A First Approach for Cost-Sensitive Classification with Linguistic Genetic Fuzzy Systems in Imbalanced Data-sets

Victoria López
Dept. of Computer Science and A.I.
University of Granada
Granada, Spain
viclm@correo.ugr.es

Alberto Fernández
Dept. of Computer Science
University of Jaén
Jaén, Spain
alberto.fernandez@ujaen.es

Francisco Herrera
Dept. of Computer Science and A.I.
University of Granada
Granada, Spain
herrera@decsai.ugr.es

***Abstract*—Classification in imbalanced domains has become one of the most relevant problems within the area of Machine Learning at the present. This problem has raised in significance due to its presence in many real applications and it occurs when the distribution of the available examples to carry out the learning process is very different between the classes (often for binary class data-sets). Usually, the underrepresented class is the concept of the most interest for the problem, being the cost derived from a misclassification of these examples much higher than that of the remaining examples.**

In this work we analyze the behaviour of a cost-sensitive learning method for Fuzzy Rule Based Classification Systems in the scenario of high imbalanced data-sets. Specifically, we focus on one representative rule learning approach for Genetic Fuzzy Systems, the Fuzzy Hybrid Genetics-Based Machine Learning algorithm.

The experimental results show how our cost-sensitive approach in this type of domains will help us to obtain very accurate solutions in shorter training times and also with a lower complexity with respect to other possibilities proposed for classification with imbalanced problems such as the use of preprocessing to rebalance the class distribution.

***Keywords*-Fuzzy Rule Based Classification Systems, Imbalanced Data-sets, Cost-Sensitive, Genetic Fuzzy Systems**

I. Introduction

In the scenario of classification problems, we distinguish imbalanced data-sets as special data-sets where the number of instances differs significantly between the classes (in a binary problem). The class imbalance is dominant in a high number of real problems including, but not limited to, telecommunications, WWW, finances, ecology, biology, medicine and so on. The positive or minority class is usually the one that has the highest interest from the learning point of view and it also implies a great cost when it is not well classified [1].

In order to address this problem, many solutions have been developed in three different levels:

- Data level solutions: the objective consists in rebalancing the class distribution by sampling the data space [2], [3]
- Algorithmic level solutions: these solutions try to adapt several classification algorithms to reinforce the learning towards the positive class [4].
- Cost-sensitive solutions: this type of solutions incorporates approaches at the data level, at the algorithmic level, or at both levels jointly, considering higher costs for the misclassification of examples of the positive class with respect to the negative class, and therefore, trying to minimize higher cost errors [5], [6], [7].

In order to design the classifier, we will use Fuzzy Rule Based Classification Systems (FRBCSs), which are a very useful tool in the field of Machine Learning since they provide interpretable models for the end user with very good accuracy results in many areas [8]. Specifically, recent works have shown that FRBCSs have a good behaviour dealing with imbalanced data-sets by means of the application of instance preprocessing techniques [9].

Given these previous issues, in this work we aim to analyze the impact of cost-sensitive learning in this type of problems, developing a cost-sensitive linguistic fuzzy learning method that allow us to obtain the Rule Base (RB) of the FRBCS. Among the different methods that have been proposed in the specialized literature for designing fuzzy rule-based systems, Genetic Fuzzy Systems (GFS) are one of the most interesting tools in order to avoid the necessity of linguistic knowledge from domain experts [10], [11]. However, one of the drawbacks of this type of methodology is that they usually require long training times due to the use of genetic algorithms, being directly dependant of the size of the training set.

In this work we have selected the Fuzzy Hybrid Genetics-Based Machine Learning (FH-GBML) algorithm [12] which has shown to be very robust for imbalanced problems when combined with the use of preprocessing [13]. In this work, we will modify the learning features of this algorithm considering the cost of the misclassification of the different examples. The results obtained with this methodology will be compared with those achieved by the standard FH-GBML algorithm and with the use of Synthetic Minority Oversampling Technique (SMOTE) [3] as preprocessing step.

In this work, we will focus on high imbalanced binary classification problems, having selected a benchmark of 22 problems from KEEL data-set repository[1] [14]. We will perform our experimental analysis focusing on the precision of the models using the Area Under the ROC curve (AUC) [15], but we will also study both the efficiency in the training stage and the interpretability of the RB measured as the number of rules. This study will be carried out using non-parametric tests to check whether there exist significant differences among the obtained results [16], [17].

This work is structured in the following way. First, Section II presents an introduction of classification with imbalanced problems, describing its features, how to address this problem, cost-sensitive learning and the metrics that are used in this framework. Next, Section III introduces the main aspects of a fuzzy learning method for building FRBCSs which will be the base of our proposal for a methodology of cost-sensitive fuzzy learning. Section IV shows the experimental study carried out. Finally, the conclusions achieved in this work are shown in Section V

II. Classification with Imbalanced Data-sets

In this section we delimit the context in which this work is content, briefly introducing the problem of imbalanced classification and the cost-sensitive learning. We finish this section describing the evaluation metrics that are used in this specific problem with respect to the most common ones in classification.

A. The problem of imbalanced classification

Learning from imbalanced data is a significant topic that has recently appeared in the context of Machine Learning [18], [19]. Specifically, we refer to imbalanced data-sets when the distribution between the classes is not uniform, being the number of examples that represents one of the classes much lower than the other, adding that the characterization of this class often has a higher practical interest. The significance of this problem relies on its presence in numerous real classification problem such as parasite detection in images [20], mine detection with radar and sonar [21] or the study of intestinal contractions [22], just citing some of them.

Standard classification algorithms from examples are often biased towards the negative class (majority class), since the rules that correctly classify a higher number of examples are selected in the learning process while increasing the considered metric (that it is often based in the percentage of well-classified examples). Hence, the instances of the positive class (minority class) are misclassified with a higher frequency than those that belong to the negative class [23]. Another important feature of this type of problems are the "small disjuncts", that is, a data concentration of one class in a small area of the problem being surrounded by examples of the contrary class [24], [25]; this type of regions are hard to detect for most of the learning algorithms. Furthermore, another main problem of imbalanced data-sets is the higher probability of overlapping between the positive and negative examples [26].

There exist different imbalance degrees between the data. In this work we will use the "imbalance ratio" (IR) [24] to distinguish among different categories. This metric is defined as the ratio between the number of examples of the negative class and the positive class. We consider that a data-set present a high degree of imbalance when its IR is higher than 9 (less than a 10% of instances of the positive class).

B. Cost-sensitive learning

Cost-sensitive learning takes into account the variable cost of a misclassification of the different classes [6], [27]. A cost matrix codifies the penalties of classifying examples of one class as a different one. Let $C(i, j)$ be the cost of predicting an instance of class i as class j; with this notation $C(+, -)$ is the cost of misclassifying a instance of the positive class as if it was negative and $C(-, +)$ is the cost of the opposite case.

When dealing with imbalanced problems it is usually of most interest to recognize the positive instances rather than the negative ones. Therefore, the cost when mistaking a positive instance is higher than the cost of mistaking a negative one ($C(+, -) > C(-, +)$). As a classical example, the reader may refer to a diagnosis problem in which it is often less dangerous to obtain a false positive than a false negative, since the patient will not obtain the treatment for his/her disease.

The cost-sensitive learning process tries to minimize the number of high cost errors and the total error of misclassification, taking into account the cost matrix during the building of the model with the aim of obtaining one with the lowest cost. Usually, for minimizing the cost the Bayes Theorem of the minimal risk is used to assign each example the class with lowest risk. Cost-sensitive learning supposes that there is a cost matrix available for the different type of errors. However, given a data-set, this matrix is not usually given [19], [28].

C. Evaluation metrics

Most of the approaches for classification in Machine Learning use some precision measure for the model as the percentage of well-classified examples (standard accuracy rate). However, this kind of metrics often lead to erroneous conclusions when we work with imbalanced data-sets since they do not take into account the proportion of examples within each class, neither the misclassification cost is included. For this reason, in this work we will use the metric denominated AUC [15], which is defined as:

$$AUC = \frac{1 + TP_{rate} - FP_{rate}}{2} \quad (1)$$

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

where TP_{rate} is the ratio of examples of the positive class that are well-classified and FP_{rate} is the ratio of examples of the negative class misclassified.

III. Cost-Sensitive Fuzzy Rule Based Classification Systems

In this section, we will introduce a cost-sensitive linguistic FRBCS that considers the variable costs of misclassification during the genetic training stage. First, we will describe the main features of FRBCS and the fuzzy learning algorithm used in this work. Then, we will describe the modifications carried out in this algorithm for adapting it for classification with imbalanced data-sets.

A. Fuzzy rule based classification systems

An FRBCS has two main components: the Inference System and the Knowledge Base. In a linguistic FRBCS, the Knowledge Base is composed of a RB, constituted by a set of fuzzy rules, and the Data Base that stores the membership functions of the fuzzy partitions associated to the input variables. If expert knowledge of the problem is not available, it is necessary to use some Machine Learning process to obtain the Knowledge Base from examples.

Any classification problem is composed by m training samples $x_p = (x_{p1}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$ from M classes where x_{pi} is the value of attribute i $(i = 1, 2, \ldots, n)$ of the p-th training sample. In this work we use fuzzy rules of the following scheme to build our FRBCSs:

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

where R_j is the label of the j-th rule, $x = (x_1, \ldots, x_n)$ is a n-dimensional vector of samples (input variables), A_{ji} is the i-th label of the antecedent, C_j is the class label, and RW_j is the rule weight [29]. We use triangular membership functions as fuzzy partitions associated to the input variables. To compute the rule weight, many alternatives have been proposed, although we have considered as a good choice the use of the heuristic method known as the Penalized Certainty Factor (PCF) [30]:

$$\text{PCF}_j = \frac{\sum_{x_p \in C_j} \mu_{A_j}(x_p) - \sum_{x_p \notin C_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{m} \mu_{A_j}(x_p)} \tag{3}$$

where x_p is the p-th example of the training set, C_j is the consequent class of rule j and $\mu_{A_j}(x_p)$ is the membership degree of the example with the antecedents of the rule.

B. The FH-GBML algorithm

Fuzzy learning methods are the basis to build a FRBCS since they are able to learn its RB avoiding the necessity of linguistic knowledge from domain experts. The algorithm used in this work is the FH-GBML method [12], which consists of a Pittsburgh approach where each rule set is handled as an individual. It also contains a Genetic Cooperative-Competitive learning (GCCL) approach (an individual represents a unique rule), which is used as a kind of heuristic mutation for partially modifying each rule set.

This method uses standard fuzzy rules with rule weights where each input variable x_i is represented by a linguistic term or label. The system defines 14 possible linguistic terms for each attribute as well as a special "do not care" set.

In the learning process, N_{pop} rule sets are created by randomly selecting N_{rule} training patterns. Then, a fuzzy rule from each of the selected training patterns is generated by probabilistically choosing an antecedent fuzzy set from the 14 candidates ($P(B_k) = \frac{\mu_{B_k}(x_{pi})}{\sum_{j=1}^{14} \mu_{B_j}(x_{pi})}$) and each antecedent fuzzy set of the generated fuzzy rule is replaced with *don't care* using a pre-specified probability. The class with the highest accumulated compatibility degree is selected as class label.

$N_{pop} - 1$ rule sets are generated by selection, crossover and mutation in the same manner as the Pittsburgh-style algorithm using as fitness the number of correctly classified training examples. Next, with a pre-specified probability, a single iteration of the GCCL-style algorithm is applied to each of the generated rule sets using the same fitness as the Pittsburgh part.

Finally, the best rule set is added to the current population in the newly generated (N_{pop} -1) rule sets to form the next population and, if the stopping condition is not satisfied, the genetic process is repeated again.

C. A cost-sensitive learning approach: FH-GBML-CS algorithm

We denote our proposal as FH-GBML-CS (Fuzzy Hybrid Genetics-Based Machine Learning Cost-Sensitive) algorithm. The main goal of FH-GBML-CS is to obtain a FRBCS that is able to consider the different costs associated to misclassification of some of its samples during the building process of the RB. To achieve that purpose an algorithmic level solution is used, modifying the original behaviour of the FH-GBML algorithm in some of its steps:

- *Adaption of the fitness function of the Pittsburgh approach*. Instead of using the number of correctly classified training examples FH-GBML-CS tries to minimize the misclassification cost: $FN_{rate} \times C(-,+) + FP_{rate} \times C(+,-)$.
- *Modifications in the computation of the rule weight*. We have adapted the PCF heuristic building the Cost-Sensitive Penalized Certainty Factor (CS-PCF) which is used in FH-GBML-CS to compute the rule weight:

$$\text{CS-PCF}_j = \frac{\sum_{x_p \in C_j} \mu_{A_j}(x_p) \times \text{Cs}_j}{\sum_{p=1}^{m} \mu_{A_j}(x_p) \times \text{Cs}_j} - \frac{\sum_{x_p \notin C_j} \mu_{A_j}(x_p) \times \text{Cs}_j}{\sum_{p=1}^{m} \mu_{A_j}(x_p) \times \text{Cs}_j} \tag{4}$$

where Cs_j is the misclassification cost of an example from class j.

- *Different class label choice for the rule*. Instead of selecting the class considering only the highest compatibility we choose the class with the highest $compatibility \times cost$.

IV. Experimental Study

In this study, our aim is to analyze the behaviour of the FH-GBML-CS algorithm in the context of data-sets with high imbalance. To do so, we will consider twenty-two data-sets from KEEL data-set repository [14] with different IR, as shown in Table I, where we denote the number of examples (#Ex.), number of attributes (#Atts.), class name of each class (positive and negative), class attribute distribution and IR. This table is in ascending order according to the IR.

Table I
Summary Description for Imbalanced Data-Sets.

Data-set	#Ex.	#Atts.	Class (+; -)	%Class(+, -)	IR
Data-sets with High Imbalance (IR higher than 9)					
Yeast2vs4	514	8	(cyt; me2)	(9.92, 90.08)	9.08
Yeast05679vs4	528	8	(me2; mit,me3,exc,vac,erl)	(9.66, 90.34)	9.35
Vowel0	988	13	(hid; remainder)	(9.01, 90.99)	10.10
Glass016vs2	192	9	(ve-win-float-proc; build-win-float-proc, build-win-non_float-proc,headlamps)	(8.89, 91.11)	10.29
Glass2	214	9	(Ve-win-float-proc; remainder)	(8.78, 91.22)	10.39
Ecoli4	336	7	(om; remainder)	(6.74, 93.26)	13.84
Yeast1vs7	459	8	(nuc; vac)	(6.72, 93.28)	13.87
Shuttle0vs4	1829	9	(Rad Flow; Bypass)	(6.72, 93.28)	13.87
Glass4	214	9	(containers; remainder)	(6.07, 93.93)	15.47
Page-blocks13vs2	472	10	(graphic; horiz.line,picture)	(5.93, 94.07)	15.85
Abalone9vs18	731	8	(18; 9)	(5.65, 94.25)	16.68
Glass016vs5	184	9	(tableware; build-win-float-proc, build-win-non_float-proc,headlamps)	(4.89, 95.11)	19.44
Shuttle2vs4	129	9	(Fpv Open; Bypass)	(4.65, 95.35)	20.5
Yeast1458vs7	693	8	(vac; nuc,me2,me3,pox)	(4.33, 95.67)	22.10
Glass5	214	9	(tableware; remainder)	(4.20, 95.80)	22.81
Yeast2vs8	482	8	(pox; cyt)	(4.15, 95.85)	23.10
Yeast4	1484	8	(me2; remainder)	(3.43, 96.57)	28.41
Yeast1289vs7	947	8	(vac; nuc,cyt,pox,erl)	(3.17, 96.83)	30.56
Yeast5	1484	8	(me1; remainder)	(2.96, 97.04)	32.78
Ecoli0137vs26	281	7	(pp,imL; cp,im,imU,imS)	(2.49, 97.51)	39.15
Yeast6	1484	8	(exc; remainder)	(2.49, 97.51)	39.15
Abalone19	4174	8	(19; remainder)	(0.77, 99.23)	128.87

To analyze the cost-sensitive approach proposed, we will compare the performance of FH-GBML-CS and two versions of FH-GBML: the standard FH-GBML algorithm [12] and the FH-GBML algorithm using SMOTE [3] as preprocessing method, considering only the 1-nearest neighbour to generate the synthetic samples, and balancing both classes to the 50% distribution.

To develop the different experiments we consider a 5-fold cross-validation model, i.e., 5 random partitions of data with a 20%, and the combination of 4 of them (80%) as training and the remaining one as test. For each data-set we consider the average results of the five partitions. Furthermore, Wilcoxon's Signed-Ranks Test [31] is used for statistical comparison of our empirical results. In all cases the level of confidence (α) will be set at 0.05.

The configuration for the different approaches is presented in Table II, which includes the parameters for the configuration of the FRBCS and the set-up of the GFS features (d stands for the dimensionality of the problem).

Table II
Configuration for the FRBCS features of the algorithms

Configuration	Parameter
FRBCS	*Conjunction operator:* Product T-norm *Rule Weight:* PCF (FH-GBML and FH-GBML+SMOTE) and CS-PCF (FH-GBML-CS) *Fuzzy Reasoning Method:* Winning Rule
GFS	*Number of fuzzy rules:* $5 \cdot d$ (max. 50 rules) *Number of rule sets:* 200 *Crossover probability:* 0.9 *Mutation probability:* $1/d$ *Number of replaced rules:* All rules except the best-one (Pittsburgh-part, elitist approach) number of rules/5 (GCCL-part) *Total number of generations:* $1,000$ *Don't care probability:* 0.5 *Probability of the application of the GCCL iteration:* 0.5

Tables III and IV show the results in performance for FH-GBML-CS and the algorithms employed for comparison, that is, FH-GBML and FH-GBML+SMOTE. Table III shows the general summary results for all the data-sets included in the study. This table includes the average AUC metric (being AUC_{Tr} the AUC over the training data-set, AUC_{Tst} the AUC over the test dataset), the average number of rules and average run time measured in seconds. Table IV shows the accuracy results for each data-set using the AUC metric and the average number of rules for the three algorithms considered in this study.

Table III
Summary results in performance (AUC metric, number of rules, run time) for FH-GBML, FH-GBML+SMOTE and FH-GBML-CS

Algorithm	AUC_{Tr}	AUC_{Tst}	#Rules	Run time (sec.)
FH-GBML	62.64	58.92	33.95	527.44
FH-GBML+SMOTE	89.89	81.77	18.73	1140.04
FH-GBML-CS	**92.13**	**82.35**	**6.89**	**342.97**

We observe that the precision obtained by FH-GBML-CS is higher than the one for FH-GBML and similar to the one for FH-GBML+SMOTE for AUC_{Tr} and AUC_{Tst}, showing the necessity to use specific techniques to deal with the imbalanced problem. This situation is represented statistically by means of a Wilcoxon test (Table V).

Table V
Wilcoxon test to compare FH-GBML-CS (CS) with FH-GBML (Base) and FH-GBML+SMOTE (SMOTE) according to the AUC metric. R^+ corresponds to FH-GBML-CS and R^- to FH-GBML or FH-GBML+SMOTE

Comparison	R^+	R^-	Hypothesis ($\alpha = 0.05$)	p-value
CS vs Base	251.0	2.0	Rejected for CS	0.000
CS vs SMOTE	124.0	129.0	Not Rejected	0.935

Table IV
DETAILED RESULTS IN PERFORMANCE (AUC METRIC, NUMBER OF RULES) FOR FH-GBML, FH-GBML+SMOTE AND FH-GBML-CS

Dataset	FH-GBML			FH-GBML+SMOTE			FH-GBML-CS		
	AUC_{Tr}	AUC_{Tst}	#Rules	AUC_{Tr}	AUC_{Tst}	#Rules	AUC_{Tr}	AUC_{Tst}	#Rules
Yeast2vs4	66.07	62.60	33.4	94.52	**92.50**	16.8	95.88	90.92	**6**
Yeast05679vs4	51.23	51.00	39.4	84.39	**79.38**	13.8	85.80	79.28	**7**
Vowel0	57.86	59.22	100	94.62	92.94	30.4	94.62	**93.99**	**6.8**
Glass016vs2	53.68	49.43	13.2	82.25	59.60	17.2	88.04	**68.76**	**7.2**
Glass2	52.91	49.23	19.4	82.02	**67.29**	18.8	88.07	64.79	**6.4**
Ecoli4	74.92	64.69	29.6	98.31	88.10	12.2	99.49	**89.13**	**6.6**
Yeast1vs7	52.08	50.00	39.4	81.08	74.06	19.8	85.68	**75.13**	**8.2**
Shuttle0vs4	100.00	94.60	48.8	100.00	**100.00**	49.8	100.00	99.80	**10**
Glass4	58.64	53.08	24.6	98.38	87.67	16.6	99.04	**90.39**	**6.6**
Page-Blocks13vs4	88.62	80.77	29.4	98.59	**98.88**	16.8	99.65	97.94	**5.8**
Abalone9-18	52.67	50.00	32.6	79.10	64.89	17.2	82.55	**80.21**	**5.6**
Glass016vs5	56.96	50.00	49.4	97.64	**86.86**	17.6	99.21	85.64	**6.4**
Shuttle2vs4	100.00	**100.00**	24.2	100.00	97.95	17.4	100.00	95.00	**10**
Yeast1458vs7	52.03	49.47	29.8	75.22	58.94	14.2	77.97	**62.49**	**5.8**
Glass5	56.96	50.00	50	96.22	**88.78**	23.6	98.93	85.79	**6.4**
Yeast2vs8	60.63	55.00	39.4	84.20	76.11	14.2	87.66	**76.32**	**7**
Yeast4	52.66	49.97	28	86.66	79.95	10.8	88.00	**82.14**	**6.4**
Yeast1289vs7	52.89	49.89	19.6	79.31	67.13	16.4	82.04	**70.43**	**6**
Yeast5	51.42	49.97	20.4	97.73	**96.74**	21.6	98.59	95.02	**6.8**
Ecoli0137vs26	82.33	74.63	40	97.31	**81.36**	15.8	99.41	78.09	**7**
Yeast6	53.53	52.72	26.4	91.11	**86.76**	12.4	92.96	83.38	**6.6**
Abalone19	50.00	50.00	40	78.95	**73.14**	18.6	83.23	66.99	**7**
Mean	62.64	58.92	33.95	89.89	81.77	18.73	92.13	**82.35**	**6.89**

On the other hand, we can see that the run time of FH-GBML-CS is similar to the original FH-GBML and much lower than the FH-GBML+SMOTE option. This fact is statistically supported with a Wilcoxon test (Table VI). These results make the FH-GBML-CS a method with an equivalent precision to the use of preprocessing with a smaller run time. The use of GFS turns the reduction of the training set into a critical issue since it influences the efficiency of the output model. Furthermore, regarding the complexity of the obtained models with each approach, the average number of rules is lower in the model obtained by FH-GBML-CS therefore obtaining more compact and interpretable models than the other methods by means of a more appropriate computation of the rule weight for this type of problems.

Table VI
WILCOXON TEST TO COMPARE FH-GBML-CS (CS) WITH FH-GBML (BASE) AND FH-GBML+SMOTE (SMOTE) ACCORDING TO THE RUN TIME (SECONDS). R^+ CORRESPONDS TO FH-GBML-CS AND R^- TO FH-GBML OR FH-GBML+SMOTE

Comparison	R^+	R^-	Hypothesis ($\alpha = 0.05$)	p-value
CS vs Base	153.0	100.0	Not Rejected	0.390
CS vs SMOTE	253.0	0.0	Rejected for CS	0.00

V. CONCLUDING REMARKS

In this work we have analyzed the behaviour of a cost-sensitive learning method for linguistic FRBCSs when using GFSs in the learning stage focusing on the scenario of high imbalanced data-sets. The experimental and statistical results show that cost-sensitive learning is a feasible alternative to solve imbalanced classification problems since it improves the base algorithm. Furthermore, cost-sensitive learning is a sufficiently competitive proposal compared to data-level solutions based in rebalancing the class distribution, with lower time requirement being this a very significant issue in the framework of GFSs. Finally, we must stress that cost-sensitive learning obtains lower complex models having a good trade-off between precision and interpretability in a model with a good compromise between the precision in both classes in a reasonable run time.

As future work, we will extend this study analyzing the performance of more algorithms and different cost-sensitive methods to modify and adapt the behaviour of the fuzzy learning algorithm in the imbalanced data-sets problem.

ACKNOWLEDGMENT

This work had been supported by the Spanish Ministry of Science and Technology under Project TIN2008-06681-C06-01.

REFERENCES

[1] C. Elkan, "The foundations of cost-sensitive learning," in *Proceedings of the 17th IEEE International Joint Conference on Artificial Intelligence (IJCAI'01)*, 2001, pp. 973–978.

[2] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behaviour of several methods for balancing machine learning training data," *SIGKDD Explorations*, vol. 6, no. 1, pp. 20–29, 2004.

[3] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligent Research*, vol. 16, pp. 321–357, 2002.

[4] B. Zadrozny and C. Elkan, "Learning and making decisions when costs and probabilities are both unknown," in *Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining (KDD'01)*, 2001, pp. 204–213.

[5] K. M. Ting, "An instance-weighting method to induce cost-sensitive trees," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, no. 3, pp. 659–665, 2002.

[6] B. Zadrozny, J. Langford, and N. Abe, "Cost-sensitive learning by cost-proportionate example weighting," in *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM'03)*, 2003, pp. 435–442.

[7] Z.-H. Zhou and X.-Y. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.

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

[9] A. Fernández, S. García, M. J. del Jesus, and F. Herrera, "A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets," *Fuzzy Sets and Systems*, vol. 159, no. 18, pp. 2378–2398, 2008.

[10] O. Cordón, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, "Ten years of genetic fuzzy systems: Current framework and new trends," *Fuzzy Sets and Systems*, vol. 141, no. 1, pp. 5–31, 2004.

[11] F. Herrera, "Genetic fuzzy systems: Taxonomy, current research trends and prospects," *Evolutionary Intelligence*, vol. 1, pp. 27–46, 2008.

[12] H. Ishibuchi, T. Yamamoto, and T. Nakashima, "Hybridization of fuzzy GBML approaches for pattern classification problems," *IEEE Transactions on System, Man and Cybernetics B*, vol. 35, no. 2, pp. 359–365, 2005.

[13] A. Fernández, M. J. del Jesús, and F. Herrera, "On the 2-tuples based genetic tuning performance for fuzzy rule based classification systems in imbalanced data-sets," *Information Sciences*, vol. 180, no. 8, pp. 1268–1291, 2010.

[14] 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*, vol. in press, 2010.

[15] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 3, pp. 299–310, 2005.

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

[17] 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*, vol. 9, pp. 2607–2624, 2008.

[18] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.

[19] Y. Sun, A. K. C. Wong, and M. S. Kamel, "Classification of imbalanced data: A review," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 23, no. 4, pp. 687–719, 2009.

[20] F. Tek, A. Dempster, and I. Kale, "Parasite detection and identification for automated thin blood film malaria diagnosis," *Computer Vision and Image Understanding*, vol. 114, no. 1, pp. 21–32, 2010.

[21] D. Williams, V. Myers, and M. Silvious, "Mine classification with imbalanced data," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 3, pp. 528–532, 2009.

[22] F. Vilario, P. Spyridonos, F. Deiorio, J. Vitria, F. Azpiroz, and P. Radeva, "Intestinal motility assessment with video capsule endoscopy: Automatic annotation of phasic intestinal contractions," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 246–259, 2010.

[23] G. M. Weiss, "Mining with rarity: a unifying framework," *SIGKDD Explorations*, vol. 6, no. 1, pp. 7–19, 2004.

[24] A. Orriols-Puig and E. Bernadó-Mansilla, "Evolutionary rule-based systems for imbalanced datasets," *Soft Computing*, vol. 13, no. 3, pp. 213–225, 2009.

[25] A. Orriols-Puig, E. Bernadó-Mansilla, D. E. Goldberg, K. Sastry, and P. L. Lanzi, "Facetwise analysis of XCS for problems with class imbalances," *IEEE Transactions on Evolutionary Computation*, vol. 13, pp. 260–283, 2009.

[26] V. García, R. Mollineda, and J. S. Sánchez, "On the k-NN performance in a challenging scenario of imbalance and overlapping," *Pattern Analysis Applications*, vol. 11, no. 3-4, pp. 269–280, 2008.

[27] P. Domingos, "Metacost: A general method for making classifiers cost-sensitive," in *Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining (KDD'99)*, 1999, pp. 155–164.

[28] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, "Cost-sensitive boosting for classification of imbalanced data," *Pattern Recognition*, vol. 40, no. 12, pp. 3358–3378, 2007.

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

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

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