V. Jovanoski, Apriori-sd: Adapting association rule learning to subgroup discovery, in: M. R. Berthold, H.-J. Lenz, E. Bradley, R. Kruse, C. Borgelt (Eds.), Advances in Intelligent Data Analysis V: 5th International Symposium on Intelligent Data Analysis, IDA 2003, Berlin, Germany, August 28-30, 2003. Proceedings, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 230–241.
[35] L. J. Eshelman, The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination, in: G. J. E. Rawlings (Ed.), Foundations of Genetic Algorithms, Morgan Kaufmann, San Francisco, 1991, pp. 265–283.
[36] R. Alcalá, J. Alcalá-Fdez, F. Herrera, A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection, IEEE Transactions on Fuzzy Systems 15 (4) (2007) 616–635.
[37] F. Herrera, M. Lozano, A. M. Sánchez, A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study, International Journal of Intelligent Systems 18 (3) (2003) 309–338.
[38] L. J. Eshelman, J. D. Schaffer, Real-coded genetic algorithms and interval-schemata, in: L. D. WHITLEY (Ed.), Foundations of Genetic Algorithms, Vol. 2 of Foundations of Genetic Algorithms, Elsevier, 1993, pp. 187 – 202.
[39] J. Alcalá-Fdez, L. Sánchez, S. García, M. Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fernández, F. Herrera, Keel: a software tool to assess evolutionary algorithms for data mining problems, Soft Computing 13 (3) (2009) 307–318.
[40] D. Sheskin, Handbook of parametric and nonparametric statistical procedures, 2nd Edition, Chapman & Hall/CRC, 2006.
[41] J. L. Hodges, E. L. Lehmann, Ranks methods for combination of independent experiments in analysis of variance, Annals of Mathematical Statistics 33 (1962) 482–497.
[42] S. Holm, A simple sequentially rejective multiple test procedure, Scandinavian Journal of Statistics 6 (1979) 65–70.

6 Generalized $C_{F_1F_2}$-integrals: from Choquet-like aggregation to ordered directionally monotone functions

- G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, H. Bustince and R. Mesiar, "Generalized $C_{F_1F_2}$-integrals: from Choquet-like aggregation to ordered directionally monotone functions", Fuzzy Sets and Systems (submitted).
  - Journal: Fuzzy Sets and Systems
  - Status: Submitted
  - Impact Factor (JCR 2016): 2.718
  - Knowledge Area:
    * Statistics & Probability: Ranking 8/124 (Q1)
    * Computer Science: Ranking 18/104 (Q1)
    * Theory & Methods: Ranking 18/104 (Q1)
    * Mathematics, Applied: Ranking 10/255 (Q1)

Generalized $C_{F_1F_2}$-integrals: from Choquet-like aggregation to ordered directionally monotone functions

Graçaliz Pereira Dimuro[a,b,*], Giancarlo Lucca[c], Benjamín Bedregal[d], Radko Mesiar[e], José Antonio Sanz[c], Humberto Bustince[c,b]

[a]*Centro de Ciência Computacionais, Universidade Federal do Rio Grande*
Av. Itália km 08, Campus Carreiros, 96201-900 Rio Grande, Brazil
[b]*Institute of Smart Cities, Universidad Pública de Navarra*
Campus Arrosadia s/n, 31006, Pamplona, Spain
[c]*Departamento of Automática y Computación, Universidad Pública de Navarra*
Campus Arrosadia s/n, 31006, Pamplona, Spain
[d]*Departamento de Informática e Matemática Aplicada, Universidade Federal do Rio Grande do Norte*
Campus Universitário s/n, 59072-970 Natal, Brazil
[e]*Slovak University of Technology, Bratislava, Slovakia*
Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Prague, Czech Republic

Abstract

This paper introduces the theoretical framework for a generalization of $C_{F_1F_2}$-integrals, a family of Choquet-like integrals used successfully in the aggregation process of the fuzzy reasoning mechanisms of fuzzy rule based classification systems. The proposed generalization, called by $gC_{F_1F_2}$-integrals, is based on the so-called pseudo pre-aggregation function pairs (F_1, F_2), which are pairs of fusion functions satisfying a minimal set of requirements in order to be either an aggregation function or just an ordered directionally increasing function satisfying the appropriate boundary conditions. We propose a dimension reduction of the input space, in order to deal with duplicated elements in the input, avoiding ambiguities in the definition of $gC_{F_1F_2}$-integrals. We study several properties of $gC_{F_1F_2}$-integrals, considering different constraints for the functions F_1 and F_2, and state under which conditions $gC_{F_1F_2}$-integrals present or not averaging behaviors. Several examples of $gC_{F_1F_2}$-integrals are presented, considering different pseudo pre-aggregation function pairs, defined on, e.g., t-norms, overlap functions, copulas that are neither t-norms nor overlap functions and other functions that are not even pre-aggregation functions.

Keywords: Aggregation functions, pre-aggregation functions, ordered directionally monotonicity, pseudo pre-aggregation function pair, Choquet Integral

1. Introduction

In 2016, Lucca et al. [1] introduced the notion of pre-aggregation function (PAF), which fulfills the boundary conditions as any aggregation function, but, instead of being an increasing function, it is just directional increasing [2]. That is, it increases along some specific ray (direction). Furthermore, the authors presented some methods to produce PAFs [3, 4]. One of them is by generalizing the Choquet integral [5] replacing the product operator by a t-norm, obtaining, under some constraints, idempotent and averaging PAFs. This approach was used in Fuzzy Rule-Based Classification System (FRBCS) [6], presenting excellent results, when the Hamacher t-norm [7] is used for the generalization, overcoming the Choquet integral and classical averaging operators in classification systems.

*Corresponding author. Phone: +55 53 32336685
Email addresses: gracaliz@gmail.com, gracaliz.pereira@unavarra.es, gracalizdimuro@furg.br (Graçaliz Pereira Dimuro), lucca.112793@e.unavarra.es (Giancarlo Lucca), bedregal@dimap.ufrn.br (Benjamín Bedregal), mesiar@math.sk (Radko Mesiar), joseantonio.sanz@unavarra.es (José Antonio Sanz), bustince@unavarra.es (Humberto Bustince)

Those excellent results motivated us to explore a more general method for constructing PAFs based on the Choquet integral. For that, instead of using just a t-norm, we replace the product operator by a fusion function F that is left 0-absorbent (i.e., $F(0, x) = 0$, for all $x \in [0, 1]$), obtaining the C_F-integrals [8]. C_F-integrals are pre-aggregation functions, which, under certain conditions, may be idempotent and/or averaging functions. This allowed to analyse sub-families of C_F-integrals having or not the averaging behavior, showing that a C_F-integral does not need to be an averaging function when used in FRBCSs, since the non-averaging obtained more accurate results than the averaging ones.

In the same line of this research, Lucca et al. [9] investigated another kind of Choquet integral that leads to aggregation functions, instead of just PAFs. For that, the product operation of the standard Choquet integral was first distributed and, then, replaced by a copula [10], obtaining the CC-integrals, which happen to be averaging aggregation functions [11]. This approach presented excellent results in classification, in particular, when the minimun t-norm was the considered copula, in which case it was called CMin-integral [12]. See also the application in [13].

Recently, Luca et al. [14] developed the concept of $C_{F_1F_2}$-integrals, which is a specific generalization of CC-integrals, based on two possibly different fusion functions F_1 and F_2 (instead of a copula C) satisfying some appropriate conditions, obtaining non-averaging Choquet-like integrals that were successfully used in the aggregation process of the fuzzy reasoning mechanisms of fuzzy rule based classification systems. Their performance was proved to be statistically equivalent to FURIA [15].

The aim of this paper is to generalize the concept of $C_{F_1F_2}$-integrals, obtaining the so-called $gC_{F_1F_2}$-integrals, presenting a solid theoretical framework which gives the basis for applications. $gC_{F_1F_2}$-integrals are obtained by distributing the product operation of the Choquet integral and, then, using a pair of fusion functions (F_1, F_2), satisfying some special conditions, we generalize the two instances of the product operation.

Then, the objectives of this paper are stated as:

1. To introduce the notion of pseudo pre-aggregation function pair (F_1, F_2), that is, a pair of fusion functions satisfying some kind of boundary conditions, directional increasingness and F_1-dominance property;
2. To introduce the notion of Choquet-like integral based on pseudo pre-aggregation function pair, called $gC_{F_1F_2}$-integrals;
3. To show under which conditions $gC_{F_1F_2}$-integrals based on pseudo pre-aggregation function pairs (F_1, F_2) are aggregation functions;
4. To show under which conditions $gC_{F_1F_2}$-integrals based on pseudo pre-aggregation function pairs (F_1, F_2) are ordered directional (OD) increasing functions [16] and satisfy the desirable boundary conditions;
5. To study when $gC_{F_1F_2}$-integrals are averaging;
6. To analyze several types of pseudo pre-aggregation function pairs (F_1, F_2), built from t-norms [7], overlap functions [17, 18, 19], copulas, and other functions that are not even PAFs, showing examples of different $gC_{F_1F_2}$-integrals.

The paper is organized as follows. In Section 2, we present the basic concepts required to understand the paper. In Section 3, we introduce the concept of pseudo pre-aggregation pairs and analyse several properties. The concept of $gC_{F_1F_2}$-integrals is introduced in Section 4. In Section 5, we discuss when $gC_{F_1F_2}$-integrals are aggregation functions, and the related properties. Section 6 studies when $gC_{F_1F_2}$-integrals are not aggregation functions, but OD monotone functions. Section 7 is the Conclusion.

2. Preliminaries

In this paper, we call any n-ary function $F : [0, 1]^n \to [0, 1]$ by a fusion function.

Definition 2.1. *[20, 21] A function $A : [0, 1]^n \to [0, 1]$ is an aggregation function whenever the following conditions hold:*

(A1) *A is increasing[1] in each argument: for each $i \in \{1, \ldots, n\}$, if $x_i \leq y$, then*

$$A(x_1, \ldots, x_n) \leq A(x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n);$$

[1]For an increasing (decreasing) function we do not mean a strictly increasing (decreasing) function.

(A2) *A satisfies the boundary conditions:* ***(i)*** $A(0,\ldots,0)=0$ *and* ***(ii)*** $A(1,\ldots,1)=1$.

An aggregation function $A:[0,1]^n\to[0,1]$ is said to be idempotent if and only if:

(ID) $\forall x\in[0,1]: A(x,\ldots,x)=x$, and

it is said to be averaging if and only if:

(AV) $\forall(x_1,\ldots,x_n)\in[0,1]^n: \min\{x_1,\ldots,x_n\}\leq A(x_1,\ldots,x_n)\leq\max\{x_1,\ldots,x_n\}$.

Observe that, since aggregation functions are increasing, the idempotent and averaging behaviors are equivalent in the context of aggregation functions.

Definition 2.2. *[10] 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) $C(x,1)=C(1,x)=x$.

Definition 2.3. *[2] Let $\vec{r}=(r_1,\ldots,r_n)$ be a real n-dimensional vector, $\vec{r}\neq\vec{0}=(0,\ldots,0)$. A function $F:[0,1]^n\to[0,1]$ is said to be $\vec{r}$-increasing if for all $\vec{x}=(x_1,\ldots,x_n)\in[0,1]^n$ and for all $c>0$ such that $\vec{x}+c\vec{r}=(x_1+cr_1,\ldots,x_n+cr_n)\in[0,1]^n$ it holds*

$$F(\vec{x}+c\vec{r})\geq F(\vec{x}).$$

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

Definition 2.4. *[22, 4] A function $PA:[0,1]^n\to[0,1]$ is said to be an n-ary pre-aggregation function (PAF) if the following conditions hold:*

(PA1) *Directional Increasingness: there exists $\vec{r}=(r_1,\ldots,r_n)\in[0,1]^n$, $\vec{r}\neq\vec{0}$, such that PA is $\vec{r}$-increasing;*

(PA2) *Boundary conditions: (i) $PA(0,\ldots,0)=0$ and (ii) $PA(1,\ldots,1)=1$.*

If F is a pre-aggregation function with respect to a vector $\vec{r}$ we just say that F is an $\vec{r}$-pre-aggregation function.

Another important concept used in this paper is the one of ordered directional (OD) monotonicity, introduced in [16]. Observe that, when one considers directional monotonicity, the direction along which monotonicity is required is the same for all $\vec{x}\in[0,1]^n$. On the contrary, OD monotone functions are functions that allow monotonicity along different directions depending on the ordinal size of the coordinates of each input $\vec{x}\in[0,1]^n$. First, we take a permutation $\sigma:\{1,\ldots,n\}\to\{1,\ldots,n\}$ to reorder the input $\vec{x}\in[0,1]^n$ in a decreasing order, obtaining $\vec{x}_\sigma\in[0,1]^n$. Then, a fusion function $F:[0,1]^n\to[0,1]$ is OD $\vec{r}$-increasing, for a real vector $\vec{r}=(r_1,\ldots,r_n)$, with $\vec{r}\neq\vec{0}$, whenever $F(\vec{x})$ is less than or equal to the values of F when applied to

$$(\vec{x}_\sigma+c\vec{r})_{\sigma^{-1}}=\vec{x}+c\vec{r}_{\sigma^{-1}}, \tag{1}$$

under the assumption that $\vec{x}_\sigma$ and $\vec{x}_\sigma+c\vec{r}$ are comonotone (i.e., either they increase or decrease at the same time).

Definition 2.5. *[16] Consider a function $F:[0,1]^n\to[0;1]$ and let $\vec{r}=(r_1,\ldots,r_n)$ be a real n-dimensional vector, $\vec{r}\neq\vec{0}$. F is said to be ordered directionally (OD) $\vec{r}$-increasing if, for each $\vec{x}\in[0,1]^n$, any permutation $\sigma:\{1,\ldots,n\}\to\{1,\ldots,n\}$ with $x_{\sigma(1)}\geq\ldots\geq x_{\sigma(n)}$, and $c>0$ such that $1\geq x_{\sigma(1)}+cr_1\geq\ldots\geq x_{\sigma(n)}+cr_n$, it holds that*

$$F(\vec{x}+c\vec{r}_{\sigma^{-1}})\geq F(\vec{x}),$$

where $\vec{r}_{\sigma^{-1}}=(r_{\sigma^{-1}(1)},\ldots,r_{\sigma^{-1}(n)})$. Similarly, one defines an ordered directionally (OD) $\vec{r}$-decreasing function.

In what follows, denote $N = \{1, \dots, n\}$, for $n > 0$.

Definition 2.6. *[5, 23] A function $\mathfrak{m} : 2^N \to [0,1]$ is said to be a fuzzy measure if, for all $X, Y \subseteq N$, the following conditions hold:*

($\mathfrak{m}$1) *Increasingness: if $X \subseteq Y$, then $\mathfrak{m}(X) \leq \mathfrak{m}(Y)$;*

($\mathfrak{m}$2) *Boundary conditions: $\mathfrak{m}(\emptyset) = 0$ and $\mathfrak{m}(N) = 1$.*

Definition 2.7. *[5] The discrete Choquet integral with respect to a fuzzy measure $\mathfrak{m}$ is the function $\mathfrak{C}_\mathfrak{m} : [0,1]^n \to [0,1]$, defined, for all of $\vec{x} = (x_1, \dots, x_n) \in [0,1]^n$, by:*

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

where $\left(x_{(1)}, \dots, x_{(n)}\right)$ is an increasing permutation on the input $\vec{x}$, that is, $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$, where $x_{(0)} = 0$ and $A_{(i)} = \{(i), \dots, (n)\}$ is the subset of indices corresponding to the $n - i + 1$ largest components of $\vec{x}$.

Whenever one distribute the product operation in Equation (2), we obtain the Choquet Integral in its expanded form:

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

Substituting the product operation in Equation (2) by a copula C, Lucca et al. [9] introduced the CC-integral, which are averaging aggregation functions:

Definition 2.8. *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 Choquet-like copula-based integral with respect to $\mathfrak{m}$ is defined as a function $\mathfrak{C}^C_\mathfrak{m} : [0,1]^n \to [0,1]$, given, for all $x \in [0,1]^n$, by*

$$\mathfrak{C}^C_\mathfrak{m}(\vec{x}) = \sum_{i=1}^{n} C\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - C\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right), \tag{4}$$

where $(x_{(1)}, \dots, x_{(n)})$ 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 $A_{(i)} = \{(i), \dots, (n)\}$ is the subset of indices of $n - i + 1$ largest components of $\vec{x}$.

Another integral that is related to fuzzy measure is the Sugeno Integral:

Definition 2.9. *The discrete Sugeno integral with respect to a fuzzy measure $\mathfrak{m}$ is the function $S_\mathfrak{m} : [0,1]^n \to [0,1]$, defined, for all of $\vec{x} = (x_1, \dots, x_n) \in [0,1]^n$, by:*

$$S_\mathfrak{m}(\vec{x}) = \max_{i=1}^{n} \left\{\min\left\{x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right\}\right\}, \tag{5}$$

where $\left(x_{(1)}, \dots, x_{(n)}\right)$ is an increasing permutation on the input $\vec{x}$, that is, $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$, where $x_{(0)} = 0$ and $A_{(i)} = \{(i), \dots, (n)\}$ is the subset of indices corresponding to the $n - i + 1$ largest components of $\vec{x}$.

3. Pseudo pre-aggregation function pairs (F_1, F_2)

In this section, we introduce the concept of pseudo pre-aggregation function pair and study some properties. In the following, consider $N = \{1, \dots, n\}$.

Definition 3.1. *Consider two bivariate functions $F_1, F_2 : [0,1]^2 \to [0,1]$. The pair (F_1, F_2) is said to be a pseudo pre-aggregation function pair whenever the following conditions hold, for all $y \in [0,1]$:*

(DI) *Directional Increasingness:* F_1 *is* $(1, 0)$*-increasing;*

(BC0) *Boundary Conditions for 0:*

- **(i)** $F_1(0, y) = F_2(0, y)$ *and*
- **(ii)** $F_1(0, 1) = 0$*;*

(BC1) *Boundary Condition for 1:* $F_1(1, 1) = 1$*;*

(DM) F_1*-Dominance (or, equivalently,* F_2*-Subordination):* $F_1 \geq F_2$*.*

Remark 3.1. *Observe that, for any pseudo pre-aggregation function pair* (F_1, F_2)*, by* **(i)** *and* **(ii)***, it holds that* $F_2(0, 1) = 0$*.*

Remark 3.2. *Whenever* (F_1, F_2) *is a pseudo pre-aggregation function pair then, for any* $F_3 : [0, 1]^2 \to [0, 1]$ *such that* $F_2 \leq F_3 \leq F_1$*, we have that* (F_1, F_3) *is also a pseudo pre-aggregation function pair. In particular,* (F_1, F_1) *is a pseudo pre-aggregation function pair.*

Definition 3.2. *A pseudo pre-aggregation function pair* (F_1, F_2) *is pairwise increasing if, for all* $x, y_1, y_2 \in [0, 1]$ *and* $h > 0$ *such that* $x + h \in [0, 1]$*, the following condition holds:*

(PI) *If* $y_2 \leq y_1$ *then* $F_1(x, y_1) - F_2(x, y_2) \leq F_1(x + h, y_1) - F_2(x + h, y_2)$*.*

Proposition 3.1. *Let* (F_1, F_2) *be a pseudo pre-aggregation function pair. If* F_2 *is* $(1, 0)$*-decreasing, then the pair* (F_1, F_2) *satisfies* ***(PI)****.*

PROOF. Since F_1 is $(1, 0)$-increasing, then, for any $h > 0$ and $x, y_1, y_2 \in [0, 1]$ such that $x + h \in [0, 1]$, it holds that $F_1(x + h, y_1) \geq F_1(x, y_1)$. On the other hand, since F_2 is $(1, 0)$-decreasing, then, for any $h > 0$ and $x, y_1, y_2 \in [0, 1]$ such that $x + h \in [0, 1]$, it holds that $-F_2(x + h, y_2) \geq -F_2(x, y_2)$. Thus, one has that $F_1(x, y_1) - F_2(x, y_2) \leq F_1(x + h, y_1) - F_2(x + h, y_2)$. □

Proposition 3.2. *For a copula* C*,* (C, C) *is a pseudo pre-aggregation function pair satisfying* ***(PI)****.*

PROOF. It is immediate that any copula C satisfies **(DI)**, **(BC0)** and **(BC1)**. Thus, (C, C) is a pseudo pre-aggregation function pair. From **(C1)**, it is immediate that (C, C) satisfies **(PI)**. □

Remark 3.3. *Observe that* ***(PI)*** *is a generalization of the 2-increasing property* ***(C1)****. In fact, for any fusion function* F*,* (F, F) *satisfies* ***(PI)*** *if and only if* F *satisfies* ***(C1)****.*

We present in Table 1 examples of functions $F : [0, 1]^2 \to [0, 1]$ satisfying **(DI)**, **(BC0)(ii)** and **(BC1)**. Those functions are, then, candidates to be combined in order to build pseudo pre-aggregation function pairs. In Table 2, we present an analysis of the Dominance property **(DM)**, taking into account the functions presented in Table 1, all of them obviously satisfying **(BC0)(i)**. In this table, considering that functions F_1 and F_2 are represented, respectively, in the lines and columns of the table, the pairs marked with “yes” satisfy **(DM)** or the F_1-dominance (equivalently, the F_2-subordination). Thus, those pairs are pseudo pre-aggregation function pairs. The pairs marked with “no” are not pseudo pre-aggregation function pairs since they do not satisfy **(DM)**.

4. Constructing Choquet-like integrals based on pseudo pre-aggregation function pairs (F_1, F_2)

In this section, we introduce a method for constructing Choquet-like integrals defined by combining the discrete Choquet integral in its expanded form (Equation (3)) with pseudo pre-aggregation function pairs (F_1, F_2), just replacing the product operation in Equation (3) by (F_1, F_2). Such Choquet-like integrals, which generalize the concept of $C_{F_1F_2}$-integrals introduced in [14], are called $gC_{F_1F_2}$-integrals.

Consider $N = \{1, \ldots, n\}$, where n is the dimension of the input vectors $\vec{x}$, that is, $\vec{x} = (x_1, \ldots, x_n) \in [0, 1]^n$. First, in order to handle repetitive elements in any input $\vec{x}$, which would lead to an ambiguous definition, we proceed to a dimension reduction, constructing a new space of dimension $k \leq n$, considering $K = \{1, \ldots, k\}$. For that, we introduced the following auxiliary definition:

Table 1: $F_i : [0,1]^2 \to [0,1]$, $i = 1, 2$, satisfying **(DI)**, **(BC0)(ii)**, **(BC1)**, for building pseudo pre-aggregation function pairs

(I) T-norms [7]	
Definition	Name/Reference
$T_M(x,y) = \min\{x,y\}$	Minimum
$T_P(x,y) = xy$	Algebraic Product
$T_{Ł}(x,y) = \max\{0, x+y-1\}$	Łukasiewicz
$T_{HP}(x,y) = \begin{cases} 0 & \text{if } x=y=0 \\ \frac{xy}{x+y-xy} & \text{otherwise} \end{cases}$	Hamacher Product
$T_{DP}(x,y) = \begin{cases} x & \text{if } y=1 \\ y & \text{if } x=1 \\ 0 & \text{otherwise} \end{cases}$	Drastic Product
(II) Overlap functions [17, 18, 19, 24]	
Definition	Name/Reference
$O_B(x,y) = \min\{x\sqrt{y}, y\sqrt{x}\}$	[17, Theorem 8], Cuadras-Augé family of copulas [25]
$O_{mM}(x,y) = \min\{x,y\}\max\{x^2,y^2\}$	[26, Ex. 3.1.(i)], [27, Ex. 4], [28, Ex. 3.1]
$O_\alpha(x,y) = xy(1+\alpha(1-x)(1-y))$, $\alpha \in [-1,0[\cup]0,1]$	[10, Apendix A (A.2.1)], [13], Farlie-Gumbel-Morgenstern copula family
$O_{Div}(x,y) = \frac{xy+\min\{x,y\}}{2}$	[10, Ap. A (A.8.7)], [9, Table 1]
$GM(x,y) = \sqrt{xy}$	Geometric Mean [29, Ex. 1]
$HM(x,y) = \begin{cases} 0 & \text{if } x=0 \text{ or } y=0 \\ \frac{2}{\frac{1}{x}+\frac{1}{y}} & \text{otherwise} \end{cases}$	Harmonic Mean [29, Ex. 1]
$S(x,y) = \sin\left(\frac{\pi}{2}(xy)^{\frac{1}{4}}\right)$	Sine [29, Ex. 1]
(III) Copulas that are neither t-norms nor overlap functions [10]	
Definition	Name/Reference
$C_F(x,y) = xy + x^2y(1-x)(1-y)$	[7, Ex. 9.5 (v)],[9, Table 1]
$C_L(x,y) = \max\{\min\{x, \frac{y}{2}\}, x+y-1\}$	[10, Ap. A (A.5.3a)], [9, Table 1]
(IV) Aggregation functions other than (I)-(III)	
Definition	Name/Reference
$AVG(x,y) = \frac{x+y}{2}$	Arithmetic Mean
$F_{RS}(x,y) = \min\left\{\frac{(x+1)\sqrt{y}}{2}, y\sqrt{x}\right\}$	
$F_{GL}(x,y) = \sqrt{\frac{x(y+1)}{2}}$	
$F_{BPC}(x,y) = xy^2$	[20, Ex. 1.80]
$F_\alpha(x,y) = \begin{cases} \alpha x & \text{if } x<y \\ \max\{\alpha x, y\} & \text{otherwise} \end{cases}$, $0<\alpha<1$	
(V) $(1,0)$-Pre-Aggregation functions	
Definition	Name/Reference
$F_{NA}(x,y) = \begin{cases} x & \text{if } x \le y \\ \min\{\frac{x}{2}, y\} & \text{otherwise} \end{cases}$	
$F_{NA2}(x,y) = \begin{cases} 0 & \text{if } x=0 \\ \frac{x+y}{2} & \text{if } 0<x\le y \\ \min\{\frac{x}{2}, y\} & \text{otherwise} \end{cases}$	
(VI) Non Pre-Aggregation functions	
Definition	Name/Reference
$F_{IM}(x,y) = \max\{1-y, x\}$	
$F_{IP}(x,y) = 1-y+xy$	

Definition 4.1. *The dimension reduction of the inputs* $\vec{x} \in [0,1]^n$ *is performed by family of* $(n \mapsto k)$*-dimension reduction functions* $R_{n\mapsto k|} : [0,1]^n \to [0,1]^k$*, defined, for each* $\vec{x} = (x_1,\dots,x_n) \in [0,1]^n$*, by:*

$$R_{n\mapsto k}(x_1,\dots,x_n) = (y_1,\dots,y_k), \tag{6}$$

such that:

(R1) $k = \mid \{x_1,\dots,x_n\} \mid$ *is the cardinality of the set* $\{x_1,\dots,x_n\}$.

Table 2: Analysis of the property **(DM)** for different candidates to pseudo pre-aggregation function pairs (F_1, F_2), satisfying **(BC0)(i)**, constructed from Table 1

	T_P	T_M	$T_Ł$	T_{DP}	T_{HP}	O_B	O_{mM}	O_α	O_{Div}	GM	HM	S	F_{RS}	C_F	C_L	F_{GL}	F_{BPC}	F_{NA}	F_α	F_{NA2}	AVG	F_{IM}	F_{IP}
T_P	yes	no	yes	yes	no	no	yes	no	no	no	no	no	no	no	no	no	yes	no	no	no	no	no	no
T_M	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	no	no	no	yes	yes	no	yes	yes	no	no	no	no	no
$T_Ł$	no	no	yes	yes	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
T_{DP}	no	no	no	yes	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
T_{HP}	yes	no	yes	yes	yes	no	yes	yes	no	no	no	no	no	yes	no	no	yes	no	no	no	no	no	no
O_B	yes	no	yes	yes	no	yes	yes	yes	no	no	no	no	no	no	no	no	yes	no	no	no	no	no	no
O_{mM}	no	no	no	yes	no	no	yes	no	no	no	no	no	no	no	no	no	yes	no	no	no	no	no	no
O_α	yes	no	yes	yes	no	no	yes	yes	no	no	no	no	no	no	no	no	yes	no	no	no	no	no	no
O_{Div}	yes	no	yes	yes	no	yes	yes	yes	yes	no	no	no	no	no	no	no	yes	no	no	no	no	no	no
GM	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	no	yes	yes	no	no	no	no	no
HM	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	no	no	yes	yes	no	yes	yes	no	no	no	no	no
S	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
F_{RS}	yes	no	yes	yes	no	yes	yes	yes	no	no	no	no	yes	no	no	no	yes	no	no	no	no	no	no
C_F	yes	no	yes	yes	no	no	yes	no	no	no	no	no	no	yes	no	no	yes	no	no	no	no	no	no
C_L	no	no	yes	yes	no	no	no	no	no	no	no	no	no	no	yes	no	yes	no	no	no	no	no	no
F_{GL}	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	no	no	no	no
F_{BPC}	no	no	no	yes	no	no	no	no	no	no	no	no	no	no	yes	no	yes	no	no	no	no	no	no
F_{NA}	no	no	no	yes	no	no	no	no	no	no	no	no	no	no	yes	no	no	yes	no	no	no	no	no
Fα	no	no	no	yes	no	no	no	no	no	no	no	no	no	no	yes	no	no	no	yes	no	no	no	no
F_{NA2}	no	no	no	yes	no	no	no	no	no	no	no	no	no	no	yes	no	no	yes	no	yes	no	no	no
AVG	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	no	yes	yes	yes	yes	yes	no	no
F_{IM}	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	no	no	no	yes	yes	no	yes	yes	yes	no	no	yes	no
F_{IP}	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	no	no	no	yes	yes	no	yes	yes	yes	no	no	yes	yes

(R2) $\{x_1, \ldots, x_n\} = \{y_1, \ldots, y_k\}$ *and*

(R3) $y_1 < \ldots < y_k$.

Note that the functions $R_{n \mapsto k}$ are well defined and, in case some components of the input $\vec{x}$ are repeated, they collapse into one single value. With this definition at hand, we denote, for each $j \in K$ and $\vec{x} \in [0,1]^n$:

$$B_j^R(\vec{x}) = \{i \in N \mid x_i = y_j\}. \tag{7}$$

Observe that, for every $\vec{x} \in [0,1]^n$, it holds that $\cup_{j=1}^k B_j^R(\vec{x})) = N$.

Definition 4.2. *Let $F_1, F_2 : [0,1]^2 \to [0,1]$ be a pair of functions such that $F_1 \geq F_2$ (i.e., F_1 dominates F_2) and F_1 is $(1,0)$-increasing, and consider a fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$. Let $R_{n \mapsto k} : [0,1]^n \to [0,1]^k$ be a $(n \mapsto k)$-dimension reduction function given in Definition 4.1. The generalized $C_{F_1F_2}$-integral based on (F_1, F_2) with respect to $\mathfrak{m}$ is defined as a function $g\mathfrak{C}_\mathfrak{m}^{(F_1,F_2)} : [0,1]^n \to [0,1]$, given, for all $\vec{x} \in [0,1]^n$, by*

$$g\mathfrak{C}_\mathfrak{m}^{(F_1,F_2)}(\vec{x}) = \min\left\{1, \sum_{j=1}^{k} F_1\left(y_j, \mathfrak{m}\left(\cup_{p=j}^k B_p^R(\vec{x})\right)\right) - F_2\left(y_{j-1}, \mathfrak{m}\left(\cup_{p=j}^k B_p^R(\vec{x})\right)\right)\right\}, \tag{8}$$

with the convention that $y_0 = 0$ and B_j^R is as defined in Equation (7).

Proposition 4.1. *Under the conditions of Definition 4.2, $g\mathfrak{C}_\mathfrak{m}^{(F_1,F_2)}$ is well defined, for any pair $F_1, F_2 : [0,1]^2 \to [0,1]$ and fuzzy measure $\mathfrak{m}$.*

PROOF. It is immediate that, for all $\vec{x}, \vec{x}' \in [0,1]^n$, whenever $g\mathfrak{C}_\mathfrak{m}^{(F_1,F_2)}(\vec{x}) \neq g\mathfrak{C}_\mathfrak{m}^{(F_1,F_2)}(\vec{x}')$, then $\vec{x} \neq \vec{x}'$. Now, consider $R_{n \mapsto k}$ and B_j^R as defined in equations (6) and (7), respectively. Then, since F_1 is $(1,0)$-increasing and $F_1 \geq F_2$, one has that:

$$\begin{aligned} F_1(y_j, \mathfrak{m}\left(\cup_{p=j}^k B_p^R(\vec{x})\right)) - F_2(y_{j-1}, \mathfrak{m}\left(\cup_{p=j}^k B_p^R(\vec{x})\right)) &\geq F_1(y_j, \mathfrak{m}\left(\cup_{p=j}^k B_p^R(\vec{x})\right)) - F_1(y_{j-1}, \mathfrak{m}\left(\cup_{p=j}^k B_p^R(\vec{x})\right)) \\ &\geq 0. \end{aligned}$$

Therefore, it holds that $g\mathfrak{C}_\mathfrak{m}^{(F_1,F_2)}(\vec{x}) \geq 0$, for all $\vec{x} \in [0,1]^n$. On the other hand, it is immediate that $g\mathfrak{C}_\mathfrak{m}^{(F_1,F_2)}(\vec{x}) \leq 1$, for all $\vec{x} \in [0,1]^n$. Thus, $g\mathfrak{C}_\mathfrak{m}^{(F_1,F_2)}$ is well defined. □

Lemma 4.1. *Consider $R_{n\mapsto k}$ and B_j^R as defined in equations (6) and (7), respectively. Then, for all $\vec{x} = (x, \dots, x) \in [0,1]^n$, it holds that $R_{n\mapsto k}(\vec{x}) = x$ and $B_1^R(\vec{x}) = N$.*

PROOF. For any $\vec{x} = (x, \dots, x) \in [0,1]^n$, we have that $\{x, \dots, x\} = \{y_1\}$ implies in $y_1 = x$ and $K = \{1\}$. It follows that $R_{n\mapsto k}(\vec{x}) = x$ and $B_1^R(\vec{x}) = \{i \in N \mid x_i = y_1 = x\} = \{1, \dots, n\} = N$. □

Proposition 4.2. *Under the conditions of Definition 4.2, for any fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$ and $F : [0,1]^2 \to [0,1]$, if $F(x,1) = x$, for all $x \in [0,1]$, then $g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}$ is idempotent.*

PROOF. Consider $R_{n\mapsto k}$ and B_j^R as defined in equations (6) and (7), respectively. Then, one has that:

$$\begin{aligned} g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}(\vec{x}) &= \min\left\{1, F(y, \mathfrak{m}(B_1^R(\vec{x}))) - F(0, \mathfrak{m}(B_1^R(\vec{x})))\right\} \quad \text{by Eq. (8)} \\ &= \min\{1, F(y, \mathfrak{m}(N) - F(0, \mathfrak{m}(N)\} \quad \text{by Lemma 4.1} \\ &= \min\{1, F(x,1) - F(0,1)\} \\ &= x. \end{aligned}$$

□

Proposition 4.3. *Under the conditions of Definition 4.2, for any fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$ and pair of functions $F_1, F_2 : [0,1]^2 \to [0,1]$, if $F_2(0,1) = 0$ and $F_1(x,1) \geq x$, for all $x \in [0,1]$, then $g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)} \geq \min$.*

PROOF. Consider $R_{n\mapsto k}$ and B_j^R as defined in equations (6) and (7), respectively. Since $F_1 \geq F_2$, one has that:

$$\begin{aligned} g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x}) &= \min\Bigg\{1, F_1(y_1, \mathfrak{m}(\cup_{p=1}^k B_p^R(\vec{x}))) - F_2(y_0, \mathfrak{m}(\cup_{p=1}^k B_p^R(\vec{x}))) + \\ &\qquad \sum_{j=2}^{k} F_1(y_j, \mathfrak{m}(\cup_{p=j}^k B_p^R(\vec{x}))) - F_2(y_{j-1}, \mathfrak{m}(\cup_{p=j}^k B_p^R(\vec{x})))\Bigg\} \\ &\geq \min\{1, F_1(y_1, \mathfrak{m}(N)) - F_2(y_0, \mathfrak{m}(N))\} \quad \text{by } \mathbf{(DM)} \\ &= \min\{1, F_1(y_1, 1) - F_2(0,1)\} \\ &\geq \min\{1, y_1 - 0\} \\ &= y_1 \\ &= \min \vec{x}. \end{aligned}$$

□

5. $gC_{F_1F_2}$-integrals as aggregation functions

In this subsection, we show that a $gC_{F_1F_2}$-integral is an aggregation function whenever (F_1, F_2) is a pre-aggregation function pair satisfying an additional condition, namely, the pairwise increasingness property **(PI)**.

Proposition 5.1. *Consider $R_{n\mapsto k}$ and B_j^R as defined in equations (6) and (7). For any fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$ and pseudo pre-aggregation function pair (F_1, F_2), $g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}$ satisfies the boundary conditions* **(A2)**.

PROOF. Consider $\vec{0} = (0, \dots, 0) \in [0,1]^n$. It follows that:

$$\begin{aligned} g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{0}) &= \min\left\{1, F_1(0, \mathfrak{m}(B_1^R(\vec{0})) - F_2(0, \mathfrak{m}(B_1^R(\vec{0}))\right\} \quad \text{by Eq. (8)} \\ &= \min\{1, F_1(0, \mathfrak{m}(N) - F_2(0, \mathfrak{m}(N)\} \quad \text{by Lemma 4.1} \\ &= \min\{1, F_1(0,1) - F_2(0,1)\} \\ &= 0 \quad \text{by } \mathbf{(BC0)}. \end{aligned}$$

Consider $\vec{1} = (1, \ldots, 1) \in [0,1]^n$. It follows that:

$$\begin{aligned} g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{1}) &= \min\left\{1, F_1(1, \mathfrak{m}(B_1^R(\vec{1}))) - F_2(0, \mathfrak{m}(B_1^R(\vec{1})))\right\} \text{ by Eq. (8)} \\ &= \min\{1, F_1(1, \mathfrak{m}(N) - F_2(1, \mathfrak{m}(N)\} \text{ by Lemma 4.1} \\ &= \min\{1, F_1(1,1) - F_2(0,1) \\ &= 1 \quad \text{by } \textbf{(BC1)}, \textbf{(BC0)} \end{aligned}$$

□

Lemma 5.1. *Consider $R_{n\mapsto k}$ and B_j^R as defined in equations (6) and (7). Let $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure and $F_1, F_2 : [0,1]^2 \to [0,1]$ a pair of functions satisfying the conditions of Definition 4.2. Then*

$$g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x}) \leq g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{z}),$$

for every $\vec{x} = (x_1, \ldots, x_n), \vec{z} = (z_1, \ldots, z_n) \in [0,1]^n$ such that $x_{(n)} \leq z_{(n)}$ and $x_{(i)} = z_{(i)}$, for all $(i) \in \{(1), \ldots, (n-1)\}$, where $(x_{(1)}, \ldots, x_{(n)})$ and $(z_{(1)}, \ldots, z_{(n)})$ are, respectively, any increasing permutations of $\vec{x}$ and $\vec{z}$.

PROOF. Consider $\vec{x} = (x_1, \ldots, x_n), \vec{z} = (z_1, \ldots, z_n) \in [0,1]^n$ such that $x_{(n)} \leq z_{(n)}$ and $x_{(i)} = z_{(i)}$, for all $(i) \in \{(1), \ldots, (n-1)\}$. Then, according to equations (6) and (7), we have that:

(i) $R_{n\mapsto k}(\vec{x}) = (y_1, \ldots, y_k)$ such that $\{x_1, \ldots, x_n\} = \{y_1, \ldots, y_k\}$, with $k \leq n$, and $y_1 < \ldots < y_k$;

(ii) $B_j^R(\vec{x}) = \{i \in N \mid x_i = y_j\}$, for $j \in K = \{1, \ldots, k\}$;

(iii) $R_{n\mapsto k}(\vec{z}) = (h_1, \ldots, h_w)$ such that $\{z_1, \ldots, z_n\} = \{h_1, \ldots, h_w\}$, with $w \leq n$, and $h_1 < \ldots < h_w$;

(iv) $B_j^R(\vec{z}) = \{i \in N \mid z_i = h_j\}$, for $j \in W = \{1, \ldots, w\}$.

One has the following cases:

$K = W$**:** In this case it holds that $y_1 = h_1 < \ldots < y_{k-1} = h_{w-1} < y_k \leq h_w$ and $B_j^R(\vec{x}) = B_j^R(\vec{z})$, for all $j \in K = W$. Since F_1 is $(1,0)$-increasing, it follows that:

$$\begin{aligned} g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x}) &= \min\left\{1, \sum_{j=1}^{k-1}\left(F_1(y_j, \mathfrak{m}(\cup_{p=j}^k B_p^R(\vec{x}))) - F_2(y_{j-1}, \mathfrak{m}(\cup_{p=j}^k B_p^R(\vec{x})))\right)\right. \\ &\qquad \left. + F_1(y_k, \mathfrak{m}(B_k^R(\vec{x}))) - F_2(y_{k-1}, \mathfrak{m}(B_k^R(\vec{x})))\right\} \\ &\leq \min\left\{1, \sum_{j=1}^{w-1}\left(F_1(h_j, \mathfrak{m}(\cup_{p=j}^w B_p^R(\vec{y}))) - F_2(h_{j-1}, \mathfrak{m}(\cup_{p=j}^w B_p^R(\vec{y})))\right)\right. \\ &\qquad \left. + F_1(h_w, \mathfrak{m}(B_w^R(\vec{y}))) - F_2(h_{w-1}, \mathfrak{m}(B_w^R(\vec{y})))\right\} \\ &= g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{z}). \end{aligned}$$

$k < w$**:** In this case it holds that $w = k+1$, $x_{(n)} = x_{(n-1)} = z_{(n-1)}$, $y_1 = h_1 < \ldots < y_k = h_{w-1} < h_w$, $B_j^R(\vec{x}) = B_j^R(\vec{z})$ for all $j \leq k-1$, $\mid B_{w-1}^R(\vec{z}) \mid = \mid B_k^R(\vec{x}) \mid -1$ and $B_w^R(\vec{z}) = \{n\}$ (that is, $\mid B_w^R(\vec{z}) \mid = 1$).

Since F_1 is $(1,0)$-increasing, it follows that:

$$\begin{aligned} g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x}) &= \min\left\{1, \sum_{j=1}^{k} \left(F_1(y_j, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{x}))) - F_2(y_{j-1}, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{x})))\right)\right\} \\ &\leq \min\left\{1, \sum_{j=1}^{w-1} \left(F_1(h_j, \mathfrak{m}(\cup_{p=j}^{w} B_p^R(\vec{z}))) - F_2(h_{j-1}, \mathfrak{m}(\cup_{p=j}^{w} B_p^R(\vec{z})))\right)\right. \\ &\qquad \left. + F_1(h_w, \mathfrak{m}(B_w^R(\vec{z}))) - F_2(h_{w-1}, \mathfrak{m}(B_w^R(\vec{z})))\right\} \\ &= g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{z}). \end{aligned}$$

Observe that the case $k > w$ never happens, since one would necessarily has $z_{(n)} = z_{(n-1)} = x_{(n-1)} < x_{(n)}$, which is a contradiction. □

Theorem 5.1. *Let $F : [0,1]^2 \to [0,1]$ such that (F, F) is a pair of functions satisfying the conditions of Definition 4.2. The pair (F, F) satisfies **(PI)** if and only if $g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}$ is increasing for each fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$.*

PROOF. ($\Rightarrow$) Suppose that (F, F) satisfies **(PI)** and consider $\vec{x} = (x_1, \ldots, x_{t-1}, x_t, x_{t+1} \ldots, x_n) \in [0,1]^n$, with $t \in \{1, \ldots, n\}$. By convention, given $\vec{x} \in [0,1]^n$, we state that $x_0 = x_{(0)} = 0$ and $x_{n+1} = x_{(n+1)} = 1$. We have the following cases:

(i) $\forall i \in N : i \neq t \to x_t \neq x_i$. In this case, considering equations (6) and (7), for each $\vec{x} \in [0,1]^n$, one has that:

- $R_{n \mapsto k}(\vec{x}) = (y_1, \ldots, y_{l-1}, y_l = x_t, y_{l+1}, \ldots, y_k)$ (with $l \in K = \{1, \ldots, k\}$, $k \leq n$, $y_0 = 0$ and $y_{k+1} = 1$) such that $\{x_1, \ldots, x_{t-1}, x_t, x_{t+1}, \ldots, x_n\} = \{y_1, \ldots, y_{l-1}, y_l = x_t, y_{l+1}, \ldots y_k\}$ and $y_1 < \ldots < y_{l-1} < y_l = x_t < y_{l+1} < \ldots < y_k$.
- $B_j^R(\vec{x}) = \{i \in N \mid x_i = y_j\}$, for $j \in K = \{1, \ldots, k\}$. In particular, one has that $B_l^R(\vec{x}) = \{t\}$.

Now, consider the following cases:

(ia) Suppose that there exists $\vec{z} \in [0,1]^n$, with $\vec{x} < \vec{z}$, such that $\vec{z} = (z_1 = x_1, \ldots, z_{t-1} = x_{t-1}, z_t, z_{t+1} = x_{t+1}, \ldots, z_t = x_n) \in [0,1]^n$ with $y_1 < \ldots < y_{l-1} < y_l = x_t < z_t < y_{l+1} < \ldots < y_k$. If $t = 1$ or $t = n$ then define $\vec{z} = (z, z_2, \ldots, z_n) \in [0,1]^n$ or $\vec{z} = (z_1, \ldots, z_{n-1}, z) \in [0,1]^n$, respectively. In this case, considering equations (6) and (7), one has that:

- $R_{n \mapsto w}(\vec{z}) = (h_1 = y_1, \ldots, h_{l-1} = y_{l-1}, h_l = z_t, h_{l+1} = y_{l+1}, \ldots, h_w = y_k)$, with $w = k \leq n$, where $h_1 = y_1 < \ldots < h_{l-1} = y_{l-1} < y_l = x_t < h_l = z_t < h_{l+1} = y_{l+1} < \ldots < h_w = y_k$.
- $B_j^R(\vec{z}) = \{i \in N \mid z_i = h_j\} = \{i \in N \mid x_i = y_j\} = B_j^R(\vec{x})$, for $j \in W = K = \{1, \ldots, w = k\}$. In particular, one has that $B_l^R(\vec{z}) = \{t\} = B_l^R(\vec{x})$.

Since the pair (F, F) satisfies **(PI)** and by Lemma 5.1, it follows that:

$$
\begin{aligned}
&g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}(\vec{x}) \\
&= \min\Bigg\{1, \sum_{j=1}^{l-1}\big(F(y_j, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{x}))) - F(y_{j-1}, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{x})))\big) + F(y_l, \mathfrak{m}(\cup_{p=l}^{k} B_p^R(\vec{x}))) \\
&\quad -F(y_{l-1}, \mathfrak{m}(\cup_{p=l}^{k} B_p^R(\vec{x}))) + F(y_{l+1}, \mathfrak{m}(\cup_{p=l+1}^{k} B_p^R(\vec{x}))) - F(y_l, \mathfrak{m}(\cup_{p=l+1}^{k} B_p^R(\vec{x}))) \\
&\quad + \sum_{j=l+2}^{k}\big(F(y_j, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{x}))) - F(y_{j-1}, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{x})))\big)\Bigg\} \\
&= \min\Bigg\{1, \sum_{j=1}^{l-1}\big(F(y_j, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{x}))) - F(y_{j-1}, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{x})))\big) + F(x_t, \mathfrak{m}(\cup_{p=l}^{k} B_p^R(\vec{x}))) \\
&\quad -F(y_{l-1}, \mathfrak{m}(\cup_{p=l}^{k} B_p^R(\vec{x}))) + F(y_{l+1}, \mathfrak{m}(\cup_{p=l+1}^{k} B_p^R(\vec{x}))) - F(x_t, \mathfrak{m}(\cup_{p=l+1}^{k} B_p^R(\vec{x}))) \\
&\quad + \sum_{j=l+2}^{k}\big(F(y_j, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{x})) - F(y_{j-1}, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{x})))\big)\Bigg\} \\
&\leq \min\Bigg\{1, \sum_{j=1}^{l-1}\big(F(h_j, \mathfrak{m}(\cup_{p=j}^{w} B_p^R(\vec{z}))) - F_2(h_{j-1}, \mathfrak{m}(\cup_{p=j}^{w} B_p^R(\vec{z})))\big) + F(z_t, \mathfrak{m}(\cup_{p=l}^{w} B_p^R(\vec{z}))) \\
&\quad -F(h_{l-1}, \mathfrak{m}(\cup_{p=l}^{w} B_p^R(\vec{z}))) + F(h_{l+1}, \mathfrak{m}(\cup_{p=l+1}^{w} B_p^R(\vec{z}))) - F(z_t, \mathfrak{m}(\cup_{p=l+1}^{w} B_p^R(\vec{z}))) \\
&\quad + \sum_{j=l+2}^{w}\big(F(h_j, \mathfrak{m}(\cup_{p=j}^{w} B_p^R(\vec{z}))) - F(h_{j-1}, \mathfrak{m}(\cup_{p=j}^{w} B_p^R(\vec{z})))\big)\Bigg\} \\
&= g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}(\vec{z})
\end{aligned}
$$

since $\mathfrak{m}(\cup_{p=l+1}^{k} B_p^R(\vec{x})) = \mathfrak{m}(\cup_{p=l+1}^{w} B_p^R(\vec{z})) \leq \mathfrak{m}(\cup_{p=l}^{k} B_p^R(\vec{x})) = \mathfrak{m}(\cup_{p=l}^{w} B_p^R(\vec{z}))$, and, by **(PI)**, it holds that

$$F(x_t, \mathfrak{m}(\cup_{p=l}^{k} B_p^R(\vec{x}))) - F(x_t, \mathfrak{m}(\cup_{p=l+1}^{k} B_p^R(\vec{x}))) \leq F(z_t, \mathfrak{m}(\cup_{p=l}^{w} B_p^R(\vec{z}))) - F(z_t, \mathfrak{m}(\cup_{p=l+1}^{w} B_p^R(\vec{z}))).$$

(ib) Now, consider that $n \geq 2$ and $\vec{z} \in [0,1]^n$, with $\vec{x} < \vec{z}$, such that $\vec{z} = (z_1 = x_1, \ldots, z_{t-1} = x_{t-1}, z_t, z_{t+1} = x_{t+1}, \ldots, z_t = x_n) \in [0,1]^n$ with $y_1 < \ldots < y_{l-1} < y_l = x_t < z_t = y_{l+1} < \ldots < y_k$. If $t = 1$ or $t = n$ then define $\vec{z} = (z, z_2, \ldots, z_n) \in [0,1]^n$ or $\vec{z} = (z_1, \ldots, z_{n-1}, z) \in [0,1]^n$, respectively. In this case, considering equations (6) and (7), one has that:

- $R_{n\mapsto w}(\vec{z}) = (h_1 = y_l, \ldots, h_{l-1} = y_{l-1}, h_l = z_t = y_{l+1}, h_{l+1} = y_{l+2}, \ldots, h_w = y_k)$, with $w = k - 1 \leq n$, where $h_1 = y_l < \ldots < h_{l-1} = y_{l-1} < y_l = x_t < h_l = z_t = y_{l+1} < h_{l+1} = y_{l+2} < \ldots < h_w = y_k$.
- $B_j^R(\vec{z}) = \{i \in N \mid z_i = h_j\}$.

Observe that, since $h_l = z_t = y_{l+1}$, with $l \in W$, then it holds that:

- $\forall j \in W : j < l \to B_j^R(\vec{z}) = B_j^R(\vec{x})$,
- $\mid B_l^R(\vec{z}) \mid = \mid B_{l+1}^R(\vec{x}) \mid +1$,
- $\forall j \in W : j > l \to B_j^R(\vec{z}) = B_{j+1}^R(\vec{x})$.
- $\mid \cup_{p=l}^{k} B_p^R(\vec{x})) \mid = \mid \cup_{p=l}^{w} B_p^R(\vec{z})) \mid$.

Since the pair (F,F) satisfies **(PI)** and by Lemma 5.1, it follows that:

$$
\begin{aligned}
& g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}(\vec{x}) \\
&= \min\Bigg\{1, \sum_{j=1}^{l-1}\big(F(y_j,\mathfrak{m}(\cup_{p=j}^{k}B_p^R(\vec{x}))) - F(y_{j-1},\mathfrak{m}(\cup_{p=j}^{k}B_p^R(\vec{x})))\big) \\
&\quad +F(y_l,\mathfrak{m}(\cup_{p=l}^{k}B_p^R(\vec{x}))) - F(y_{l-1},\mathfrak{m}(\cup_{p=l}^{k}B_p^R(\vec{x}))) \\
&\quad +F(y_{l+1},\mathfrak{m}(\cup_{p=l+1}^{k}B_p^R(\vec{x}))) - F(y_l,\mathfrak{m}(\cup_{p=l+1}^{k}B_p^R(\vec{x}))) \\
&\quad +F(y_{l+2},\mathfrak{m}(\cup_{p=l+2}^{k}B_p^R(\vec{x}))) - F(y_{l+1},\mathfrak{m}(\cup_{p=l+2}^{k}B_p^R(\vec{x}))) \\
&\quad + \sum_{j=l+3}^{k}\big(F(y_j,\mathfrak{m}(\cup_{p=j}^{k}B_p^R(\vec{x}))) - F(y_{j-1},\mathfrak{m}(\cup_{p=j}^{k}B_p^R(\vec{x})))\big)\Bigg\} \\
&= \min\Bigg\{1, \sum_{j=1}^{l-1}\big(F(y_j,\mathfrak{m}(\cup_{p=j}^{k}B_p^R(\vec{x}))) - F(y_{j-1},\mathfrak{m}(\cup_{p=j}^{k}B_p^R(\vec{x})))\big) \\
&\quad +F(x_t,\mathfrak{m}(\cup_{p=l}^{k}B_p^R(\vec{x}))) - F(y_{l-1},\mathfrak{m}(\cup_{p=l}^{k}B_p^R(\vec{x}))) \\
&\quad +F(y_{l+1},\mathfrak{m}(\cup_{p=l+1}^{k}B_p^R(\vec{x}))) - F(x_t,\mathfrak{m}(\cup_{p=l+1}^{k}B_p^R(\vec{x}))) \\
&\quad +F(y_{l+2},\mathfrak{m}(\cup_{p=l+2}^{k}B_p^R(\vec{x}))) - F(y_{l+1},\mathfrak{m}(\cup_{p=l+2}^{k}B_p^R(\vec{x}))) \\
&\quad + \sum_{j=l+3}^{k}\big(F(y_j,\mathfrak{m}(\cup_{p=j}^{k}B_p^R(\vec{x}))) - F(y_{j-1},\mathfrak{m}(\cup_{p=j}^{k}B_p^R(\vec{x})))\big)\Bigg\} \\
&\leq \min\Bigg\{1, \sum_{j=1}^{l-1}\big(F(h_j=y_j,\mathfrak{m}(\cup_{p=j}^{w}B_p^R(\vec{z}))) - F(h_{j-1}=y_{j-1},\mathfrak{m}(\cup_{p=j}^{w}B_p^R(\vec{z})))\big) \\
&\quad +F(h_l=z_t=y_{l+1},\mathfrak{m}(\cup_{p=l}^{w}B_p^R(\vec{z}))) - F(h_{l-1}=y_{l-1},\mathfrak{m}(\cup_{p=l}^{w}B_p^R(\vec{z}))) \\
&\quad +F(h_{l+1}=y_{l+2},\mathfrak{m}(\cup_{p=l+1}^{w}B_p^R(\vec{x}))) - F(h_l=z_t=y_{l+1},\mathfrak{m}(\cup_{p=l+1}^{w}B_p^R(\vec{x}))) \\
&\quad + \sum_{j=l+2}^{w}\big(F(h_j,\mathfrak{m}(\cup_{p=j}^{w}B_p^R(\vec{z}))) - F(h_{j-1},\mathfrak{m}(\cup_{p=j}^{w}B_p^R(\vec{z})))\big)\Bigg\} \\
&= g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}(\vec{z})
\end{aligned}
$$

since $\cup_{p=l+1}^{k}B_p^R(\vec{x}) \subset \cup_{p=l}^{k}B_p^R(\vec{x}) = \cup_{p=l}^{w}B_p^R(\vec{z})$, and, then, by **(PI)**, it holds that

$$
\begin{aligned}
& F(x_t,\mathfrak{m}(\cup_{p=l}^{k}B_p^R(\vec{x}))) - F(x_t,\mathfrak{m}(\cup_{p=l+1}^{k}B_p^R(\vec{x}))) \\
&\quad < \quad F(h_l=z_t=y_{l+1},\mathfrak{m}(\cup_{p=l}^{w}B_p^R(\vec{z}))) - F(y_{l+1},\mathfrak{m}(\cup_{p=l+1}^{k}B_p^R(\vec{x}))).
\end{aligned}
$$

(ic) Now consider $l \in \{1,\ldots,k-3\}$, and $\vec{z} = (z_1 = x_1,\ldots,z_{t-1} = x_{t-1}, z_t, z_{t+1} = x_{t+1},\ldots,x_n) \in [0,1]^n$, such that $\vec{x} < \vec{z}$, with $y_1 < \ldots < y_{l-1} < y_l = x_t < y_{l+1} < \ldots < z_t < \ldots < y_k$. If $t = 1$ or $t = n$ then define $\vec{z} = (z, z_2,\ldots,z_n) \in [0,1]^n$ or $\vec{z} = (z_1,\ldots,z_{n-1},z) \in [0,1]^n$, respectively. In this case, considering equations (6) and (7), one has that:

- $R_{n\mapsto w}(\vec{z}) = (h_1,\ldots,h_{l-1},h_l = z_t,h_{l+1},\ldots,h_w)$, with $w \leq n$, where $h_1 < \ldots < h_{l-1} < h_l = z_t < h_{l+1} < \ldots < h_w$ and $\{x_1 = z_1,\ldots,x_{t-1} = z_{t-1}, z_t, x_{t+1} = z_{t+1},\ldots,z_n = x_n\} = \{h_1,\ldots,h_{l-1},h_l = z,h_{l+1},\ldots h_w\}$.
- $B_j^h = \{i \mid z_i = h_j\}$, for $j \in W = \{1,\ldots,w\}$.

Consider $r \in \{2, \ldots, w-l-1\}$. Suppose that $y_l = x_t < y_{l+1} < \ldots < y_{k-r} < z_t < y_{k-r+2}$. Then, by **(ia)** and **(ib)**, it follows that:

$$g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x}) \leq g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{s}_1) \leq \ldots \leq g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{s}_{n-r-l}) \leq g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{z}),$$

where, for $i = 1, \ldots, n-r-l$, $\vec{s}_i = (x_1, \ldots, x_{t-1}, y_{l+i}, x_{t+1}, \ldots, x_n)$.

(id) Suppose the same conditions of case **(ic)**, but for $\vec{z} = (z_1 = x_1, \ldots, z_{t-1} = x_{t-1}, z_t, z_{t+1} = x_{t+1}, \ldots, x_n) \in [0,1]^n$, such that $z_t = y_j$, for some $y_j > y_{l+1}$, that is, $y_1 < \ldots < x_t = y_l < y_{l+1} < \ldots < z_t = y_j < \ldots < y_k$. In this case, considering equations (6) and (7), one has that:

- $R_{n\mapsto w}(\vec{z}) = (h_1 = y_1, \ldots, h_{l-1} = y_{j-1}, h_l = z_t = y_j, h_{l+1} = y_{j+1}, \ldots, h_w)$, with $w < k$, where $h_1 < \ldots < h_{l-1} < h_l = z_t < h_{l+1} < \ldots < h_w$ and $\{x_1 = z_1, \ldots, x_{t-1} = z_{t-1}, z_t, x_{t+1} = z_{t+1}, \ldots, z_n = x_n\} = \{h_1, \ldots, h_{l-1}, h_l = z_t, h_{l+1}, \ldots h_w\}$.
- $B_j^h = \{i \mid z_i = h_j\}$, for $j \in W = \{1, \ldots, w\}$.

Consider $r \in \{2, \ldots, w-l-1\}$. Suppose that $y_l = x_t < y_{l+1} < \ldots < y_{k-r} < z_t = k-r+1 < y_{k-r+2}$. Then, considering **(ib)**, the proof is analogous to **(ic)**.

(ii) $\exists i \in N, i \neq t,\ s.t.\ x_t = x_i$. In this case, we have the same subcases **(ia)**-**(id)**, and the proofs are analogous.

($\Leftarrow$) We prove the contrapositive. Suppose that the pair (F, F) does not satisfy **(PI)**. Then, there exist $a, b, c, d \in [0,1]$ such that $a \leq b$, $c \leq d$ and $F(a,d) - F(a,c) > F(b,d) - F(b,c)$. Observe that $a \neq 1$ and $c \neq 1$. Let $\mathfrak{m} : 2^N \to [0,1]$ be such that $\mathfrak{m}(\{n-1, n-2\}) = d$ and $\mathfrak{m}(\{n-1\}) = c$. Then, for $\vec{x} = (0, \ldots, 0, a, 1)$ and $\vec{z} = (0, \ldots, 0, b, 1)$, we have that $k = 3$ and $\vec{x} \leq \vec{z}$. Consider $\vec{y} = (0, a, 1)$ and $\vec{h} = (0, b, 1)$. Then, one has that:

$$\begin{aligned}
g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}(\vec{x}) &= \min\big\{1, F(0, \mathfrak{m}(\cup_{p=1}^{3} B_p^R(\vec{x}))) - F(0, \mathfrak{m}(\cup_{p=1}^{k} B_p^R(\vec{x}))) + F(a, \mathfrak{m}(\cup_{p=2}^{3} B_p^R(\vec{x}))) \\
&\quad -F(0, \mathfrak{m}(\cup_{p=2}^{3} B_p^R(\vec{x}))) + F(1, \mathfrak{m}(B_3^R(\vec{x}))) - F(a, \mathfrak{m}(B_3^R(\vec{x})))\big\} \\
&= \min\{1, F(a, \mathfrak{m}(\{n-2, n-1\})) - F(0, \mathfrak{m}(\{n-2, n-1\})) + F(1, \mathfrak{m}(\{n-1\})) \\
&\quad -F(a, \mathfrak{m}(\{n-1\}))\} \\
&= \min\{1, F(a,d) - F(0,d) + F(1,c) - F(a,c)\} \\
&> \min\{1, F(b,d) - F(0,d) + F(1,c) - F(b,c)\} \\
&= \min\{1, F(b, \mathfrak{m}(\{n-2, n-1\})) - F(0, \mathfrak{m}(\{n-2, n-1\})) + F(1, \mathfrak{m}(\{n-1\})) \\
&\quad -F(b, \mathfrak{m}(\{n-1\}))\} \\
&= \min\big\{1, F(0, \mathfrak{m}(\cup_{p=1}^{3} B_p^R(\vec{x}))) - F(0, \mathfrak{m}(\cup_{p=1}^{k} B_p^R(\vec{x}))) + F(b, \mathfrak{m}(\cup_{p=2}^{3} B_p^R(\vec{x}))) \\
&\quad -F(0, \mathfrak{m}(\cup_{p=2}^{3} B_p^R(\vec{x}))) + F(1, \mathfrak{m}(B_3^R(\vec{x}))) - F(b, \mathfrak{m}(B_3^R(\vec{x})))\big\} \\
&= g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}(\vec{z})
\end{aligned}$$

Therefore, $g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}$ is not increasing for each fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$. □

Corollary 5.1. *For any fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$ and pseudo pre-aggregation function pair (F, F) satisfying **(PI)**, $g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}$ is an aggregation function.*

PROOF. It follows from Proposition 5.1 and Theorem 5.1. □

Observe that if for some pseudo pre-aggregation function pair (F, F) and fuzzy measure $\mathfrak{m}$ we have that $g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}$ is not increasing (and, thus, it is not an aggregation function) then (F, F) does not satisfy **(PI)**. Nevertheless, this does not mean that, for some other fuzzy measure $\mathfrak{m}'$, $g\mathfrak{C}_{\mathfrak{m}'}^{(F,F)}$ would not be an aggregation function.

Example 5.1. *Let $F : [0,1]^2 \to [0,1]$ be the function defined by*

$$F(x,y) = \begin{cases} 0 & \text{if } x = 0 \ \vee\ y = 0; \\ \frac{x+y}{2} & \text{if } 0 < x \leq y; \\ x & \text{otherwise.} \end{cases}$$

*Clearly, F is (1,0)-increasing, $F(0,1)=0$ and $F(1,1)$, and therefore (F,F) is a pseudo pre-aggregation pair (in fact, F is an aggregation function). But, (F,F) does not satisfy **(PI)**. In fact, one has that*

$$F(0.3,0.7)-F(0.3,0.5)=0.5-0.4=0.1>0=1-1=F(1,0.7)-F(1,0.5).$$

Hence, by Theorem 5.1, for some fuzzy measure $\mathfrak{m}$, $g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}$ is not increasing. In particular, by the proof of this Theorem, $g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}$ is not increasing for any fuzzy measure $\mathfrak{m}$ such that $1>\mathfrak{m}(\{n-2,n-1\})>\mathfrak{m}(\{n-1\})$. However, for the fuzzy measure $\mathfrak{m}_\perp : 2^N \to [0,1]$, defined by:

$$\mathfrak{m}_\perp(X)=\begin{cases}1 & \text{if } X=N;\\ 0 & \text{otherwise,}\end{cases}$$

one has that $g\mathfrak{C}_{\mathfrak{m}_\perp}^{(F,F)} : [0,1]^n \to [0,1]$ is the aggregation function, defined, for all $\vec{x}=(x_1,\ldots,x_n)\in[0,1]^n$, by:

$$g\mathfrak{C}_{\mathfrak{m}_\perp}^{(F,F)}(\vec{x})=\begin{cases}0 & \text{if } y_1=0\\ \frac{y_1+1}{2} & \text{otherwise,}\end{cases}$$

where $y_1=x_t$, such that $x_t\le x_i$, for all $i\in\{1,\ldots,n\}$.

Notice that Theorem 5.1 requires a pseudo pre-aggregation function pair (F_1,F_2) satisfying **(PI)** and additionally that $F_1=F_2$ in order to guarantee that $g\mathfrak{C}_{\mathfrak{m}_\perp}^{(F_1,F_2)}$ is increasing. The following example shows that there exist pseudo pre-aggregation function pairs (F_1,F_2), with $F_1\neq F_2$, satisfying **(PI)** such that $g\mathfrak{C}_{\mathfrak{m}_\perp}^{(F_1,F_2)}$ is not increasing.

Example 5.2. *Consider the pseudo pre-aggregation function pair (T_P,F_{BPC}), where T_P is the product t-norm and F_{BPC} is an aggregation function (which is neither a t-norm, overlap function nor a copula), as defined in Tables 1 and 2. Observe that T_P dominates F_{BPC}. Moreover, this pair satisfies **(PI)**. In fact, for all $x,y_1,y_2\in[0,1]$ and $h>0$ such that $x+h\in[0,1]$, if $y_2\le y_1$, it holds that:*

$$\begin{aligned}T_P(x+h,y_1)-F_{BPC}(x+h,y_2) &= (x+h)y_1-(x+h)y_2^2 \quad \text{by Table 1}\\ &= xy_1-xy_2^2+h(y_1-y_2^2)\\ &= T_P(x,y_1)-F_{BPC}(x,y_2)+h(y_1-y_2^2) \quad \text{by Table 1}\\ &\ge T_P(x,y_1)-F_{BPC}(x,y_2),\end{aligned}$$

since $h(y_1-y_2^2)\ge 0$. However, $g\mathfrak{C}_{\mathfrak{m}}^{(T_P,F_{BPC})}$ is not increasing. In fact, consider $\vec{x}=(0.6,0.4,0.6,0.5,0.4,0.6,0.7)$ and $\vec{z}=(0.6,0.4,0.6,0.6,0.4,0.6,0.7)$, that is, $\vec{x}<\vec{z}$. Then, $k=4$ and $w=3$, and:

- $R_{n\mapsto k}(\vec{x})=(0.4,0.5,0.6,0.7)$ *and* $R_{n\mapsto w}(\vec{z})=(0.4,0.6,0.7)$;
- $B_1^R(\vec{x})=\{2,5\}$, $B_2^R(\vec{x})=\{4\}$, $B_3^R(\vec{x})=\{1,3,6\}$ *and* $B_4^R(\vec{x})=\{7\}$;
- $B_1^R(\vec{z})=\{2,5\}$, $B_2^R(\vec{z})=\{1,3,4,6\}$ *and* $B_3^R(\vec{x})=\{7\}$.

Suppose that the fuzzy measure $\mathfrak{m}:2^N\to[0,1]$ is such that: $\mathfrak{m}(\{1,2,3,4,5,6,7\})=1$, $\mathfrak{m}(\{1,3,4,6,7\})=0.8$, $\mathfrak{m}(\{1,3,6,7\})=0.5$ and $\mathfrak{m}(\{7\})=0.2$. Then one has that:

$$\begin{aligned}g\mathfrak{C}_{\mathfrak{m}}^{(T_P,F_{BPC})}(\vec{x}) &= \min\left\{1,\sum_{j=1}^{4}\left(T_P(y_j,\mathfrak{m}(\cup_{p=j}^k B_p^R(\vec{x})))-F_{BPC}(y_{j-1},\mathfrak{m}(\cup_{p=j}^k B_p^R(\vec{x})))\right)\right\}\\ &= \{1, T_P(0.4,\mathfrak{m}(\{1,2,3,4,5,6,7\}))-F_{BPC}(0,\mathfrak{m}(\{1,2,3,4,5,6,7\}))\\ &\qquad +T_P(0.5,\mathfrak{m}(\{1,3,4,6,7\}))-F_{BPC}(0.4,\mathfrak{m}(\{1,3,4,6,7\}))\\ &\qquad +T_P(0.6,\mathfrak{m}(\{1,3,6,7\}))-F_{BPC}(0.5,\mathfrak{m}(\{1,3,6,7\}))\\ &\qquad +T_P(0.7,\mathfrak{m}(\{7\}))-F_{BPC}(0.6,\mathfrak{m}(\{7\}))\}\\ &= \min\{1,0.4\cdot 1-0\cdot(1)^2+0.5\cdot 0.8-0.4\cdot(0.8)^2+0.6\cdot 0.5-0.5\cdot(0.5)^2+0,7\cdot 0.2\\ &\qquad -0.6\cdot(0.2)^2\}\\ &= 0.835\end{aligned}$$

and

$$
\begin{aligned}
g\mathfrak{C}_{\mathfrak{m}}^{(T_P,F_{BPC})}(\vec{z}) &= \min\left\{1, \sum_{j=1}^{3}\left(T_P(y_j, \mathfrak{m}(\cup_{p=j}^{w} B_p^R(\vec{z}))) - F_{BPC}(y_{j-1}, \mathfrak{m}(\cup_{p=j}^{k} B_p^R(\vec{z})))\right)\right\} \\
&= \{1, T_P(0.4, \mathfrak{m}(\{1,2,3,4,5,6,7\})) - F_{BPC}(0, \mathfrak{m}(\{1,2,3,4,5,6,7\})) \\
&\qquad +T_P(0.6, \mathfrak{m}(\{1,3,4,6,7\})) - F_{BPC}(0.4, \mathfrak{m}(\{1,3,4,6,7\})) \\
&\qquad +T_P(0.7, \mathfrak{m}(\{7\})) - F_{BPC}(0.6, \mathfrak{m}(\{7\}))\} \\
&= \min\{1, 0.4 \cdot 1 - 0 \cdot (1)^2 + 0.6 \cdot 0.8 - 0.4 \cdot (0.8)^2 + 0,7 \cdot 0.2 - 0.6 \cdot (0.2)^2\} \\
&= 0.74.
\end{aligned}
$$

Thus, $g\mathfrak{C}_{\mathfrak{m}}^{(T_P,F_{BPC})}(\vec{x}) > g\mathfrak{C}_{\mathfrak{m}}^{(T_P,F_{BPC})}(\vec{z})$ *and* $g\mathfrak{C}_{\mathfrak{m}}^{(T_P,F_{BPC})}$ *is not an aggregation function, since it is not increasing.*

Now we present an example of a pseudo pre-aggregation function pair (F, F) satisfying **(PI)** (then, fulfilling all the requirements of Theorem 5.1), thus generating an aggregation function $g\mathfrak{C}_{\mathfrak{m}_\perp}^{(F,F)}$.

Example 5.3. *Consider the pseudo pre-aggregation function pair* (F_{IP}, F_{IP})*, where* F_{IP} *is not even a pre-aggregation function, as defined in Tables 1 and 2. This pair satisfies* ***(PI)****. In fact, for all* $x, y_1, y_2 \in [0,1]$ *and* $h > 0$ *such that* $x + h \in [0,1]$*, if* $y_2 \leq y_1$*, it holds that:*

$$
\begin{aligned}
F_{IP}(x+h, y_1) - F_{IP}(x+h, y_2) &= 1 - y_1 + (x+h)y_1 - (1 - y_2 + (x+h)y_2) \text{ by Table 1} \\
&= (1 - y_1 + xy_1) - (1 - y_2 + xy_2) + h(y_1 - y_2) \\
&= F_{IP}(x, y_1) - F_{IP}(x, y_2) + h(y_1 - y_2) \text{ by Table 1} \\
&\geq F_{IP}(x, y_1) - F_{IP}(x, y_2),
\end{aligned}
$$

since $h(y_1 - y_2) \geq 0$*. Thus, from Corollary 5.1,* $g\mathfrak{C}_{\mathfrak{m}}^{(F_{IP},T_{IP})}$ *is an aggregation function, for any fuzzy measure* $\mathfrak{m} : 2^N \to [0,1]$.

Corollary 5.2. *For any fuzzy measure* $\mathfrak{m} : 2^N \to [0,1]$ *and pseudo pre-aggregation function pair* (F, F) *satisfying* ***(PI)****,* $g\mathfrak{C}_{\mathfrak{m}}^{(F,F)}$ *is an averaging aggregation function if and only if* $F(x,1) = x$*, for all* $x \in [0,1]$.

PROOF. It follows from Corollary 5.1 and Proposition 4.2. □

Example 5.4. *Consider the pseudo pre-aggregation function pair* (F_{BPC}, F_{BPC})*, where* F_{BPC} *is an aggregation function (which is neither a t-norm, overlap function nor a copula), as defined in Tables 1 and 2. This pair satisfies* ***(PI)****. In fact, for all* $x, y_1, y_2 \in [0,1]$ *and* $h > 0$ *such that* $x + h \in [0,1]$*, then, whenever* $y_2 \leq y_1$*, it holds that:*

$$
\begin{aligned}
F_{BPC}(x+h, y_1) - F_{BPC}(x+h, y_2) &= (x+h)y_1^2 - (x+h)y_2^2 \text{ by Table 1} \\
&= xy_1^2 - xy_2^2 + h(y_1^2 - y_2^2) \\
&= F_{BPC}(x, y_1) - F_{BPC}(x, y_2) + h(y_1^2 - y_2^2) \text{ by Table 1} \\
&\geq F_{BPC}(x, y_1) - F_{BPC}(x, y_2),
\end{aligned}
$$

since $h(y_1^2 - y_2^2) \geq 0$*. Therefore, since* $F_{BPC}(x,1) = x$*, then, from Corollary 5.2, it follows that* $g\mathfrak{C}_{\mathfrak{m}}^{(F_{BPC},F_{BPC})}$ *is an averaging aggregation function, for any fuzzy measure* $\mathfrak{m} : 2^N \to [0,1]$.

Corollary 5.3. *For any fuzzy measure* $\mathfrak{m} : 2^N \to [0,1]$ *and copula* C*,* $g\mathfrak{C}_{\mathfrak{m}}^{(C,C)}$ *is an averaging aggregation function.*

PROOF. It follows from Corollary 5.2 and Proposition 3.2. □

Remark 5.1. *Considering equations (4) and (8), by an easy calculation it is possible to check that, whenever $F_1 = F_2 = C$, for a copula C, one has that:*

$$
\begin{aligned}
g\mathfrak{C}_{\mathfrak{m}}^{(C,C)}(\vec{x}) &= \min\left\{1, \sum_{j=1}^{k} C\left(y_j, \mathfrak{m}\left(\cup_{p=j}^{k} B_p^R(\vec{x})\right)\right) - C\left(y_{j-1}, \mathfrak{m}\left(\cup_{p=j}^{k} B_p^R(\vec{x})\right)\right)\right\} \\
&= \sum_{i=1}^{n} C\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - C\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right) \qquad (9) \\
&= \mathfrak{C}_{\mathfrak{m}}^{C}(\vec{x}),
\end{aligned}
$$

which is, in fact, the CC-Integral used in classification problems in [9]. In [30, Theorem 1], Radko and Stupnanová showed that the CC-Integral is a C-based universal integral I_m^C, for a fuzzy measure $\mathfrak{m}$ and copula C. Additionally, from [30, Corollary 2], for any fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$ and copula $C : [0,1]^2 \to [0,1]$, one has that $g\mathfrak{C}_{\mathfrak{m}}^{(C,C)}$ is an OMA[2] operator and vice-versa.

Remark 5.2. *Observe that, by Remark 5.2, whenever $F_1 = F_2 = C$, for a copula C, it is not necessary to make the dimension reduction to deal with duplicated elements.*

Example 5.5. *Consider the pseudo pre-aggregation pair (T_M, T_M), where T_M is the minimum t-norm. Then, for any fuzzy measure $\mathfrak{m} : 2^N \to [0,1]$, $g\mathfrak{C}_{\mathfrak{m}}^{(T_M,T_M)}$ is an averaging aggregation function, since (T_M, T_M) satisfies **PI** and $T_M(x,1) = x$. Moreover, by [30, Corollary 1], $g\mathfrak{C}_{\mathfrak{m}}^{(T_M,T_M)}$ is a Sugeno Integral [32]. Observe that, by Remark 5.2, since $F_1 = F_2 = T_M$, we do not need to worry about the duplicated components in the input $\vec{x}$, so that we can just consider that $K = N$ in Definition 4.1. In fact, consider $\vec{x} \in [0,1]^n$ and let $(x_{(1)}, \ldots, x_{(n)})$ be an increasing permutation on the input $\vec{x}$, and $A_{(i)} = \{(i), \ldots, (n)\}$ be the subset of indices of the $n - i + 1$ largest components of $\vec{x}$. It follows that:*

$$
\begin{aligned}
g\mathfrak{C}_{\mathfrak{m}}^{(T_M,T_M)}(\vec{x}) &= \min\left\{1, \sum_{i=1}^{n} \min\left\{x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right\} - \min\left\{x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right\}\right\}, \\
&= \min\left\{1, \sum_{i=1}^{n} \begin{cases} x_{(i)} - x_{(i-1)} & \text{if } x_{(i)} \leq \mathfrak{m}\left(A_{(i)}\right) \\ \mathfrak{m}\left(A_{(i)}\right) - x_{(i-1)} & \text{if } x_{(i)} > \mathfrak{m}\left(A_{(i)}\right) \ \wedge \ x_{(i-1)} \leq \mathfrak{m}\left(A_{(i)}\right) \\ 0 & \text{otherwise.} \end{cases}\right\}
\end{aligned}
$$

Suppose that for some $k \in \{1, \ldots, n\}$, it holds that $x_{(k)} > \mathfrak{m}\left(A_{(k)}\right)$, but $x_{(k-1)} \leq \mathfrak{m}\left(A_{(k)}\right)$. Then it holds that:

$$
\begin{aligned}
g\mathfrak{C}_{\mathfrak{m}}^{(T_M,T_M)}(\vec{x}) &= \min\left\{1, \sum_{i=1}^{n} \begin{cases} x_{(i)} - x_{(i-1)} & \text{if } x_{(i)} \leq \mathfrak{m}\left(A_{(i)}\right) \\ \mathfrak{m}\left(A_{(i)}\right) - x_{(i-1)} & \text{if } x_{(i)} > \mathfrak{m}\left(A_{(i)}\right) \ \wedge \ x_{(i-1)} \leq \mathfrak{m}\left(A_{(i)}\right) \\ 0 & \text{otherwise.} \end{cases}\right\} \\
&= \min\Big\{1, (x_{(1)} - x_{(0)}) + (x_{(2)} - x_{(1)}) + \ldots + (x_{(k-1)} - x_{(k-2)}) + (\mathfrak{m}\left(A_{(k)}\right) - x_{(k-1)}) \\
&\qquad + \underbrace{0 + \ldots + 0}_{n-k}\Big\} \\
&= \min\left\{1, \mathfrak{m}\left(A_{(k)}\right)\right\} \\
&= \mathfrak{m}\left(A_{(k)}\right)
\end{aligned}
$$

Otherwise, one has the following possibilites:

[2] An aggregation function $A = [0,1]^n \to [0,1]$ is an Ordered Modular Average (OMA) operator if it is commutative, idempotent, and comonotone modular.[31]

(i) *For all $k \in \{1, \ldots, n\}$, it holds that $x_{(k)} \leq \mathfrak{m}\left(A_{(k)}\right)$. In this case, one has that:*

$$\begin{aligned} g\mathfrak{C}_{\mathfrak{m}}^{(T_M,T_M)}(\vec{x}) &= \min\left\{1, \sum_{i=1}^{n} \begin{cases} x_{(i)} - x_{(i-1)} & \text{if } x_{(i)} \leq \mathfrak{m}\left(A_{(i)}\right) \\ \mathfrak{m}\left(A_{(i)}\right) - x_{(i-1)} & \text{if } x_{(i)} > \mathfrak{m}\left(A_{(i)}\right) \wedge x_{(i-1)} \leq \mathfrak{m}\left(A_{(i)}\right) \\ 0 & \text{otherwise.} \end{cases} \right\} \\ &= \min\left\{1, (x_{(1)} - x_{(0)}) + \ldots + (x_{(n)} - x_{(n-1)}\right) \\ &= \min\left\{1, x_{(n)}\right\} \\ &= x_{(n)} \end{aligned}$$

(ii) *For all $k \in \{1, \ldots, n\}$ such that $x_{(k)} > \mathfrak{m}\left(A_{(k)}\right)$ it holds that $x_{(k-1)} > \mathfrak{m}\left(A_{(k)}\right)$. In this case, one has that:*

$$\begin{aligned} g\mathfrak{C}_{\mathfrak{m}}^{(T_M,T_M)}(\vec{x}) &= \min\left\{1, \sum_{i=1}^{n} \begin{cases} x_{(i)} - x_{(i-1)} & \text{if } x_{(i)} \leq \mathfrak{m}\left(A_{(i)}\right) \\ \mathfrak{m}\left(A_{(i)}\right) - x_{(i-1)} & \text{if } x_{(i)} > \mathfrak{m}\left(A_{(i)}\right) \wedge x_{(i-1)} \leq \mathfrak{m}\left(A_{(i)}\right) \\ 0 & \text{otherwise.} \end{cases} \right\} \\ &= \min\Big\{1, (x_{(1)} - x_{(0)}) + (x_{(2)} - x_{(1)}) + \ldots + (x_{(k-1)} - x_{(k-2)}) \\ &\qquad + \underbrace{0 + \ldots + 0}_{n-k+1}\Big\} \\ &= \min\left\{1, x_{(k-1)}\right\} \\ &= x_{(k-1)}. \end{aligned}$$

Then, it follows that:

$$\begin{aligned} g\mathfrak{C}_{\mathfrak{m}}^{(T_M,T_M)}(\vec{x}) &= \begin{cases} \mathfrak{m}\left(A_{(k)}\right) & \text{if } \exists k \in \{1, \ldots, n\} : x_{(k)} > \mathfrak{m}\left(A_{(k)}\right) \wedge x_{(k-1)} \leq \mathfrak{m}\left(A_{(k)}\right) \\ x_{(n)} & \text{if } \forall k \in \{1, \ldots, n\} : x_{(k)} \leq \mathfrak{m}\left(A_{(k)}\right) \\ x_{(k-1)} & \text{if } \forall k \in \{1, \ldots, n\} : x_{(k)} > \mathfrak{m}\left(A_{(k)}\right) \wedge x_{(k-1)} > \mathfrak{m}\left(A_{(k)}\right) \end{cases} \\ &= \max_{i=1}^{n}\left\{\min\left\{x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right\}\right\} \\ &= S_{\mathfrak{m}}(\vec{x}), \end{aligned}$$

where $S_{\mathfrak{m}}$ is the Sugeno integral. Observe that C_{T_M,T_M}-integral is the $CMin$-integral analysed in [12].

6. $gC_{F_1F_2}$-integrals as OD monotone functions

In the previous section, we presented the requirements for $gC_{F_1F_2}$-integrals to be aggregation functions, showing that there exist pseudo pre-aggregation function pairs that do not fulfill such requirements, and, therefore, the corresponding $gC_{F_1F_2}$-integrals are not aggregation functions. However, under some constraints, $gC_{F_1F_2}$-integrals are OD increasing functions satisfying **(A2)**, presenting, thus, some desirable conditions to play the role of "aggregation operators" in applications (see, for example, [14]). In this section we prove such properties of $gC_{F_1F_2}$-integrals.

First, notice that, in order to study the directional increasingness feature of our integrals, it is necessary to compatibilize the dimension reduction process, which should be performed in both input $\vec{x} = (x_1, \ldots, x_n) \in [0,1]^n$ and direction vector $\vec{r} = (r_1, \ldots, r_n) \in \mathbb{R}^n$, $\vec{r} \neq \vec{0}$, reducing both vectors to the same dimension $k \leq n$. It is easy to see that this compatible dimension reduction is possible if it holds that:

$$\forall i, l \in \{1, \ldots, n\}, i < l : x_i = x_l \Rightarrow r_i = r_l \ \vee r_l = 0. \tag{10}$$

Example 6.1. *There are different vectors $\vec{r} \in \mathbb{R}^n$ that satisfy (10) for all $\vec{x} \in [0,1]^n$. For example, consider the vectors $(w, \ldots, w)$ and $(w, 0, \ldots, 0)$, with $w \neq 0$. However, the vector $(w, 0, 0, w', 0)$, with $w, w' \neq 0$ does not satisfy (10) for some $\vec{x} \in [0,1]^n$. Take, for example, $\vec{x} = (0.2, 0.3, 0.5, 0.5, 0.6)$. Observe that $x_3 = x_4 = 0.5$ but $r_3 \neq r_4$ and $r_4 = w' \neq 0$.*

It follows that:

Proposition 6.1. *Let $\mathbb{R}^n_{\vec{x}}$ be the set of non null vectors $\vec{r} \in \mathbb{R}^n$ satisfying (10), for a given $\vec{x} \in [0,1]^n$. Then, for each $\vec{x} \in [0,1]^n$, $\vec{r} \in \mathbb{R}^n_{\vec{x}}$ if and only if $\vec{r} = (w, 0, \ldots, 0) \in \mathbb{R}^n$ or $\vec{r} = (w, \ldots, w) \in \mathbb{R}^n$, with $w \neq 0$.*

PROOF. ($\Rightarrow$) Suppose that $\vec{r} \in \mathbb{R}^n_{\vec{x}}$, for all $\vec{x} = (x_1, \ldots, x_n) \in [0,1]^n$, and $\vec{r} = (w_1, \ldots, w_n)$, with $\vec{r} \neq \vec{0}$, such that there exist $i, j \in \{1, \ldots, n\}$ with $w_i \neq w_j$ and there exists $h \in \{2, \ldots, n\}$ with $w_h \neq 0$. Then, take $\vec{x} = (x_1, \ldots, x_n) \in [0,1]^n$ such that $x_i = x_j = x_h$. Since $x_i = x_j$ then, by (10), considering that $w_i \neq w_j$, it holds that $w_j = 0$. Now, since $x_j = x_h$, then, by (10), considering that $w_h \neq 0$, then $w_j = w_h$, which is a contradiction with $w_j = 0$. Then, one concludes that either $w_i = w_j$, for all $i, j \in \{1, \ldots, n\}$, or $w_h = 0$, for all $h \in \{2, \ldots, n\}$. ($\Leftarrow$) It is immediate. □

The dimension reduction of such direction vectors $\vec{r}$ can be done as follows:

Definition 6.1. *Let $R_{n \mapsto k} : [0,1]^n \to [0,1]^k$ be $(n \mapsto k)$-dimension reduction function, as defined in Equation (6). The direction $(n \mapsto k)$-dimension reduction function is performed by the function $S_{n \mapsto k} : \{(w, 0, \ldots, 0), (w, \ldots, w) \in \mathbb{R}^n \mid w \neq 0\} \to \{(w, 0, \ldots, 0), (w, \ldots, w) \in \mathbb{R}^k \mid w \neq 0\}$, defined by:*

$$S_{n \mapsto k}((w, 0, \ldots, 0)) = (w, 0, \ldots, 0) \tag{11}$$

$$S_{n \mapsto k}((w, \ldots, w)) = (w, \ldots, w). \tag{12}$$

Then we have the following results:

Lemma 6.1. *Consider $\vec{r} = (w, 0, \ldots, 0) \in \mathbb{R}^n$, $w \neq 0$. Let $R_{n \mapsto k}$ and $S_{n \mapsto k}$ be as defined in equations (6) and (11), respectively, and denote $S_{n \mapsto k}((w, 0, \ldots, 0)) = (w, 0, \ldots, 0) = (s_1, \ldots, s_n)$. Let $\sigma_K : \{1, \ldots, k\} \to \{1, \ldots, k\}$ be a permutation in decreasing order defined, for all $j \in K = \{1, \ldots, k\}$, as*

$$\sigma_K(j) = (k - j + 1), \tag{13}$$

(i.e., $\sigma_K(1) = (k)$, $\sigma_K(2) = (k-1)$, …, $\sigma_K(k) = (1)$). Then, for all $c > 0$ such that $y_{\sigma_K(1)} + cw \in [0,1]$, if

$$1 \geq y_{\sigma_K(1)} + cw > y_{\sigma_K(2)} > \ldots > y_{\sigma_K(k)}, \tag{14}$$

then, for any $\vec{z} = \vec{y} + c\vec{s}_{\sigma_K^{-1}}$, where $\vec{s}_{\sigma_K^{-1}} = (s_{\sigma_K^{-1}(1)}, \ldots, s_{\sigma_K^{-1}(k)})$, it holds that $z_{(j)} = y_j + cs_{k-j+1}$, that is, $z_{(k)} = y_k + cw$ and $z_{(j)} = y_j$, for all $j \in \{1, \ldots, k-1\}$.

PROOF. For all $\vec{x} \in [0,1]^n$ and respective $\vec{y} \in [0,1]^k$, since $y_1 < \ldots < y_k$, then $y_{\sigma_K(1)} = y_n > \ldots > y_{\sigma_K(k)} = y_1$. Considering $\vec{r} = (w, 0, \ldots, 0) \in \mathbb{R}^n$, with $w \neq 0$, and its respective $\vec{s} = (w, 0, \ldots, 0) \in \mathbb{R}^k$, suppose that, for all $c > 0$ the inequality (14) holds (i.e., $\vec{y}_{\sigma_K}$ and $\vec{y}_{\sigma_K} + c\vec{s}$ are comotone, and either they increase or decrease at the same time). Then, for any $\vec{z} = \vec{y} + c\vec{s}_{\sigma_K^{-1}}$, where $\vec{s}_{\sigma_K^{-1}} = (s_{\sigma_K^{-1}(1)}, \ldots, s_{\sigma_K^{-1}(k)})$, as the same as in Equation (1), it holds that $\vec{z}_{\sigma_K} = (\vec{y} + c\vec{s}_{\sigma_K^{-1}})_{\sigma_K} = \vec{y}_{\sigma_K} + c\vec{s}$, and, thus, by the inequality (14), it holds that

$$1 \geq z_{\sigma_K(1)} = y_{\sigma_K(1)} + cs_1 > \ldots > z_{\sigma_K(k)} = y_{\sigma_K(k)} + cs_k,$$

that is,

$$1 \geq z_{\sigma_K(1)} = y_{\sigma_K(1)} + cw > z_{\sigma_K(2)} = y_{\sigma_K(2)} > \ldots > z_{\sigma_K(k)} = y_{\sigma_K(k)}.$$

This means that $z_{\sigma_K(k)} = y_{\sigma_K(k)} + cw$ and, for all $j \in \{1, \ldots, k-1\}$, $z_{\sigma_K(j)} = y_{\sigma_K(j)}$. From Equation (13), it holds that:

$$z_{(k)} = z_{\sigma_K^{-1}\sigma_K(k)} = y_{\sigma_K^{-1}\sigma_K(k)} + cs_{\sigma_K^{-1}(k)} = y_{(k)} + cs_1 = y_k + cw$$

and, for all $j \in \{1, \ldots, k-1\}$,

$$z_{(j)} = z_{\sigma_K^{-1}\sigma_K(j)} = y_{\sigma_K^{-1}\sigma_K(j)} + cs_{\sigma_K^{-1}(j)} = y_{(j)} + cs_{k-j+1} = y_j + cs_{k-j+1} = y_j,$$

where $(\cdot) : \{1, \ldots, k\} \to \{1, \ldots, k\}$ is a permutation in an increasing order with $z_{(1)} < \ldots < z_{(k)}$.

Theorem 6.1. *Let $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure and $F_1, F_2 : [0,1]^2 \to [0,1]$ be fusion functions satisfying the conditions of Definition 4.2. Consider $\vec{r} = (w, 0, \ldots, 0) \in \mathbb{R}^n$, with $w > 0$. Then $g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}$ is OD $\vec{r}$-increasing.*

PROOF. Let $R_{n\mapsto k}$, B_j^R and $S_{n\mapsto k}$ be as defined in equations (6), (7) and (11), respectively. Let $\sigma_N : \{1,\ldots,n\} \to \{1,\ldots,n\}$ be any permutation such that, for all $\vec{x} \in [0,1]^n$, with

$$x_{\sigma_N(1)} \geq \ldots \geq x_{\sigma_N(k)}, \tag{15}$$

and for all $c > 0$, such that $1 \geq x_{\sigma_N(1)} + cw \geq x_{\sigma_N(2)} \geq \ldots \geq x_{\sigma_N(k)}$, where $\vec{r}_{\sigma_N^{-1}} = (r_{\sigma_N^{-1}(1)}, \ldots, r_{\sigma_N^{-1}(k)}) \in \mathbb{R}$. Clearly, one can consider the permutation in the decreasing order $\sigma_N : \{1,\ldots,n\} \to \{1,\ldots,n\}$ defined in terms of the permutation in the increasing order $(\cdot) : \{1,\ldots,n\} \to \{1,\ldots,n\}$ as $\sigma_N(1) = (n)$, $\sigma_N(2) = (n-1)$, ..., $\sigma_N(n) = (1)$, that is, $\sigma_N(j) = (n - j + 1)$, with $j \in \{1,\ldots,n\}$. Then, one has that $x_{(1)} \leq \ldots \leq x_{(n)}$, $x_{\sigma_N(1)} \geq \ldots \geq x_{\sigma_N(n)}$ and $\vec{r}_{\sigma_N^{-1}} = (0,\ldots,0,w)$.

Due to the dimension reduction, for each $\vec{x} \in [0,1]^n$, consider its respective $\vec{y} = (y_1,\ldots,y_k) \in [0,1]^k$ and $\vec{s} = (w,0,\ldots,0) \in \mathbb{R}^k$, obtained from $\vec{r}$. Let $\sigma_K : \{1,\ldots,k\} \to \{1,\ldots,k\}$ be a permutation defined as the restriction of σ_N to K, such that $\{x_{\sigma_N(1)},\ldots,x_{\sigma_N(n)}\} = \{y_{\sigma_K(1)},\ldots,y_{\sigma_K(k)}\}$ and $y_{\sigma_K(1)} > \ldots > y_{\sigma_K(k)}$. Observe that, after the dimension reduction, for any $\vec{y} \in [0,1]^k$ with respect to a $\vec{x} \in [0,1]^n$ satisfying (15), and, for all $c > 0$, it holds that $1 \geq y_{\sigma_K(1)} + cw > 1_{\sigma_K(2)} > \ldots > y_{\sigma_K(k)}$, with $\vec{s}_{\sigma_K^{-1}} = (s_{\sigma_K^{-1}(1)},\ldots,s_{\sigma_K^{-1}(k)}) = (0,\ldots,0,w) \in \mathbb{R}^k$.

Clearly, when considering σ_N defined in terms of the permutation in the increasing order $(\cdot)$, we have that $\sigma_K(1) = (k)$, $\sigma_K(2) = (k-1)$, ..., $\sigma_K(k) = (1)$, that is, $\sigma_K(j) = (k - j + 1)$, with $j \in \{1,\ldots,k\}$. Then, one has that $y_{(1)} = y_1 < \ldots < y_{(k)} = y_k$ and $y_{\sigma_K(1)} > \ldots > y_{\sigma_K(k)}$. Then, from Lemma 6.1, it follows that:

$$\begin{aligned}
g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x} + c\vec{r}_{\sigma_N^{-1}}) &= \min\Big\{1, F_1(y_k + cw, \mathfrak{m}(B_k^R(\vec{x}))) - F_2(y_{k-1}, \mathfrak{m}(B_k^R(\vec{x}))) \\
&\qquad + \sum_{j=1}^{k-1} F_1\left(y_j, \mathfrak{m}\left(\cup_{p=j}^k B_p^R(\vec{x})\right)\right) - F_2\left(y_{j-1}, \mathfrak{m}\left(\cup_{p=j}^k B_p^R(\vec{x})\right)\right)\Big\} \\
&\geq \min\Big\{1, F_1(y_k, \mathfrak{m}(B_k^R(\vec{x}))) - F_2(y_{k-1}, \mathfrak{m}(B_k^R(\vec{x}))) \\
&\qquad + \sum_{j=1}^{k-1} F_1\left(y_j, \mathfrak{m}\left(\cup_{p=j}^k B_p^R(\vec{x})\right)\right) - F_2\left(y_{j-1}, \mathfrak{m}\left(\cup_{p=j}^k B_p^R(\vec{x})\right)\right)\Big\} \\
&= g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}(\vec{x}),
\end{aligned}$$

since F_1 is $(1,0)$-increasing. Thus, $g\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}$ is OD $(w,0,\ldots,0)$-increasing, for $w > 0$.

Corollary 6.1. *Let $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure and (F_1, F_2) be a pseudo pre-aggregation function pair. Consider $\vec{r} = (w,0,\ldots,0) \in \mathbb{R}^n$, with $w > 0$. Then $\mathfrak{C}_{\mathfrak{m}}^{(F_1,F_2)}$ is an OD $\vec{r}$-increasing function satisfying the boundary conditions **(A2)**.*

PROOF. It follows from Proposition 5.1 and Theorem 6.1.

7. Conclusion

In this paper, we introduced the $gC_{F_1F_2}$-integrals, either aggregation or OD monotone functions based on pseudo pre-aggregation pairs for the generalization of the $C_{F_1F_2}$-integrals. We have stated under which conditions $C_{F_1F_2}$-integrals are (averaging) aggregation or OD pre-aggregation functions. In summary, the main features of $C_{F_1F_2}$-integrals in relation to our previous approaches related to the generalizations of the Choquet integral are:

1. The pseudo pre-aggregation pairs (F_1, F_2) used for building $gC_{F_1F_2}$-integrals satisfy a few number of constraint, less than, for example a pair of copulas (C, C) of the CC-integrals [9], and we still have an aggregation function or, at least, an OD monotone function satisfying boundary conditions;

2. The obtained aggregation or OD monotone function need not to be neither averaging nor idempotent to present excellent results in classification (see [14]).

In future work, we will study our generalizations in an interval-valued context, following the approach in [33, 34, 35].

Acknowledgment

This work is supported by CNPq (Proc. 306970/2013-9, 233950/2014-1, 307781/2016-0), the Spanish Ministry of Science and Technology (under project TIN2016-77356-P (AEI/FEDER, UE)), grant APVV-14-0013 and Caixa and Fundación Caja Navarra.

References

[1] G. Lucca, J. Sanz, G. Pereira Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, H. Bustince Sola, Pre-aggregation functions: construction and an application, IEEE Transactions on Fuzzy Systems 24 (2) (2016) 260–272. doi:10.1109/TFUZZ.2015.2453020.

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

[3] H. Bustince, J. A. Sanz, G. Lucca, G. P. Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, G. Ochoa, Pre-aggregation functions: Definition, properties and construction methods, in: Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on, IEEE, 2016, pp. 294–300.

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

[5] G. Choquet, Theory of capacities, Annales de l'Institut Fourier 5 (1953–1954) 131–295.

[6] H. Ishibuchi, T. Nakashima, M. Nii, Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining, Springer-Verlag, Berlin, 2004.

[7] E. P. Klement, R. Mesiar, E. Pap, Triangular Norms, Kluwer Academic Publisher, Dordrecht, 2000.

[8] G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, H. Bustince, R. Mesiar, Cf-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems, Information Sciences 435 (2018) 94 – 110. doi:https://doi.org/10.1016/j.ins.2017.12.029.

[9] G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, M. J. Asiain, M. Elkano, H. Bustince, CC-integrals: Choquet-like copula-based aggregation functions and its application in fuzzy rule-based classification systems, Knowledge-Based Systems 119 (2017) 32 – 43. doi:http://dx.doi.org/10.1016/j.knosys.2016.12.004.

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

[11] G. Lucca, J. A. Sanz, G. P. Dimuro, B. Bedregal, J. Fernández, H. Bustince, Analyzing the behavior of a cc-integral in a fuzzy rule-based classification system, in: 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE, Los Alamitos, 2017, pp. 1–6. doi:10.1109/FUZZ-IEEE.2017.8015579.

[12] G. P. Dimuro, G. Lucca, J. A. Sanz, H. Bustince, B. Bedregal, Cmin-integral: A choquet-like aggregation function based on the minimum t-norm for applications to fuzzy rule-based classification systems, in: V. Torra, R. Mesiar, B. D. Baets (Eds.), Aggregation Functions in Theory and in Practice, Springer International Publishing, Cham, 2018, pp. 83–95. doi:10.1007/978-3-319-59306-7_9.

[13] G. Lucca, G. P. Dimuro, V. Mattos, B. Bedregal, H. Bustince, J. A. Sanz, 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, 2015, pp. 1–8. doi:10.1109/FUZZ-IEEE.2015.7337911.

[14] G. Lucca, G. P. Dimuro, J. Fernandéz, H. Bustince, B. n Bedregal, J. A. Sanz, Improving the performance of fuzzy rule-based classification systems using a new non-averaging generalization of cc-integrals named $c_{F_1F_2}$-integrals(submitted).

[15] J. Hühn, E. Hüllermeier, Furia: an algorithm for unordered fuzzy rule induction, Data Mining and Knowledge Discovery 19 (3).

[16] H. Bustince, E. Barrenechea, M. Sesma-Sara, J. Lafuente, G. P. Dimuro, R. Mesiar, A. Kolesárová, Ordered directionally monotone functions. justification and application, IEEE Transactions on Fuzzy Systems (2017) 1(In press, corrected proof). doi:10.1109/TFUZZ.2017.2769486.

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

[18] B. C. Bedregal, G. P. Dimuro, H. Bustince, E. Barrenechea, New results on overlap and grouping functions, Information Sciences 249 (2013) 148–170.

[19] G. P. Dimuro, B. Bedregal, On residual implications derived from overlap functions, Information Sciences 312 (2015) 78 – 88.

[20] G. Beliakov, A. Pradera, T. Calvo, Aggregation Functions: A Guide for Practitioners, Springer, Berlin, 2007.

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

[22] G. Lucca, J. Sanz, G. Pereira Dimuro, B. Bedregal, R. Mesiar, A. Kolesárová, H. Bustince Sola, Pre-aggregation functions: construction and an application, IEEE Transactions on Fuzzy Systems 24 (2) (2016) 260–272. doi:10.1109/TFUZZ.2015.2453020.

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

[24] J. Qiao, B. Q. Hu, On multiplicative generators of overlap and grouping functions, Fuzzy Sets and Systems 332 (2018) 1 – 24.

[25] R. B. Nelsen, An introduction to copulas, Vol. 139 of Lecture Notes in Statistics, Springer, New York, 1999.

[26] G. P. Dimuro, B. Bedregal, Archimedean overlap functions: The ordinal sum and the cancellation, idempotency and limiting properties, Fuzzy Sets and Systems 252 (2014) 39 – 54.
[27] G. P. Dimuro, B. Bedregal, Additive generators of overlap functions, in: H. Bustince, J. Fernandez, R. Mesiar, T. Calvo (Eds.), Aggregation Functions in Theory and in Practice, Vol. 228 of Advances in Intelligent Systems and Computing, Springer, Berlin, 2013, pp. 167–178.
[28] G. P. Dimuro, B. Bedregal, H. Bustince, M. J. Asiáin, R. Mesiar, On additive generators of overlap functions, Fuzzy Sets and Systems 287 (2016) 76 – 96, theme: Aggregation Operations.
[29] M. Elkano, M. Galar, J. Sanz, A. Fernández, E. Barrenechea, F. Herrera, H. Bustince, 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) (2015) 1562–1580.
[30] R. Mesiar, A. Stupnanová, A note on cc-integral(to appear).
[31] R. Mesiar, A. Mesiarova-Zemánková, The ordered modular averages, IEEE Transactions on Fuzzy Systems 19 (1) (2011) 42–50.
[32] M. Sugeno, Theory of fuzzy integrals and its applications, Ph.D. thesis, Tokyo Institute of Technology (1974).
[33] G. P. Dimuro, B. R. C. Bedregal, R. H. S. Reiser, R. H. N. Santiago, Interval additive generators of interval t-norms, in: W. Hodges, R. de Queiroz (Eds.), Proceedings of the 15th International Workshop on Logic, Language, Information and Computation, WoLLIC 2008, Edinburgh, no. 5110 in Lecture Notes in Artificial Intelligence, Springer, Berlin, 2008, pp. 123–135.
[34] B. C. Bedregal, G. P. Dimuro, R. H. N. Santiago, R. H. S. Reiser, On interval fuzzy S-implications, Information Sciences 180 (8) (2010) 1373 – 1389.
[35] J. Qiao, B. Q. Hu, On interval additive generators of interval overlap functions and interval grouping functions, Fuzzy Sets and Systems 323 (2017) 19 – 55.

References

[AAFH07] Alcalá R., Alcalá-Fdez J., y 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.

[AFAH11] Alcalá-Fdez J., Alcalá R., y 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.

[AFHHP07] Alcala-Fdez J., Herrera F., Hernández F. A., y Peregrín A. (09 2007) Increasing fuzzy rules cooperation based on evolutionary adaptive inference systems22: 1035–1064.

[AFS06] Alsina C., Frank M. J., y Schweizer B. (2006) *Associative Functions: Triangular Norms and Copulas*. World Scientific Publishing Company, Singapore.

[AFSG+09] Alcalá-Fdez J., Sánchez L., García S., Jesus M., Ventura S., Garrell J., Otero J., Romero C., Bacardit J., Rivas V., Fernández J., y Herrera F. (2009) Keel: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing* 13(3): 307–318.

[Alp10] Alpaydin E. (2010) *Introduction to Machine Learning*. The MIT Press, 2nd edition.

[AS94] Agrawal R. y Srikant R. (1994) Fast algorithms for mining association rules in large databases. En *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB '94, páginas 487–499. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

[BBF+13] Barrenechea E., Bustince H., Fernandez J., Paternain D., y Sanz J. A. (2013) Using the Choquet integral in the fuzzy reasoning method of fuzzy rule-based classification systems. *Axioms* 2(2): 208–223.

[BBSS+17] Bustince H., Barrenechea E., Sesma-Sara M., Lafuente J., Dimuro G. P., Mesiar R., y Kolesárová A. (2017) Ordered directionally monotone functions. justification and application. *IEEE Transactions on Fuzzy Systems* página 1. (In press, corrected proof).

[BDBB13] Bedregal B. C., Dimuro G. P., Bustince H., y Barrenechea E. (2013) New results on overlap and grouping functions. *Information Sciences* 249: 148–170.

[BFKM15] Bustince H., Fernandez J., Kolesárová A., y Mesiar R. (2015) Directional monotonicity of fusion functions. *European Journal of Operational Research* 244(1): 300–308.

[BFM+10] Bustince H., Fernandez J., Mesiar R., Montero J., y Orduna R. (2010) Overlap functions. *Nonlinear Analysis: Theory, Methods & Applications* 72(3-4): 1488–1499.

[BFOS84] Breiman L., Friedman J., Olshen R., y Stone C. (1984) *Classification and Regression Trees*. Chapman and Hall (Wadsworth and Inc.).

[BPC07] Beliakov G., Pradera A., y Calvo T. (2007) *Aggregation Functions: A Guide for Practitioners*. Springer, Berlin.

[BPM+12] Bustince H., Pagola M., Mesiar R., Hüllermeier E., y Herrera F. (2012) Grouping, overlaps, and generalized bientropic functions for fuzzy modeling of pairwise comparisons. *IEEE Transactions on Fuzzy Systems* 20(3): 405–415.

[BSGR03] Barandela R., Sánchez J., Garci´a V., y Rangel E. (2003) Strategies for learning in class imbalance problems. *Pattern Recognition* 36(3): 849 – 851.

[BSL+16] Bustince H., Sanz J. A., Lucca G., Dimuro G. P., Bedregal B., Mesiar R., Kolesárová A., y Ochoa G. (2016) Pre-aggregation functions: Definition, properties and construction methods. En *Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on*, páginas 294–300. IEEE.

[BVGV+03] Baesens B., Van Gestel T., Viaene S., Stepanova M., Suykens J., y Vanthienen J. (Jun 2003) Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society* 54(6): 627–635.

[CABFO06] Crockett K., A. Bandar Z., Fowdar J., y O'Shea J. (05 2006) Genetic tuning of fuzzy inference within fuzzy classifier systems23.

[CBM07] Crockett K., Bandar Z., y Mclean D. (July 2007) On the optimization of t-norm parameters within fuzzy decision trees. En *2007 IEEE International Fuzzy Systems Conference*, páginas 1–6.

[CCdJH05] Casillas J., Cordon O., del Jesus M. J., y Herrera F. (Feb 2005) Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction. *IEEE Transactions on Fuzzy Systems* 13(1): 13–29.

[CCJH01] Casillas J., Cordón O., Jesus M. D., y Herrera F. (2001) Genetic feature selection in a fuzzy rule-based classification system learning process for high-dimensional problems. *Information Sciences* 136(1): 135 – 157. Recent Advances in Genetic Fuzzy Systems.

[CdJH98] Cordon O., del Jesus M. J., y Herrera F. (1998) Analyzing the reasoning mechanisms in fuzzy rule based classification systems. *Mathware and Soft Computing* 5(2-3): 321 – 332.

[CdJH99] Cordón O., del Jesus M. J., y Herrera F. (1999) A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning* 20(1): 21 – 45.

[CFCH01] Chapman W. W., Fizman M., Chapman B. E., y Haug P. J. (2001) A comparison of classification algorithms to automatically identify chest x-ray reports that support pneumonia. *Journal of Biomedical Informatics* 34(1): 4 – 14.

[CH01] Cordón O. y Herrera F. (2001) Hybridizing genetic algorithms with sharing scheme and evolution strategies for designing approximate fuzzy rule-based systems. *Fuzzy Sets and Systems* 118(2): 235 – 255.

[Cho54] Choquet G. (1953–1954) Theory of capacities. *Annales de l'Institut Fourier* 5: 131–295.

[CHV01] Cordon O., Herrera F., y Villar P. (Aug 2001) Generating the knowledge base of a fuzzy rule-based system by the genetic learning of the data base. *IEEE Transactions on Fuzzy Systems* 9(4): 667–674.

[Coh60] Cohen J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20(1): 37–46.

[CV95] Cortes C. y Vapnik V. (1995) Support vector networks. *Machine Learning* 20: 273–297.

[CYP96a] Chi Z., Yan H., y Pham T. (1996) *Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition*. World Scientific Publishing Co., Inc., River Edge, NJ, USA.

[CYP96b] Chi Z., Yan H., y Pham T. (1996) *Fuzzy Algorithms: With Applications to Image Processing and Pattern Recognition*. World Scientific, Singapore.

[DB13] Dimuro G. P. y Bedregal B. (2013) Additive generators of overlap functions. En Bustince H., Fernandez J., Mesiar R., y Calvo T. (Eds.) *Aggregation Functions in Theory and in Practice*, volumen 228 of *Advances in Intelligent Systems and Computing*, páginas 167–178. Springer, Berlin.

[DB14] Dimuro G. P. y Bedregal B. (2014) Archimedean overlap functions: The ordinal sum and the cancellation, idempotency and limiting properties. *Fuzzy Sets and Systems* 252: 39 – 54.

[DB15] Dimuro G. P. y Bedregal B. (2015) On residual implications derived from overlap functions. *Information Sciences* 312: 78 – 88.

[DBB+16a] Dimuro G. P., Bedregal B., Bustince H., Fernandez J., Lucca G., y Mesiar R. (2016) New results on pre-aggregation functions. En *Uncertainty Modelling in Knowledge Engineering and Decision Making, Proceedings of the 12th International FLINS Conference (FLINS 2016)*, volumen 10 of *World Scientific Proceedings Series on Computer Engineering and Information Science*, páginas 213–219. World Scientific, Singapura.

[DBB+16b] Dimuro G. P., Bedregal B., Bustince H., Asiáin M. J., y Mesiar R. (2016) On additive generators of overlap functions. *Fuzzy Sets and Systems* 287: 76 – 96. Theme: Aggregation Operations.

[DHS00] Duda R. O., Hart P. E., y Stork D. G. (2000) *Pattern Classification (2Nd Edition)*. Wiley-Interscience.

[DLS+18] Dimuro G. P., Lucca G., Sanz J. A., Bustince H., y Bedregal B. (2018) *CMin-Integral: A Choquet-Like Aggregation Function Based on the Minimum t-Norm for Applications to Fuzzy Rule-Based Classification Systems*, páginas 83–95. Springer International Publishing, Cham.

[EGS+15] Elkano M., Galar M., Sanz J., Fernández A., Barrenechea E., Herrera F., y 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.

[EGSB16] Elkano M., Galar M., Sanz J., y 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.

[ES93] Eshelman L. J. y Schaffer J. D. (1993) Real-coded genetic algorithms and interval-schemata. En WHITLEY L. D. (Ed.) *Foundations of Genetic Algorithms*, volumen 2 of *Foundations of Genetic Algorithms*, páginas 187 – 202. Elsevier.

[Esh91] Eshelman L. J. (1991) The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination. En Rawlings G. J. E. (Ed.) *Foundations of Genetic Algorithms*, páginas 265–283. Morgan Kaufmann, San Francisco.

[FGH11] Fernández A., García S., y Herrera F. (2011) Addressing the classification with imbalanced data: Open problems and new challenges on class distribution. En Corchado E., Kurzyński M., y Woźniak M. (Eds.) *Hybrid Artificial Intelligent Systems*, páginas 1–10. Springer Berlin Heidelberg, Berlin, Heidelberg.

[FIS36] FISHER R. A. (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics* 7(2): 179–188.

[GFB+12] Galar M., Fernandez A., Barrenechea E., Bustince H., y Herrera F. (July 2012) A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42(4): 463–484.

[GMMP09] Grabisch M., Marichal J., Mesiar R., y Pap E. (2009) *Aggregation Functions.* Cambridge University Press, Cambridge.

[GP98] Gonzalez A. y Perez R. (1998) Completeness and consistency conditions for learning fuzzy rules. *Fuzzy Sets and Systems* 96(1): 37 – 51.

[GPGOF07] García-Pedrajas N., García-Osorio C., y Fyfe C. (2007) Nonlinear boosting projections for ensemble construction. *Journal of Machine Learning Research* 8: 1–33.

[GS15] Goel N. y Sehgal P. (2015) Fuzzy classification of pre-harvest tomatoes for ripeness estimation - an approach based on automatic rule learning using decision tree. *Applied Soft Computing* 36: 45 – 56.

[GSP+14] Galar M., Sanz J., Pagola M., Bustince H., y Herrera F. (July 2014) A preliminary study on fingerprint classification using fuzzy rule-based classification systems. En *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, páginas 554–560.

[Her08] Herrera F. (Mar 2008) Genetic fuzzy systems: taxonomy, current research trends and prospects. *Evolutionary Intelligence* 1(1): 27–46.

[HH09] Hühn J. y Hüllermeier E. (2009) Furia: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery* 19(3).

[HL05] Huang J. y Ling C. X. (March 2005) Using auc and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering* 17(3): 299–310.

[HL09] Herrera F. y Lozano M. (2009) *Fuzzy Evolutionary Algorithms and Genetic Fuzzy Systems: A Positive Collaboration between Evolutionary Algorithms and Fuzzy Systems*, páginas 83–130. Springer Berlin Heidelberg, Berlin, Heidelberg.

[HLS03] Herrera F., Lozano M., y Sánchez A. M. (2003) A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study. *International Journal of Intelligent Systems* 18(3): 309–338.

[HM95] Homaifar A. y McCormick E. (May 1995) Simultaneous design of membership functions and rule sets for fuzzy controllers using genetic algorithms. *IEEE Transactions on Fuzzy Systems* 3(2): 129–139.

[IN01] Ishibuchi H. y Nakashima T. (2001) Effect of rule weights in fuzzy rule-based classification systems. *Fuzzy Systems, IEEE Transactions on* 9(4): 506–515.

[INN05] Ishibuchi H., Nakashima T., y Nii M. (2005) *Classification and Modeling with Linguistic Information Granules, Advanced Approaches to Linguistic Data Mining*. Advanced Information Processing. Springer, Berlin.

[INYT94] Ishibuchi H., Nozaki K., Yamamoto N., y Tanaka H. (1994) Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms. *Fuzzy Sets and Systems* 65(2-3): 237 – 253.

[JBP$^+$13] Jurio A., Bustince H., Pagola M., Pradera A., y Yager R. (2013) Some properties of overlap and grouping functions and their application to image thresholding. *Fuzzy Sets and Systems* 229: 69 – 90.

[Kar91] Karr C. (Febrero 1991) Genetic algorithms for fuzzy controllers. *AI Expert* 6(2): 26–33.

[KLJ03] Kavšek B., Lavrač N., y Jovanoski V. (2003) *Advances in Intelligent Data Analysis V: 5th International Symposium on Intelligent Data Analysis, IDA 2003, Berlin, Germany, August 28-30, 2003. Proceedings*, Capítulo: APRIORI-SD: Adapting Association Rule Learning to Subgroup Discovery, páginas 230–241. Springer Berlin Heidelberg, Berlin, Heidelberg.

[KMP00] Klement E. P., Mesiar R., y Pap E. (2000) *Triangular Norms*. Kluwer Academic Publisher, Dordrecht.

[Kun05] Kuncheva L. I. (2005) Diversity in multiple classifier systems. *Information Fusion* 6(1): 3 – 4. Diversity in Multiple Classifier Systems.

[LDB$^+$16] Lucca G., Dimuro G. P., Bedregal B., Bustince H., y Sanz J. A. (2016) A proposal for tuning the alpha parameter in a copula function applied in fuzzy rule-based classification systemspáginas 367–372.

[LDM+15] Lucca G., Dimuro G. P., Mattos V., Bedregal B., Bustince H., y Sanz J. A. (2015) A family of Choquet-based non-associative aggregation functions for application in fuzzy rule-based classification systems. En *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, páginas 1–8. IEEE, Los Alamitos.

[LSD+15] Lucca G., Sanz J. A., Dimuro G. P., Bedregal B., Mesiar R., Kolesárová A., y Bustince H. (2015) The notion of pre-aggregation function. En *Modeling Decisions for Artificial Intelligence*, páginas 33–41. Springer International Publishing.

[LSD+17a] Lucca G., Sanz J. A., Dimuro G. P., Bedregal B., Fernández J., y Bustince H. (July 2017) Analyzing the behavior of a cc-integral in a fuzzy rule-based classification system. En *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, páginas 1–6.

[LSD+17b] Lucca G., Sanz J. A., Dimuro G. P., Bedregal B., Asiain M. J., Elkano M., y 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.

[LSD+18a] Lucca G., Sanz J. A., Dimuro G. P., Bedregal B., Bustince H., y Mesiar R. (2018) Cf-integrals: A new family of pre-aggregation functions with application to fuzzy rule-based classification systems. *Information Sciences* 435: 94 – 110.

[LSD+18b] Lucca G., Sanz J., Dimuro G. P., Bedregal B., y Bustince H. (2018) *Analyzing the Behavior of Aggregation and Pre-aggregation Functions in Fuzzy Rule-Based Classification Systems with Data Complexity Measures*, páginas 443–455. Springer International Publishing, Cham.

[LSPD+16] Lucca G., Sanz J., Pereira Dimuro G., Bedregal B., Mesiar R., Kolesárová A., y Bustince Sola H. (April 2016) Pre-aggregation functions: construction and an application. *IEEE Transactions on Fuzzy Systems* 24(2): 260–272.

[Mam74] Mamdani E. H. (December 1974) Application of fuzzy algorithms for control of simple dynamic plant. *Electrical Engineers, Proceedings of the Institution of* 121(12): 1585–1588.

[Men42] Menger K. (1942) Statistical metrics. *Proceedings of the National Academy of Sciences* 28(12): 535–537.

[MSM94] Murofushi T., Sugeno M., y Machida M. (1994) Non-monotonic fuzzy measures and the Choquet integral. *Fuzzy Sets and Systems* 64(1): 73 – 86.

[Nel99] Nelsen R. B. (1999) *An introduction to copulas*, volumen 139 of *Lecture Notes in Statistics*. Springer, New York.

[OM99] Opitz D. y Maclin R. (12 1999) Popular ensemble methods: An empirical study11.

[PBP+16] Paternain D., Bustince H., Pagola M., Sussner P., Kolesárová A., y Mesiar R. (2016) Capacities and overlap indexes with an application in fuzzy rule-based classification systems. *Fuzzy Sets and Systems* 305: 70 – 94.

[Qui93] Quinlan J. (1993) *C4.5: Programs for Machine Learning*. Morgan Kauffman.

[Sal15] Salahinejad M. (2015) Application of classification models to identify solvents for single-walled carbon nanotubes dispersion. *RSC Adv.* 5: 22391–22398.

[Sam75] Sambuc R. (1975) *Fonctions ϕ-floues. Application l'aide au diagnostic en pathologie thyroidienne*. PhD thesis, Université Marseille, Marseille.

[SBH+15] Sanz J. A., Bernardo D., Herrera F., Bustince H., y Hagras H. (Aug 2015) 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* 23(4): 973–990.

[SD11] Sanz Delgado J. A. (2011) *Sistemas de clasificación basados en reglas difusas lingüísticas utilianco conjunto difusos intervalorados y ajuste de la ignorancia*. PhD thesis, Pamplona.

[SFB+16] Sanz J. A., Fernandes A. M., Barrenechea E., Silva S., Santos V., Gonçalves N., Paternain D., Jurio A., y Melo-Pinto P. (2016) Lamb muscle discrimination using hyperspectral imaging: Comparison of various machine learning algorithms. *Journal of Food Engineering* 174(Supplement C): 92 – 100.

[SFBH13] Sanz J., Fernández A., Bustince H., y 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.

[SH09] Sedighiani K. y HashemiKhabir S. (May 2009) Constructing interpretable genetic fuzzy rule-based system for breast cancer diagnostic. En *2009 WRI Global Congress on Intelligent Systems*, volumen 1, páginas 441–446.

[SS11] Schweizer B. y Sklar A. (2011) *Probabilistic Metric Spaces*. Dover Publications.

[Sug74] Sugeno M. (1974) *Theory of Fuzzy Integrals and its Applications*. PhD thesis.

[Thr91] Thrift P. (1991) Fuzzy logic synthesis with genetic algorithms. En Belew R. y Booker L. (Eds.) *Proceedings of the Fourth International Conference on Genetic Algorithms*, páginas 509–513. Morgan Kaufman, San Mateo, CA.

[Top08] Topouzelis K. N. (2008) Oil spill detection by sar images: Dark formation detection, feature extraction and classification algorithms. *Sensors* 8(10): 6642–6659.

[TS85] Takagi T. y Sugeno M. (Jan 1985) Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics* SMC-15(1): 116–132.

[TSK05] Tan P.-N., Steinbach M., y Kumar V. (2005) *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

[Tun09] Tung A. K. H. (2009) *Rule-based Classification*, páginas 2459–2462. Springer US, Boston, MA.

[Uno11] Unold O. (2011) Diagnosis of cardiac arrhythmia using fuzzy immune approach. En Dobnikar A., Lotrič U., y Šter B. (Eds.) *Adaptive and Natural Computing Algorithms*, páginas 265–274. Springer Berlin Heidelberg, Berlin, Heidelberg.

[VRT+15] Vitoriano B., Rodríguez J. T., Tirado G., Martín-Campo F. J., Ortuño M. T., y Montero J. (2015) Intelligent decision-making models for disaster management. *Human and Ecological Risk Assessment: An International Journal* 21(5): 1341–1360.

[YK97] Yager R. R. y Kacprzyk J. (Eds.) (1997) *The Ordered Weighted Averaging Operators: Theory and Applications*. Kluwer Academic Publishers, Norwell, MA, USA.

[YKB11] Yager R. R., Kacprzyk J., y Beliakov G. (2011) *Recent Developments in the Ordered Weighted Averaging Operators: Theory and Practice*. Springer Publishing Company, Incorporated, 1st edition.

[Zad65] Zadeh L. A. (1965) Fuzzy sets. *Information and Control* 8(3): 338–353.

[Zad75] Zadeh L. A. (1975) The concept of a linguistic variable and its application to approximate reasoning - I. *Information Sciences* 8(3): 199–249.

Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa