Improving the Performance of Fuzzy Rule Based Classification Systems for Highly Imbalanced Data-Sets Using an Evolutionary Adaptive Inference System[⋆]

Alberto Fernández[1,⋆⋆], María José del Jesus[2], and Francisco Herrera[1]

[1] Dept. of Computer Science and A.I., University of Granada
{alberto,herrera}@decsai.ugr.es
[2] Dept. of Computer Science, University of Jaén, Spain
mjjesus@ujaen.es

Abstract. In this contribution, we study the influence of an Evolutionary Adaptive Inference System with parametric conjunction operators for Fuzzy Rule Based Classification Systems. Specifically, we work in the context of highly imbalanced data-sets, which is a common scenario in real applications, since the number of examples that represents one of the classes of the data-set (usually the concept of interest) is usually much lower than that of the other classes.

Our experimental study shows empirically that the use of the parametric conjunction operators enables simple Fuzzy Rule Based Classification Systems to enhance their performance for data-sets with a high imbalance ratio.

Keywords: Fuzzy Rule-Based Classification Systems, Inference System, Parametric Conjunction Operators, Genetic Fuzzy Systems, Imbalanced Data-Sets.

1 Introduction

This work is focused on the use of Fuzzy Rule Based Classification Systems (FRBCSs), for classification with imbalanced data-sets [1]. This problem occurs when one or more classes are represented by few examples (known as positive classes), whereas the others are described by many instances (negative classes).

Standard classification algorithms are usually biased towards the majority classes trying to maximize the overall accuracy and their performance is poor on highly imbalanced data-sets. In our previous study on the topic [2], we studied different configurations for FRBCS in order to determine the most suitable model in this framework. 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) [3].

[⋆⋆] Corresponding author. Tel.:+34-958-240598; Fax: +34-958-243317

Our aim now is to improve the fuzzy system behaviour by means of the application of parametrized expressions for the conjunction connectors in the Inference System (IS), while maintaining the original interpretability associated to fuzzy systems [4]. This approach is usually called Adaptive Inference System (AIS) and it has shown very good results in fuzzy modelling [5,6]. The idea is to perform a tuning of parameters with a Rule Base (RB) previously established, using Genetic Algorithms (GAs) as a tool to evolve the connector parameters resulting on an Evolutionary Adaptive Inference System (EAIS). Thus, we will increase the performance of the fuzzy models to make them competitive with C4.5 [7], a decision tree algorithm that presents a good behaviour in imbalanced data-sets [8,9].

We will develop an experimental study with twenty-two data-sets from UCI repository with a high imbalance ratio (IR). Multi-class data sets are modified to obtain two-class non-balanced problems, defining the joint of one or more classes as positive and the joint of one or more classes as negative. To evaluate our results we have applied the Area Under the Curve (AUC) metric [10] carrying out some non-parametric tests [11,12] with the aim to show the significance in the performance improvements obtained with the EAIS model.

In order to do that, the contribution 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 [13]. Next, Section 4 introduces the EAIS with the parametric conjunction operators. In Section 5 we include our experimental analysis in highly imbalanced data-sets. Finally, in Section 6 some concluding remarks are pointed out.

2   Imbalanced Data-Sets in Classification

Learning from imbalanced data is an important topic that has recently appeared in the Machine Learning community [1]. This problem is very representative since it appears in a variety of real-world applications including, but not limited to, medical applications, finance, telecommunications, biology and so on.

In this framework, the class distribution is not uniform, resulting on a high number of examples belonging to one or more classes and a few from the rest. The minority classes are usually associated to the most interesting concepts from the point of view of learning and, due to that fact, the cost derived from a misclassification of one of the examples of these classes is higher than that of the majority classes. In this work we will focus on binary problems where there are just one positive and negative class.

Standard classifier algorithms have a bias towards the majority class, since the rules that predicts the higher number of examples are positively weighted during the learning process in favour of the accuracy metric. Consequently, the instances that belongs to the minority class are misclassified more often than those belonging to the majority class. Nevertheless, the most important issues

of this type of problems are the apparition of small disjuncts and the overlapping between the examples of the positive and the negative classes.

In our previous work on this topic [2], 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 [3]. According to this, we will employ in this contribution the SMOTE algorithm in order to deal with the problem of imbalanced data-sets.

We will use the IR [8] as a threshold to categorize the different imbalanced scenarios, which is defined as the ratio of the number of instances of the majority class and the minority class. We consider that a data-set presents a high degree of imbalance when its IR is higher than 9 (less than 10% of positive instances).

Regarding the empirical measure, instead of using accuracy, a more correct metric is considered. This is due to the fact that accuracy can lead to erroneous conclusions, since it doesn't take into account the proportion of examples for each class. Because of this, in this work we use the AUC metric [10], which can be defined as

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

where TP_{rate} is the percentage of positive cases correctly classified as belonging to the positive class and FP_{rate} is the percentage of negative cases misclassified as belonging to the positive class.

3   Fuzzy Rule Based Classification Systems and Linguistic Rule Generation Method

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 p-th training pattern.

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

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

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

Fuzzy learning methods are the basis to build a FRBCS. The algorithm used in this work is the method proposed in [13], that we have called the Chi et al.'s rule generation. 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.

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.

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

4   Evolutionary Adaptive Inference System

The conjunction operator is employed in the IS to compute the activation strength of the if-part for all rules in the RB with an input pattern. This operator is suitable to be parameterized in order to adapt the IS showing a considerable improvement in the accuracy of linguistic Fuzzy Systems [5,6].

We will use Dubois t-norm, because it is more efficiently computed and since it has obtained a better behaviour than other parametric t-norms [5]. The expression for the computation of this parametric t-norm is shown below:

$$T_{Dubois}(x, y, \alpha) = \frac{x \cdot y}{\max(x, y, \alpha)} \quad (0 \leq \alpha \leq 1) \tag{3}$$

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.

In this work we will use a model that considers individual parameters α_i for every rule of the KB, having a local tuning mechanism of the behavior of the IS in order to obtain a good accuracy level, because of the high degree of freedom of this model. We will consider the use of a specific GA to design the proposed learning method, the CHC algorithm. The CHC algorithm is a GA that presents a good trade-off between 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:* Each chromosome will be composed by R genes, being R the number of rules in the RB, having a value between 0 and 1.
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, the AUC measure.
3. *Initial Gene Pool:* 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:* We consider the Parent Centric BLX (PCBLX) operator with incest prevention mechanism. In this manner, two parents are crossed if their hamming distance divided by 2 is above a predetermined threshold, L. Since we consider a real coding scheme, we have to transform each gene considering a Gray Code (binary code) with a fixed number of bits per gene ($BITSGENE$). In this way, the threshold value is initialized as:

$$L = (\#Genes \cdot BITSGENE)/4.0$$

   L is decremented by $BITSGENE$ when there are no new individuals in the next generation.
5. *Restarting approach:* When the threshold value L is lower than zero, all the chromosomes are regenerated at random within the interval [0.0, 1.0]. Furthermore, the best global solution found is included in the population to increase the convergence of the algorithm.

5   Experimental Study

The objective of this work is to analyse the behaviour achieved by FRBCSs using the EAIS model. With this aim, we will perform an empirical study using a large collection of highly imbalanced data-sets to support our analysis. We will also include the C4.5 decision tree [7], since this method has a good behaviour in the framework of imbalanced data-sets [8,9].

We have selected twenty-two data-sets with different IR from UCI repository. The data is summarized in Table 1, showing the number of examples (#Ex.), number of attributes (#Atts.), class name of each class (minority and majority), class attribute distribution and IR.

In the remaining of this section, we will first present the experimental framework and all the parameters employed in this study. Then, we will illustrate how the EAIS model enhances the behaviour of the base FRBCS model also contrasting its performance with C4.5.

5.1   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

Table 1. Summary Description for Imbalanced Data-Sets

Data-set	#Ex.	#Atts.	Class (min., maj.)	%Class(min.; maj.)	IR
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

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 [3] for all our experiments, considering only the 1-nearest neighbour to generate the synthetic samples, and balancing both classes to the 50% distribution.

We will carry out a rigorous comparison of the performance obtained by the algorithms used in this work. We consider the use of non-parametric tests due to the fact that the initial conditions that guarantee the reliability of the parametric tests are not satisfied in a single data-set analysis [11,12]. We will use Wilcoxon's Signed-Ranks Test [14], which computes the ranking of both algorithms (noted as R^+ and R^-) with respect to the absolute differences between their performance in each data-set (average ranks are assigned in case of ties). The level of confidence (α) will be set at 0.05 (95% of confidence).

We will employ the following configuration for the FRBCS: 3 fuzzy labels, product T-norm as conjunction operator (for the base FRBCS), together with the Penalized Certainty Factor approach [15] for the rule weight and FRM of the winning rule. Finally, the values that have been considered for the parameters of the CHC algorithm are the following ones: 50 individuals for the population size, $5,000 \cdot number\ of\ variables$ as number of evaluations and 30 bits per gene for the Gray codification (incest prevention mechanism).

5.2   Analysis of the Behaviour of the EAIS

The results for the Chi et al.'s algorithm, the Chi method with EAIS and C4.5 are shown in Table 2. We observe that the EAIS improves the performance for the FRBCS in many data-sets and achieves the best global result among the algorithms used in the analysis.

Table 2. Detailed results table for the Chi FRBCS, basic approach and with EAIS (parametric conjunction operator), and C4.5 for highly imbalanced data-sets

Data-set	Chi		Chi+EIAS		C4.5	
	AUC_{Tr}	AUC_{Tst}	AUC_{Tr}	AUC_{Tst}	AUC_{Tr}	AUC_{Tst}
Yeast2vs4	89.68 ± 1.30	87.36 ± 5.16	92.99 ± 1.91	86.31 ± 5.29	98.14 ± 0.88	85.88 ± 8.78
Yeast05679vs4	82.65 ± 1.38	79.17 ± 5.66	86.38 ± 1.39	80.75 ± 5.92	95.26 ± 0.94	76.02 ± 9.36
Vowel0	98.57 ± 0.18	98.39 ± 0.60	99.86 ± 0.18	98.28 ± 1.57	99.67 ± 0.48	94.94 ± 4.95
Glass016vs2	62.71 ± 2.15	54.17 ± 6.82	73.52 ± 6.50	63.69 ± 15.38	97.16 ± 1.86	60.62 ± 12.66
Glass2	66.54 ± 2.18	55.30 ± 14.48	77.39 ± 2.97	59.71 ± 17.82	95.71 ± 1.51	54.24 ± 14.01
Ecoli4	94.06 ± 1.49	91.51 ± 7.21	98.62 ± 0.42	92.15 ± 7.45	97.69 ± 1.96	83.10 ± 9.90
Yeast1vs7	82.00 ± 2.34	80.63 ± 6.61	86.82 ± 1.71	73.93 ± 12.07	93.51 ± 2.20	70.03 ± 1.46
Shuttle0vs4	100.00 ± 0.00	99.12 ± 1.14	100.00 ± 0.00	99.12 ± 1.14	99.99 ± 0.02	99.97 ± 0.07
Glass4	95.27 ± 0.91	85.70 ± 12.92	97.82 ± 0.73	86.94 ± 12.10	98.44 ± 2.29	85.08 ± 9.35
Page-Blocks13vs2	93.68 ± 2.41	92.05 ± 4.73	97.38 ± 0.58	93.94 ± 4.42	99.75 ± 0.21	99.55 ± 0.47
Abalone9-18	70.23 ± 2.25	64.70 ± 10.73	78.08 ± 4.68	62.52 ± 12.67	95.31 ± 4.44	62.15 ± 4.96
Glass016vs5	90.57 ± 4.12	79.71 ± 23.29	98.57 ± 0.56	86.29 ± 22.71	99.21 ± 0.47	81.29 ± 24.44
Shuttle2vs4	95.00 ± 4.71	90.78 ± 7.80	98.98 ± 2.28	96.38 ± 6.98	99.90 ± 0.23	99.17 ± 1.86
Yeast1458vs7	71.25 ± 3.52	64.65 ± 5.92	80.50 ± 1.05	58.79 ± 4.48	91.58 ± 2.78	53.67 ± 2.09
Glass5	94.33 ± 1.23	83.17 ± 11.12	98.23 ± 0.59	86.10 ± 12.45	99.76 ± 0.40	88.29 ± 13.31
Yeast2vs8	78.61 ± 2.61	77.28 ± 10.36	81.71 ± 1.41	77.18 ± 10.23	91.25 ± 1.84	80.66 ± 11.22
Yeast4	83.58 ± 0.93	83.15 ± 2.96	89.63 ± 1.31	83.92 ± 4.12	91.01 ± 2.64	70.04 ± 5.65
Yeast1289vs7	74.70 ± 1.79	77.12 ± 6.50	84.65 ± 2.45	74.51 ± 3.91	94.65 ± 1.13	68.32 ± 6.16
Yeast5	94.68 ± 1.28	93.58 ± 5.11	98.16 ± 0.14	93.30 ± 4.61	97.77 ± 1.45	92.33 ± 4.72
Ecoli0137vs26	93.96 ± 5.63	81.90 ± 20.49	98.27 ± 1.02	82.08 ± 20.58	96.78 ± 3.28	81.36 ± 21.68
Yeast6	88.48 ± 2.38	88.09 ± 9.82	92.73 ± 1.79	87.94 ± 8.03	92.42 ± 3.54	82.80 ± 12.77
Abalone19	71.44 ± 1.82	63.94 ± 9.32	76.76 ± 3.26	58.23 ± 8.12	85.44 ± 2.49	52.02 ± 4.41
Mean	85.09 ± 2.12	80.52 ± 8.58	90.32 ± 1.68	**81.00 ± 9.18**	95.93 ± 1.68	78.25 ± 8.38

In order to support this conclusion, we carry out a Wilcoxon test (shown in Table 3) comparing the Chi et al.'s method versus the EAIS. We observe that the EAIS achieves a higher ranking, which implies the robustness of this approach. When comparing with C4.5 (also in Table 3) we must emphasize that the use of the EAIS outperforms C4.5, whereas the simple FRBCS is not enough to obtain significant differences.

Table 3. Wilcoxon's test to compare the basic Chi method (R^+) against the Chi approach with EAIS (parametric conjunction operator) (R^-)

Comparison	R^+	R^-	Hypothesis ($\alpha = 0.05$)	p-value
Chi+AIS vs. Chi	143.5	109.5	Not Rejected	0.566
Chi vs. C4.5	176.0	77.0	Not Rejected	0.108
Chi+AIS vs. C4.5	205.0	48.0	Rejected for Chi+AIS	0.011

In brief, the EAIS is a good design option in order to improve the accuracy in linguistic FRBCSs for imbalanced data-sets with a high IR. Furthermore, 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.

6   Conclusions

The aim of this contribution was to study the goodness of the EAIS model to improve the performance of simple FRBCS in the framework of imbalanced data-sets with a high IR. The idea is to maintain the initial interpretability of the fuzzy model by changing the computation of the conjunction operator.

Our experimental analysis have shown that the EAIS approach obtains very good results in highly imbalanced data-sets in contrast with the base FRBCS. Furthermore, we observe that the Chi et al. method by itself do not outperforms the results of the C4.5 decision tree, but this algorithm is highly improved with the application of the EAIS model, obtaining significant differences with respect to C4.5 in this case.

References

1. Chawla, N.V., Japkowicz, N., Kolcz, A.: Editorial: special issue on learning from imbalanced data sets. SIGKDD Explorations 6(1), 1–6 (2004)
2. Fernández, A., García, S., del Jesus, M.J., Herrera, F.: 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 (2008)
3. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligent Research 16, 321–357 (2002)
4. Crockett, K.A., Bandar, Z., Mclean, D., O'Shea, J.: On constructing a fuzzy inference framework using crisp decision trees. Fuzzy Sets and Systems 157, 2809–2832 (2006)
5. Alcalá-Fdez, J., Herrera, F., Márquez, F.A., Peregrín, A.: Increasing fuzzy rules cooperation based on evolutionary adaptive inference systems. International Journal of Intelligent Systems 22(9), 1035–1064 (2007)
6. Márquez, F.A., Peregrín, A., Herrera, F.: Cooperative evolutionary learning of fuzzy rules and parametric aggregation connectors for mamdani linguistic fuzzy systems. IEEE Transactions on Fuzzy Systems 15(6), 1162–1178 (2007)
7. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo (1993)
8. Orriols-Puig, A., Bernadó-Mansilla, E.: Evolutionary rule-based systems for imbalanced datasets. Soft Computing 13(3), 213–225 (2009)
9. Su, C.T., Hsiao, Y.H.: An evaluation of the robustness of MTS for imbalanced data. IEEE Transactions on Knowledge Data Engineering 19(10), 1321–1332 (2007)
10. Huang, J., Ling, C.X.: Using AUC and accuracy in evaluating learning algorithms. IEEE Transactions on Knowledge and Data Engineering 17(3), 299–310 (2005)
11. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30 (2006)
12. García, S., Fernández, A., Luengo, J., Herrera, F.: A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability. Soft Computing 13(10), 959–977 (2009)
13. Chi, Z., Yan, H., Pham, T.: Fuzzy algorithms with applications to image processing and pattern recognition. World Scientific, Singapore (1996)
14. Sheskin, D.: Handbook of parametric and nonparametric statistical procedures. Chapman & Hall/CRC, Boca Raton (2006)
15. Ishibuchi, H., Yamamoto, T.: Rule weight specification in fuzzy rule-based classification systems. IEEE Transactions on Fuzzy Systems 13, 428–435 (2005)