On the combination of pairwise and granularity learning for improving fuzzy rule based classification systems: GL-FARCHD-OVO

Pedro Villar[1], Alberto Fernández[2], and Francisco Herrera[3]

[1] Dep. of Software Engineering, University of Granada, Granada, Spain
pvillarc@ugr.es,
[2] Dep. of Computer Science, University of Jaén, Jaén, Spain
alberto.fernandez@ujaen.es
[3] Dep. of Computer Science and Artificial Intelligence, University of Granada, Granada, Spain
herrera@decsai.ugr.es

Abstract. Fuzzy rule-based systems constitute a very spread tool for classification problems, but several proposals may decrease its performance when dealing with multi-class problems. Among existing approaches, the FARC-HD algorithm has excelled as it has shown to achieve accurate and compact classifiers, even in the context of multi-class problems. In this work, we aim to go one step further to improve the behavior of the former algorithm by means of a "divide-and-conquer" approach, via binarization in a one-vs-one scheme. Besides, we will contextualize each binary classifier by adapting the data base for each subproblem by means of a granularity learning process to adapt the number of fuzzy labels per variable. Our experimental study, using several data-sets from KEEL data-set repository, shows the goodness of the proposed methodology.

Keywords: multi-class classification, fuzzy association rules, One-vs-One decomposition, genetic glgorithms, granularity learning

1 Introduction

Fuzzy rule-based systems is a commonly used tool for classification problems. An advantage of fuzzy systems is the interpretability of the generated model, especially when fuzzy sets are represented by linguistic labels [12]. We will make use of linguistic fuzzy association rule-based classification systems, being one of the most robust method proposed in this field the FARC-HD algorithm (Fuzzy Association Rule-based Classification model for High Dimensional problems) [1], that obtains models with high prediction ability and low complexity. This method is based on fuzzy association rules for classification, where the antecedent of the rule is a combination of fuzzy labels and the consequent is a class label. The possible values for the antecedents are sets of fuzzy labels of linguistic partitions defined over the attribute's universe of discourse. FARC-HD considers a fuzzy Data Base (DB) with a fixed number of labels for all fuzzy partitions. In the last step of the algorithm the DB definition is adjusted by a lateral tuning, but maintaining the predefined granularity level for variable. This is the usual way to proceed in the majority of fuzzy systems learning methods, i.e. it is necessary to select a priori a number of labels for all fuzzy partitions. However, the granularity level has a significant influence on the Fuzzy Systems performance as it has been analyzed in [7]. Some learning methods proposed in fuzzy modeling and fuzzy classification include the granularity level learning [8, 22].

Classification problems with more than two classes (multi-class problems) are known to present more difficulties than binary-class problems. A robust solution to cope with the former problem is to use a decomposition approach [13]. 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 [16]. This method divides the original problem by confronting all pairs of classes against them. Then, an independent 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 [13, 14].

The main purpose of this contribution is to improve the performance of FARC-HD in multi-class problems, using an OVO scheme built with fuzzy association classifiers generated by learning an appropriate granularity level for each attribute in the FARC-HD method. We think that the optimal granularity level for learning each pair of classes will be different and so, looking for a good number of labels for each binary classifier can contribute to obtain better prediction ability in this decomposition scheme (OVO) for multi-class problems. To do so, we employ an approach to derive the Fuzzy classifier that involves the use of two different (and independent) learning processes, in which a DB definition process wraps the FARC-HD algorithm. We use a Genetic Algorithm (GA) for the granularity learning. A similar learning scheme was performed in [22] to design Fuzzy Rule-Based Classification Systems 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 learning process mentioned for FARC-HD, we will contrast the obtained results with the FARC-HD algorithm with the application of an OVO decomposition using the Cohen's kappa measure, that equilibrates the performance in each individual class, independently of the number of examples of every class. We have selected a collection of multi-class data-sets from KEEL data-set repository [2] for developing our experimental analysis. Furthermore, we will perform a statistical analysis using non-parametric tests [15] to find significant differences among the obtained results.

This paper is organized as follows. First, Section 2 introduces the preliminary concepts used in this paper. Next, in Section 3 we will describe our proposal, an OVO strategy for FARC-HD designed using a GA for granularity learning. Then, Section 4 describes the experimental study. Finally, in Section 5, some conclusions will be pointed out.

2 Preliminaries

This section provides a description of the working procedure of FARC-HD joint with some basic concepts about fuzzy association rules (Section 2.1). Next, a brief introduction of the OVO scheme is given (Section 2.2). Finally, we present the metric of performance used in this paper, i.e. the kappa metric (Section 2.3).

2.1 FARC-HD learning method

In this paper we have make use of a robust fuzzy model known as FARC-HD [1]. This algorithm is based on association discovery, a commonly used technique in data mining for extract interesting knowledge from large data-sets by means of finding relationships between the different items in a database. The integration between association discovery and classification leads to precise and interpretable models. FARC-HD is aimed at obtaining an accurate and compact fuzzy rule-based classifier with a low computational cost.

The usual data set of classification examples used for learning a Classifier 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 the Fuzzy Association rules used in FARC-HD: R_i : $If\ A\ then\ Class = C_k$, where R_i is the label of the ith rule, A is a set of labels of the attribute's fuzzy partitions and C_k is a class label. We use triangular membership functions as antecedent fuzzy sets as it is used in FARC-HD. In short, this method is based on the following three stages:

Stage 1. *Fuzzy association rule extraction for classification*: A search tree is employed to list all possible frequent fuzzy item sets and to generate fuzzy association rules for classification, limiting the depth of the branches in order to find a small number of short (i.e., simple) fuzzy rules.
Stage 2. *Candidate rule pre-screening*: Afterwards the rule generation, the size of the rule set can be too large to be interpretable by the end user. Therefore, a pre-selection of the most interesting rules is carried out by means of a "subgroup discovery" mechanism based on an improved weighted relative accuracy measure (wWRAcc').
Stage 3. *Genetic rule selection and lateral tuning*: Finally, in order to obtain a compact and accurate set of rules within the context of each problem, an evolutionary process will be carried out in a combination for the selection of the rules with a tuning of membership function.

2.2 One-vs-One decomposition

The most common approaches for decomposition a multi-class problem into a binary-class problem are OVO [16] and OVA (One-vs-All) [5]. The former learns a binary classifier for each possible pair of classes, whereas the latter constructs a binary classifier considering each single class and all the other classes joined. OVA approaches are easier to apply and they have shown to be an interesting scheme for one-class classifiers [9, 18]. However, in a general framework the goodness of the OVO scheme versus the former have been experimentally proven [13]. 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:

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

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 [13]. The simplest aggregation is the voting strategy, where each classifier contributes with a vote for its 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 [17] as it will be explained in section 3.2.

2.3 Performance metric: Cohen's kappa index

Cohen's kappa is an alternative measure to *classification rate*, since it compensates for random hits [6, 4]. In contrast to classification rate, kappa evaluates the portion of hits that can be attributed to the classifier itself (i.e., not to mere chance), relative to all the classifications that cannot be attributed to chance alone. An easy way of computing Cohen's kappa is by making use of the resulting confusion matrix (Table 1) in a classification task. With the expression (2), we can obtain Cohen's kappa:

Table 1. Confusion Matrix for an n-class problem

Correct Class	Predicted Class				
	C_1	C_2	$\ldots$	C_m	Total
C_1	h_{11}	h_{12}	$\ldots$	h_{1m}	T_{r1}
C_2	h_{12}	h_{22}	$\ldots$	h_{2m}	T_{r2}
$\vdots$			$\ddots$		$\vdots$
C_m	h_{1m}	h_{2m}	$\ldots$	h_{mm}	T_{rm}
Total	T_{c1}	T_{c2}	$\ldots$	T_{cm}	T

$$kappa = \frac{n\sum_{i=1}^{m} h_{ii} - \sum_{i=1}^{m} T_{ri}T_{ci}}{n^2 - \sum_{i=1}^{m} T_{ri}T_{ci}}, \quad (2)$$

where h_{ii} is the cell count in the main diagonal (the number of true positives for each class), n is the number of examples, m is the number of class labels, and T_{ri}, T_{ci} are the rows' and columns' total counts, respectively ($T_{ri} = \sum_{j=1}^{m} h_{ij}$,

$T_{ci} = \sum_{j=1}^{m} h_{ji}$). Cohen's kappa ranges from -1 (total disagreement) through 0 (random classification) to 1 (perfect agreement). Being a scalar, it is less expressive than the ROC curves applied to binary-class cases. However, for multi-class problems, kappa is a very useful, yet simple, meter for measuring a classifier's classification rate while compensating for random successes.

3 OVO strategy using FARC-HD with granularity learning: GL-FARCHD-OVO

In this section, we describe the proposed method for learning the Fuzzy association rule-based model 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 GL-FARCHD-OVO (Granularity Learning for FARC-HD in an OVO scheme).

3.1 Genetic Algorithm for granularity learning in FARC-HD

Any optimization/search algorithm can be used for our learning approach. In our case, we have considered a GA, and more specifically, a integer-coded CHC algorithm [10] 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 individual, first, the fuzzy partitions are built considering the number of labels codified in the chromosome. Uniform partitions with triangular membership functions are chosen as in FARC-HD.

As we employ a GA for determining a good granularity level for variable, we need to run the FARC-HD algorithm in the evaluation of each chromosome, that also includes a GA in the last tuning stage. In order to decrease the computational cost of the proposed method, only the two first two stages of FARC-HD are executed in the chromosome evaluation. The last stage of FARC-HD is executed once, over the best granularity level configuration found by the GA of our proposal. Next, we describe the components of the GA for granularity learning.

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

Initial Gene Pool. The initial population is composed of two parts. In the first group all the chromosomes have the same granularity in all its variables. This group is composed of v chromosomes, with v being the cardinality of the significant term set, in our case $v = 6$, corresponding to the six possibilities for the number of labels, $\{2 \ldots 7\}$. 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.

Evaluation of the chromosome, composed of three steps:

1. Define the 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).
2. Generate the fuzzy association rules by running the FARC-HD method using the DB obtained. We must remark that only the two first stages of FARC-HD are executed, providing a rule base composed of "interesting" rules.
3. Calculate the fitness value, that it is the kappa index of the rule base obtained in the previous step over the training data set.

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

Crossover This operator combines two chromosomes of the population to generate their offspring. The standard crossover operator in one point is applied, which works 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.

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:

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

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 and used as a template from generate at random new chromosomes by randomly changing the 35% of the genes.

3.2 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)* [11] 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.

4 Experimental Study

We have used twenty multi-class data-sets from KEEL data-set repository[4] [2]. In order to correct the data-set shift [20], situation in which the training data set and the test data set do not follow the same distribution, we do not use the commonly used cross-validation scheme. We will employ a recently published partitioning procedure called Distribution Optimally Balanced Cross Validation [19] with five different partitions for each data-set. Table 2 summarizes the characteristics of these data-sets: number of examples, number of attributes and number of classes. 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, shuttle 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 2. Summary description of data-sets.

Data-set	**#Ex.**	**#Atts.**	**#Cl.**	**Data-set**	**#Ex.**	**#Atts.**	**#Cl.**
balance	625	4	3	page-blocks	548	10	5
contraceptive	1473	9	3	autos	159	25	6
hayes-roth	132	4	3	shuttle	5800	9	7
iris	150	4	3	glass	214	9	7
newThyroid	215	5	3	satimage	643	36	7
tae	151	5	3	segment	2310	19	7
thyroid	720	21	3	ecoli	336	7	8
wine	178	13	3	penbased	1100	16	10
vehicle	846	18	4	yeast	1484	8	10
cleveland	297	13	5	vowel	990	13	11

We will analyze the influence of granularity learning by means of a comparison between the performance of GL-FARCHD-OVO and the original FARC-HD method used in an OVO strategy. The original fitness function of the GA performed in the third stage of FARC-HD has been modified changing the accuracy rate for the kappa index. The configuration and parameters for FARC-HD are the ones suggested in its seminal paper [1] and they are presented in Table 3 being "Conjuction operator" the operator used to compute the compatibility degree of the example with the antecedent of the rule. FARC-HD needs also a predefined number of labels for all the fuzzy partitions, we have used 5 as granularity level, as suggested in [1]. In the execution of the two first stages of FARC-HD performed in the GA for learning the granularity level, we have used the same parameters except the depth of the trees, that it reduced to 2, in order to go down the computational cost of the GA proposed. We remark that the

[4] `http://www.keel.es/dataset.php`

final step of GL-FARCHD-OVO is the execution of the stage 3 of FARC-HD over the best individual found by the GA.

Table 3. Parameters of FARC-HD

Conjunction operator:	Product T-norm	Parameter K of the prescreening:	2
Fuzzy Reasoning Method:	Additive combination	Maximum evaluations:	15000
Minimum Support:	0.05	Population size:	50
Maximum Confidence:	0.8	Parameter alpha:	0.15
Depth of the trees:	3	Bits per gen:	30

The specific parameters setting for the GA of GL-FARCHD-OVO are 60 individuals and $100 \cdot N$ number of evaluations, being N the number of variables. In order to carry out the comparison of the classifiers appropriately, non-parametric tests should be considered, according to the recommendations made in [15]. We will use the Wilcoxon paired signed-rank test [21] to perform comparisons between the two algorithms executed.

Table 4 shows the results in performance (using the performance metric) for GL-FARCHD-OVO and FARCHD-OVO, being *tra* the kappa index over the training data-set and *tst* the kappa index over the test data-set. The highest performance value for each test data-set is stressed in boldface. As it can be observed, the values obtained by GL-FARCHD-OVO are higher than the obtained for FACHD-OVO, showing the influence of the granularity level in the behavior of the classifier regarding to the classical way to proceed (with a predefined number of labels, the same for all the attributes).

Table 4. Experimental results in training and test with the kappa metric

	FARCHD-OVO		GL-FARCHD-OVO			FARCHD-OVO		GL-FARCHD-OVO	
Dataset	tra	tst	tra	tst	Dataset	tra	tst	tra	tst
balance	0.846	0.682	0.808	**0.710**	page-blocks	0.774	0.554	0.787	**0.563**
contraceptive	0.468	0.268	0.425	**0.287**	autos	0.986	0.708	0.984	**0.737**
hayes	0.826	0.663	0.868	**0.672**	shuttle	0.827	0.824	0.994	**0.990**
iris	0.975	0.920	0.988	**0.930**	glass	0.850	**0.571**	0.830	0.560
newthyroid	0.993	0.861	0.998	**0.900**	satimage	0.832	0.717	0.849	**0.751**
tae	0.697	0.337	0.680	**0.344**	segment	0.941	0.920	0.959	**0.936**
thyroid	0.530	0.368	0.485	**0.401**	ecoli	0.921	**0.771**	0.926	0.769
wine	1.000	0.906	1.000	**0.932**	penbased	0.990	0.899	0.989	**0.903**
vehicle	0.811	**0.636**	0.820	0.600	yeast	0.590	0.478	0.585	**0.484**
cleveland	0.936	**0.325**	0.858	0.311	vowel	0.979	0.918	0.988	**0.924**

	FARCHD-OVO		GL-FARCHD-OVO	
Average	tra	tst	tra	tst
	0.839	0.666	0.841	**0.685**

In order to validate these results, we show the ranking on precision of the different models. Table 4 presents the results obtained in by applying Wilcoxon test. The p-value obtained shows significative differences between our proposed method (GL-FARCHD-OVO) and FARCHD-OVO.

Table 5. Results obtained by the Wilcoxon test for algorithm GL-FARCHD-OVO

VS	R^+	R^-	p-value	Hypothesis
FARCHD-OVO	172.0	38.0	0.010688	Rejected for GL-FARCHD-OVO

5 Conclusions

This contribution has described a learning process for multi-class problems following the OVO decomposition strategy that aggregates the outputs of the binary classifiers obtained for each pair of classes. We have used FARC-HD as learning method to build the classifiers. A stationary GA based on the well-known CHC algorithm is used for granularity learning. Our proposal uses a divide-and-conquer strategy and aims at finding a good granularity level for each pair of classes that outperform the prediction ability of the classifier and it is compared with an OVO scheme using the original FARC-HD algorithm, that is, considering a fixed granularity level. The proposed method obtains better results in performance rate in the majority of data-sets considered, showing significative differences according the non-parametric statistical test. 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.

Acknowledgments

This work was partially supported by the Spanish Ministry of Science and Technology under project TIN-2012-33856 and the Andalusian regional projects P10-TIC-06858 and P11-TIC-7765.

References

1. Alcalá-Fdez, J., Alcalá, R., Herrera, F.: 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)
2. Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: 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)
3. Allwein, E.L., Schapire, R.E., Singer, Y.: Reducing multiclass to binary: A unifying approach for margin classifiers. Journal of Machine Learning Research 1, 113–141 (2000)
4. Ben-David, A.: A lot of randomness is hiding in accuracy. Engineering Applications of Artificial Intelligence 20, 875–885 (2007)
5. Clark, P., Boswell, R.: Rule induction with cn2: Some recent improvements. In: Kodratoff, Y. (ed.) EWSL. Lecture Notes in Computer Science, vol. 482, pp. 151–163. Springer (1991)
6. Cohen, J.A.: Coefficient of agreement for nominal scales. Educational and Psychological Measurement pp. 37–46 (1960)

7. Cordón, O., Herrera, F., Villar, P.: 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)
8. Cordón, O., Herrera, F., Villar, P.: 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)
9. Cyganek, B.: One-class support vector ensembles for image segmentation and classification. Journal of Mathematical Imaging and Vision 42(2-3), 103–117 (2012)
10. Eshelman, L.J.: Foundations of Genetic Algorithms, chap. The CHC adaptive search algorithm: How to have safe search when engaging in nontraditional genetic recombination, pp. 265–283. Morgan Kaufman (1991)
11. Fernandez, A., Calderon, M., Barrenechea, E., Bustince, H., Herrera, F.: 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)
12. Gacto, M.J., Alcalá, R., Herrera, F.: Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. Information Sciences 181(20), 4340–4360 (2011)
13. Galar, M., Fernández, A., Barrenechea, E., Bustince, H., Herrera, F.: 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)
14. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F.: Dynamic classifier selection for one-vs-one strategy: Avoiding non-competent classifiers. Pattern Recognition 46(12), 3412–3424 (2013)
15. García, S., Herrera, F.: An extension on "statistical comparisons of classifiers over multiple data sets" for all pairwise comparisons. Journal of Machine Learning Research 9, 2607–2624 (2008)
16. Hastie, T., Tibshirani, R.: Classification by pairwise coupling. The Annals of Statistics 26(2), 451–471 (1998)
17. Hüllermeier, E., Brinker, K.: Learning valued preference structures for solving classification problems. Fuzzy Sets and Systems 159(18), 2337–2352 (2008)
18. Krawczyk, B., Wozniak, M., Cyganek, B.: Clustering-based ensembles for one-class classification. Information Sciences 264, 182–195 (2014)
19. Moreno-Torres, J.G., Raeder, T., Aláiz-Rodríguez, R., Chawla, N.V., Herrera, F.: A unifying view on dataset shift in classification. Pattern Recognition 45(1), 521–530 (2012)
20. Moreno-Torres, J.G., Sáez, J.A., Herrera, F.: 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)
21. Sheskin, D.: Handbook of parametric and nonparametric statistical procedures. Chapman & Hall/CRC (2006)
22. Villar, P., Fernández, A., Carrasco, R.A., Herrera, F.: 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)