Analyzing the performance of different fuzzy measures with generalizations of the Choquet integral in classification problems

Giancarlo Lucca
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
Email: giancarlo.lucca@furg.com

José Antonio Sanz
Dpt. de Automática y Computación, ISC
Universidad Publica de Navarra
Pamplona, Spain
Email: joseantonio.sanz@unavarra.es

Graçaliz P. Dimuro
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
Email: gracalizdimuro@furg.br

Eduardo Nunes Borges
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
Email: eduardoborges@furg.br

Helida Santos
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
Email:helida@furg.br

Humberto Bustince
Dpt. de Automática y Computación, ISC
Universidad Publica de Navarra
Pamplona, Spain
Email: bustince@unavarra.es

***Abstract*—Fuzzy Rule Based Classification Systems are an useful tool to deal with classification problems. In these systems, one fundamental point is the manner of how the available information about the problem is aggregated. The mechanism responsible to perform the aggregation is the Fuzzy Reasoning Method (FRM). Recently, some FRMs using generalizations of the Choquet integral to perform the aggregation were proposed in the literature. Since these generalizations are defined considering a specific fuzzy measure, in this paper we apply different fuzzy measures in the generalizations that presented the best performance in each study in the literature. We analyze how the performance is affected according to each fuzzy measure.**

I. INTRODUCTION

An effective way to deal with classification problems [1], [2] is by using Fuzzy Rule-Based Classification Systems (FRBCSs) [3]. An important component of a FRBCS is the Fuzzy Reasoning Method (FRM) [4], since it is responsible for classifying new examples. A key point in the FRM is the way in which the information given by the fired fuzzy rules is aggregated. The aggregation function is the component that differs the FRMs, e.g., the Winning Rule (WR), which applies the maximum, the Additive Combination (AC), which applies the normalized addition, and the arithmetic mean [5].

In [6], it was introduced a FRM considering the Choquet integral [7] as the aggregation function. Since the Choquet integral is defined with a fuzzy measure [7], [8], the authors provided an experimental study considering different fuzzy measures, namely, the Cardinality (Uniform), Dirac's, Weighted mean (Wmean), Ordered Weighted Averaging (OWA) and Power measures. The Power measure, which achieved the best performance, is defined em terms of an exponent q, which was adapted for each class by using a genetic algorithm.

After that, Lucca et al. [9] proposed the concept of pre-aggregation functions, whose one of the construction methods is based on a generalization of the Choquet integral by t-norms [10] (i.e., the C_T-integrals). In the experimental study, six different generalizations were used, combined with the same measures used in [6]. Again, the best results were achieved by the Power measure with the exponent q learnt genetically.

Lucca et al. [11] generalized C_T-integrals by using a fusion function F satisfying some requirements, obtaining the C_F-integrals. The Choquet integral in its expanded form was generalized by copulas [12], resulting in the CC-integrals [13]. CC-integrals were generalized by a pair of fusion functions F_1 and F_2 under some constraints, obtaining the $C_{F_1F_2}$-integrals [14] and the more general $gC_{F_1F_2}$-integrals [15]. In all such works only the Power measure was adopted.

Having this in mind, we have the following question: "Was the Power measure the best choice to be used in the generalizations of the Choquet integral (C_T-integrals, CC-integrals, C_F-integrals and $C_{F_1F_2}$-integrals)?. Then, the aim of this paper is to analyse the performance of the systems when applying different fuzzy measures. For that, we consider the generalizations with best performance in the works [9], [11], [13], [14] and all the fuzzy measures used in [6]. The experimental study was conducted considering 33 different datasets, which are available in KEEL[1] dataset repository [16].

The paper is organized as follows. In Sect. II, we present some necessary basic concepts. In Sect. III, we present the usage of the adopted generalizations of the Choquet integral in FRMs. Section IV introduces the experimental framework. The obtained results are in Sect. V. Section VI is the Conclusion.

II. PRELIMINARIES

In this section, we introduce the necessary background to develop the paper. Consider $N = \{1, \ldots, n\}$.

[1]https://www.keel.es

Definition 1. *[17] A function $A : [0,1]^n \rightarrow [0,1]$ is an aggregation function if: **(A1)** A is increasing in each argument: $\forall i \in \{1,\ldots,n\} : x_i \leq y \Rightarrow A(x_1,\ldots,x_n) \leq A(x_1,\ldots,x_{i-1},y,x_{i+1},\ldots,x_n)$; **(A2)** A satisfies the boundary conditions: **(i)** $A(0,\ldots,0)=0$; **(ii)** $A(1,\ldots,1)=1$.*

Examples of aggregation functions are the t-norms $T : [0,1]^2 \rightarrow [0,1]$, which are commutative and associative functions with right neutral element (i.e., $\forall x \in [0,1] : T(x,1)=x$) [10], overlap functions, which are continuous positive t-norms [18] and the copulas $C : [0,1]^2 \rightarrow [0,1]$, which have 0 as absorbent element, 1 as neutral element and, for all $x, x', y, y' \in [0,1]$ with $x \leq x'$ and $y \leq y'$, $C(x,y)+C(x',y') \geq C(x,y')+C(x',y)$ [12].

Definition 2. *[19] Let $\vec{r} = (r_1,\ldots,r_n)$ be a real n-dimensional vector, $\vec{r} \neq \vec{0}$. A function $PA : [0,1]^n \rightarrow [0,1]$ is directionally increasing with respect to $\vec{r}$ ($\vec{r}$-increasing, for short) if for all $(x_1,\ldots,x_n) \in [0,1]^n$ and $c > 0$ such that $(x_1+cr_1,\ldots,x_n+cr_n) \in [0,1]^n$ it holds that*

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

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

Definition 3. *Let $\vec{r} = (r_1,\ldots,r_n)$ be a real n-dimensional vector, $\vec{r} \neq \vec{0}$. A function $PA : [0,1]^n \rightarrow [0,1]$ is said to be an n-ary $\vec{r}$-pre-aggregation function if: **(PA1)** F is $\vec{r}$-increasing; **(PA2)** F satisfies thr boundary conditions: **(A2) (i)** and **(ii)**.*

A. Fuzzy Measures

A fundamental point of this paper is the fuzzy measure. In the context of aggregation functions, fuzzy measures are used to evaluate the relationship among the elements to be aggregated, representig the importance of a coalition.

Definition 4. *A function $\mathfrak{m} : 2^N \rightarrow [0,1]$ 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 $\mathfrak{m}(N)=1$.*

The fuzzy measures considered in this paper are the same that were used in [6], and are the following, for $A \subseteq N$:

- Cardinality of uniform measure:

$$\mathfrak{m}(A) = \frac{|A|}{n} \tag{2}$$

- Dirac's measure: For a previously fixed $i \in N$,

$$\mathfrak{m}(A) = \begin{cases} 1 & \text{if } i \in A \\ 0 & \text{if } i \notin A. \end{cases} \tag{3}$$

- Weighted mean (Wmean): Take an arbitrary vector of weights $(w_1,\ldots,w_n) \in [0,1]^n$ such that $\sum_{i=1}^n w_i = 1$ and consider the the fuzzy measure: $m(\{1\}) = w_1,\ldots,m(\{n\}) = w_n$. For $|A| > 1$, one defines:

$$\mathfrak{m}(A) = \sum_{i \in A} m(\{i\}) \tag{4}$$

- Ordered Weighted Averaging (OWA): 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 considers an OWA operator. For $|A| > 1$, one defines:

$$\mathfrak{m}(A) = \sum_{i \in A} m(\{i\}) \tag{5}$$

- Power Measure (PM):

$$\mathfrak{m}(A) = \left(\frac{|A|}{n}\right)^q, \text{ with } q > 0. \tag{6}$$

B. The Choquet integral

In what follows, consider a fuzzy measure $\mathfrak{m} : 2^N \rightarrow [0,1]$, $A \subseteq N$ and an increasing permutation on the input $\vec{x} \in [0,1]^n$, namely, $(x_{(1)},\ldots,x_{(n)})$, i.e., $0 \leq x_{(1)} \leq \ldots \leq x_{(n)}$, 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 $\vec{x}$.

The Choquet integral is a special type of aggregation function since it takes into consideration the importance of groups of criteria, offering flexibility for modeling aggregations.

Definition 5. *Let $\mathfrak{m} : 2^N \rightarrow [0,1]$ be a fuzzy measure. The discrete Choquet integral (CI) is the function $\mathfrak{C}_\mathfrak{m} : [0,1]^n \rightarrow [0,1]$, defined, for all of $\vec{x} = (x_1,\ldots,x_n) \in [0,1]^n$, by:*

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

Using the distributivity property of the product, Eq. (7) can be also written in what we call the expanded form (Eq. 8):

$$\mathfrak{C}_\mathfrak{m}(\vec{x}) = \sum_{i=1}^{n} \left(x_{(i)} \cdot \mathfrak{m}\left(A_{(i)}\right) - x_{(i-1)} \cdot \mathfrak{m}\left(A_{(i)}\right)\right). \tag{8}$$

C. Generalizations of the Choquet integral

The CI has been generalized to perform different aggregation-like operators, for application in the FRM of a fuzzy classifier. The first generalization is by t-norms, resulting in a family of pre-aggregation functions C_T-integrals [9]:

Definition 6. *[9] Let $T : [0,1]^2 \rightarrow [0,1]$ be a t-norm. The C_T-integral is the function $\mathfrak{C}_\mathfrak{m}^{\mathbb{T}} : [0,1]^n \rightarrow [0,n]$, defined by*

$$\mathfrak{C}_\mathfrak{m}^{T}(\vec{x}) = \sum_{i=1}^{n} T\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right). \tag{9}$$

The second generalization is related to the expanded form (Eq. (8)) of the CI, generalized by copulas [12], resulting in aggregation functions called CC-integrals [13], [20].

Definition 7. *[13] Let $C : [0,1]^2 \rightarrow [0,1]$ be a copula. The CC-integral is the function $\mathfrak{C}_\mathfrak{m}^{C} : [0,1]^n \rightarrow [0,1]$, given by*

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

Considering functions F satisfying some specific conditions, the resultant of generalizing the CI (Eq. (7)) by such Fs are pre-aggregation functions called C_F-integrals [11].

Definition 8. *[11] Let $F : [0,1]^2 \rightarrow [0,1]$ be a funtion satisfying: (i) $\forall y \in [0,1] : F(0,y) = 0$; (ii) $\forall x \in [0,1] : F(x,1) =$*

x. The C_F-integral is the function $\mathfrak{C}^F_{\mathfrak{m}} : [0,1]^n \to [0,1]$, defined by

$$\mathfrak{C}^F_{\mathfrak{m}}(\vec{x}) = \min\left\{1, \sum_{i=1}^{n} F\left(x_{(i)} - x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\}. \quad (11)$$

Finally, the expanded form (Eq. (8)) of the CI was generalized by a pair of functions under some constraints, resulting non-averaging ordered directionally monotone functions called $C_{F_1F_2}$-integrals [14].

Definition 9. *[14] Let $\mathfrak{m}$ be a symmetric fuzzy measure and $F_1, F_2 : [0,1]^2 \to [0,1]$ satisfying: (i) $\forall x, y \in [0,1] : F_1(x,y) \geq F_2(x,y)$; (ii) F_1 is $(1,0)$-increasing. A $C_{F_1F_2}$-integral is defined as a function $\mathfrak{C}^{(F_1,F_2)}_{\mathfrak{m}} : [0,1]^n \to [0,1]$, given by*

$$\mathfrak{C}^{(F_1,F_2)}_{\mathfrak{m}}(\vec{x}) = \quad (12)$$
$$\min\left\{1, x_{(1)} + \sum_{i=2}^{n} F_1\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - F_2\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right)\right\}.$$

For the considered generalizations, we present in Table I the functions that achieved the best results in each family. In this table, $x = (x_{(i)} - x_{(i-1)})$ represents the difference of the elements to be aggregated and $y = \mathfrak{m}(A_{(i)})$ is related to the fuzzy measure. Furthermore, since the C_F-integrals may be either averaging (AVG) (i.e., the results are limited by the maximum and the minimum of the elements to be aggregated) or non-averaging (N–AVG), in this table, we selected the best functions of each case. We also highlight that the $C_{F_1F_2}$-integral considered in this study is a non-averaging function.

TABLE I: Generalizations of the CI considered in this study

Generalization	Function	Equation
C_T-integrals	Hamacher t-norm	$T_{HP}(x,y) = \begin{cases} 0 & \text{if } x = y = 0 \\ \frac{xy}{x+y-xy} & \text{otherwise} \end{cases}$
CC-integrals	Copula of the minimum	$C_{\min}(x,y) = \min\{x,y\}$
C_F-integrals$_{AVG}$	$(0,1)$-pre-aggregation F_{NA}	$F_{NA}(x,y) = \begin{cases} x & \text{if } x \leq y \\ \min\{\frac{x}{2}, y\} & \text{otherwise} \end{cases}$
C_F-integrals$_{N-AVG}$	$(0,1)$-pre-aggregation F_{NA2}	$F_{NA2}(x,y) = \begin{cases} 0 & \text{if } x = 0 \\ \frac{x+y}{2} & \text{if } 0 < x \leq y \\ \min\{\frac{x}{2}, y\} & \text{otherwise.} \end{cases}$
$C_{F_1F_2}$-integral	$F_1 =$ overlap function GM $F_2 =$ aggregation function F_{BPC}	$GM(x,y) = \sqrt{x,y}$ $F_{BPC}(x,y) = xy^2$

III. Applying generalizations of the CI in the FRM

In this section, we introduce the application of generalizations of the CI to cope with classification problems [1], [2]. So, consider classification as a problem composed by 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 i-th attribute and $y_p \in \mathbb{C} = \{C_1, \ldots, C_M\}$ is the label of the class of the p-th training example, where M is the number of classes.

To tackle this kind of problems we consider the usage of FRBCSs. Specifically, we have into account the same fuzzy classifier used in the original generalizations, that is, the Fuzzy Association Rule-based Classification method for High-Dimensional problems (FARC-HD) [21]. The fuzzy rules used by this algorithm have the following form:

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

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 [22], which, in this case, is computed using the certainty factor.

The generalizations of the CI were applied in the FRM of this classifier to obtain the information associated with each class of the problem. Precisely, 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:

$$class = \arg \max_{k=\{1,\ldots,M\}} (\mathbb{A}_k(\mu_{A_j}(x_p) * RW_j \mid Class(R_j)))$$
$$\text{with, } j = 1, ..., L,$$

where $\mathbb{A}$ is the considered generalization of the CI used to aggregate the information for a class k (See Table I), μ_{A_j} is the matching degree of the example x_p with the antecedent of the j-th fuzzy rule, RW_j is its rule weight and L is the number of fuzzy rules in the system.

We must stress out that is at this point that we combine the considered generalizations with different fuzzy measures. Regarding the Power Measure, that used a parameter q, we adopted the methodology that was proposed in [6], which used a genetic algorithm to adapt this parameter for each class.

IV. Experimental Framework

In this section, we present the datasets considered in the study and the configuration of the proposal.

A. Considered datasets

In this paper, to analyze the performance of our proposal, we consider 33 different datasets selected from the KEEL dataset repository [16]. The properties of the selected datasets are summarized in Table II, 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.

For each dataset, we have performed a 5-fold cross-validation technique, that is, the dataset is splitted into five random partitions. 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. The average result of the five testing partitions is the output of the algorithm.

TABLE II: Summary of the properties of the datasets considered in this study.

Id.	Dataset	#Ex.	#Atts.	#Class	Id.	Dataset	#Ex.	#Atts.	#Class
App	Appendiciticis	106	7	2	Pen	Penbased	10,992	16	10
Bal	Balance	625	4	3	Pho	Phoneme	5,404	5	2
Ban	Banana	5300	2	2	Pim	Pima	768	8	2
Bnd	Bands	365	19	2	Rin	Ring	740	20	2
Bup	Bupa	345	6	2	Sah	Saheart	462	9	2
Cle	Cleveland	297	13	5	Sat	Satimage	6,435	36	7
Con	Contraceptive	1473	9	3	Seg	Segment	2,310	19	7
Eco	Ecoli	336	7	8	Shu	Shuttle	58,000	9	7
Gla	Glass	214	9	6	Son	Sonar	208	60	2
Hab	Haberman	306	3	2	Spe	Spectfheart	267	44	2
Hay	Hayes-Roth	160	4	3	Tit	Titanic	2,201	3	2
Ion	Ionosphere	351	33	2	Two	Twonorm	740	20	2
Iri	Iris	150	4	3	Veh	Vehicle	846	18	4
Led	led7digit	500	7	10	Win	Wine	178	13	3
Mag	Magic	1,902	10	2	Wis	Wisconsin	683	11	2
New	Newthyroid	215	5	3	Yea	Yeast	1,484	8	10
Pag	Pageblocks	5,472	10	5					

B. *Configuration of the proposal*

In this subsection we present the configuration of methods used in this paper. We stress out that they are the same used in the original generalizations of the CI.

We have used the set up suggested by the authors of FARC-HD, which is as follows:

- Conjunction operator: Product t-norm.
- Number of linguistic labels per variable: 5 (modeled by triangular shaped membership functions).
- Minimum support: 0.05.
- Threshold for the confidence: 0.8.
- Maximum depth of the search tree: 3.

With respect to the parameters of the genetic algorithm, we consider:

- Number of individuals in the population: 50.
- Number of evaluations: 20.000.
- Bits for gene in the gray codification: 30.

V. Experimental Results

In this section, we present and discuss the results obtained by the standard CI and its generalizations (See Table I), combined with five different fuzzy measures presented in subsection II-A.

The achieved results are shown in Table III, where the rows are related to the accuracy obtained in the considered dataset and the columns are related with the aggregations (the standard CI (Choquet), CC-min, C_T-integrals, C_F-integrals, which is divided in averaging (AVG) and non-averaging (N-AVG), and $C_{F_1F_2}$-integrals, taking into account different fuzzy measures, namely, Cardinality (Card), Dirac, Weighted Mean (Wmean), Ordered Weighted Average (OWA) and the Power Measure (PM). In this table, we highlight in **boldface** the largest accuracy mean obtained among the considered generalization and the five measures. On the other hand, the smaller result is <u>underlined</u>. Finally, we emphasise with * the best global result, that is, the highest accuracy obtained for each dataset.

In a first and general analysis, considering the obtained mean, it is noticeable that PM is the fuzzy measure that achieved the highest accuracy mean in all considered generalizations (the best method is the $C_{F_1F_2}$-integral using the PM fuzzy measure). This affirmation acts jointly with the results presented in [6] and it reinforces the idea that the usage of this fuzzy measure in the remainder generalizations was a good choice. We believe that the reason of this fuzzy measure performing so well with any generalization, is due to the fact that it is genetically learnt for each class of the problem. The lowest accuracy mean was obtained with the OWA (four times) and the Dirac (two times) fuzzy measures, respectively. Another interesting result is that the obtained mean with the Dirac fuzzy measure are quite similar, around 78%. This behavior also appeared in [6] and we believe that is due to the fact that this fuzzy measure only return 0's and 1's.

An important point that can be observed is related to the usage of non-averaging generalizations. In all combinations of functions and fuzzy measures, the results obtained by the non-averaging generalizations are superior than the averaging ones. In fact, it is noticeable that the accuracy mean obtained by the $C_{F_1F_2}$-integrals is the largest one in the study.

In order to give a statistical support of the previous observations, we have conducted some hypothesis validation techniques [23]. We have used the aligned Friedman rank test [24] to detect statistical differences among a group of results. Furthermore, we also have applied the post-hoc Holm's test [25] in order to find the method that is able to reject the null hypothesis with respect to the best method obtained with the align Friedman rank test. Finally, we have computed the adjusted p-value (APV), which considers that multiple comparisons are made.

In Table IV, we present the rankings obtained in the aligned Friedman, where we have compared for each generalization (shown in columns) all the fuzzy measures (rows). For each generalization the fuzzy measure obtaining the lowest rank is considered the best one and consequently is used as control method for the Holm's test. We present between brackets the APV computed by the Holm's test. We <u>underline</u> the APV if there are differences in favor to the control variable, considering a confidence level of 90%.

TABLE IV: Statistical results using the align Friedman rank test and Holm's post hoc test.

	Choquet	CC-min	C_T-integrals	C_F-integrals$_{AVG}$	C_F-integrals$_{N-AVG}$	$C_{F_1F_2}$-integrals
PM	61.34 (-)	54.48 (-)	57.75 (-)	56.66 (-)	60.87 (-)	50.39 (-)
Card	62.57 (0.91)	75.36 (<u>0.07</u>)	68.28 (0.56)	76.62 (0.17)	90.54 (<u>0.03</u>)	71.59 (<u>0.07</u>)
Wmean	80.27 (0.21)	79.71 (<u>0.06</u>)	70.34 (0.56)	75.37 (0.17)	88.92 (<u>0.03</u>)	75.43 (<u>0.06</u>)
Dirac	93.59 (<u>0.01</u>)	98.53 (<u>0.00</u>)	114.34 (<u>0.00</u>)	111.10 (<u>0.00</u>)	79.09 (0.12)	104.87 (<u>0.00</u>)
OWA	117.21 (<u>0.00</u>)	106.90 (<u>0.00</u>)	104.25 (<u>0.00</u>)	95.22 (<u>0.00</u>)	95.56 (<u>0.01</u>)	112.69 (<u>0.00</u>)

For all generalizations, PM is the fuzzy measure considered as control method. Moreover, it presents statistical differences when compared to Dirac and OWA (except when the generalization is the C_F-integrals$_{N-AVG}$). When using both the $C_{F_1F_2}$-integral and the CC-min, PM presents differences in relation to the remainder fuzzy measures. A similar behavior happens with the C_F-integrals$_{N-AVG}$ method as the APVs are low. In the tow remainder generalization PM is not statistically different form Card and Wmean.

In order to present a more complete analysis, in Table V, we have performed another Friedman rank test, comparing the method used as the control variables in Table IV.

TABLE III: Results achieved in test by considering different fuzzy measures.

Dataset	Choquet Card	Choquet Dirac	Choquet Wmean	Choquet OWA	Choquet PM	CC-min Card	CC-min Dirac	CC-min Wmean	CC-min OWA	CC-min PM	C_T-integrals Card	C_T-integrals Dirac	C_T-integrals Wmean	C_T-integrals OWA	C_T-integrals PM
App	**86.80***	80.17	82.08	83.03	80.13	**85.89**	80.17	83.98	83.03	85.84	**85.89**	80.17	85.84	84.85	82.99
Bal	78.24	78.24	78.08	78.88	**82.40**	77.60	78.24	78.56	76.16	**81.60**	80.96	78.24	81.12	80.80	**82.72**
Ban	84.45	84.09	83.85	84.55	**86.32***	83.43	84.09	83.85	83.42	**84.30**	84.19	84.09	84.47	83.23	**85.96**
Bnd	**69.34**	65.97	69.10	67.38	68.56	65.79	65.97	70.22	66.63	**71.06**	69.96	65.97	67.99	68.56	**72.13***
Bup	64.35	64.06	64.35	62.90	**66.96**	63.77	64.06	59.71	**65.22**	61.45	65.80	64.06	65.51	**66.67**	65.80
Cle	**57.57**	55.56	57.56	53.20	55.58	**57.58**	55.56	56.24	56.59	54.88	56.90	55.56	**57.92**	56.21	55.58
Con	49.83	50.24	50.10	50.64	**51.26**	51.93	50.24	52.07	50.44	**52.61**	50.64	50.24	51.26	50.51	**53.09**
Eco	78.28	77.70	**79.46**	76.20	76.51	**79.17**	77.70	77.39	74.10	77.09	79.17	77.70	76.49	74.12	**80.07**
Gla	**65.90**	64.98	63.54	63.09	64.02	65.44	64.98	64.04	66.82	**69.17***	64.47	64.98	64.02	**67.74**	63.10
Hab	**74.50***	71.23	72.54	73.19	72.52	73.84	71.23	73.52	71.23	**74.17**	**72.87**	71.23	72.21	71.57	72.21
Hay	**81.00**	78.69	77.98	78.75	79.49	79.49	78.69	80.26	78.01	**81.74***	**79.49**	78.69	**79.49**	**79.49**	**79.49**
Ion	87.76	88.61	**90.89***	86.61	90.04	88.33	88.61	87.47	**89.47**	88.89	89.18	88.61	**90.04**	87.19	89.18
Iri	**94.00**	93.33	93.33	92.00	91.33	92.00	**93.33**	92.67	91.33	92.67	**93.33**	**93.33**	**93.33**	92.00	93.33
Led	68.20	68.00	**68.40**	67.60	68.20	**68.80**	68.00	67.80	67.40	68.40	69.00	68.20	**69.40**	68.40	68.60
Mag	79.02	77.86	**80.55**	79.49	78.86	79.02	77.86	79.02	79.44	**79.81**	**80.65**	77.86	80.02	80.13	79.76
New	94.42	93.02	93.95	91.63	**94.88**	93.49	93.02	**95.35**	92.56	93.95	94.88	93.02	94.42	91.16	**95.35**
Pag	94.16	**94.52**	94.34	94.34	94.16	94.34	94.52	**94.88**	94.52	93.97	94.34	**94.52**	94.34	94.34	94.34
Pen	89.45	89.91	88.18	85.00	**90.55**	89.91	89.91	90.36	86.73	**91.27**	90.18	89.91	90.55	87.36	**90.82**
Pho	**83.14**	82.33	82.51	81.98	82.98	82.98	82.33	**83.09**	82.72	82.94	83.33*	82.33	82.25	82.72	**83.83**
Pim	73.44	73.04	73.44	72.39	**73.95**	**76.43**	73.04	73.44	74.08	74.21	**76.30**	73.04	75.12	73.69	74.87
Rin	85.81	84.59	85.68	81.89	**90.95**	85.95	84.59	86.49	84.59	**87.97**	87.43	84.59	88.78	86.49	**88.78**
Sah	69.47	68.82	67.74	67.09	**69.69**	70.12	68.82	69.48	**70.98**	70.78	69.68	68.82	**71.21**	69.25	70.77
Sat	77.92	78.85	79.00	78.84	**79.47**	79.16	78.85	**80.56***	78.07	79.01	79.47	78.85	79.78	79.00	**80.40**
Seg	92.60	91.04	92.12	92.21	**93.46***	92.47	91.04	**92.81**	92.03	92.25	93.25	91.04	92.86	92.03	**93.33**
Shu	97.93	**98.57***	97.70	97.10	97.61	97.70	**98.57***	98.30	96.41	98.16	97.06	**98.57**	96.78	96.92	97.20
Son	76.48	76.48	74.07	75.01	**77.43**	76.46	76.48	74.53	73.59	**76.95**	78.90	76.48	76.49	73.59	**79.34**
Spe	79.01	77.88	**79.73**	77.11	77.88	75.24	77.88	73.38	**79.39**	78.99	75.26	77.88	77.53	**79.00**	76.02
Tit	78.87	**79.06***	78.87	78.87	78.87	78.87	**79.06***	78.87	78.87	78.87	78.87	**79.06**	78.87	78.87	78.87
Two	80.54	82.30	82.03	81.35	**84.46**	84.59	82.30	82.03	83.51	**85.14**	82.70	82.30	85.41	**88.78**	85.27
Veh	67.96	64.66	67.73	66.55	**68.44**	69.27	64.66	66.19	65.60	**69.86**	69.03	64.66	**69.86**	66.07	68.20
Win	94.37	**96.06**	94.40	93.25	93.79	94.37	**96.06**	95.51	93.75	93.83	95.51	96.06	93.81	94.37	**96.63**
Wis	95.90	95.90	95.76	95.17	**97.22**	95.61	95.90	**96.05**	94.29	95.90	95.76	95.90	**97.07**	96.05	96.78
Yea	56.94	**57.21**	56.00	56.47	55.73	56.87	57.21	57.08	**57.28**	57.01	56.00	**57.21**	56.00	56.13	56.53
Mean	79.02	78.27	78.64	77.69	**79.20**	78.97	78.27	78.64	78.13	**79.54**	79.41	78.28	79.40	78.71	**79.74**

Dataset	C_F-integrals$_{AVG}$ Card	C_F-integrals$_{AVG}$ Dirac	C_F-integrals$_{AVG}$ Wmean	C_F-integrals$_{AVG}$ OWA	C_F-integrals$_{AVG}$ PM	C_F-integrals$_{N-AVG}$ Card	C_F-integrals$_{N-AVG}$ Dirac	C_F-integrals$_{N-AVG}$ Wmean	C_F-integrals$_{N-AVG}$ OWA	C_F-integrals$_{N-AVG}$ PM	$C_{F_1F_2}$-integrals Card	$C_{F_1F_2}$-integrals Dirac	$C_{F_1F_2}$-integrals Wmean	$C_{F_1F_2}$-integrals OWA	$C_{F_1F_2}$-integrals PM
App	84,03	80,17	82,12	**84,85**	82,99	83,94	84,85	**85,89**	83,07	85,84	84,89	84,85	84,94	82,16	**86,80**
Bal	**83,20**	78,24	80,32	83,20	82,56	88,00	87,68	88,00	88,32	**88,64**	88,48	85,12	87,52	87,68	**89,12***
Ban	83,92	84,09	83,92	82,64	**86,09**	83,23	84,13	82,42	83,21	**84,60**	84,04	83,26	84,60	82,75	**84,79**
Bnd	64,96	65,97	66,90	67,13	**69,40**	69,64	68,32	68,81	70,25	**70,48**	69,35	69,12	69,15	69,41	**71,30**
Bup	63,48	64,06	63,48	62,61	**67,83***	62,61	64,93	62,90	**65,22**	64,64	65,80	63,77	62,61	60,00	**66,96**
Cle	54,89	55,56	57,58	56,57	**57,92**	56,88	**58,90***	55,54	57,60	56,55	**57,88**	56,89	55,54	55,90	56,22
Con	51,86	50,24	**53,43**	51,26	52,27	53,77	54,51	54,18	**54,72**	53,16	53,50	54,38	52,82	53,02	**54,72***
Ecp	78,29	77,70	77,40	74,42	**78,88**	80,37	80,68	80,98	**82,15**	80,08	**83,64***	83,06	82,74	80,97	81,86
Gla	64,50	64,98	**65,88**	65,43	64,51	65,94	**66,83**	65,91	66,36	**66,83**	62,64	64,99	67,32	63,57	**68,25**
Hab	71,88	71,23	73,19	71,23	**73,51**	71,55	**71,88**	71,55	70,58	71,87	71,54	71,89	69,93	71,21	**72,53**
Hay	78,69	78,69	78,69	**79,49**	78,72	77,95	77,21	**80,23**	79,52	79,43	75,73	**78,72**	77,98	**78,72**	78,66
Ion	90,03	88,61	90,04	89,76	**90,60**	88,32	**90,32**	89,75	89,47	89,75	**90,32**	89,46	**90,32**	89,75	88,33
Iri	**94,00**	93,33	**94,00**	93,33	93,33	**95,33***	94,00	94,00	94,00	94,00	94,00	**94,00**	94,00	92,67	94,00
Led	68,00	68,00	**69,20**	68,40	68,60	**70,00**	**70,00**	69,80	68,80	69,80	**70,60***	68,20	70,40	69,40	70,00
Mag	**80,91***	77,86	79,81	79,28	80,02	79,23	79,39	78,97	78,81	**79,70**	80,18	79,65	80,65	79,23	**80,86**
New	**93,49**	93,02	92,09	92,09	93,49	**96,74**	95,81	95,81	93,95	96,28	96,74	93,49	**97,21***	93,02	96,74
Pag	**94,52**	94,52	94,52	94,52	93,97	**95,07**	94,89	94,52	94,70	94,15	95,07	94,71	94,52	94,52	**95,25***
Pen	90,00	89,91	91,27	88,45	**91,45**	91,73	92,18	91,55	91,00	**92,91***	92,82	92,00	92,82	90,27	**92,91***
Pho	**83,44**	82,33	82,68	82,94	82,86	81,25	**81,92**	81,00	81,09	81,44	81,64	**81,75**	81,44	81,31	81,42
Pim	75,26	73,04	75,38	74,74	**75,64**	74,74	75,51	74,60	**76,43***	75,52	74,87	74,09	73,96	**75,38**	**75,38**
Rin	89,59	84,59	88,78	86,76	**90,27**	87,43	88,11	**90,27**	86,89	89,86	91,49	87,84	90,68	87,57	**91,89***
Sah	68,39	68,82	67,97	**70,77**	68,61	69,68	69,04	**70,34**	69,90	70,12	69,91	**71,85***	70,55	71,64	71,43
Sat	79,47	78,85	**79,63**	76,68	78,54	80,09	79,47	79,32	78,39	**80,41**	**79,94**	78,38	79,63	78,38	79,47
Seg	**92,60**	91,04	92,38		92,55	92,34	92,12	91,90	92,08	**92,42**	92,42	92,81	93,07	92,25	**93,29**
Shu	97,10	**98,57***	96,83	96,55	96,78	96,32	96,97	96,92	96,32	**97,15**	95,91	**97,93**	95,86	96,64	96,83
Son	77,89	76,48	**79,84**	75,01	78,85	**83,68**	83,66	82,71	81,25	83,21	**85,57***	78,87	81,75	81,29	85,15
Spe	76,38	77,88	77,13	**78,64**	78,26	79,76	**80,13**	78,28	79,39	79,77	77,90	77,50	**81,27***	79,02	79,39
Tit	78,87	**79,06***	78,87	78,87	78,87	78,87	78,87	78,87	78,87	78,87	78,87	78,87	78,87	78,87	78,87
Two	82,97	82,30	**85,81**	85,68	83,92	91,08	90,81	90,54	90,95	**92,57***	91,62	86,89	**92,43**	90,81	92,30
Veh	**68,20**	64,66	66,79	67,85	67,97	68,91	69,51	**70,10***	67,97	68,08	**69,98**	67,61	69,03	69,74	68,20
Win	95,49	**96,06**	93,81	92,11	96,03	95,49	92,11	96,05	95,51	**96,08**	95,51	91,56	**97,17***	90,41	95,48
Wis	**96,34**	95,90	95,75	95,32	96,34	96,63	96,05	**96,92**	96,34	96,78	95,90	96,63	96,63	96,34	**96,78**
Yea	57,48	57,21	**57,75**	57,07	56,40	**58,83**	56,80	58,22	57,89	57,08	57,35	56,27	**59,10***	58,22	58,56
Mean	79,10	78,27	79,19	78,66	**79,64**	80,16	80,23	80,21	80,03	**80,55**	80,49	79,59	80,50	79,46	**81,02***

TABLE V: Statistical results using the align Friedman rank test and Holm's post hoc test only in the considered control variables of Table IV.

	PM
$C_{F_1F_2}$-integrals	59.93 (-)
C_F-integrals$_{N-AVG}$	76.10 (0.25)
C_T-integrals	104.81 (<u>0.00</u>)
C_F-integrals$_{AVG}$	113.28 (<u>0.00</u>)
CC-min	115.56 (<u>0.00</u>)
Choquet	127.28 (<u>0.00</u>)

We can notice that the $C_{F_1F_2}$-integral is considered as control variable, without statistical differences against the non-averaging C_F-integral . We stress out that both methods are non-averaging. Whereas the remainder methods, which are averaging, are statistically improved by the $C_{F_1F_2}$-integral. This affirmation goes in favor of the previous results, since the $C_{F_1F_2}$-integral are even equivalent to state-of-the-art fuzzy classifiers. Finally, considering that the non-averaging approaches achieve the lowest rank among the best generalizations, and they obtain the largest accuracy mean, we reinforce the idea that using aggregations functions that are not bounded by the maximum is a good option to deal with classification problems.

VI. Conclusion

In this paper we combined different generalizations of the CI that are present in the literature, with different fuzzy measures, by considering 33 different datasets.

From the obtained results, we can conclude that the Power Measure (PM) is the measure that achieves the best performance, independently of the considered aggregation. The Cardinality and Wmean fuzzy measures also presented good results in some cases. On the other hand, we have empirically proved that the Dirac and OWA, in general, presented the worst performance. We also showed that the non-averaging generalization of the CI obtained a superior accuracy mean for any fuzzy measure, which reinforces previous results about the usage of averaging and non-averaging approaches in the FRM.

Acknowledgment

Supported by CNPq (Proc. 307781/2016-0, 305882/2016-3), FAPERGS (Proc. 17/2551-0000872-3) and the Spanish Ministry of Science and Technology (project TIN2016-77356-P (AEI/FEDER, UE)).

References

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

[2] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000.

[3] H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and Modeling with Linguistic Information Granules, Advanced Approaches to Linguistic Data Mining*, ser. Advanced Information Processing. Berlin: Springer, 2005.

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

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

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

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

[8] T. Murofushi, M. Sugeno, and M. Machida, "Non-monotonic fuzzy measures and the Choquet integral," *Fuzzy Sets and Systems*, vol. 64, no. 1, pp. 73 – 86, 1994.

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

[10] E. P. Klement, R. Mesiar, and E. Pap, *Triangular Norms*. Dordrecht: Kluwer Academic Publisher, 2000.

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

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

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

[14] G. Lucca, G. P. Dimuro, J. Fernandez, H. Bustince, B. Bedregal, and J. A. Sanz, "Improving the performance of fuzzy rule-based classification systems based on a nonaveraging generalization of CC-integrals named $C_{F_1F_2}$-integrals," *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 1, pp. 124–134, Jan 2019.

[15] G. P. Dimuro, G. Lucca, B. Bedregal, R. Mesiar, J. A. Sanz, C.-T. Lin, and H. Bustince, "Generalized $C_{F_1F_2}$-integrals: from Choquet-like aggregation to ordered directionally monotone functions," 2018, (submitted).

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

[17] G. Beliakov, A. Pradera, and T. Calvo, *Aggregation Functions: A Guide for Practitioners*. Berlin: Springer, 2007.

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

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

[20] G. Lucca, G. P. Dimuro, V. Mattos, B. Bedregal, H. Bustince, and 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)*. Los Alamitos: IEEE, 2015, pp. 1–8.

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

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

[23] S. García, A. Fernández, J. Luengo, and F. Herrera, "A study of statistical techniques and performance measures for genetics–based machine learning: Accuracy and interpretability," *Soft Computing*, vol. 13, no. 10, pp. 959–977, 2009.

[24] J. L. Hodges and E. L. Lehmann, "Ranks methods for combination of independent experiments in analysis of variance," *Annals of Mathematical Statistics*, vol. 33, pp. 482–497, 1962.

[25] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian Journal of Statistics*, vol. 6, pp. 65–70, 1979.