Expert Systems with Applications 36 (2009) 9805–9812

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

On the influence of an adaptive inference system in fuzzy rule based classification systems for imbalanced data-sets ☆

Alberto Fernández [a,*], María José del Jesus [b], Francisco Herrera [a]

[a] Department of Computer Science and A.I., University of Granada, Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain
[b] Department of Computer Science, University of Jaén, Spain

ARTICLE INFO

Keywords:
Fuzzy rule-based classification systems
Inference mechanism
Parametric conjunction operators
Genetic fuzzy systems
Imbalanced data-sets

ABSTRACT

Classification with imbalanced data-sets supposes a new challenge for researches in the framework of data mining. This problem appears when the number of examples that represents one of the classes of the data-set (usually the concept of interest) is much lower than that of the other classes. In this manner, the learning model must be adapted to this situation, which is very common in real applications.

In this paper, we will work with fuzzy rule based classification systems using a preprocessing step in order to deal with the class imbalance. Our aim is to analyze the behaviour of fuzzy rule based classification systems in the framework of imbalanced data-sets by means of the application of an adaptive inference system with parametric conjunction operators.

Our results shows empirically that the use of the this parametric conjunction operators implies a higher performance for all data-sets with different imbalanced ratios.

1. Introduction

Fuzzy rule based classification systems (FRBCSs) (Ishibuchi, Nakashima, & Nii, 2004) are a very useful tool in the ambit of machine learning, since they provide an interpretable model for the end user. There are many real applications in which the FRBCS have been employed, including anomaly intrusion detection (Tsang, Kwong, & Wang, 2007), cloud cover estimation from satellite imagery (Ghosh, Pal, & Das, 2006) and image processing (Nakashima, Schaefer, Yokota, & Ishibuchi, 2007). In most of these areas the data used is highly skewed, i.e. the number of instances of one class is much lower than the instances of the other classes. This situation is known as the imbalanced data-set problem, and it has been recently identified as one important problem in data mining (Chawla, Japkowicz, & Kolcz, 2004).

Most learning algorithms obtain a high predictive accuracy over the majority class, but predict poorly over the minority class (Weiss, 2004). Furthermore, the examples in the minority class can be treated as noise and they might are completely ignored by the classifier. In fact, there are studies that show that most classification methods lose their classification ability when dealing with imbalanced data (Japkowicz & Stephen, 2002; Phua, Alahakoon, & Lee, 2004). In this manner, many recent studies are focused on developing new approaches in this area (Hong, Chen, & Harris, 2007; Lee, Tsai, Wu, & Yang, 2008; Su, Chen, & Yih, 2006).

The use of the appropriate conjunction connectors in the Inference System can improve the fuzzy system behaviour by using parametrized expressions, while maintaining the original interpretability associated to fuzzy systems (Crockett, Bandar, Fowdar, & O'Shea, 2006; Crockett, Bandar, Mclean, & O'Shea, 2006; Wu & Mendel, 2004). This approach is usually called Adaptive Inference System (AIS) and it has shown very good results in fuzzy modelling (Alcalá-Fdez, Herrera, Márquez, & Peregrín, 2007; Márquez, Peregrín, & Herrera, 2007).

Our aim in this paper is to analyze the influence of the AIS for FRBCSs in the framework of imbalanced data-sets. We start from the analysis performed in Fernández, García, del Jesus, and Herrera (2008), where we studied different configurations for FRBCS in order to determine the most suitable model for imbalanced data-sets. Furthermore, we showed the necessity to apply a re-sampling procedure; specifically, we found a very good behaviour in the case of the "Synthetic Minority Over-Sampling Technique" (SMOTE) (Chawla, Bowyer, Hall, & Kegelmeyer, 2002).

We will present a postprocessing study on the tuning of parameters with a previously established Rule Base (RB), using Genetic Algorithms (GAs) as a tool to evolve the connector parameters. We will develop an experimental study with 33 data-sets from UCI repository with different imbalance ratios. Data-sets with more than two classes have been modified by taking one against the others or by contrasting one class with another. To evaluate

☆ Supported by the Spanish Ministry of Science and Technology under Projects TIN-2005-08386-C05-01 and TIN-2005-08386-C05-03.

* Corresponding author. Tel.: +34 958 240598; fax: +34 958 243317.
E-mail addresses: alberto@decsai.ugr.es (A. Fernández), mjjesus@ujaen.es (M.J. del Jesus), herrera@decsai.ugr.es (F. Herrera).

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.02.048

our results we have applied the geometric mean metric (Barandela, Sánchez, García, & Rangel, 2003; Kubat, Holte, & Matwin, 1998) which aims to maximize the accuracy of both classes. We have also made use of some non-parametric tests (Demšar, 2006; García, Fernández, Luengo, & Herrera, in press) with the aim to show the significance in the performance improvements obtained with the AIS model.

In order to do that, the paper is organized as follows: Section 2 presents an introduction on the class imbalance problem, including the description of the problem, proposed solutions, and proper measures for evaluating classification performance in the presence of the imbalance data-set problem. In Section 3, we describe the fuzzy rule learning methodology used in this study, the Chi et al. rule generation method (Chi, Yan, & Pham, 1996), and introduces the AIS with the parametric conjunction operators and the evolutionary algorithm that tunes these parameters. In Section 4, we include our experimental analysis in imbalanced data-sets with different degrees of imbalance. Finally, in Section 5 some concluding remarks are pointed out.

2. Imbalanced data-sets in classification

In this section, we will first introduce the problem of imbalanced data-sets. Then we will describe the preprocessing technique we have applied in order to deal with the imbalanced data-sets: the SMOTE algorithm. Finally, we will present the evaluation metrics for this kind of classification problem.

2.1. The problem of imbalanced data-sets

Learning from imbalanced data is an important topic that has recently appeared in the machine learning community. When treating with imbalanced data-sets, one or more classes might be represented by a large number of examples while the others are represented by only a few.

We focus on the two class imbalanced data-sets, where there is only one positive and one negative class. We consider the positive class as the one with the lowest number of examples and the negative class the one with the highest number of examples. Furthermore, in this work we use the imbalance ratio (IR) (Orriols-Puig & Bernadó-Mansilla, 2009), defined as the ratio of the number of instances of the majority class and the minority class, to organize the different data-sets according to their IR.

The problem of imbalanced data-sets is extremely significant because it is implicit in most real world applications, such as fraud detection (Fawcett & Provost, 1997), text classification (Tan, 2005), risk management (Huang, Hung, & Jiau, 2006), medical diagnosis (Mazurowski et al., 2008) and classification of weld flaws (Liao, 2008) among others.

In classification, this problem (also named the "class imbalance problem") will cause a bias on the training of classifiers and will result in the lower sensitivity of detecting the minority class examples. In fact, the main handicap on imbalanced data-sets is the overlapping between the examples of the positive and the negative class, because of the difficulty of most learning algorithms to detect those small disjuncts (Weiss & Provost, 2003). This fact is depicted in Fig. 1.

For this reason, a large number of approaches have been previously proposed to deal with the class imbalance problem. These approaches can be categorized into two groups: the internal approaches that create new algorithms or modify existing ones to take the class imbalance problem into consideration (Barandela et al., 2003; Hung & Huang, 2008; Xu, Chow, & Taylor, 2007) and external approaches that preprocess the data in order to diminish the effect cause by their class imbalance (Batista, Prati, & Monard, 2004; Estabrooks, Jo, & Japkowicz, 2004). Furthermore, cost-sensitive learning solutions incorporating both the data and algorithmic level approaches assume higher misclassification costs with samples in the minority class and seek to minimize the high cost errors (Domingos, 1999; Sun, Kamel, Wong, & Wang, 2007).

The internal approaches have the disadvantage of being algorithm specific, while external approaches are independent of the classifier used and are, for this reason, more versatile. Furthermore, in our previous work on this topic (Fernández et al., 2008) we analyzed the cooperation of some preprocessing methods with FRBCSs, showing a good behaviour for the oversampling methods, specially in the case of the SMOTE methodology (Chawla et al., 2002).

According to this, we will employ in this paper the SMOTE algorithm in order to deal with the problem of imbalanced data-sets. This method is detailed in the next subsection.

2.2. Preprocessing imbalanced data-sets. The SMOTE algorithm

As mentioned before, applying a preprocessing step in order to balance the class distribution is a positive solution to the imbalance data-set problem (Batista et al., 2004). Specifically, in this work we have chosen an oversampling method which is a reference in this area: the SMOTE algorithm (Chawla et al., 2002).

In this approach, the minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbours. Depending upon the amount of over-sampling required, neighbours from the k nearest neighbours are randomly chosen. This process is illustrated in Fig. 2, where x_i is the selected

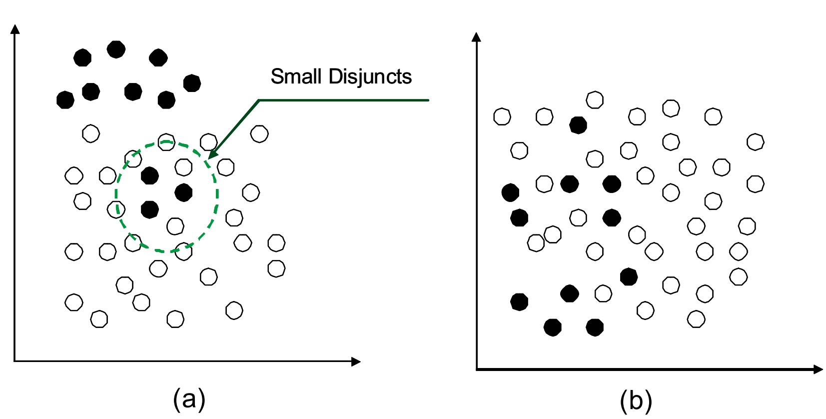

Fig. 1. Example of the imbalance between classes: (a) small disjuncts; (b) overlapping between classes.

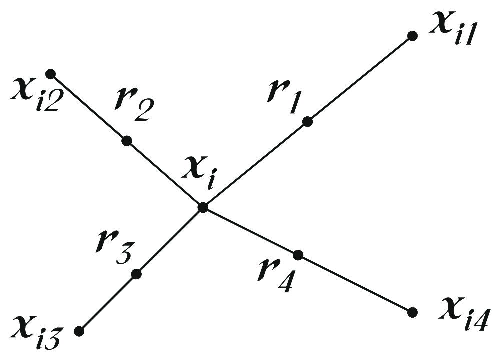

Fig. 2. An illustration on how to create the synthetic data points in the SMOTE algorithm.

point, x_{i1} to x_{i4} are some selected nearest neighbours and r_1 to r_4 the synthetic data points created by the randomized interpolation. The implementation employed in this work uses only one nearest neighbour using the euclidean distance, and balance both classes to the 50% distribution.

Synthetic samples are generated in the following way: Take the difference between the feature vector (sample) under consideration and its nearest neighbour. Multiply this difference by a random number between 0 and 1, and add it to the feature vector under consideration. This causes the selection of a random point along the line segment between two specific features. This approach effectively forces the decision region of the minority class to become more general. An example is detailed in Fig. 3.

In short, its main idea is to form new minority class examples by interpolating between several minority class examples that lie together. Thus, the overfitting problem is avoided and causes the decision boundaries for the minority class to spread further into the majority class space.

2.3. Evaluation in imbalanced domains

The measures of the quality of classification are built from a confusion matrix (shown in Table 1) which records correctly and incorrectly recognized examples for each class.

Traditionally, accuracy is the most commonly used measure for empirical evaluation. However, for classification with imbalanced data-sets, this metric may lead to erroneous conclusions since the minority class has very little impact on accuracy as compared to the majority class (Weiss, 2004). For example, a classifier that obtains an accuracy of 90% in a data-set with an IR value of 9, might not be accurate if it does not cover correctly any minority class instance

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \tag{1}$$

Because of this, instead of using accuracy, more correct metrics are considered. Two common measures, sensitivity and specificity (2) and (3), approximate the probability of the positive (negative) label being true. In other words, they assess the effectiveness of the algorithm on a single class.

$$\text{sensitivity} = \frac{TP}{TP + FN} \tag{2}$$

$$\text{specificity} = \frac{TN}{FP + TN} \tag{3}$$

In this paper, we consider both classes (positive and negative) to be equivalent in importance. In this manner, both sensitivity and specificity are expected to be high simultaneously and thus, the selected metric is the geometric mean of the true rates (Barandela et al., 2003; Kubat et al., 1998), which measures the balanced performance of a learning algorithm between these two classes, and can be defined as:

$$GM = \sqrt{\frac{TP}{TP + FN} \cdot \frac{TN}{FP + TN}} \tag{4}$$

```
Consider a sample (6,4) and let (4,3) be its nearest neighbour.
(6,4) is the sample for which k-nearest neighbours are being
identified (4,3) is one of its k-nearest neighbours.
Let: f1_1 = 6 f2_1 = 4, f2_1 - f1_1 = -2
f1_2 = 4 f2_2 = 3, f2_2 - f1_2 = -1
The new samples will be generated as
f1',f2') = (6,4) + rand(0-1) * (-2,-1)
rand(0-1) generates a random number between 0 and 1.
```

Fig. 3. Example of the SMOTE application.

Table 1
Confusion matrix for a two-class problem.

	Positive prediction	Negative prediction
Positive class	True positive (TP)	False negative (FN)
Negative class	False positive (FP)	True negative (TN)

3. Fuzzy rule based classification systems: linguistic rule generation method and adaptive inference system

Any classification problem 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 pth training pattern.

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

$$\text{Rule } R_j: \text{ If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \text{ then Class} = C_j \text{ with } RW_j, \tag{5}$$

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. We use triangular membership functions as antecedent fuzzy sets.

In the following subsections we will first describe the rule generation procedure used in this paper and then we will introduce the AIS and the evolutionary algorithm used to adjust the parameters of the conjunction operator.

3.1. Linguistic rule generation method: Chi et al. approach

We have employed a simple learning method in order to generate the RB for the FRBCS. Specifically we have selected the method proposed in Chi et al. (1996), that we have named the Chi et al.'s rule generation, which is just an extension of the well known Wang and Mendel algorithm (Wang & Mendel, 1992) to classification problems.

To generate the fuzzy RB this FRBCSs design method determines the relationship between the variables of the problem and 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 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.

We must remark 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.

3.2. Adaptive inference system

In this section, we first analyze the AIS and we justify the use of the Dubois parametric t-norm as conjunction operator. Then we present the evolutionary algorithm used to adapt the parameters of the conjunction operator.

3.2.1. Adaptive components in the inference system

Considering a new pattern $x_p = (x_{p1}, \ldots, x_{pn})$ and an RB composed of L fuzzy rules, the steps of the inference system are the following (Cordón, del Jesus, & Herrera, 1999):

(1) *Matching degree*. To calculate the *strength of activation of the if-part for all rules in the RB with the pattern* x_p, using a conjunction operator (usually a t-norm)

$$\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})), \quad j = 1, \ldots, L \tag{6}$$

(2) *Association degree*. To compute the *association degree of the pattern* x_p *with the M classes according to each rule in the RB*. When using rules with the form of (5) this association degree only refers to the consequent class of the rule (i.e. $k = C_j$)

$$b_j^k = h(\mu_{A_j}(x_p), RW_j^k), \quad k = 1, \ldots, M, \; j = 1, \ldots, L \tag{7}$$

We model function h as the product t-norm in all cases.

(3) *Pattern classification soundness degree for all classes*. We use an aggregation function that combines the positive degrees of association calculated in the previous step

$$Y_k = f(b_j^k, j = 1, \ldots, L \text{ and } b_j^k > 0), \quad k = 1, \ldots, M. \tag{8}$$

(4) *Classification*. We apply a decision function F over the soundness degree of the system for the pattern classification for all classes. This function will determine the class label l corresponding to the maximum value

$$F(Y_1, \ldots, Y_M) = l \quad \text{such that } Y_l = \{\max(Y_k), \; k = 1, \ldots, M\} \tag{9}$$

The conjunction operator (function T in Step 1) is suitable to be parameterized in order to adapt the inference system. In fact, the model based on the tuning of the inference system has shown a considerable improvement in the accuracy of linguistic fuzzy systems (Alcalá-Fdez et al., 2007; Márquez et al., 2007). Table 2 exemplifies three classical parametric t-norms (Mizumoto, 1989) that can be used to model the adaptive conjunction operator.

The effect of the parameter in the adaptive conjunction is sometimes equivalent to one of the well-known mechanisms to modify the linguistic meaning of the rule structure, the use of linguistic modifiers (Liu, Chen, & Tsao, 2001), as shown in Fig. 4. We must point out that the effect of the adaptive t-norm playing the role of conjunction operator does not modify the shape of the inferred fuzzy set, maintaining the original interpretability of the fuzzy labels.

Table 2
Adaptive t-norms.

Name	Expression	Domain
t-norm		
Dubois	$T_{\text{Dubois}}(x, y, \alpha) = \frac{x \cdot y}{\max(x, y, \alpha)}$	$(0 \leqslant \alpha \leqslant 1)$
Dombi	$T_{\text{Dombi}}(x, y, \alpha) = \frac{1}{1 + \sqrt[\alpha]{(\frac{1-x}{x})^\alpha + (\frac{1-y}{y})^\alpha}}$	$(\alpha > 0)$
Frank	$T_{\text{Frank}}(x, y, \alpha) = \log_\alpha[1 + \frac{(\alpha^x - 1)\cdot(\alpha^y - 1)}{\alpha - 1}]$	$(\alpha > 0)$, $(\alpha \neq 1)$

Two models of AIS can be considered depending on the amount of parameters they use:

- A single parameter α to tune globally the behavior of the AIS.
- Individual parameters α_i for every rule of the KB, having a local tuning mechanism of the behavior of the inference system for every rule.

The model used in this paper is based on the results obtained in Alcalá-Fdez et al. (2007), Márquez et al. (2007), where the authors learn the conjunctive connector for every rule separately and obtains the highest accuracy because of its high degree of freedom. Furthermore, we will use Dubois t-norm, not only because it is more efficiently computed, but also because it has obtained a better behaviour than other parametric t-norms (Alcalá-Fdez et al., 2007).

We must note that Dubois t-norm achieves like a minimum when $\alpha = 0$ and like algebraic product $\alpha = 1$. When $0 < \alpha < 1$, it continues performing like minimum excepting when every match with antecedents are below α, that takes values between minimum and product, being similar to a concentration effect. Thus, Dubois t-norm connects with minimum in those cases when the matches with antecedents are more significant, while the rest are connected with a value between minimum and product.

3.2.2. Evolutionary adaptive inference system

GAs has been widely used to derive fuzzy systems (Cordón, Gomide, Herrera, Hoffmann, & Magdalena, 2004; Herrera, 2008). In this work, we will consider the use of a specific GA to design the proposed learning method, the CHC algorithm (Eshelman, 1991). The CHC algorithm is a GA that presents a good trade-off between

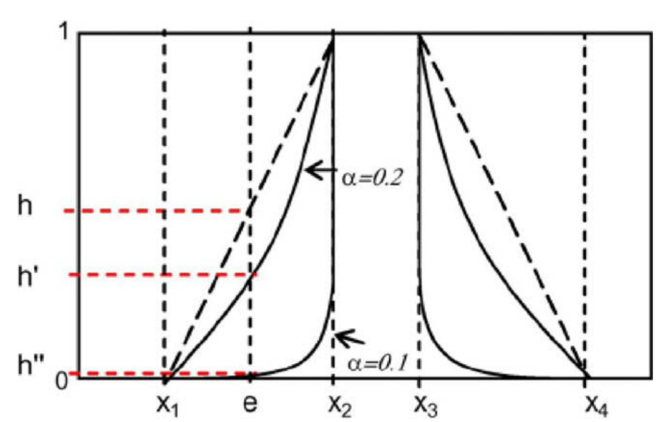

Fig. 4. Graphical representation of the antecedent linguistic modification produced by different values of Dombi t-norm.

diversity and convergence, being a good choice in problems with complex search spaces.

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 of the next population (with M being the population size). Furthermore, no mutation is applied during the recombination phase, Instead, when the population converges or the search stops making progress, the population is re-initialized.

The components needed to design this process are explained below. They are: coding scheme, initial gene pool, chromosome evaluation, crossover operator (together with an incest prevention) and restarting approach.

(1) *Coding scheme:* Since we are using one parameter for every fuzzy rule, each chromosome will be composed by R genes, being R the number of rules in the RB. Also, we are using a real-coding version of the CHC, so each gene will take a value between 0 and 1, that is, the domain for the α value in the Dubois t-norm.
(2) *Chromosome evaluation:* The fitness function must be in accordance with the framework of imbalanced data-sets. Thus, we will use, as presented in Section 2.3, the geometric mean of the true rates, defined in (4) as:

$$GM = \sqrt{\frac{TP}{TP + FN} \cdot \frac{TN}{FP + TN}}$$

(3) *Initial gene pool:* Remember from the previous section that Dubois parametric t-norm behaves like minimum or product t-norm when $\alpha = 0$ and $\alpha = 1$, respectively. For this reason, we will initialize one chromosome with all its genes at 0 to model the minimum t-norm and another chromosome with all genes at 1 to model the product t-norm. The remaining individuals of the population will be generated at random in the interval $[0, 1]$.
(4) *Crossover operator:* The BLX-α crossover $(\alpha = 0.5)$ is employed in order to recombine the parent's genes. The incest prevention mechanism works as follows: two parents are crossed if their Hamming distance divided by 2 is above a predetermined threshold, L.The Hamming distance is computed by translating the real-coded genes into strings and by taking into account whether each character is different or not. For that purpose we will use a Gray Code with a fixed number of bits per gene (*BITSGENE*), that is determined by the system expert. The initial threshold is set to $L = (\#Genes \cdot BITSGENE)/4.0$ where L is the length of the string and $\#Genes$ stands for the total length of the chromosome. When no offspring is inserted into the new population, the threshold is reduced by 1 (*BITSGENE* in this case).
(5) *Restarting approach:* Since no mutation is performed, to get away from local optima a restarting mechanism is considered (Eshelman, 1991) when the threshold value L is lower than zero. In this case, all the chromosomes are generated at random within the interval $[0, 1]$. Furthermore, the best global solution found is included in the population to increase the convergence of the algorithm.

4. Experimental study

In this section, we will show empirically the good behaviour achieved by FRBCSs when using the parametric conjunction operator, using a large amount of imbalanced data-sets to support our analysis.

We will employ all data-sets to perform a global study disregard the degree of imbalance, but we will also section the study by using the IR to distinguish among three classes of imbalanced data-sets to contrast the performance in each imbalance scenario.

Specifically, we distinguish among data-sets with a *low imbalance* when the instances of the positive class are between 25% and 40% of the total instances (IR between 1.5 and 3), data-sets with a *medium imbalance* when the number of the positive instances is between 10% and 25% of the total instances (IR between 3 and 9), and data-sets with a *high imbalance* where there are no more than 10% of positive instances in the whole data-set compared to the negative ones (IR higher than 9).

We have selected 33 data-sets with different IR from UCI repository. The data is summarized in Table 3, showing the number of examples (#Ex.), number of attributes (#Atts.), class name of each class (minority and majority), class attribute distribution and IR. This table is ordered by the IR, from data-sets with low imbalance to highly imbalanced data-sets.

In the remaining of this section, we will first present the experimental framework and all the parameters employed in this study. Then, we will perform a comparative analysis between the base FRBCS and the use of the AIS model (parametric conjunction operator), in order to show the improvement obtained with this model.

4.1. Experimental set-up

To develop the different experiments we consider a *5-folder 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.

We must emphasize that, in order to reduce the effect of imbalance, we have employed the SMOTE preprocessing method (Chawla et al., 2002) for all our experiments, considering only the 1-nearest neighbour to generate the synthetic samples, and balancing both classes to the 50% distribution.

Statistical analysis needs to be carried out in order to find significant differences among the results obtained by the studied methods. We consider the use of non-parametric tests, according to the recommendations made in Demšar (2006), García et al. (in press), where it is presented a set of simple, safe and robust non-parametric tests for statistical comparisons of classifiers. For pair-wise comparisons we will use Wilcoxon's Signed-Ranks Test (Sheskin, 2006; Wilcoxon, 1945) and in all cases the level of confidence (α) will be set at 0.05 (95% of confidence).

We will employ the following configuration for the FRBCS approach:

- Number of fuzzy labels: 3 and 5 labels.
- Conjunction operator to compute the compatibility degree of the example with the antecedent of the rule: product t-norm.
- Rule weight: penalized certainty factor (Ishibuchi & Yamamoto, 2005).
- Conjunction operator between the compatibility degree and the rule weight: Product t-norm.
- Fuzzy reasoning method: winning rule.

We have selected this FRBCS model as it achieved a good performance in our former studies on imbalanced data-sets (Fernández, García, del Jesus, & Herrera, 2007; Fernández et al., 2008). We will use both 3 and 5 labels per variable because it is not clear what level of granularity must be employed for the FRBCS.

Finally, we indicate the values that have been considered for the parameters of the CHC algorithm:

- Population Size: 50 individuals.
- Number of evaluations: $5000 \cdot$ *number of variables*.

Table 3
Summary description for imbalanced data-sets.

Data-set	#Ex.	#Atts.	Class (min., maj.)	% Class (min., maj.)	IR
Data-sets with low imbalance (1.5–3 IR)					
Glass2	214	9	(build-window-non_float-proc, remainder)	(35.51, 64.49)	1.82
EcoliCP-IM	220	7	(im, cp)	(35.00, 65.00)	1.86
Wisconsin	683	9	(malignant, benign)	(35.00, 65.00)	1.86
Pima	768	8	(tested-positive, tested-negative)	(34.84, 66.16)	1.90
Iris1	150	4	(Iris-Setosa, remainder)	(33.33, 66.67)	2.00
Glass1	214	9	(build-window-float-proc, remainder)	(32.71, 67.29)	2.06
Yeast2	1484	8	(NUC, remainder)	(28.91, 71.09)	2.46
Vehicle2	846	18	(Saab, remainder)	(28.37, 71.63)	2.52
Vehicle3	846	18	(bus, remainder)	(28.37, 71.63)	2.52
Vehicle4	846	18	(Opel, remainder)	(28.37, 71.63)	2.52
Haberman	306	3	(Die, survive)	(27.42, 73.58)	2.68
Data-sets with medium imbalance (3–9 IR)					
GlassNW	214	9	(non-window glass, remainder)	(23.83, 76.17)	3.19
Vehicle1	846	18	(van, remainder)	(23.64, 76.36)	3.23
Ecoli2	336	7	(im, remainder)	(22.92, 77.08)	3.36
New-thyroid3	215	5	(hypo, remainder)	(16.89, 83.11)	4.92
New-thyroid2	215	5	(hyper, remainder)	(16.28, 83.72)	5.14
Ecoli3	336	7	(pp, remainder)	(15.48, 84.52)	5.46
Segment1	2308	19	(brickface, remainder)	(14.26, 85.74)	6.01
Glass7	214	9	(headlamps, remainder)	(13.55, 86.45)	6.38
Yeast4	1484	8	(ME3, remainder)	(10.98, 89.02)	8.11
Ecoli4	336	7	(iMU, remainder)	(10.88, 89.12)	8.19
Page-blocks	5472	10	(remainder, text)	(10.23, 89.77)	8.77
Data-sets with high imbalance (higher than 9 IR)					
Vowel0	988	13	(hid, remainder)	(9.01, 90.99)	10.10
Glass3	214	9	(Ve-win-float-proc, remainder)	(8.78, 91.22)	10.39
Ecoli5	336	7	(om, remainder)	(6.74, 93.26)	13.84
Glass5	214	9	(containers, remainder)	(6.07, 93.93)	15.47
Abalone9-18	731	8	(18, 9)	(5.65, 94.25)	16.68
Glass6	214	9	(tableware, remainder)	(4.20, 95.80)	22.81
YeastCYT-POX	482	8	(POX, CYT)	(4.15, 95.85)	23.10
Yeast5	1484	8	(ME2, remainder)	(3.43, 96.57)	28.41
Yeast6	1484	8	(ME1, remainder)	(2.96, 97.04)	32.78
Yeast7	1484	8	(EXC, remainder)	(2.49, 97.51)	39.16
Abalone19	4174	8	(19, remainder)	(0.77, 99.23)	128.87

- Bits per gene for the Gray codification (for incest prevention): 30 bits.

4.2. Empirical analysis

The first part of this study will be oriented to determine the granularity level of the fuzzy partitions, between 3 and 5 labels. In this manner, Table 4 presents the results with all imbalanced data-sets for the Chi FRBCS and in Table 5 we show the statistical analysis performed with a Wilcoxon's test.

There are no significant differences between both models, but since we obtain a higher ranking when using 3 labels per variable, we will employ this configuration in the study of the AIS. This analysis is shown in Table 6, where we include the results for all data-sets and for the three types of imbalanced data-sets proposed in the beginning of the experimental study.

Table 4
Average results table for the Chi FRBCS with 3 and 5 labels per variable.

Algorithm	GM_{Tr}	GM_{Tst}
Chi3	84.95 ± 1.48	$\mathbf{80.48 \pm 6.24}$
Chi5	90.24 ± 0.94	79.57 ± 6.00

Table 5
Wilcoxon's test to compare Chi with 3 labels per variable (R^+) against Chi with 5 labels per variable (R^-) in all imbalanced data-sets.

Comparison	R^+	R^-	Hypothesis ($\alpha = 0.05$)	*p*-Value
Chi3 vs. Chi5	352	209	Not Rejected	0.201

Table 6
Average results table for the Chi FRBCS with 3 labels per variable, basic approach and with AIS (parametric conjunction operator), for the different degrees of imbalance.

	Chi3		Chi3+AIS	
	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}
All data-sets	84.95 ± 1.48	80.48 ± 6.24	90.22 ± 1.19	$\mathbf{82.06 \pm 6.48}$
Low imbalance	80.22 ± 1.14	75.38 ± 4.09	85.89 ± 1.00	$\mathbf{77.71 \pm 4.01}$
Medium imbalance	90.78 ± 1.42	87.21 ± 5.72	95.30 ± 0.75	$\mathbf{89.50 \pm 4.93}$
High imbalance	83.85 ± 1.88	78.85 ± 8.90	89.46 ± 1.81	$\mathbf{78.96 \pm 10.49}$

Table 7
Wilcoxon's test to compare the basic Chi method (R^+) against the Chi approach with AIS (parametric conjunction operator) (R^-) in imbalanced data-sets.

Comparison	R^+	R^-	Hypothesis ($\alpha = 0.05$)	*p*-Value
All data-sets				
Chi3 vs. Chi3+AIS	107.5	453.5	Rejected for Chi3+AIS	0.002
Data-sets with low imbalance				
Chi3 vs. Chi3+AIS	6.0	60.0	Rejected for Chi3+AIS	0.016
Data-sets with medium imbalance				
Chi3 vs. Chi3+AIS	5.5	60.5	Rejected for Chi3+AIS	0.017
Data-sets with high imbalance				
Chi3 vs. Chi3+AIS	28	38	Not Rejected	0.657

Table 8
Detailed results table for the Chi FRBCS with 3 labels per variable, basic approach and with AIS (parametric conjunction operator).

Dataset	Chi3		Chi3 + AIS	
	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}
Data-sets with low imbalance				
EcoliCP-IM	95.49 ± 1.82	92.27 ± 5.93	98.52 ± 0.70	96.54 ± 5.04
Haberman	66.21 ± 2.86	58.91 ± 6.03	71.03 ± 2.29	58.33 ± 5.93
Iris1	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
Pima	72.31 ± 1.32	66.80 ± 5.93	82.37 ± 1.11	68.34 ± 5.84
Vehicle3	88.10 ± 1.22	85.54 ± 3.36	97.44 ± 0.49	91.32 ± 3.19
Wisconsin	98.07 ± 0.18	88.91 ± 2.13	98.86 ± 0.12	89.04 ± 1.68
Yeast2	68.33 ± 0.68	67.69 ± 1.91	76.13 ± 0.65	71.76 ± 1.70
Glass1	66.57 ± 1.08	64.06 ± 3.51	69.31 ± 2.09	64.14 ± 3.44
Glass2	75.37 ± 1.49	64.90 ± 6.91	85.52 ± 1.80	72.70 ± 5.72
Vehicle2	76.47 ± 1.00	70.92 ± 4.34	82.32 ± 0.94	71.04 ± 5.06
Vehicle4	75.52 ± 0.92	69.22 ± 4.89	83.28 ± 0.84	71.60 ± 6.56
Mean	80.22 ± 1.14	75.38 ± 4.09	85.89 ± 1.00	**77.71 ± 4.01**
Data-sets with medium imbalance				
Ecoli2	87.92 ± 2.30	85.28 ± 9.77	94.05 ± 0.40	83.04 ± 12.46
GlassNW	94.05 ± 1.69	85.83 ± 3.04	95.65 ± 1.43	88.33 ± 3.37
New-thyroid2	92.32 ± 3.35	87.44 ± 8.11	99.79 ± 0.19	93.40 ± 6.22
New-thyroid3	94.70 ± 1.43	89.81 ± 10.77	99.07 ± 0.97	96.49 ± 4.23
Page-blocks	80.60 ± 0.92	79.91 ± 4.29	84.29 ± 0.73	82.38 ± 4.16
Segment1	95.45 ± 0.28	94.99 ± 0.45	98.73 ± 0.19	97.21 ± 0.88
Vehicle1	88.23 ± 0.51	86.41 ± 3.06	96.36 ± 0.76	88.54 ± 2.83
Ecoli3	89.66 ± 1.24	88.01 ± 5.45	92.44 ± 0.84	89.42 ± 4.56
Yeast4	91.37 ± 0.80	90.13 ± 4.09	95.00 ± 0.37	90.29 ± 2.21
Ecoli4	89.24 ± 1.03	87.58 ± 4.08	95.42 ± 1.32	91.09 ± 3.96
Glass7	95.04 ± 2.04	83.87 ± 9.82	97.53 ± 1.02	84.30 ± 9.33
Mean	90.78 ± 1.42	87.21 ± 5.72	95.30 ± 0.75	**89.50 ± 4.93**
Data-sets with high imbalance				
Abalone9-18	69.80 ± 2.25	63.93 ± 11.00	77.88 ± 4.81	59.18 ± 15.57
Abalone19	70.39 ± 1.46	62.96 ± 8.27	75.91 ± 3.19	55.15 ± 12.56
Ecoli5	94.04 ± 1.47	91.27 ± 7.43	98.61 ± 0.43	91.90 ± 7.69
Glass3	58.00 ± 4.89	47.67 ± 10.16	75.65 ± 4.96	56.37 ± 18.01
Yeast5	83.44 ± 0.93	82.99 ± 3.10	89.60 ± 1.28	83.71 ± 4.32
Vowel0	98.56 ± 0.18	98.37 ± 0.61	99.86 ± 0.18	98.26 ± 1.58
YeastCYT-POX	75.66 ± 3.47	72.75 ± 14.99	79.70 ± 1.74	72.66 ± 14.88
Glass5	95.15 ± 0.96	84.96 ± 13.80	97.80 ± 0.75	86.09 ± 13.24
Glass6	94.15 ± 1.31	81.56 ± 12.65	98.21 ± 0.60	84.57 ± 14.33
Yeast6	94.67 ± 1.28	93.41 ± 5.35	98.14 ± 0.15	93.16 ± 4.77
Yeast7	88.43 ± 2.47	87.50 ± 10.55	92.71 ± 1.80	87.50 ± 8.38
Mean	83.85 ± 1.88	78.85 ± 8.90	89.46 ± 1.81	**78.96 ± 10.49**
All data-sets				
Mean	84.95 ± 1.48	80.48 ± 6.24	90.22 ± 1.19	**82.06 ± 6.48**

Our results clearly show that the use of the parametric conjunction operator implies a higher performance for the FRBCS in imbalanced data-sets. The null hypothesis for the Wilcoxon's test in all imbalanced data-sets (Table 7) has been rejected with a very small *p*-value, which supports our conclusion with a high degree of confidence.

Our interest is now focused on the behaviour of the parametric conjunction operator in the different imbalanced scenarios. In order to perform a detailed comparative study, Table 8 shows the results for the Chi basic approach and with parametric conjunction operator for every single data-set, to contrast the performance and robustness achieved for each model.

(1) **Data-sets with low imbalance:** The Chi method with the parametric conjunction operator approach obtains very good results in this case. In every single case the parametric conjunction operator improves the results of the basic Chi algorithm, except in the Haberman data-set, in which there is a very small difference, and in the Iris1, where there is a tie.
(2) **Data-sets with medium imbalance:** The same conclusion is extracted in this case, in which the parametric conjunction operator outperforms in all data-sets not including Ecoli2.
(3) **Data-sets with high imbalance:** Now the null hypothesis of Wilcoxon's test is not rejected, although the use of the parametric conjunction connector implies a higher ranking when comparing with the basic Chi approach. Regarding the results in each data-set, there are high differences in the Abalone9-18 and Abalone19 data-sets, which diminish the ranking of the AIS approach. Nevertheless, we can see that we obtain better results in most of the cases, following the same behaviour as in the previous imbalanced scenarios.

5. Conclusions

Our objective in this paper was to analyze the behaviour of FRBCSs in the framework of imbalanced data-sets, using an AIS whose parameters are learnt by GAs.

Our empirical results have shown the goodness of this approach. This conclusion has been supported with a high degree of confidence for all types imbalanced data-sets, which allows us to emphasize the robustness of this methodology, disregard the IR.

References

Alcalá-Fdez, J., Herrera, F., Márquez, F. A., & Peregrín, A. (2007). Increasing fuzzy rules cooperation based on evolutionary adaptive inference systems. *International Journal of Intelligent Systems, 22*(9), 1035–1064.

Barandela, R., Sánchez, J. S., García, V., & Rangel, E. (2003). Strategies for learning in class imbalance problems. *Pattern Recognition, 36*(3), 849–851.

Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behaviour of several methods for balancing machine learning training data. *SIGKDD Explorations, 6*(1), 20–29.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligent Research, 16*, 321–357.

Chawla, N. V., Japkowicz, N., & Kolcz, A. (2004). Editorial: Special issue on learning from imbalanced data sets. *SIGKDD Explorations, 6*(1), 1–6.

Chi, Z., Yan, H., & Pham, T. (1996). *Fuzzy algorithms with applications to image processing and pattern recognition*. World Scientific.

Cordón, O., del Jesus, M. J., & Herrera, F. (1999). A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning, 20*(1), 21–45.

Cordón, O., Gomide, F., Herrera, F., Hoffmann, F., & Magdalena, L. (2004). Ten years of genetic fuzzy systems: Current framework and new trends. *Fuzzy Sets and Systems, 141*(1), 5–31.

Crockett, K. A., Bandar, Z., Fowdar, J., & O'Shea, J. (2006). Genetic tuning of fuzzy inference within fuzzy classifier systems. *Expert Systems, 23*(2), 63–82.

Crockett, K. A., Bandar, Z., Mclean, D., & O'Shea, J. (2006). On constructing a fuzzy inference framework using crisp decision trees. *Fuzzy Sets and Systems, 157*, 2809–2832.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*, 1–30.

Domingos, P. (1999). Metacost: A general method for making classifiers cost sensitive. *Advances in Neural Networks, International Journal of Pattern Recognition and Artificial Intelligence*, 155–164.

Eshelman, L. J. (1991). *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.

Estabrooks, A., Jo, T., & Japkowicz, N. (2004). A multiple resampling method for learning from imbalanced data sets. *Computational Intelligence, 20*(1), 18–36.

Fawcett, T., & Provost, F. J. (1997). Adaptive fraud detection. *Data Mining and Knowledge Discovery, 1*(3), 291–316.

Fernández, A., García, S., del Jesus, M. J., Herrera, F. (2007). A study on the use of the fuzzy reasoning method based on the winning rule vs. voting procedure for classification with imbalanced data sets. In *Ninth international work-conference on artificial neural networks (IWANN07). Lecture notes on computer science* (Vol. 4507, pp. 375–382). Springer-Verlag.

Fernández, A., García, S., del Jesus, M. J., & Herrera, F. (2008). A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. *Fuzzy Sets and Systems, 159*(18), 2378–2398.

García, S., Fernández, A., Luengo, J., Herrera, F. (in press). A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability. Soft Computing. doi:10.1007/s00500-008-0392-y.

Ghosh, A., Pal, N., & Das, J. (2006). A fuzzy rule based approach to cloud cover estimation. *Remote Sensing of Environment, 100*(4), 531–549.

Herrera, F. (2008). Genetic fuzzy systems: Taxonomy, current research trends and prospects. *Evolutionary Intelligence, 1*, 27–46.

Hong, X., Chen, S., & Harris, C. J. (2007). A kernel-based two-class classifier for imbalanced data sets. *IEEE Transactions on Neural Networks, 18*(1), 28–41.

Huang, Y. M., Hung, C. M., & Jiau, H. C. (2006). Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem. *Nonlinear Analysis: Real World Applications, 7*(4), 720–747.

Hung, C.-M., & Huang, Y.-M. (2008). Conflict-sensitivity contexture learning algorithm for mining interesting patterns using neuro-fuzzy network with decision rules. *Expert Systems with Applications, 34*, 159–172.

Ishibuchi, H., Nakashima, T., & Nii, M. (2004). *Classification and modeling with linguistic information granules: Advanced approaches to linguistic data mining*. Springer-Verlag.

Ishibuchi, H., & Yamamoto, T. (2005). Rule weight specification in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems, 13*, 428–435.

Japkowicz, N., & Stephen, S. (2002). The class imbalance problem: A systematic study. *Intelligent Data Analysis, 6*(5), 429–450.

Kubat, M., Holte, R. C., & Matwin, S. (1998). Machine learning for the detection of oil spills in satellite radar images. *Machine Learning, 30*(2–3), 195–215.

Lee, C.-I., Tsai, C.-J., Wu, T.-Q., & Yang, W.-P. (2008). An approach to mining the multi-relational imbalanced database. *Expert Systems with Applications, 34*, 3021–3032.

Liao, T. W. (2008). Classification of weld flaws with imbalanced class data. *Expert Systems with Applications, 35*, 1041–1052.

Liu, B.-D., Chen, C.-Y., & Tsao, J.-Y. (2001). Design of adaptive fuzzy logic controller based on linguistic-hedge concepts and genetic algorithms. *IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics, 31*(1), 32–53.

Márquez, F. A., Peregrín, A., & Herrera, F. (2007). Cooperative evolutionary learning of fuzzy rules and parametric aggregation connectors for Mamdani linguistic fuzzy systems. *IEEE Transactions on Fuzzy Systems, 15*(6), 1162–1178.

Mazurowski, M., Habas, P., Zurada, J., Lo, J., Baker, J., & Tourassi, G. (2008). Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance. *Neural Networks, 21*(2–3), 427–436.

Mizumoto, M. (1989). Pictorial representations of fuzzy connectives, Part I: Cases of t-norms, t-conorms and averaging operators. *Fuzzy Sets and Systems, 31*, 217–242.

Nakashima, T., Schaefer, G., Yokota, Y., & Ishibuchi, H. (2007). A weighted fuzzy classifier and its application to image processing tasks. *Fuzzy Sets and Systems, 158*, 284–294.

Orriols-Puig, A., & Bernadó-Mansilla, E. (2009). Evolutionary rule-based systems for imbalanced datasets. *Soft Computing, 13*(3), 213–225.

Phua, C., Alahakoon, D., & Lee, V. (2004). Minority report in fraud detection: Classification of skewed data. *SIGKDD Explorations Newsletter, 6*(1), 50–59.

Sheskin, D. (2006). *Handbook of parametric and nonparametric statistical procedures* (2nd ed.). Chapman & Hall/CRC.

Su, C.-T., Chen, L.-S., & Yih, Y. (2006). Knowledge acquisition through information granulation for imbalanced data. *Expert Systems with Applications, 31*, 531–541.

Sun, Y., Kamel, M. S., Wong, A. K. C., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition, 40*, 3358–3378.

Tan, S. (2005). Neighbor-weighted k-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications, 28*(4), 667–671.

Tsang, C., Kwong, S., & Wang, H. (2007). Genetic-fuzzy rule mining approach and evaluation of feature selection techniques for anomaly intrusion detection. *Pattern Recognition, 40*(9), 2373–2391.

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

Weiss, G. M. (2004). Mining with rarity: A unifying framework. *SIGKDD Explorations, 6*(1), 7–19.

Weiss, G., & Provost, F. (2003). Learning when training data are costly: The effect of class distribution on tree induction. *Journal of Artificial Intelligence Research, 19*, 315–354.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics, 1*, 80–83.

Wu, H., & Mendel, J. M. (2004). On choosing models for linguistic connector words for Mamdani fuzzy logic systems. *IEEE Transactions on Fuzzy Systems, 12*(1), 29–44.

Xu, L., Chow, M. Y., & Taylor, L. S. (2007). Power distribution fault cause identification with imbalanced data using the data mining-based fuzzy classification E-algorithm. *IEEE Transactions on Power Systems, 22*(1), 164–171.