

# Study of $n$ -dimensional overlap functions in Fuzzy Rule-Based Classification Systems

Mikel Elcano<sup>1</sup> Mikel Galar<sup>1</sup> Jose Sanz<sup>1</sup> Humberto Bustince<sup>1</sup>

<sup>1</sup>Departamento de Automatica y Computacion, Universidad Publica de Navarra, Spain

## Abstract

In a previous work we proposed to enhance the performance of FARC-HD fuzzy classifier in multi-class classification problems using decomposition strategies. This synergy was further improved by introducing  $n$ -dimensional overlap functions in the learning algorithm and the inference of FARC-HD instead of the product t-norm.

In this work, we extend this methodology to FURIA algorithm aiming at analyzing the behavior of the new Fuzzy Reasoning Method with overlap functions, using both decomposition strategies and the baseline FURIA. We also compare the results obtained by this new combination with those obtained in the synergy presented in our previous work for FARC-HD.

**Keywords:** Fuzzy Rule-Based Classification Systems, Decomposition strategies, Overlap functions, Aggregations, One-vs-One, Multi-classification

## 1. Introduction

There are multiple techniques to solve classification problems in the field of pattern recognition. Among them, Fuzzy Rule-Based Classification Systems (FRBCSs) [1] offer very good performance whilst providing interpretable models by using linguistic labels in their rules.

In classification, two different types of problems can be differentiated depending on the number of classes considered: binary (two classes) and multi-class problems (more than two classes). Normally, facing a multi-class problem is more difficult due to the higher complexity of the definition of decision boundaries. One way of addressing multi-class problems is applying decomposition strategies [2, 3], which divide the original problem into easier-to-solve binary ones that are faced by independent binary classifiers called *base classifiers*.

Among decomposition strategies [3], *One-vs-One* (OVO) and *One-vs-All* (OVA) are the most commonly used ones. In OVO scheme the original problem is divided into as many binary sub-problems as possible pairs of classes, whereas in OVA as many sub-problems as classes in the original one are considered. In both strategies each binary problem is solved by an independent base classifier. When classifying a new instance, all base classifiers are queried and their outputs are combined to make the final

decision. This technique usually works better than addressing the problem directly [2, 4, 5].

In our previous work [6], we proposed to apply decomposition strategies to improve the performance of FARC-HD fuzzy classifier [7] in multi-class problems. We showed that the outputs provided by FARC-HD are not suitable for the aggregation process carried out in decomposition techniques, due to the usage of the product to model the conjunction. The reason is that, when aggregating small values, the values returned by the product have a low variation and tend quickly to zero. This effect is even more accentuated when the number of arguments increases, penalizing those rules with more antecedents.

In order to solve these problems, the concept of  $n$ -dimensional overlap function was introduced aiming at modeling the conjunction in the fuzzy rules of FARC-HD. These functions produce outputs with a higher variation and without depending on the number of arguments, providing a significant improvement when applying decomposition strategies.

In this work, we extend this methodology to another state-of-the-art fuzzy classifier such as FURIA [8], since it is one of the most accurate fuzzy classifiers in the literature and its rule structure is clearly different from that of FARC-HD. In this manner, we study the behavior of  $n$ -dimensional overlap functions when using FURIA algorithm. To do so, we have adapted the Fuzzy Reasoning Method (FRM) of FURIA to apply these functions. We have tested this adaptation when facing directly multi-class problems as well as when it is applied as base classifier in OVO and OVA schemes.

In order to achieve well-founded conclusions, we carry out an empirical study considering twenty numerical datasets from the KEEL dataset repository [9] and we support the results obtained using non-parametric statistical tests, as suggested in the specialized literature [10]. In this study, we analyze the influence of  $n$ -dimensional overlap functions in the baseline FURIA and when it is used as base classifier for both OVO and OVA decomposition strategies. More specifically, we study the effect of these functions on both the final performance and the rule base. All the results obtained by this new combination have been compared against those obtained with FARC-HD.

The structure of this contribution is as follows. A brief explanation of FURIA and FARC-HD fuzzy

classifiers is shown in Section 2. Section 3 describes OVO and OVA decomposition strategies, along with the two aggregation strategies for OVO considered in this work. In Section 4 we recall the concept of  $n$ -dimensional overlap functions and we describe the adaptation of the FRM performed to use these functions. The experimental framework and the analysis of the results obtained are presented in Section 5. Finally, in Section 6 we draw the conclusions.

## 2. Fuzzy Rule-Based Classification Systems

In order to generate the knowledge base, a fuzzy rule learning algorithm is applied using a training set  $\mathcal{D}_T$  composed of  $P$  labeled examples  $x_p = (x_{p1}, \dots, x_{pn})$ ,  $p = \{1, \dots, P\}$ , where  $x_{pi}$  is the value of the  $i$ -th attribute ( $i = \{1, 2, \dots, n\}$ ) of the  $p$ -th training example. Each example belongs to a class  $y_p \in \mathbb{C} = \{C_1, C_2, \dots, C_m\}$ , where  $m$  is the number of classes of the problem.

In the rest of this section we describe the two FRBCSs considered in this work, that is, FARC-HD (Section 2.1) and FURIA (Section 2.2).

### 2.1. FARC-HD

FARC-HD (Fuzzy Association Rule-based Classification model for High-Dimensional problems) [7] is a fuzzy association rule-based classifier. The rule structure used by FARC-HD is the following:

$$\begin{aligned} \text{Rule } R_j : & \text{ If } x_1 \text{ is } A_{j1} \text{ and } \dots \text{ and } x_{n_j} \text{ is } A_{jn_j} \\ & \text{ then Class} = C_j \text{ with } RW_j \end{aligned} \quad (1)$$

where  $R_j$  is the label of the  $j$ -th rule,  $x = (x_1, \dots, x_n)$  is a vector representing the example,  $A_{ji} \in \mathbb{L}_i$  is a linguistic label modeled by a triangular membership function (where  $\mathbb{L}_i = \{L_{i1}, \dots, L_{il}\}$  is the set of linguistic labels for the  $i$ -th antecedent, being  $l$  the number of linguistic labels in this set),  $C_j$  is the class label and  $RW_j$  is the rule weight computed using the certainty factor defined in [11].

The learning algorithm of FARC-HD is composed of three steps:

1. *Fuzzy association rule extraction for classification*: Fuzzy rules are obtained building a search tree for each class. The number of linguistic terms in the antecedents is limited by the maximum depth of the tree.
2. *Candidate rule pre-screening*: The most interesting fuzzy rules are selected from the rule base obtained in the previous stage applying a pattern weighting scheme.
3. *Genetic rule selection and lateral tuning*: An evolutionary algorithm is used both to tune the lateral position of membership functions and to select the most accurate rules from the rule base generated in the previous steps.

In order to classify a new example  $x_p = (x_{p1}, \dots, x_{pn})$ , FARC-HD applies the following Fuzzy Reasoning Method:

1. *Matching degree*. The strength of activation of the antecedent part for all rules in the rule base with the example  $x_p$  is computed.

$$\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \dots, \mu_{A_{jn}}(x_{pn})) \quad (2)$$

2. *Association degree*. The association degree of the example  $x_p$  with each rule in the rule base is computed.

$$b_j(x_p) = \mu_{A_j}(x_p) \cdot RW_j \quad (3)$$

3. *Confidence degree*. The confidence degree for each class is computed.

$$\text{conf}_c(x_p) = \sum_{R_j \in RB; C_j = c} b_j(x_p), \quad c = 1, 2, \dots, m \quad (4)$$

4. *Classification*. The class that obtains the highest confidence degree is predicted.

$$\text{Class} = \arg \max_{c=1, \dots, m} (\text{conf}_c(x_p)) \quad (5)$$

### 2.2. FURIA

FURIA (Fuzzy Unordered Rule Induction Algorithm) [8] is an extension of RIPPER algorithm [12] that makes use of fuzzy rules instead of conventional rules. The rule structure in FURIA is as follows:

$$\begin{aligned} \text{Rule } R_j : & \text{ If } x_1 \text{ is } A_{j1}^I \text{ and } \dots \text{ and } x_{n_j} \text{ is } A_{jn_j}^I \\ & \text{ then Class} = C_j \text{ with } RW_j \end{aligned} \quad (6)$$

where  $A_{ji}^I$  is a trapezoidal membership function corresponding to the variable  $i$  defined as  $A_{ji}^I = (\phi_{ji}^{s,L}, \phi_{ji}^{c,L}, \phi_{ji}^{c,U}, \phi_{ji}^{s,U})$ , being  $(\phi_{ji}^{c,L}, \phi_{ji}^{c,U})$  and  $(\phi_{ji}^{s,L}, \phi_{ji}^{s,U})$  the lower and upper bounds of the core and the support, respectively. In this case,  $RW_j$  is computed using the  $m$ -estimate for  $m = 2$  [13]. Note that, in contrast to FARC-HD, each trapezoidal membership function is specific to each antecedent.

The learning algorithm of FURIA is composed of the following stages:

1. *Learn a rule set for each class using RIPPER algorithm*. This phase is divided into the building and the optimization phase.
2. *Fuzzification of rules generated by RIPPER*. The interval representing each antecedent is replaced by a trapezoidal membership function (Eq. (6)) applying a greedy algorithm. To do so, the quality of each fuzzification is computed in terms of *rule purity*. We must remark that the rule purity is obtained by adding the membership degrees of positive and negative instances to the antecedent, and thus no t-norm is involved in this step.

When classifying a new example, FURIA applies the same FRM as FARC-HD (Eq. (2)-(5)), but using trapezoidal membership functions ( $\mu_{A_{ji}^r}$ ) instead of triangular ones. If the example is not covered by any rule, a rule generalization process (*stretching*) is performed.

### 3. Decomposition strategies

Decomposition strategies [3] divide the original multi-class problem into binary problems, which are faced by independent base classifiers. These strategies are valid both for classifiers that only work with two classes and those with multi-class support. Even in the latter case, better results are obtained when decomposition strategies are applied [2].

#### 3.1. One-Versus-One (OVO) strategy

OVO strategy divides the original  $m$  class problem into  $m(m-1)/2$  binary sub-problems (all possible pairs of classes). Each binary problem is faced by an independent base classifier which distinguishes a pair of classes  $\{C_i, C_j\}$ . In order to classify a new example, all base classifiers are queried and each one returns a pair of confidence degrees  $r_{ij}, r_{ji} \in [0, 1]$  in favor of classes  $C_i$  and  $C_j$ , respectively. These outputs are stored in the *score-matrix*  $R$ :

$$R = \begin{pmatrix} - & r_{12} & \cdots & r_{1m} \\ r_{21} & - & \cdots & r_{2m} \\ \vdots & & & \vdots \\ r_{m1} & r_{m2} & \cdots & - \end{pmatrix} \quad (7)$$

Since each binary sub-problem is addressed by an independent base classifier, we need to normalize the score-matrix  $R$  to obtain a new one,  $\hat{R}$ , in which all confidence degrees are within the same range of values:

$$\hat{r}_{ij} = \begin{cases} \frac{r_{ij}}{r_{ij} + r_{ji}} & \text{if } r_{ij} \neq 0 \text{ or } r_{ji} \neq 0 \\ 0.5 & \text{if } r_{ij} = r_{ji} = 0 \end{cases} \quad (8)$$

Finally the outputs of base classifiers are aggregated and the predicted class is obtained.

In this work we consider two well-studied aggregation methods in the literature:

- *Voting strategy* (VOTE) [14]. Each base classifier gives a vote for its predicted class. The class having the largest number of votes is given as output:

$$Class = arg \max_{i=1, \dots, m} \sum_{1 \leq j \neq i \leq m} s_{ij} \quad (9)$$

where  $s_{ij}$  is 1 if  $\hat{r}_{ij} > \hat{r}_{ji}$  and 0 otherwise.

- *Non-Dominance criteria* (ND) [15]. The score-matrix is considered as a fuzzy preference relation. Then the non-dominance degree is computed, being the winning class the one with the

highest value:

$$Class = arg \max_{i=1, \dots, m} \left\{ 1 - \max_{j=1, \dots, m} r'_{ji} \right\} \quad (10)$$

where  $R'$  is the strict score-matrix (after normalization).

#### 3.2. One-Versus-All (OVA) strategy

OVA decomposition divides an  $m$  class problem into  $m$  binary sub-problems. Each binary sub-problem is addressed by an independent base classifier which distinguishes one of the classes from the remaining ones, learning the model using all examples of the training set. In this manner, the examples of the class to be distinguished are considered as positives, whereas the rest are labeled as negatives. When classifying a new example, all base classifiers are queried and each one returns a confidence degree  $r_i \in [0, 1]$  in favor of the class  $C_i$ . These outputs are stored in the *score-vector*  $R$ :

$$R = (r_1, \dots, r_i, \dots, r_m) \quad (11)$$

For the same reason as in OVO, we need to normalize the score-vector with respect to the confidences obtained by each classifier for the negative class (stored in another score-vector  $\bar{R}$ ). Once we obtain both vectors, the normalized score-vector ( $\hat{R}$ ) is computed as follows.

$$\hat{r}_i = \frac{r_i}{r_i + \bar{r}_i} \quad (12)$$

Finally, the class with the highest confidence will be predicted.

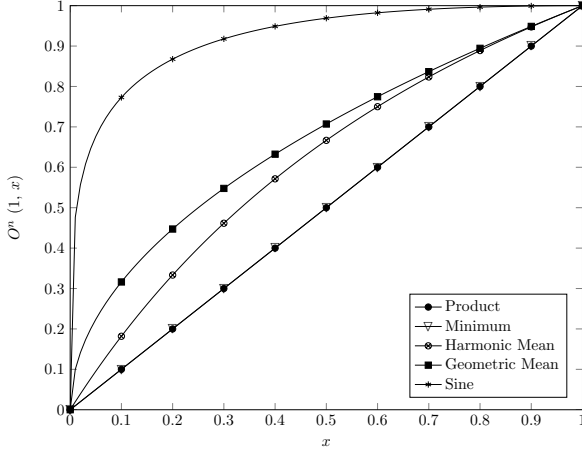
### 4. Adapting the FRM of FURIA to model the conjunction with $n$ -dimensional overlap functions

In this work we study the behavior of  $n$ -dimensional overlap functions in FURIA and we compare it with that obtained in FARC-HD.

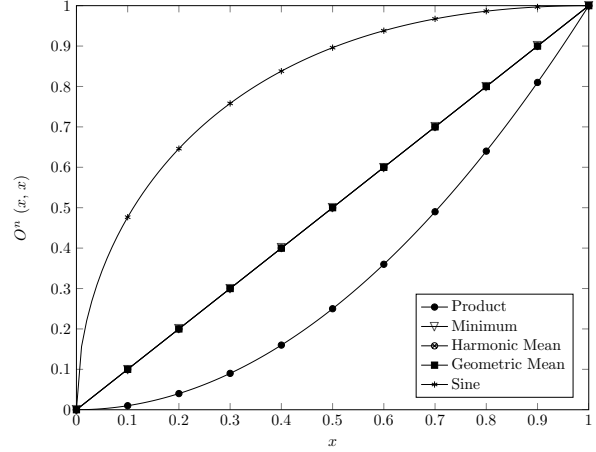
First, we recall the definition of  $n$ -dimensional overlap function.

**Definition 1** [6] *A  $n$ -dimensional function  $O : [0, 1]^n \rightarrow [0, 1]$  with  $n \geq 2$  is a  $n$ -dimensional overlap function if the following properties hold:*

1.  $O$  is symmetric.
2.  $O(x_1, \dots, x_n) = 0$  if and only if  $\prod_{i=1}^n x_i = 0$ .
3.  $O(x_1, \dots, x_n) = 1$  if and only if  $\prod_{i=1}^n x_i = 1$ .
4.  $O$  is increasing.
5.  $O$  is continuous in each of the variables.



(a) Considering  $(1, x)$  as input.



(b) Considering  $(x, x)$  as input.

Figure 1: Values returned by the different overlap functions.

Since  $n$ -dimensional overlap functions fulfill similar properties to those of t-norms, we can apply overlap functions to model the conjunction, as shown in our previous work [6]. In this contribution, we propose to adapt the FRM of FURIA using these functions. Due to the fact that FURIA applies the same FRM as FARC-HD, the adaptation carried out in FURIA is the same as that performed in FARC-HD. In this manner, we replace the product used in the computation of matching (2) and association (3) degrees by a  $n$ -dimensional overlap function:

- *Matching degree:*

$$\mu_{A_j^I}(x_p) = O\left(\mu_{A_{j_1}^I}(x_{p1}), \mu_{A_{j_2}^I}(x_{p2}), \dots, \mu_{A_{j_{n_j}}^I}(x_{pn_j})\right) \quad (13)$$

- *Association degree:*

$$b_j(x_p) = O\left(\mu_{A_j^I}(x_p), RW_j\right) \quad (14)$$

We must remark that, contrary to FARC-HD, this adaptation does not affect the learning process of FURIA, because it does not make use of the inference in the learning process.

In this work we have considered five different  $n$ -dimensional overlap functions:

- **Product (PROD):** The returned value is the product of input values. It recovers the original case of FARC-HD and FURIA.

$$O(x_1, \dots, x_n) = \prod_{i=1}^n x_i \quad (15)$$

- **Minimum (MIN):** Returns the minimum of input values. This is a t-norm as well, but

unlike the product, the minimum satisfies the idempotency property.

$$O(x_1, \dots, x_n) = \min(x_1, \dots, x_n) \quad (16)$$

- **Harmonic mean (HM):** The returned value is the harmonic mean of input values if all of them are different than zero and 0 otherwise.

$$O(x_1, x_2, \dots, x_n) = \begin{cases} \frac{n}{\frac{1}{x_1} + \dots + \frac{1}{x_n}} & \text{if } x_i \neq 0, \text{ for all } i = 1, \dots, n \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

- **Geometric mean (GM):** Returns the geometric mean of input values.

$$O(x_1, x_2, \dots, x_n) = \sqrt[n]{\prod_{i=1}^n x_i} \quad (18)$$

- **Sine (SIN):** This overlap function returns higher values than means.

$$O(x_1, \dots, x_n) = \sin\left(\frac{\pi}{2} \left(\prod_{i=1}^n x_i\right)^\alpha\right) \quad (19)$$

where  $\alpha \leq \frac{1}{2n}$ . In the experiments carried out in Section 5, we take  $\alpha = \frac{1}{2n}$ .

Among the considered overlap functions, first we have the product and the minimum as representative of t-norms, where the minimum satisfies the idempotency property. Then, we consider the harmonic and geometric means, which return higher values than t-norms and they also satisfy the idempotency. Finally, we have the SIN, whose outputs are greater than those of the means.

In Fig. (1a) we can observe the values returned by each overlap function when aggregating a value with 1, whereas Fig. (1b) shows the values returned when aggregating a value with itself.

## 5. Experimental study

The two main objectives of the experimental study are the following:

1. To analyze the performance of  $n$ -dimensional overlap functions in both baseline FURIA and decomposition strategies, comparing it with that obtained in FARC-HD.
2. To compare the impact of  $n$ -dimensional overlap functions on the rule base of FURIA and FARC-HD.

In the remainder of this section we present the experimental framework (Section 5.1), the analysis of the performance of each method (Section 5.2), and the impact on the rule base (Section 5.3).

### 5.1. Experimental framework

In order to carry out the experimental study we have considered twenty datasets selected from the KEEL dataset repository [9]. In Table 1, we present a summary of the features of all datasets, indicating for each one the number of examples (#Ex.), number of attributes (#Atts.), number of numerical (#Num.) and nominal (#Nom.) attributes, and the number of classes (#Class.).

Table 1: Summary of the features of the datasets used in the experimental study.

Id.	Dataset	#Ex.	#Atts.	#Num.	#Nom.	#Class.
aut	autos	159	25	15	10	6
bal	balance	625	4	4	0	3
cle	cleveland	297	13	13	0	5
con	contraceptive	1473	9	6	3	3
eco	ecoli	336	7	7	0	8
gla	glass	214	9	9	0	7
hay	hayes-roth	132	4	4	0	3
iri	iris	150	4	4	0	3
new	newthyroid	215	5	5	0	3
pag	pageblocks	548	10	10	0	5
pen	penbased	1100	16	16	0	10
sat	satimage	643	36	36	0	7
seg	segment	2310	19	19	0	7
shu	shuttle	2175	9	9	0	5
tae	tae	151	5	3	2	3
thy	thyroid	720	21	21	0	3
veh	vehicle	846	18	18	0	4
vow	vowel	990	13	13	0	11
win	wine	178	13	13	0	3
yea	yeast	1484	8	8	0	10

We have used a *5-fold stratified cross-validation model*, i.e., we randomly split the dataset into five partitions of data, each one containing 20% of the examples, and we employed a combination of four of them (80%) to train the system and the remaining one to test it. Additionally, in each partition we consider three different seeds for the execution of each method. In order to correct the dataset shift (when the training and test data do not follow the same distribution [16]), we have used the *Distribution Optimally Balanced Cross Validation* procedure [17], instead of the commonly used cross-validation.

In order to support the quality of the methods we apply non-parametric statistical tests [10]. More specifically, we use the Wilcoxon rank test to compare two methods, the Friedman aligned ranks test to check whether there are statistical differences among a group of methods and the Holm post-hoc test to find the algorithms that reject the null hypothesis of equivalence against the selected control method.

The configuration used for FARC-HD and FURIA are the ones recommended by the authors:

- FARC-HD: 5 labels for each fuzzy partition, the maximum tree depth is 3, minimum support of 0.05, minimum confidence of 0.8, populations formed by 50 individuals, 30 bits per gene for the Gray codification and a maximum of 20000 evaluations.
- FURIA: 2 optimizations and 3 folds.

### 5.2. Analysis of the performance of $n$ -dimensional overlap functions

Table 2 shows the average accuracy rate obtained in testing by FARC-HD and FURIA. As we can observe, the results obtained by each baseline classifier along with those achieved when they are used with both OVA scheme and the two aggregation strategies of OVO model ( $OVO^{ND}$  and  $OVO^{VOTE}$ ) are presented. We show the performance of the five overlap functions (PROD, MIN, HM, GM, SIN) for each method, where the result of the best overlap function is highlighted in bold-face.

Table 2: Average accuracy rate obtained in testing by each method.

	FARC-HD					FURIA				
	PROD	MIN	HM	GM	SIN	PROD	MIN	HM	GM	SIN
<i>Baseline</i>	<b>80.37</b>	80.17	80.11	79.89	79.98	<b>80.56</b>	80.55	80.46	80.49	80.12
<i>OVA</i>	79.92	80.27	<b>80.48</b>	80.13	79.97	80.39	80.38	80.40	80.36	<b>80.41</b>
$OVO^{ND}$	81.45	81.88	<b>82.18</b>	82.13	81.46	81.97	<b>81.98</b>	81.92	81.90	81.67
$OVO^{VOTE}$	81.52	82.03	<b>82.26</b>	82.25	81.71	82.37	<b>82.39</b>	82.37	82.34	82.12

Table 3: Aligned Friedman and Holm tests to compare the different overlaps in FARC-HD, OVA and OVO.

	FARC-HD	OVA	$OVO^{ND}$	$OVO^{VOTE}$
PROD	<b>43.80</b>	57.90 (0.128)	55.23 (0.327)	56.53 (0.269)
MIN	48.63 (0.967)	51.72 (0.282)	49.03 (0.708)	49.77 (0.672)
HM	50.22 (0.967)	<b>38.23</b>	<b>40.52</b>	<b>40.95</b>
GM	56.25 (0.699)	48.95 (0.282)	45.65 (0.708)	43.65 (0.768)
SIN	53.60 (0.856)	55.70 (0.170)	62.08 ( <u>0.075</u> )	61.60 ( <u>0.097</u> )

In order to study the differences among the overlap functions used in this work, we carry out the Aligned Friedman test and the Holm post-hoc test, whose results are shown in Tables 3 and 4. These are grouped in columns according to the method used to perform the comparison and in rows according to the considered overlap function. The

Table 4: Aligned Friedman and Holm tests to compare the different overlaps in FURIA, OVA and OVO.

	FURIA	OVA	OVO <sup>ND</sup>	OVO <sup>VOTE</sup>
PROD	<b>40.17</b>	51.00 (1.000)	42.65 (0.871)	43.75 (1.000)
MIN	43.10 (0.750)	51.75 (1.000)	<b>40.90</b>	<b>42.77</b>
HM	50.90 (0.727)	50.00 (1.000)	48.05 (0.871)	44.50 (1.000)
GM	48.97 (0.727)	54.73 (1.000)	53.80 (0.479)	50.52 (1.000)
SIN	69.35 ( <u>0.006</u> )	<b>45.03</b>	67.10 ( <u>0.018</u> )	70.95 ( <u>0.008</u> )

first column corresponds to the baseline classifier execution using each overlap function, whereas the second one shows the different overlap functions over OVA model. The rest of columns correspond to OVO model using ND and VOTE aggregation strategies, respectively. The value of each cell is the rank obtained with the Friedman aligned-rank test that compares the different overlap functions for the same method. The value shown in brackets is the adjusted p-value obtained by the Holm test using as control method the one having the smallest rank in the same column, which is shown in bold-face. The adjusted p-value is underlined when there are statistical differences ( $\alpha = 0.05$ ).

Next, we analyze the behavior of  $n$ -dimensional overlap functions in baseline FARC-HD and FURIA, as well as when decomposition strategies are applied in each of them.

#### • FARC-HD

- *Baseline*: Table 2 shows that the five overlap functions obtain similar results when executing the baseline FARC-HD algorithm. This situation is confirmed by the results obtained in the Aligned Friedman test shown in Table 3, where there are no statistical differences among these functions. The reason is that FARC-HD is able to maintain the classification accuracy when using overlap functions, due to the fact that they are involved in all stages of the learning process and the generated rules are general enough to retain the discrimination capability.
- *OVO and OVA models*: looking at Tables 2 and 3, we observe that the greater the overlap function is, the better the results obtained are (even though the GM is greater than the HM, both of them return similar values). However, the SIN does not provide a good performance as it can return a value which is greater than all the input values, losing part of the discrimination capability. Thus, the overlap functions that obtain the best results in almost all cases are those returning the highest values preserving the idempotence property. As we can observe, this algorithm is able to take advantage of the confidences provided by overlap functions, since the classification accuracy is maintained when using these functions in the baseline FARC-HD, as shown in [6].

#### • FURIA

- *Baseline*: according to Tables 2 and 4, the results obtained by all overlap functions are similar when using FURIA (except with the SIN). This is because overlap functions are not involved in the learning algorithm of FURIA, since all rules are generated by RIPPER and the subsequent fuzzification process does not apply any t-norm. Therefore, the rules generated when using different overlap functions are the same. Moreover, the highly adjusted trapezoidal membership functions (whose optimization is not performed using t-norms) used by FURIA return high membership degrees, and hence the values returned by different overlap functions when aggregating large values are more similar (see Fig. 1a and 1a).
- *OVO and OVA models*: Tables 2 and 4 show that both OVO and OVA models obtain similar results when using different overlap functions. Although FURIA maintains the classification accuracy when using these functions, the confidences provided by all of them are very similar due to the highly adjusted trapezoidal membership functions, as previously explained. The exception is the SIN, due to the same reasons as in the case of FARC-HD.

#### 5.3. Impact of $n$ -dimensional overlap functions on the rule base

Table 5 shows the average number of antecedents by rule and the average number of rules for each baseline FRBCS, OVA and OVO models (using as base classifier the same FRBCS).

As we can observe in Table 5, in the case of FARC-HD, the usage of a greater overlap function implies a growing trend in the number of rules, which is necessary to maintain the discrimination capability. On the other side, the number of antecedents is similar in all overlap functions, even though there is an upward trend when using greater overlap functions.

In FURIA, however, overlap functions are not involved in the learning process and consequently the rules generated are the same for all overlap functions, as shown in Table 5.

Regarding the differences between baseline and decomposition strategies, Table 5 clearly shows that the rule base becomes simpler when decomposition strategies are applied. Since these strategies divide the original problem into easier-to-solve binary sub-problems, a lower number of rules and antecedents is needed to solve each sub-problem. In the same manner, the rule base in OVO is simpler than in OVA because OVO scheme considers only the examples of two classes while OVA takes into account all examples in the training set, increasing the complexity of the decision boundary definition.

Table 5: Average number of rules and antecedents.

		avg. rules					avg. antecedents				
		PROD	MIN	HM	GM	SIN	PROD	MIN	HM	GM	SIN
FARC-HD	Baseline	32.67	35.70	40.15	41.11	46.30	2.34	2.38	2.44	2.44	2.47
	OVA	13.03	14.26	16.09	16.64	18.35	1.76	1.79	1.84	1.84	1.86
	OVO	8.55	9.72	11.28	11.72	12.58	1.61	1.63	1.66	1.66	1.69
FURIA	Baseline	16.54	16.54	16.54	16.54	16.54	2.76	2.76	2.76	2.76	2.76
	OVA	7.95	7.95	7.95	7.95	7.95	2.05	2.05	2.05	2.05	2.05
	OVO	4.50	4.50	4.50	4.50	4.50	1.58	1.58	1.58	1.58	1.58

## 6. Conclusions

The motivation of this work arises from the significant improvement found in FARC-HD when combining  $n$ -dimensional overlap functions and decomposition strategies. Specifically, we have extended this methodology to FURIA with the aim of understanding the behavior of these functions in a different FRBCS.

We have shown that FURIA is not able to take advantage of the benefits provided by  $n$ -dimensional overlap functions. Contrary to FARC-HD, although FURIA preserves the classification accuracy when applying overlap functions, the highly adjusted fuzzy intervals return membership degrees that are likely to be 0 or close to 1. This implies that the differences among the values returned by different overlap functions are smaller, and hence the behavior of all overlap functions is similar in FURIA.

Besides the performance, we have analyzed the effect of  $n$ -dimensional overlap functions on the rule base of FURIA. We have shown that the rule base of this FRBCS is not affected by the usage of these functions, since they are not involved in its learning process.

## References

- [1] H. Ishibuchi, T. Nakashima, and M. Nii. *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining*. Springer-Verlag, 2004.
- [2] M. Galar, A. Fernández, E. Barrenechea, H. Bustince, and F. Herrera. An overview of ensemble methods for binary classifiers in multi-class problems: Experimental study on one-vs-one and one-vs-all schemes. *Pattern Recognition*, 44(8):1761 – 1776, 2011.
- [3] A.C. Lorena, A.C. Carvalho, and J.M. Gama. A review on the combination of binary classifiers in multiclass problems. *Artificial Intelligence Review*, 30(1-4):19–37, 2008.
- [4] M. Galar, A. Fernández, E. Barrenechea, and F. Herrera. Empowering difficult classes with a similarity-based aggregation in multi-class classification problems. *Information Sciences*, 264:135–157, 2014.
- [5] J. A. Sáez, M. Galar, J. Luengo, and F. Herrera. Analyzing the presence of noise in multi-class problems: alleviating its influence with the one-vs-one decomposition. *Knowledge and Information Systems*, 38(1):179–206, 2014.
- [6] M. Elkanó, 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*, 2014.
- [7] 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*, 19(5):857–872, 2011.
- [8] J. Hühn and E. Hüllermeier. FURIA: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery*, 19(3):293–319, 2009.
- [9] J. Alcalá-Fdez, A. Fernández, J. Luengo, J. Derrac, S. García, L. Sánchez, and F. Herrera. KEEL data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing*, 17:2-3:255–287, 2011.
- [10] 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 Comput.*, 13(10):959–977, 2009.
- [11] H. Ishibuchi, T. Yamamoto, and T. Nakashima. Hybridization of fuzzy GBML approaches for pattern classification problems. *IEEE Transactions on System, Man and Cybernetics B*, 35(2):359–365, 2005.
- [12] W. W. Cohen. Fast effective rule induction. presented at the 12th Int. Conf. Mach. Learn., Lake Tahoe, CA, USA, 1995.
- [13] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical recipes in FORTRAN: the art of scientific computing*. Cambridge University Press, 1992.
- [14] J.H. Friedman. Another approach to polychotomous classification. Technical report, Department of Statistics, Stanford University, 1996.
- [15] A. Fernández, M. Calderón, E. Barrenechea, H. Bustince, and F. Herrera. Solving multi-class

problems with linguistic fuzzy rule based classification systems based on pairwise learning and preference relations. *Fuzzy Sets and Systems*, 161(23):3064–3080, 2010.

- [16] V. López, A. Fernández, and F. Herrera. On the importance of the validation technique for classification with imbalanced datasets: Addressing covariate shift when data is skewed. *Information Sciences*, 257:1–13, 2014.
- [17] J.G. Moreno-Torres, J.A. Saez, and F. Herrera. Study on the impact of partition-induced dataset shift on k-fold cross-validation. *IEEE Transactions on Neural Networks and Learning Systems*, 23(8):1304–1312, 2012.