Improving the OVO performance in fuzzy rule-based classification systems by the genetic learning of the granularity level

Pedro Villar*, Alberto Fernández†, Rosana Montes*, Ana María Sánchez* and Francisco Herrera‡
*Department of Software Engineering
University of Granada, Granada, Spain
Email: pvillarc,rmontes,amlopez@ugr.es
†Dept. of Computer Science
University of Jaén, Jaén, Sapin
Email: alberto.fernandez@ujaen.es
‡Department of Computer Science & Artificial Intelligence
University of Granada, Granada, Spain
Email: herrera@decsai.ugr.es

***Abstract*—This contribution proposes a genetic learning process for designing the knowledge base of Fuzzy Rule-Based classification Systems (FRCBSs), that will be used as binary classifiers in a One-vs-One decomposition for multi-class problems.**

A Genetic Algorithm is designed to adapt the number of fuzzy labels per variable for each classifier in order to improve the accuracy rate of a multi-class classifier. The genetic learning process evolves granularity levels and needs a fuzzy rules generation method for generating the whole knowledge base of the FRBCS.

Several data-sets from UCI repository are used in the experimental study and we compare our proposal with the standard way to design FRBCS using the rule generation method chosen with and without One-vs-One decomposition.

I. Introduction

Classification problems with more than two classes (multi-class problems) are known to present more difficulties than binary-class problems. One robust solution to cope with the former problem is to use a decomposition approach[1][2]. Its main strategy is to reduce the multi-class problem to several binary-class problems[3], where the One-vs-One (OVO) technique is widely used[21]. This method divides the original problem by confronting all pairs of classes against them. Then, an indepedent classifier is built for each pair of classes and it is necessary to combine the outputs of these classifiers to obtain the final predicted class label for a given instance[2][4]. Therefore, the way the decision process is carried out has a strong influence in the classification performance[2].

We develop an experimental analysis in the context of multi-class classification. We will make use of linguistic Fuzzy Rule Based Classification Systems (FRBCSs), that constitute a very spread tool for classification problems. An advantage of the FRBCSs is the interpretability of the generated model[5]. An FRBCS presents two main components: the Inference System and the Knowledge Base (KB). The KB is composed of two parts: the Rule Base (RB), constituted by the collection of fuzzy rules, and the Data Base (DB), that includes the membership functions of the fuzzy partitions associated to each linguistic variables. The composition of the KB directly depends on the problem being solved. If there is no expert information about the problem, it is necessary to perform an automatic learning process to derive the KB from examples. There are some proposals of decomposition techniques for dealing multi-class problems with FRBCSs [6][7][8].

The majority of algorithms for learning the KB of an FRBCS, considers a previously defined DB and only derives the RB. Generally, the DB is built by choosing a number of linguistic terms for all the variables (an odd number between 3 and 7 is a typical decision) and considering uniform fuzzy partitions. However, the granularity level has a significant influence on the FRBCS performance as it has been analyzed in [9]. The number of labels of each partition can be viewed as a sort of context information. A fuzzy partition with too many linguistic terms, probably will have unnecessary terms, that is, they can contribute nothing, probably decreasing the interpretability of the model. Additionally, in some case they may cause confusion also hindering the discrimination ability of the classifier. In other cases, it would be convenient to add new linguistic terms to appropriately differentiate the values of the variable. Some methods for the KB learning in fuzzy modeling and fuzzy classification include the granularity level learning [10][11][12][13][14].

The main purpose of this paper is to improve the performance of an OVO scheme built with FRBCSs by learning an appropriate granularity level for each fuzzy partition. To do so, we employ an approach to derive the whole KB that involves the use of two different (and independent) learning processes, in which a DB definition process wraps a RB learning one. Specifically, we use a Genetic Algorithm (GA)[15] for the granularity learning and a classical fuzzy rules generation algorithm, the Chi et al.'s method[16] for the RB derivation. A similar KB learning scheme was performed in [10][11] for regression problems and in [14] to design more interpretable FRBCSs for binary-class problems with imbalanced data-sets.

In order to illustrate the good performance of the proposed scheme of an OVO strategy with the KB learning process mentioned, we will compare the obtained results with the Chi et al.'s algorithm and with the application of OVO decomposition using Chi et al.'s method as method for generating the binary classifiers.

We have selected a collection of multi-class data-sets from KEEL data-set repository[1] [17] for developing our experimental analysis. Furthermore, we will perform a statistical analysis using non-parametric tests [18], [19], [20] to find significant differences among the obtained results.

This paper is organized as follows. First, Section II introduces the preliminary concepts of the OVO scheme and FRBCSs used in this paper. Next, in Section III we will describe our proposal, an OVO strategy with FRBCSs designed using a GA for granularity learning. The next section describes the experimental study. Finally, in Section V, some conclusions will be pointed out.

II. Preliminaries

This section introduces the OVO scheme, some basic concepts about FRBCS and describes the fuzzy rule learning algorithm used in our work.

A. One-vs-One decomposition

The most common approaches for decomposition a multi-class problem into a binary-class problem are OVO [21] and OVA [22]. The former learns a binary classifier for each posible pair of classes, whereas the latter constructs a binary classifier considering each single class and all the other classes joined. OVO has shown a better behavior than OVA in a general scenario[2], and it has been established by default in several widely used software tools [23], [24], [25].

OVO divides a m-class problem into $m(m-1)/2$ independent binary subproblems by contrasting all classes among them, each of which is learnt by a single classifier. In the classification stage, the input instance is presented to all classifiers, so that each one of them outputs a confidence degree r_{ij} and $r_{ji} \in [0,1]$ in favor of their couple of classes C_i and C_j (usually $r_{ji} = 1 - r_{ij}$). Then, these confidence degrees are set within a score-matrix:

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

It is necessary an additional phase to combine the confidence degrees of each single classifier. Different aggregation methods have been proposed in order to determine the final class [2]. The simplest aggregation is the voting strategy, where each classifier contributes with a vote for its predicted class. The class with the largest number of votes is the final predicted class. However, in our case we aim to benefit from the characteristics of fuzzy classifiers to make use of the framework of fuzzy preference relations for classification [26] as it will be explained in section III-B.

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

B. Fuzzy Rule Based Classification Systems

The usual data set of classification examples used for learning a FRBCS consists of m training patterns $x_p = (x_{p1}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$ from M classes where x_{pi} is the ith attribute value ($i = 1, 2, \ldots, n$) of the p-th training pattern.

In this work we use fuzzy rules of the following form for our FRBCSs:

$$\begin{array}{l} \text{Rule } R_j: \text{ If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \\ \text{then Class} = C_j \text{ with } RW_j \end{array} \qquad (2)$$

where R_j is the label of the jth rule, $x = (x_1, \ldots, x_n)$ is an n-dimensional pattern vector, A_{ji} is an antecedent fuzzy set, C_j is a class label, and RW_j is the rule weight[27]. We use triangular MFs as antecedent fuzzy sets.

C. Fuzzy rules generation algorithm

In our KB learning method, it is necessary to use a RB derivation method. As mentioned in the previous section we will use the Chi et al.'s rule generation algorithm[16]. This method is an extension of the well-known Wang and Mendel method [28] to classification problems. To generate the RB, this method establishes an association between the space of the features and the space of the classes by means of the following steps:

1) *Establishment of the linguistic partitions.* Once the domain of variation of each feature A_i is determined, the fuzzy partitions are computed.
2) *Generation of a fuzzy rule for each example* $x_p = (x_{p1}, \ldots, x_{pn}, C_p)$. To do this it is necessary:
   - 2.1 To compute the matching degree $\mu(x_p)$ of the example to the different fuzzy regions using a conjunction operator (usually modeled with a minimum or product T-norm).
   - 2.2 To assign the example x_p to the fuzzy region with the greatest membership degree.
   - 2.3 To generate a rule for the example, whose antecedent is determined by the selected fuzzy region and whose consequent is the label of class of the example.
   - 2.4 To compute the rule weight.

Notice that rules with the same antecedent can be generated during the learning process. If they have the same class in the consequent we just remove one of the duplicated rules, but if they have a different class only the rule with the highest weight is kept in the RB.

III. OVO strategy using FRBCSs with granularity learning

In this section, we describe the proposed method for learning the FRBCS KB of each binary classifier that form the set of classifiers of the OVO scheme and the aggregation method used for compute the final class prediction. We denote our proposal as CHC-GL-OVO (CHC for Granularity Learning in OVO scheme)

A. Genetic Algorithm for learning the KB

Any optimization/search algorithm can be used for our learning approach. In our case, we have considered a GA [15], and more specifically, a binary-coded CHC algorithm [29] as a robust model in accordance with its tradeoff between exploration and exploitation.

The individuals of the GA codify the granularity level of each feature. For evaluating every chromosome, a FRBCS is generated. First, the DB is built considering the number of labels codified. Uniform partitions with triangular membership functions are considered due to its simplicity. Next, we use an efficient method that derives the fuzzy classification rules and then the whole KB is obtained. We must recall from the previous section that the RB learning algorithm used in this work is the method proposed in [16], that we have called the Chi et al.'s rule generation method.

Several RB learning methods, including the Chi et al.'s algorithm, tends to generate too many rules when the number of labels per variable is high, thus leading sometimes to a certain overfitting to the training data-set used for the learning process. In order to avoid that problem, our GA try to learn compact KBs by penalizing the FRBCSs with high number of rules as it will be explained in Section III-A3.

The basic structure of the proposed GA is presented in figure 1. Next, we describe the components of the GA integrated in CHC-GL-OVO.

```
Initialize population
Evaluate initial population
While ((number_evaluations < MAX_Evaluations) and
(Number_restart_processes_done < 3)) do
Begin
        Selection
        Crossover
        Evaluation
        Elitism
        If (no individuals can be selected for recombination)
                then Begin
                Threshold - -
                If Threshold < 0 then Begin
                        Restart
                        Number_restart_processes_done ++
                        End
                End
                else Number_restart_processes_done = 0
End
```

Fig. 1. CHC algorithm scheme

1) Coding scheme: An integer coding approach is considered, with a chromosome length equal to the number of features in the data set. Each value stands for the number of fuzzy partitions to be used in each input variable. In this contribution, the possible values considered are taken from the set $\{2,\ldots,7\}$.

If g_i is the value that represents the granularity of variable i, a graphical representation of the chromosome is shown next:

$$C = (g_1, g_2, \ldots, g_N)$$

2) Initial Gene Pool: The initial population is composed of two parts. The generation of the initial gene pool is described next:

- In the first group all the chromosomes have the same granularity in all its variables. This group is composed of $\#val$ chromosomes, with $\#val$ being the cardinality of the significant term set, in our case $\#val = 6$, corresponding to the six possibilities for the number of labels, $2\ldots7$. For these six possible granularity levels, one individual is created.
- The second part is composed for the remaining chromosomes, and all of their components are randomly selected among the possible values.

3) evaluating the chromosome: There are three steps that must be done to evaluate each individual:

- Define the whole DB using the granularity level encoded in the chromosome. For all the features, a uniform fuzzy partition with triangular membership functions is built considering the specific number of labels of the variable (g_i).
- Generate the fuzzy rules by running the the Chi et al.'s method using the DB obtained.
- Calculate the value of the evaluation function: The usual value for the chromosome fitness in this type of genetic learning is to choose an accuracy measure over the training data-set, like the accuracy rate. However, as mentioned before, we will lightly penalize FRBCSs with a high number of rules in order to avoid the possible overfitting. To do that, once the RB has been generated and its accuracy rate (Acc) over the training set has been calculated, the fitness function to be minimized is:

  $$F_C = \omega_1 \cdot (1 - Acc) + \omega_2 \cdot N_R$$

  being N_R the number of rules of the FRBCS and $\omega_1 \in [0, 1]$.
  In order to normalize these two values, we calculate ω_2 taking two values as a base: the Acc of the FRBCS obtained with the RB generation method considering the DB with the maximum number of labels in all the variables (7 in our case) and the number of rules of this RB (N_rules):

  $$\omega_2 = \alpha_{\omega_2} \cdot \frac{Acc_{max\_g}}{N\_rules_{max\_g}}$$

  with $\alpha_{\omega_2} = 1 - \omega_1$

4) Selection: This genetic model makes use of a mechanism of "Selection of Populations". M parents and their corresponding offspring are put together to select the best M individuals to take part in the next population (with M being the population size).

5) Crossover: This operator allows one to combine two chromosomes of the population to generate their offspring. The standard crossover operator in one point is applied. This operator performs as follows. A crossover point p is randomly generated (the possible values for p are $\{2,\ldots,N\}$) and the two parents are crossed at the p-th variable.

6) Incest prevention: It promotes diversity among solutions (which is important to properly search the whole search space). Two parents are crossed if their distance divided by 2 is above a predetermined threshold T, which is initially computed as $N/4$ being N the length of the chromosome. If no individuals are recombined, then the threshold value is reduced by one.

If C_1 and C_2 are the two chromosome to recombine:

$$C_1 = (g_1, g_2, \ldots, g_N)$$

$$C_2 = (h_1, h_2, \ldots, h_N)$$

The distance measure used in this paper ($Dist$)is calculated by:

$$Dist = \sum abs(g_i - h_i) \quad i : 1..N$$

7) Restarting approach: The mutation operator is replaced by this mechanism in order to get away from local optima. When the threshold value T is zero, the best chromosome is maintained (elitist scheme) and used as a temple from generate at random new chromosomes by randomly changing the 35% of the genes.

B. Aggregation method for the OVO decomposition

As mentioned in the previous section, we make use of the fuzzy preference relations for aggregating the outputs of each binary classifier. In this scheme, the classification problem is translated into a decision making problem for determining the final predicted class among all predictions for the binary classifiers. Specifically, in this paper we consider the use of a maximal *Non-Dominance Criterion (ND)* [6] for the final decision process. This method predicts the class which is less dominated by all the remaining classes:

$$Class = \arg \max_{i=1,\ldots,m} \left\{ 1 - \sup_{j \in C} r'_{ji} \right\} \quad (3)$$

where r'_{ji} corresponds to the normalized and strict score-matrix.

IV. Experimental Study

In this section, we will first provide details of the multi-class problems chosen for the experimentation (subsection IV-A). Then, we will introduce the algorithms selected for comparison and the configuration parameters (subsection IV-B). Next, we will describe the statistical tests applied to compare the results obtained along the experimental study (subsection IV-C). Finally, we show the results obtained for all the methods and the statistical analysis (subsection IV-D).

A. Data-sets

We have used nineteen data-sets from KEEL data-set repository [17], so that the same data partitions can used by other researchers. In order to correct the data-set shift [30], [31], [32], situation in which the training data set and the test data set do not follow the same distribution, we do not use the the commonly used cross-validation scheme. We will employ a recently published partitioning procedure called Distribution Optimally Balanced Cross Validation [33].

Table I summarizes the characteristics of these data-sets. There are different imbalance ratios, from totally balanced data-sets to highly imbalanced ones, besides the different number of classes. Some of the largest data-sets (page-blocks, penbased, satimage and thyroid) were stratified sampled at 10% in order to reduce the computational time required for training. In the case of missing values (autos and cleveland), we removed those instances from the data-set before doing the partitions.

TABLE I. Summary description of data-sets.

Data-set	#Ex.	#Atts.	#Num.	#Nom.	#Cl.
Balance	625	4	4	0	3
Contraceptive	1473	9	9	0	3
Hayes-roth	132	4	4	0	3
Iris	150	4	4	0	3
NewThyroid	215	5	5	0	3
Tae	151	5	5	0	3
Thyroid	720	21	21	0	3
Wine	178	13	13	0	3
Vehicle	846	18	18	0	4
Cleveland	297	13	13	0	5
Page-blocks	548	10	10	0	5
Autos	159	25	15	10	6
Glass	214	9	9	0	7
Satimage	643	36	36	0	7
Segment	2310	19	19	0	7
Ecoli	336	7	7	0	8
Penbased	1100	16	16	0	10
Yeast	1484	8	8	0	10
Vowel	990	13	13	0	11

B. Algorithms of comparison and parameters

We will analyze the influence of granularity learning by means of a comparison between the performance of CHC-GL-OVO and two other related methods:

- The original Chi et al.'s method [16], that needs of the existence of a previous definition for the DB, normally uniform fuzzy partitions with the same number of labels in all the variables. Therefore, it is necessary to choose a granularity level, being the usual values employed for any standard FRBCS approach in the specialized literature 3, 5 and 7 labels per variable. the best results were obtained with three variables. So, we use that value for building the FRBCS. We call this method Chi-G3.
- The OVO strategy using the Chi et al.'s algorithm for determining every FRBCS (denoted Chi-OVO). In this case, we also consider three labels per feature and the

same aggregation method than our proposal, the *Non-Dominance Criterion* explained in section III-B.

The configuration for the FRBCSs is presented in Table II being "Conjuction operator" the operator used to compute the compatibility degree of the example with the antecedent of the rule and the operator used to compute the compatibility degree and the rule weight.

TABLE II. CONFIGURATION FOR THE FRBCS

Conjunction operator:	Product T-norm
Rule Weight:	Penalized Certainty Factor [27]
Fuzzy Reasoning Method:	Winning Rule

The specific parameters setting for the GA of CHC-GL-OVO is listed below, being N the number of variables:

- Number of evaluations: $500 \cdot N$
- Population Size: 60 individuals
- Fitness function weights: ($\omega_1 = 0.8$, $\alpha_{\omega_2} = 0.2$)

C. Statistical tests for performance comparison

In order to carry out the comparison of the classifiers appropriately, non-parametric tests should be considered, according to the recommendations made in [18], [19]. In this contribution, we will consider the Friedman Aligned test for both computing the ranking of the algorithms according to its performance, and the p-value that determines significant differences among the results. Then, we will proceed with a Holm non-parametric statistical procedure for $1 \cdot n$ comparisons, obtaining the adjusted p-value (APV) associated with each comparison, which represents the lowest level of significance of a hypothesis that results in a rejection. Additionally, in order to perform comparisons between two algorithms, we will use the Wilcoxon paired signed-rank test [34]. Any interested reader can find additional information on the thematic website http://sci2s.ugr.es/sicidm/, where software for the application of the statistical tests is provided.

D. Experimental Analysis

Table III shows the results in performance (using the accuracy metric) for CHC-GA-OVO and the algorithms employed for comparison, that is, Chi-G3 and Chi-OVO, being Tr the accuracy over the training data-set and Tst the accuracy over the test data-set.

As it can be observed, the prediction ability obtained by CHC-GA-OVO is higher than the obtained for the other methods, showing the significative influence of the granularity level in the behavior of the classifier regarding to the classical way to proceed in both possibilities, with and without an OvO decomposition. It can be seen that there are data-sets in which Chi-OvO obtains results clearly worse than Chi-G3 (thyroid, page-blocks, balance) and data-sets with the opposite behavior, Chi-OVO clearly better than Chi-G3 (satimage, yeast, vowel, cleveland). Note that CHC-GA-OVO obtains similar results to the best of the other two algorithms in both cases, showing the robustness of the method.

In order to validate these results, we show the ranking on precision of the different models by means of the procedure described in subsection IV-C. Table IV-D shows the P-values obtained in by applying post hoc methods over the results of Friedman Alligned procedure.

TABLE IV. POST HOC COMPARISON TABLE FOR $\alpha = 0.05$ (FRIEDMAN ALLIGNED)

i	algorithm	$z = (R_0 - R_i)/SE$	p	Holm
2	Chi-G3	3.757887	0.000171	0.025
1	Chi-OVO	1.944914	0.051785	0.05

Next, we perform a sign test and a Wilcoxon test for detecting significant differences between the results of CHC-GL-OVO and the other two approaches. The results of these tests are shown in Table IV-D where, by columns, it is represented the current comparison, the number of wins, ties and loses for the CHC-GL-OVO method versus the standard FRBCS approaches, the sum of the ranks for CHC-GL-OVO and the other methods respectively, and the p-values obtained, first by the sign test, and second by the Wilcoxon test.

TABLE V. SIGN AND WILCOXON TESTS TO COMPARE CHC-GL-OVO [R^+] WITH THE OTHER METHODS [R^-] REGARDING THE ACCURACY RATE METRIC

VS	R^+	R^-	Exact P-value	Asymptotic P-value
Chi-OVO	146.0	25.0	0.006576	0.007894
Chi-G3	170.0	20.0	0.0014114	0.002379

V. CONCLUSIONS

This contribution has described a learning process for multi-class problems following the One-vs-One decomposition strategy that aggregates the outputs of the binary classifiers obtained for each pair of classes. We have used FRBCSs as classifiers. A stationary GA based on the well-known CHC algorithm is used for granularity learning, which is combined with an efficient fuzzy classification rule generation method to obtain the complete KB of each binary-class FRBCS.

Our proposal try to find a good granularity level that outperform the prediction ability of the classifier and it is compared with the classical way to design FRBCS, that is, without granularity learning, with and without OVO decomposition. The proposed method obtains better results in accuracy rate than the classical approaches in the majority od data-sets considered, showing significative differences according the non-parametric statistical test.

We must remark that one advantage of the learning process proposed is that the GA can be combined with any fuzzy rules generation method. We have used a basic algorithm for efficiency but more accurate ones can be used, or another more suitable for a specific data-set.

In future works, we will try to adjust the learning process in order to improve the results and to decrease the computational time of the GA.

ACKNOWLEDGMENT

The authors would like to thank...

TABLE III. EXPERIMENTAL RESULTS IN TRAINING AND TEST WITH THE STANDARD ACCURACY METRIC. FROM THE LEFTMOST TO THE RIGHTMOST COLUMN WE SHOW THE RESULTS FOR THE STANDARD CHI ET AL.'S ALGORITHM WITH 3 LABELS (CHI-G3), THE PAIRWISE LEARNING APPROACH (CHI-OVO) AND OUR PROPOSED OVO GRANULARITY LEARNING APPROACH (CHC-GL-OVO). THE HIGHEST PERFORMANCE VALUE PER DATASET IS STRESSED IN BOLDFACE.

Dataset	Chi-G3-tra	Chi-G3-tst	Chi-OVO-tra	Chi-OVO-tst	CHC-GL-OVO-tra	CHC-GL-OVO-tst
autos	91.99	61.09	97.66	64.81	99.68	**65.62**
balance	91.56	**90.24**	84.84	80.18	91.16	88.63
cleveland	92.17	38.39	94.95	**53.88**	94.94	52.22
contraceptive	51.93	40.05	59.18	46.37	77.16	**48.20**
ecoli	75.83	72.39	84.00	78.07	90.63	**80.51**
glass	66.24	59.02	73.38	59.86	84.94	**63.02**
hayes	78.75	64.97	91.41	64.38	88.12	**68.07**
iris	93.67	93.33	96.33	**96.00**	96.67	94.67
newthyroid	85.93	84.65	95.35	**93.02**	94.30	**93.02**
page-blocks	92.06	**91.98**	79.17	79.06	90.94	89.67
penbased	98.24	**97.85**	98.50	98.05	94.14	90.50
satimage	48.32	48.28	74.41	71.98	93.54	**77.14**
segment	87.10	86.19	92.93	**91.08**	92.14	90.74
tae	61.44	54.18	64.60	57.12	76.50	**57.88**
thyroid	92.97	92.13	53.07	52.55	93.38	**92.24**
vehicle	66.11	61.36	73.23	62.43	94.39	**66.79**
vowel	55.73	53.23	92.70	89.49	96.72	**93.03**
wine	98.59	92.15	98.59	91.52	98.03	**93.84**
yeast	29.68	28.98	57.26	55.21	62.47	**57.62**
Z_AVG	76.75	68.97	82.19	72.90	89.99	**77.02**

REFERENCES

[1] 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.

[2] M. Galar, A. Fernandez, 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] E. L. Allwein, R. E. Schapire, and Y. Singer, *Reducing multiclass to binary: A unifying approach for margin classifiers*, Journal of Machine Learning Research 1: 113-141, 2000.

[4] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, *Dynamic classifier selection for one-vs-one strategy: Avoiding non-competent classifiers* Pattern Recognition 46(12): 3412-3424, 2013.

[5] H. Ishibuchi, T. Nakashima and M. Nii, *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining*, Springer–Verlag, 2004.

[6] A. Fernandez, M. Calderon, E. Barrenechea, H. Bustince, Humberto 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.

[7] S. Elhag, A. Fernandez. A. Bawakid, S. Alshomrani and F. Herrera, *On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on Intrusion Detection Systems*, Expert Systems with Applications 42(1): 193–202, 2015.

[8] M. Elkano, M. Galar, J. Sanz, A. Fernandez, 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, 2015. In press

[9] O. Cordón, F. Herrera, and P. Villar, *Analysis and guidelines to obtain a good uniform fuzzy partition granularity for fuzzy rule–based systems using simulated annealing*, International Journal of Approximate Reasoning 25(3): 187–215, 2000.

[10] O. Cordón, F. Herrera, and P. Villar, *Generating the knowledge base of a fuzzy rule–based system by the genetic learning of the data base*, IEEE Transactions on Fuzzy Systems 9(4): 667–674, 2001.

[11] O. Cordón, F. Herrera, L. Magdalena, and P. Villar, *"A genetic learning process for the scaling factors, granularity and contexts of the fuzzy rule–based system data base*, Information Sciences 136: 85–107, 2001.

[12] E. Zhou and A. Khotanzad, *Fuzzy classifier design using genetic algorithms*, Pattern Recognition 40(12): 3401–3414, 2007.

[13] I. Walter and F. Gomide, *Genetic fuzzy systems to evolve interaction strategies in multiagent systems*, International Journal of Intelligent Systems 22(9): 971–991, 2007.

[14] P. Villar, A. Fernández, R. Carrasco, F. Herrera, *Feature Selection and Granularity Learning in Genetic Fuzzy Rule-Based Classification Systems for Highly Imbalanced Data-Sets*, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 20(3): 369–397, 2012.

[15] D.E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, 1989.

[16] Z. Chi, H. Yan, and T. Pham, *Fuzzy algorithms with applications to image processing and pattern recognition*, World Scientific, 1996.

[17] 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 Multi-Valued Logic and Soft Computing, 17(2-3): 255–287, 2011.

[18] J. Demšar, *Statistical comparisons of classifiers over multiple data sets*, Journal of Machine Learning Research 7: 1–30, 2006.

[19] S. García and F. Herrera, *An extension on statistical comparisons of classifiers over multiple data sets for all pairwise comparisons*, Journal of Machine Learning Research 9: 2677–2694, 2008.

[20] 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 13(10): 959–977, 2009.

[21] T. Hastie and R. Tibshirani, *Classification by pairwise coupling*, The Annals of Statistics 26(2): 451–471, 1998.

[22] P. Clark and R. Boswell, *Rule Induction with CN2: Some Recent Improvements*, in: EWSL 1991, Lecture Notes in Computer Science 482: 151-163, 1991.

[23] J. Alcala–Fdez, L. Sanchez, S. García, M.J. del Jesus, S. Ventura, J.M. Garrell, J. Otero, C. Romero, J. Bacardit, V.M. Rivas, J. Fernández and F. Herrera, *KEEL: a software tool to assess evolutionary algorithms for data mining problems*, Soft Computing, 13(3): 307–318, 2009.

[24] C.-C. Chang and C.-J. Lin, *LIBSVM: A library for support vector*

machines, ACM Transactions on Intelligent Systems and Technology 2: 1–27, 2011. http://www.csie.ntu.edu.tw/~cjlin/libsvm

[25] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I. H. Witten, *The WEKA Data Mining Software: An Update*, SIGKDD Explorations 11(1): 10–18, 2009.

[26] E. Hüllermeier and K. Brinker, *Learning valued preference structures for solving classification problems*, Fuzzy Sets and Systems 159(18): 2337–2352, 2008.

[27] H. Ishibuchi and T. Yamamoto, *Rule weight specification in fuzzy rule–based classification systems*, IEEE Transactions on Fuzzy Systems 13: 428–435, 2005.

[28] L. X. Wang and J. M. Mendel, *Generating fuzzy rules by learning from examples*,IEEE Transactions on System, Man and Cybernetics 25(2): 353–361, 1992.

[29] L. J. Eshelman, *The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination*, in Foundations of Genetic Algorithms: 265–283, ed. G. Rawlin, Morgan Kaufman, 1991.

[30] J. G. Moreno-Torres, T. Raeder, R. Aláiz-Rodríguez, N. V. Chawla and F. Herrera, *A unifying view on dataset shift in classification*, Pattern Recognition 45(1): 521–530, 2012

[31] J. Quionero, M. Sugiyama, A. Schwaighofer and N. D. Lawrence, *Dataset Shift in Machine Learning*, The MIT Press, 2009.

[32] V. Lopez, A. Fernandez, 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.

[33] J. G. Moreno-Torres and J. A. Sáez 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): 1304–1313, 2012.

[34] D. Sheskin, *Handbook of parametric and nonparametric statistical procedures*, Chapman & Hall/CRC, 2006.