A proposal for tuning the α parameter in a copula function applied in fuzzy rule-based classification systems

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

Graçaliz P. Dimuro
Centro de Ciências Computacionais
Universidade Federal do Rio Grande
Rio Grande, Brazil
Email: gracaliz@gmail.com

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

José Antonio Sanz
and Humberto Bustince
Dpt. de Automática y Computación,
Institute of Smart Cities
Universidad Publica de Navarra
Pamplona, Spain
Email: {joseantonio.sanz, bustince}@unavarra.es

***Abstract*—In this paper, we use the concept of extended Choquet integral generalized by a copula function, as proposed by Lucca et al.. More precisely, the copula considered in their study uses a variable α, with different fixed values for testing its behavior. In this contribution we propose a modification of this method assigning a value to this α parameter using a genetic algorithm in order to find the value that best fits it for each class. Specifically, this new proposal is applied in the Fuzzy Reasoning Method (FRM) of Fuzzy Rule-Based Classification Systems (FRBCSs). Finally, we compare the results provided by our new approach against the best solution proposed by Lucca et al. (that uses an fixed value for the variable α). From the obtained results it can be concluded that the genetic learning of the α parameter is statistically superior than the fixed one. Therefore, we demonstrate that our genetic method can be used as an alternative for this function.**

I. INTRODUCTION

Fuzzy Rule-Based Classification Systems (FRBCSs) [1] are a powerful tool and widely used in classifications problems. Since this kind of systems have a good classification rate and provide interpretable models by using linguist labels in their antecedent rules, FRBCSs have been applied in severals real world problems, including industry [2], health [3], economy [4] and many others.

Lucca et al. [5] proposed the concept of extended Choquet integral, by distributing the fuzzy measure in this function and generalizing this extended function by a copula function, which 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 [6], therefore, introducing a family of Choquet-based non-associative aggregation functions.

These functions were applied in the Fuzzy Reasoning Method (FRM) of FRBCSs, in order to improve the quality of the system. This copula function is directly related to an α variable, so, they tested in this function 5 different fixed values for this variable, showing that the copula function using an $\alpha = 0.1$, present a good classification rate in the study.

In this paper, based on the idea presented by Lucca et al. [5], we proposed a new way to deal with the usage of this variable α, using a genetic algorithm to learn the most suitable value of this alpha parameter for each class. However, instead of using the Chi et al. [7] algorithm to accomplish the fuzzy rule learning process, we consider the usage of the FARC-HD [8], since it is one of the most accurate fuzzy classifier nowadays [8], [9], [10], [11].

In order to demonstrate the quality of our approach, we have selected 30 datasets that are available in KEEL[1] database repository [12]. We analyze the behavior of our genetic method with respect to the best value for the alpha parameter presented by Lucca et al. in [5]. Finally, our conclusions are supported by the Wilcoxon signed-rank test [13].

The paper is organized as follows. Section II presents some preliminary concepts that are necessary to develop the paper. In section III the proposed FRM is presented, along with the approach to use the genetic learned value for the parameter α. We explain the experimental framework, the results achieved in test by the application of the generalized Choquet integral in FRBCSs and the analyze of these results in Section IV. The main conclusions are drawn in Section V.

---

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

II. PRELIMINARIES

This section aims at introducing the background necessary to understand the paper. One important class of fuzzy operators are the *aggregation operators* [14], [15].

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

(A1) A is increasing[2] in each argument: for each $i \in \{1,\ldots,n\}$, if $x_i \leq y$, then $A(x_1,\ldots,x_n) \leq A(x_1,\ldots,x_{i-1},y,x_{i+1},\ldots,x_n)$;

(A2) A satisfies the Boundary conditions: $A(0,\ldots,0) = 0$ and $A(1,\ldots,1) = 1$.

Definition 2: A bivariate function $C : [0,1]^2 \to [0,1]$ is a copula [6] 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) $C(x,1) = C(1,x) = x$.

An example of copula functions are the family of non-associative bivariate functions $C_\alpha : [0,1]^2 \to [0,1]$, defined, for any $\alpha \in [-1,0[\cup]0,1]$, by:

$$C_\alpha(x,y) = xy(1+\alpha(1-x)(1-y)). \qquad (1)$$

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

Definition 3: A function $\mathfrak{m} : 2^N \to [0,1]$ is a discrete fuzzy measure [16], [17] if, for all $X, Y \subseteq N$, it satisfies 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 $\mathfrak{m}(N) = 1$.

In this paper, we consider the power measure $\mathfrak{m}_{Card\_GA} : 2^N \to [0,1]$, defined, for $X \subseteq N$, by:

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

where the exponent q is learned genetically.[3] We consider the discrete Choquet integral, related to discrete fuzzy measures:

Definition 4: [14, Def. 1.74] Let $\mathfrak{m} : 2^N \to [0,1]$ be a discrete fuzzy measure. The discrete Choquet integral for $\mathfrak{m}$ is defined as a function $\mathfrak{C}_{\mathfrak{m}} : [0,1]^n \to [0,1]$, given by

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

where $(x_{(i)},\ldots,x_{(n)})$ is a 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.

Observe that the Eq. (3) can be also written as:

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

which we call the Choquet Integral in its expanded form.

The Choquet integral combines the inputs in such a way that not only the importance of individual inputs or of their magnitude are taken into account, but also the importance of their groups (or coalitions in which it takes part), allowing to assign importance to all possible groups of criteria. [19].

Definition 5: Let $\mathfrak{m} : 2^N \to [0,1]$ be a fuzzy measure and $C_\alpha : [0,1]^2 \to [0,1]$ be the family of copula functions defined in Eq. (1), for $\alpha \in [-1,0[\cup]0,1]$. The family of discrete C_α-Choquet integrals with respect to $\mathfrak{m}$ is defined as the function $\mathfrak{C}_{\mathfrak{m}}^{C_\alpha} : [0,1]^n \to [0,1]$, given, for all $x \in [0,1]^n$, by

$$\mathfrak{C}_{\mathfrak{m}}^{C_\alpha}(x) = \sum_{i=1}^{n} C_\alpha\left(x_{(i)}, \mathfrak{m}\left(A_{(i)}\right)\right) - C_\alpha\left(x_{(i-1)}, \mathfrak{m}\left(A_{(i)}\right)\right), \qquad (5)$$

where $(x_{(i)},\ldots,x_{(n)})$ is a increasing permutation on the input x and $A_{(i)} = \{(i),\ldots,(n)\}$ is the subset of indices of $n-i+1$ largest components of x.

Then, substituting Eq. (1) in Eq. (5), one has that the definition of the family of discrete C_α-Choquet integrals with respect to $\mathfrak{m}$ can be also given by:

$$\begin{aligned}\mathfrak{C}_{\mathfrak{m}}^{C_\alpha}(x) \;=\; & \sum_{i=1}^{n}(x_i\mathfrak{m}(A_i)(1+\alpha(1-x_i)(1-\mathfrak{m}(A_i))) \qquad (6)\\ & -x_{i-1}\mathfrak{m}(A_i)(1+\alpha(1-x_{i-1})(1-\mathfrak{m}(A_i)))).\end{aligned}$$

For more informations see [5].

III. A PROPOSAL OF FUZZY REASONING METHOD USING THE CHOQUET INTEGRAL GENERALIZED BY COPULAS AND GENETIC ALGORITHMS

This section is aimed at introducing our proposal. Before than, we start by describing the main concepts of FRBCSs, the FRM using the generalized Choquet integral and finally, the genetic approach for the α variable. For the following consider that, a classification problem consists of m training examples $\mathrm{x}_p = (x_{p1},\ldots,x_{pn},y_p)$, with $p = 1,\ldots,m$, where x_{pi}, with $i = 1,\ldots,n$, is the value of the ith variable and $y_p \in \mathbb{Y} = \{Y_1,\ldots,Y_M\}$ is the label of the class of the pth training example. Among the different approaches presented in the literature, we use the FRBCSs. These kind of system have two main components:

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

$$\text{Rule } R_j: \text{ If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \text{ then } x \text{ in } Y_j^k \text{ with } RW_j,$$

where R_j is the label of the jth rule, A_{ji} is an antecedent fuzzy set modeling a linguistic term, Y_j^k is the label of the consequent fuzzy set Y^k modeling the class associated to the rule R_j, $k \in \{1,\ldots,M\}$, $RW_j \in [0,1]$ is the rule weight [20].

2) *Fuzzy Reasoning Method*: used to classify examples using the information available in the knowledge base. We must stress out that we consider the same FRM used in [5].

---

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

[3] The choice for this fuzzy measure was based on the results obtained by Barrenechea et al. [18], who introduced an evolutionary algorithm to define the most suitable exponential cardinality definition to be used for each class.

The Fuzzy Association Rule-based Classification model for High Dimensional problems (FARC-HD) is the fuzzy classifier considered in this paper. As mentioned before, we consider the usage of this classifier instead of the Chi et al. algorithm to accomplish the task of the fuzzy rule learning process, since it provides a better performance. The main components of this classifier are the following ones:

- Fuzzy association rule extraction for classification: This step is aimed to obtain the fuzzy rule base, to do so, a search tree [21] is constructed for each class. The confidence and support is calculated for each class considering each item (the item is a linguistic label). The fuzzy rules are generated by the most frequent itemset and the number of linguistic terms for each rule is limited to the depth of the tree.
- Candidate rule prescreening: This step considers a weighting scheme [22] to preselect the better generated rules.
- Genetic rule selection and lateral tuning: This stage uses an evolutionary algorithm to perform the lateral tuning of the fuzzy sets [23] and select the best rules that were generated in the previous steps.

A. A fuzzy reasoning method using the Choquet integral generalized by copulas

We present in this subsection the combination of the extended Choquet integral by copulas and the FRM of FRBCSs. For the following consider that $x_p = (x_{p1}, \ldots, x_{pn})$ 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 FRM used in this paper is the same used in [5] and consist in four steps:

1) *Matching degree:* it 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 $T : [0,1]^2 \to [0,1]$, for $j = 1, \ldots, L$:

$$\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})). \quad (7)$$

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

$$b_j^k(x_p) = \mu_{A_j}(x_p) \cdot RW_j^k, \quad (8)$$

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

3) *Example classification soundness degree for all classes:* in this step, we use the proposed family of C_α-Choquet integrals (Eq. (6)), with an α genetically learnt in $\in [-1, 0[\ \cup \ ]0, 1]$ and the power measure $\mathfrak{m}_{Card\_GA}$ (Eq. (2)), in order to combine the positive association degrees calculated in the previous step, obtaining the classification soundness degrees:

$$S_k(x_p) = \mathfrak{C}^{C_\alpha}_{\mathfrak{m}_{Card\_GA}}\left(b_1^k(x_p), \ldots, b_L^k(x_p)\right), \quad (9)$$

with $k = 1, \ldots, M$. Since, whenever $b_i^k(x_p) = 0$, for all $\alpha \in [-1, 0[\cup]0, 1]$, it holds that:

$$\begin{aligned} &\mathfrak{C}^{C_\alpha}_{\mathfrak{m}_{Card\_GA}}(b_1^k(x_p), \ldots, b_L^k(x_p)) = \\ &\quad \mathfrak{C}^{C_\alpha}_{\mathfrak{m}_{Card\_GA}}(b_1^k(x_p), \\ &\quad \ldots, b_{j-1}^k(x_p), b_{j+1}^k(x_p), \ldots, b_L^k(x_p)), \end{aligned}$$

for practical reasons, only $b_j^k > 0$ are considered in Eq. (9).

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

$$F(S_1, \ldots, S_M) = arg \max_{k=1,\ldots,M}(S_k). \quad (10)$$

B. A Genetic Approach to learn the α parameter for each class

In this paper, we propose the usage of a value for the variable α which is genetically learnt. Therefore, this section is aimed to explain the main features of the genetic process used to learn it.

Barrenechea et al., proposed in [18], the power measure, a measure that raises the standard cardinality to an exponent q using an genetic algorithm (CHC [24]) to adapt this exponent for each class (Eq. (2)). Therefore, the main idea in this paper is to use the same model, but also applying this method to learn the α parameter for each class of the problem. That is, for each class of the problem, we will learn the most suitable value that model the interactions among the fired rules and the considered class. The configuration of the evolutionary model as presented in the following:

1) Coding Scheme: The chromosome is divided in three main parts. The first one considers the genes related to the the tuning of lateral position of the membership functions [23] and have as many genes as the number of linguistic labels considered in the data set, moreover, having a range in [0, 1]; The second part has one gene per number of class and it is used to learn the q exponent related to the power measure [18], having a range in [0, 2]; The third part consists in the genes used to the α parameter, also using one gene for each class and having a range [-1, 1]. As the defined in the function (Eq. (1)) this alpha value must be a value in the range $\in [-1, 0[\cup]0, 1]$, therefore, if the selected value is 0, we assign "0.1" to α parameter, since it is the best solution achieved by Lucca et al. [5].
2) Chromosome Evaluation: The fitness function considered in this paper is the standard accuracy.
3) Initial Gene Pool: Our population is composed by 50 individuals, and the initial pool is obtained with the first individual having all genes with "0.5" to perform the tuning of the lateral position of the membership functions, "1.0" (which is the value for the cardinality measure [18]) to the q exponent and "0.1" to the α

variable. The remainder of the individuals are randomly initialized in the ranks of each gene.

4) Crossover Operator: We consider the usage of the Parent Centric BLX (PCBLX) crossover operator [25]. Two parents are crossed if their hamming distance divide by 2 is superior than a predetermined threshold L. Moreover, we consider the usage of the Gray Code where each gene is converted to binary code with a fixed number of bits for each gene (BITSGENE)[4]

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

where #Genes is the total length of the chromosome.

5) Restarting Approach: Aiming to increase the convergence of the algorithm, when the threshold value is smaller than zero, we consider the usage of the best chromosome (elitist scheme) and reset all the population considering the range appropriate for each gene. If there is no increment in the best solution, after three restarts the process the process is finished.

IV. EXPERIMENTAL FRAMEWORK AND RESULTS

In this section, firstly we present the 30 real world classification problems selected from the KEEL dataset repository. Furthermore, in the subsection IV-B, we present the achieved results in test by the FRM generalized by our method, along with an analysis of these obtained results.

A. Experimental Framework

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

As proposed in [18], [9], [27], we adopt the 5-fold cross-validation model, in other words, a dataset is splitted in five random partitions, where each partition have 20% of the examples, and a combination of four of them is used for training and the remainder one is used for testing. This process is repeated five times by using a different partition to test the created system each time. In order to measure the quality of each partition, the accuracy rate is calculated, that is, we divide the number of correctly classified examples divided by the total number of examples for each partition. Then, as the final result of the algorithm we consider the average of the achieved accuracy in this five partitions.

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

- Conjunction operator: product t-norm
- Rule weight: Certain factor
- Five linguistic labels per class
- Minimum support: 0.05

TABLE I
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

- Confidence limit: 0.8
- Depth of the tree: 3

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

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

B. Analysis of the results provided by the genetic learning of the α parameter

This subsection present the results achieved in test by the FRM generalized by the copula function C_α genetically learnt ($C_{\alpha Gen}$) along with the best fixed alpha value presented in [5], that is, $C_{\alpha=0.1}$. The results achieved in testing by these approaches are presented in Table II by columns along with the standard deviation, also the best result achieved among the different datasets is highlighted in **boldface**.

From the results shown in the Table II, it is possible to observe that our approach using an parameter α, which is genetically learnt, presents a good performance, since it achieves

---

[4] For more information see [26], [18].

TABLE II
RESULTS IN TESTING PROVIDED BY BOTH, THE COPULA THAT USES A FIXED VALUE AND THE ONE WHICH USES THE GENETIC APPROACH.

Dataset	$C_{\alpha=0.1}$	$C_{\alpha Gen}$
appendicitis	80.26 ± 8.21	**82.08** ± 6.19
balance	81.76 ± 1.54	**83.04** ± 1.73
banana	85.98 ± 1.15	**86.02** ± 2.30
bupa	64.93 ± 4.96	**66.96** ± 3.30
cleveland	52.86 ± 2.27	**57.90** ± 1.47
ecoli	**77.39** ± 5.97	76.80 ± 6.39
glass	63.55 ± 1.97	**65.43** ± 3.74
haberman	**73.18** ± 4.25	72.85 ± 4.85
hayes-roth	**80.23** ± 7.04	**80.23** ± 7.04
iris	93.33 ± 4.71	**94.00** ± 4.35
magic	79.33 ± 3.15	**79.81** ± 2.72
newthyroid	93.02 ± 2.85	**93.02** ± 3.68
pageblocks	**94.34** ± 1.21	93.61 ± 1.47
phoneme	82.66 ± 1.50	**82.83** ± 1.43
pima	74.21 ± 3.76	**74.99** ± 3.34
titanic	**78.87** ± 1.48	**78.87** ± 1.48
wine	93.27 ± 5.03	**94.94** ± 4.13
wisconsin	**96.49** ± 1.30	96.20 ± 1.58
yeast	**57.21** ± 2.47	56.94 ± 1.45
ring	88.51 ± 3.54	**89.46** ± 3.43
segment	**93.16** ± 0.79	92.90 ± 0.42
twonorm	**83.92** ± 3.49	83.92 ± 2.63
vehicle	**69.38** ± 2.95	69.03 ± 2.53
bands	**69.14** ± 6.25	68.81 ± 4.17
penbased	**90.91** ± 1.96	90.27 ± 1.31
satimage	78.69 ± 2.16	**80.24** ± 2.71
saheart	**70.33** ± 4.09	68.61 ± 2.16
spectfheart	**78.64** ± 5.29	78.23 ± 7.50
shuttle	97.43 ± 1.05	**97.52** ± 1.01
vowel	66.36 ± 2.80	**68.38** ± 3.72
Mean	79.65	**80.13**

TABLE III
WILCOXON TEST TO COMPARE THE DIFFERENT ALPHAS IN THE COPULA FUNCTION C_{α}.

Comparison	R^{+}	R^{-}	p-value
$C_{\alpha Gen}$ vs. $C_{\alpha=0.1}$	306	159	0.096

a good mean, superior than that presented by the copula function that uses an fixed alpha parameter. Furthermore, if we look closer, it is possible to observe that $C_{\alpha Gen}$ achieves a better classification rate in 16 datasets under this study. Obviously, $C_{\alpha=0.1}$ achieved a better performance than the genetic methodology in 12 datasets and they tie twice (Titanic and Hayes-roth datasets). In order to support the quality of this study, we present an appropriate statistical study.

More specifically, we carried out a pair-wise comparison between the genetic approach ($C_{\alpha Gen}$) versus the fixed approach ($C_{\alpha=0.1}$) using the Wilcoxon signed-rank test [13]. Table III present the results of this comparisons, where R^{+} indicates the ranks obtained by $C_{\alpha Gen}$ and R^{-} represents the ranks achieved by $C_{\alpha=0.1}$.

According to the obtained statistical results presented in Table III, we can affirm with a high level of confidence, that there are statistical differences among these two generalizations. We must highlight that our genetic approach achieves a bigger rank and mean, thus, reinforcing the results obtained in the statistical test.

V. CONCLUSION

In this paper, we introduce an genetic approach to learn the parameter α used in a generalized Choquet integral. The study is defined in a similar way of the one presented by Lucca et al. [5].

We applied the generalized Choquet integral based on the copula function $C_{\alpha=0.1}$ and $C_{\alpha Gen}$ in the FRM of FRBCSs. Moreover, we have to highlight that this genetic approach allows to conclude that this generalization is statistically superior than the one that uses a fixed value in the α parameter. Therefore, the proposed methodology in this work is an alternative to the function presented by Lucca et al. [5], whereas an improvement in the quality of the new FRM was presented.

Future work is concerned with the usage of different copula functions and also the study of the properties satisfied by the generalized Choquet integral.

ACKNOWLEDGMENT

The authors would like to thank the 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 NSF (TIN2013-40765-P).

REFERENCES

[1] 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.

[2] S. R. Samantaray, K. El-Arroudi, G. Joos, and I. Kamwa, "A fuzzy rule-based approach for islanding detection in distributed generation," *Power Delivery, IEEE Transactions on*, vol. 25, no. 3, pp. 1427–1433, 2010.

[3] J. A. Sanz, M. Galar, A. Jurio, A. Brugos, M. Pagola, and H. Bustince, "Medical diagnosis of cardiovascular diseases using an interval-valued fuzzy rule-based classification system," *Applied Soft Computing*, vol. 20, pp. 103 – 111, 2014.

[4] J. Sanz, D. Bernardo, F. Herrera, H. Bustince, and H. Hagras, "A compact evolutionary interval-valued fuzzy rule-based classification system for the modeling and prediction of real-world financial applications with imbalanced data," *Fuzzy Systems, IEEE Transactions on*, vol. 23, no. 4, pp. 973–990, Aug 2015.

[5] 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.

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

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

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

[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] M. Elkano, M. Galar, J. Sanz, and 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," *Information Sciences*, vol. 332, pp. 94–114, 2016.

[11] M. Elkano, M. Galar, J. Sanz, A. Fernández, E. Barrenechea, F. Herrera, and 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*, vol. 23, no. 5, pp. 1562–1580, 2015.

[12] J. Alcal'a-Fdez, L. S'anchez, S. Garc'ia, M. Jesus, S. Ventura, J. Garrell, J. Otero, C. Romero, J. Bacardit, V. Rivas, J. Fern'andez, 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.

[13] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics*, vol. 1, pp. 80–83, 1945.

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

[15] G. Mayor and E. Trillas, "On the representation of some aggregation functions," in *Proceedings of IEEE International Symposium on Multiple-Valued Logic*. Los Alamitos: IEEE, 1986, pp. 111–114.

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

[17] 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.

[18] 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.

[19] M. Grabisch and C. Labreuche, "A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid," *Annals of Operations Research*, vol. 175, no. 1, pp. 247–286, 2010.

[20] 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.

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

[22] B. Kavšek, N. Lavrač, and V. Jovanoski, *Advances in Intelligent Data Analysis V: 5th International Symposium on Intelligent Data Analysis, IDA 2003, Berlin, Germany, August 28-30, 2003. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, ch. APRIORI-SD: Adapting Association Rule Learning to Subgroup Discovery, pp. 230–241.

[23] R. Alcala, J. Alcala-Fdez, and F. Herrera, "A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection," *IEEE Transactions on Fuzzy Systems*, vol. 15, no. 4, pp. 616–635, 2007.

[24] L. J. Eshelman, "The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination," in *Foundations of Genetic Algorithms*, G. J. E. Rawlings, Ed. San Francisco: Morgan Kaufmann, 1991, pp. 265–283.

[25] F. Herrera, M. Lozano, and A. M. S'anchez, "A taxonomy for the crossover operator for real-coded genetic algorithms: An experimental study," *International Journal of Intelligent Systems*, vol. 18, no. 3, pp. 309–338, 2003.

[26] J. A. Sanz, A. Fern'andez, H. Bustince, and F. Herrera, "Improving the performance of fuzzy rule-based classification systems with interval-valued fuzzy sets and genetic amplitude tuning," *Information Sciences*, vol. 180, no. 19, pp. 3674 – 3685, 2010.

[27] J. Sanz, A. Fernández, H. Bustince, and F. Herrera, "IVTURS: A linguistic fuzzy rule-based classification system based on a new interval-valued fuzzy reasoning method with tuning and rule selection," *IEEE Transactions on Fuzzy Systems*, vol. 21, no. 3, pp. 399–411, 2013.