An Analysis of the Rule Weights and Fuzzy Reasoning Methods for Linguistic Rule Based Classification Systems Applied to Problems with Highly Imbalanced Data Sets

Alberto Fernández[1], Salvador García[1], Francisco Herrera[1],
and María José del Jesús[2]

[1] Dept. of Computer Science and A.I., University of Granada
{alberto,salvagl,herrera}@decsai.ugr.es
http://sci2s.ugr.es

[2] Dept. of Computer Science, University of Jaén, Spain
mjjesus@ujaen.es
http://wwwdi.ujaen.es

Abstract. In this contribution we carry out an analysis of the rule weights and Fuzzy Reasoning Methods for Fuzzy Rule Based Classification Systems in the framework of imbalanced data-sets with a high imbalance degree. We analyze the behaviour of the Fuzzy Rule Based Classification Systems searching for the best configuration of rule weight and Fuzzy Reasoning Method also studying the cooperation of some pre-processing methods of instances. To do so we use a simple rule base obtained with the Chi (and co-authors') method that extends the well-known Wang and Mendel method to classification problems.

The results obtained show the necessity to apply an instance pre-processing step and the clear differences in the use of the rule weight and Fuzzy Reasoning Method.

Finally, it is empirically proved that there is a superior performance of Fuzzy Rule Based Classification Systems compared to the 1-NN and C4.5 classifiers in the framework of highly imbalanced data-sets.

Keywords: Fuzzy Rule Based Classification Systems, Over-sampling, Imbalanced Data-sets, rule weight, Fuzzy Reasoning Method.

1 Introduction

In the last years the data-set imbalance problem has demanded more attention by researchers in the field of classification [3]. This problem occurs when the number of instances of one class overwhelms the others. In this contribution we focus on the two class imbalanced data-sets, where there are only one positive and one negative class. We consider the positive class as the one with the lower number of examples.

We may distinguish between three degrees of imbalance: a *low imbalance degree* when the instances of the positive class are between the 25 and 40% of

F. Masulli, S. Mitra, and G. Pasi (Eds.): WILF 2007, LNAI 4578, pp. 170–178, 2007.

the total instances, a *medium imbalance degree* when the number of the positive instances is between the 10 and 25% of the total instances and a *high imbalance degree* where there are no more than the 10% of positive instances in the whole data-set compared to the negative ones.

To solve the imbalance data-set problem there are two main types of solutions: solutions at the data level which is achieved balancing the class distribution and solutions at the algorithmic level, for example adjusting the cost per class.

In this work we study the performance of the Fuzzy Rule Based Classification Systems (FRBCSs) [8] in the field of high imbalanced data-sets. In order to deal with the class imbalance problem we analyze the cooperation of some preprocessing methods of instances.

Our aim is to locate the best configuration of rule weight and Fuzzy Reasoning Method (FRM) for imbalanced data-sets with a high degree of imbalance. To do so we use a simple rule base obtained with the Chi (and co-authors') method [4] that extends the well-known Wang and Mendel method [11] to classification problems. We use triangular membership functions with five labels per variable.

Finally we will compare our results with the classic KNN and C4.5 classifiers to prove the higher performance of our model.

In order to do that, this contribution is organized as follows. In Section 2 we introduce the FRBCS, the inductive learning algorithm used and the rule weights and FRMs. Then in Section 3 we propose some preprocessing techniques for imbalanced data-sets. Section 4 shows the experimental study carried out with seven different data-sets. Finally, in Section 5 we present some conclusions about the study done.

2 Fuzzy Rule Based Classification Systems

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:

$$\text{Rule } R_j : \text{ If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn} \text{ then Class} = C_j \text{ with } RW_j \quad (1)$$

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 a rule weight. As antecedent fuzzy sets we use triangular fuzzy sets with 5 partitions per variable.

To generate the fuzzy Rule Base we use the method proposed in [4] that extends the Wang and Mendel method [11] to classification problems. This FRBCS 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 determined the domain of variation of each feature A_i, 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, which antecedent is determined by the selected fuzzy region and with the label of class of the example in the consequent.
   - 2.4 To compute the rule weight.

In the following of this section we will first explain the use of rule weights for fuzzy rules and the different types of weights analyzed in this work and then we will introduce the two FRMs employed: classification via the winning rule or via a voting procedure.

2.1 Rule Weights for Fuzzy Rules

Rule weights are used in FRBCSs in order to improve their performance [7]. In the literature different mechanisms have been employed to compute the rule weight. In [9] we can find some heuristics methods for rule weight specification, where the most common one is the Certainty Factor (CF) [5]:

$$CF_j^I = \frac{\sum_{x_p \in ClassC_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{m} \mu_{A_j}(x_p)} \tag{2}$$

In addition, in [9] another definition for the rule weight is proposed:

$$CF_j^{IV} = CF_j^I - \frac{\sum_{x_p \notin ClassC_j} \mu_{A_j}(x_p)}{\sum_{p=1}^{m} \mu_{A_j}(x_p)} \tag{3}$$

This new definition may be named as the Penalized CF.

A third rule weight is used in this work. In [10], Mansoori et al., using weighting functions, modify the compatibility degree of patterns to improve the classification accuracy. Their approach specifies a positive pattern (i.e. pattern with the true class) from the covering subspace of each fuzzy rule as splitting pattern and uses its compatibility grade as threshold. All patterns having compatibility grade above this threshold are positive so any incoming pattern for this subdivision should be classified as positive. When using rule weights, the weighting function for R_j is computed as:

$$\mu_{A_j}^{out}(x_p) \cdot CF_j = \begin{cases} \mu_{A_j}^{in}(x_p) \cdot CF_j & \text{if } \mu_{A_j}^{in}(x_p) < n_j \\ \left(\frac{p_j - n_j \cdot CF_j}{m_j - n_j}\right) \cdot \mu_{A_j}^{in}(x_p) - \left(\frac{p_j - m_j \cdot CF_j}{m_j - n_j}\right) \cdot n_j & \text{if } n_j \leqslant \mu_{A_j}^{in}(x_p) < m_j \\ CF_j \cdot \mu_{A_j}^{in}(x_p) - CF_j \cdot m_j + p_j & \text{if } \mu_{A_j}^{in}(x_p) \geqslant m_j \end{cases} \tag{4}$$

where $\mu_{A_j}^{out}(x_p){\cdot}CF_j$ is the output degree of association used in the FRM, $\mu_{A_j}^{in}(x_p)$ is the compatibility grade of x_p, CF_j is the rule weight and the parameters n_j, m_j, p_j are obtained as:

$$\begin{aligned} n_j &= t_j\sqrt{\tfrac{2}{1+CF_j^2}} \\ m_j &= \{t_j \cdot (CF_j+1) - (CF_j-1)\}/\sqrt{2CF_j^2+2} \\ p_j &= \{t_j \cdot (CF_j-1) - (CF_j+1)\}/\sqrt{2CF_j^2+2} \end{aligned} \quad (5)$$

where t_j is the compatibility grade threshold for Rule R_j. For more details of this proposal please refer to [10]

2.2 Fuzzy Reasoning Methods

We study the performance of two different FRMs for classifying new patterns by the rule set. We use the classic fuzzy reasoning method or maximum matching, where every new pattern is classified as the consequent class of a single winner rule which is determined as

$$\mu_{A_w}(x_p) \cdot CF_w = max\{\mu_{A_q}(x_p) \cdot CF_q | R_q \in S\} \quad (6)$$

where S is the set of fuzzy rules of the form in (1) and $x_p = (x_{p1}, \ldots, x_{pn})$ is the pattern example. We also use a weighted vote method or additive combination method where each fuzzy rule casts a vote for its consequent class. The total strength of the vote for each class is computed as follows:

$$V_{Class_h}(x_p) = \sum_{R_q \in S; C_q = h} \mu_{A_q}(x_p) \cdot CF_q, h = 1, 2, \ldots, M. \quad (7)$$

The pattern x_p is classified as the class with maximum total strength of the vote.

3 Preprocessing Imbalanced Data-Sets

In order to deal with the imbalanced data-set problem we can distinguish between two kind of solutions: those applied at the data level such as instance selection and those applied at the algorithmic level. In this work we evaluate different instance selection based on oversampling and hybrid techniques to adjust the class distribution in the training data. Specifically we have chosen the following methods which have been studied in [2]:

- Oversampling methods:
  - **Random Over-Sampling.** Is a non-heuristic method that aims to balance class distribution through the random replication of minority class examples.
  - **"Synthetic Minority Over-Sampling Technique (Smote)".** 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.

- Hybrid methods: Oversampling + Undersampling:
  - **"Smote + Tomek links".** In order to create better-defined class clusters, Tomek links may be applied to the over-sampled training set as a data cleaning method. Instead of removing only the majority class examples that form Tomek links, examples from both classes are removed.
  - **"Smote + ENN".** After applying the Smote mechanism, ENN is used to remove examples from both classes. Any example that is misclassified by its three nearest neighbors is removed from the training set.

For a further explanation please refer to [2]. The preprocessing methods chosen are the ones based on oversampling because they are proved to provide a good performance for imbalanced data-sets when using FRBCSs [6].

4 Experimental Study

In this section we introduce our experimentation framework. First we will describe the data-sets we have chosen for this work and all the parameters used. Then we will present our results and we will make a brief analysis according to our objetives.

4.1 Data-Sets and Parameters

In this study we have considered seven data sets from UCI with a high imbalance degree. Table 1 summarizes the data employed in this study and shows, for each data set the number of examples (#Examples), number of attributes (#Atts.), class name of each class (majority and minority) and class attribute distribution.

Table 1. Data Sets summary descriptions

Data set	**#Examples**	**#Atts.**	**Class (min., maj.)**	**%Class(min.,maj.)**
Abalone9-18	731	8	(18, 9)	(5.65,94.25)
Abalone19	4174	8	(19, remainder)	(0.77,99.23)
EcoliMO	336	7	(MO, remainder)	(6.74,93.26)
Glass	214	9	(Ve-win-float-proc, remainder)	(8.78,91.22)
Vowel0	988	13	(hid, remainder)	(9.01,90.99)
YeastCYT-POX	482	8	(POX,CYT)	(4.15,95.85)
Yeast5	1484	8	(ME5, remainder)	(3.4,96.6)

In order to develop a comparative study, we use a five fold cross validation approach, that is, five partitions where the 80% is used for training and the 20% for test. For each data-set we consider the average results of the five partitions.

We consider the following parameters for the Chi et al. algorithm:

- Membership Function: Linear triangular membership function.
- Number of labels per fuzzy partition: 5 labels.
- Computation of the compatibility degree: Minimum and Product T-norm.

- Combination of compatibility degree and rule weight: Product T-norm.
- Inference method: Classic method (winning rule) and additive combination among rules classification degree per class (voting procedure).

Table 2. Percentage of classes after balancing for the seven data-sets

Balancing Method	% Positives	% Negatives
None (Original Data-Sets)	5.5	94.5
RandomOverSampling	50.0	50.0
SMOTE	50.0	50.0
SMOTE-TomekLinks	45.78	54.22
SMOTE-ENN	46.1	53.9

In Table 2 the percentages of examples for each class after balancing are shown, together with the original percentage.

As we are in the imbalanced data-set field a properly evaluation measure must be used. We employ the geometric mean metric (8), suggested in [1] where acc^+ is the accuracy classification on the positive instances, and acc^- the accuracy on the negative ones.

$$GM = \sqrt{acc^+ \cdot acc^-} \tag{8}$$

4.2 Results and Analysis

Our study is oriented to compare and find the best configuration for FRBCSs in the framework of highly imbalanced data-sets. In this section we present the average results for the FRBCSs obtained by the Chi et al. method for the different rule weight, T-norm and FRM used. Then we demonstrate the necessity to apply a preprocessing step to transform the data into a more balanced set and we analyze the best configuration (T-norm, rule weight and FRM) found. Finally we make a comparison between the classic algorithms and the FRBCSs used in this work in order to prove the validity of this model.

In Table 3 we show a comparative of the average results obtained with the FRBCS method (Chi et al.) with the C4.5 decision tree model and 1-NN algorithms for the 7 data-sets chosen in this work.

The following information is showed by columns:

- The first colum “Weight” is the rule weight used in the FRBCS. CF stands for the classic Certainty Factor, P-CF stands for the Penalized CF and M-CF stands for the Mansoori weighting system.
- Inside column “t-norm” we note if the results correspond to minimum or product T-norm.
- In the third column “FRM” we distinguish between each type of FRM, where WR stands for the Winning Rule method and AC stands for the Additive Combination method
- Finally in the rest of the columns the average results for the geometric mean in training (GM_{Tr}) and test (GM_{Tst}) are showed for each type of preprocessing method, where None indicates that the data-set employed in the experiment is the original one (without preprocessing).

Table 3. Global comparison of the average results for FRBCSs with different Rule Weigths, T-norm and FRM. Including the results for C4.5 and 1-NN algorithms.

			None		RandomOS		SMOTE		SMOTE-TL		SMOTE-ENN	
Weight	T-norm	FRM	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}
CF	Minimum	WR	51.03	40.21	84.25	73.59	83.52	74.4	83.06	*75.17*	83.01	75.05
CF	Product	WR	53.96	40.51	85.74	73.1	84.56	74.68	84.17	75.07	84.09	*75.17*
P-CF	Minimum	WR	45.83	37.0	83.7	75.07	82.98	*75.11*	82.32	74.73	82.12	74.79
P-CF	Product	WR	51.72	39.76	84.48	73.99	83.78	**76.32**	83.36	76.13	83.28	76.09
M-CF	Minimum	WR	45.87	34.9	81.54	69.78	80.77	70.21	80.54	*71.25*	80.9	71.01
M-CF	Product	WR	50.69	35.93	83.85	70.98	82.55	71.5	82.16	*72.17*	82.25	72.14
CF	Minimum	AC	46.39	36.8	74.54	60.99	81.4	66.31	81.16	*68.18*	81.44	67.95
CF	Product	AC	51.25	40.06	81.01	64.84	83.77	70.72	83.43	*70.93*	83.76	70.22
P-CF	Minimum	AC	43.15	34.43	83.08	72.39	82.4	72.99	81.84	*73.11*	81.65	73.01
P-CF	Product	AC	50.24	36.89	84.3	72.76	83.7	74.3	83.15	74.68	83.09	*75.17*
M-CF	Minimum	AC	30.55	21.58	79.36	68.57	79.44	69.3	79.32	69.75	79.35	*69.9*
M-CF	Product	AC	36.41	21.43	80.87	68.42	79.61	68.48	79.01	*69.12*	79.1	68.88
C4.5	—	—	59.07	40.52	99.57	62.89	93.46	60.27	93.47	*68.16*	93.58	67.86
1-NN	—	—	58.69	57.45	58.69	57.45	97.56	67.6	97.17	67.95	96.96	*68.39*

We focus our analysis on the generalization capacity via the test partition. In **bold** the best results for test are stressed. In underline the best results in columns, that is, for each preprocessing method, are marked. In *italics* we may observe the best results in rows, that is, for the different configurations for FRBCSs and for C4.5 and 1-NN.

As we can see in Table 3 there is a huge difference in the results when we apply a preprocessing mechanism to balance the data comparing with the results without preprocessing. The performance achieved with the FRBCS with the original data-sets is roughly 30 points below the performance in the case of preprocessed data-sets via oversampling, which confirms the necessity to transform the data-sets into a more balanced format. Since there is a clear over-fitting with Random-Oversampling, we select as most appropriate the methods based in the SMOTE family.

The use of the Penalized CF obtains the best results for both types of T-norms and FRMs. Specifically the highest performance is achieved in the case of Penalized CF with product T-norm and FRM of the Winning Rule.

Regarding the use of the T-norms applied in this work we found that in most cases the product T-norm is more effective. Comparing the two FRMs applied in this work, it is clear that the one based in the Winning Rule is much better than the one based in the Additive Combination.

When we compare our results obtained with FRBCSs with the classic reference algorithms (C4.5 and 1-NN) we can conclude that, for highly imbalanced data-sets, our methods outperforms these algorithms. We can observe a high over-fitting in the classic algorithms, with a difference in almost 30 points between the training and test results.

5 Concluding Remarks

In this work we have analyzed the performance of the FRBCSs searching for the best configuration of rule weight and FRM in the framework of highly imbalanced data-sets. Also we have studied the cooperation of some pre-processing methods of instances.

Our results shown the necessity of using pre-processing instances methods to improve the balance between classes before the use of the FRBCS method. We have found a kind of mechanism (SMOTE) that provides very good results as a preprocessing technique for FRBCSs. It helps fuzzy methods (Chi et al. in this case) to became a very competitive model in high imbalanced domains.

We have also studied the differences for the most appropriate configuration for rule weight and FRM in highly imbalanced data-sets, concluding that the Penalized CF is the most accurate for the rule weight and the Winning Rule is the best selection for the FRM.

Finally we have found a superior behaviour of the FRBCSs against the classic algorithms 1-NN and C4.5. In this way we can see fuzzy methods as a promising technique in the highly imbalanced data-sets framework.

Acknowledgements

Supported by Spanish Projects TIN-2005-08386-C05-01 & TIC-2005-08386-C05-03.

References

1. Barandela, R., Sánchez, J.S., García, V., Rangel, E.: Strategies for learning in class imbalance problems. Pattern Recognition 36(3), 849–851 (2003)
2. Batista, G.E.A.P.A., Prati, R.C., Monard, M.C.: A study of the behaviour of several methods for balancing machine learning training data. SIGKDD Explorations 6(1), 20–29 (2004)
3. Chawla, N.V., Japkowicz, N., Kolcz, A.: Editorial: special issue on learning from imbalanced data sets. SIGKDD Explorations 6(1), 1–6 (2004)
4. Chi, Z., Yan, H., Pham, T.: Fuzzy algorithms with applications to image processing and pattern recognition. World Scientific, Singapore (1996)
5. Cordón, O., del Jesús, M.J., Herrera, F.: A proposal on reasoning methods in fuzzy rule-based classification systems. International Journal of Approximate Reasoning 20(1), 21–45 (1999)
6. del Jesús, M.J., Fernández, A., García, S., Herrera, F.: A First Study on the Use of Fuzzy Rule Based Classification Systems for Problems with Imbalanced Data Sets. FSCS06, Magdeburg , Germany, pp. 63–72 (2006)
7. Ishibuchi, H., Nakashima, T.: Effect of Rule Weights in Fuzzy Rule-Based Classification Systems. IEEE Transactions on Fuzzy Systems 9(4), 506–515 (2001)
8. Ishibuchi, H., Nakashima, T., Nii, M.: Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining. Springer, Heidelberg (2004)
9. Ishibuchi, H., Yamamoto, T.: Rule Weight Specification in Fuzzy Rule-Based Classification Systems. IEEE Trans. on Fuzzy Systems 13, 428–435 (2005)

10. Mansoori, E.G., Zolghadri, M.J., Katebi, S.D.: A Weigthing Function for Improving Fuzzy Classification Systems Performance. Fuzzy Sets and Systems 158(5), 583–591 (2007)
11. Wang, L.X., Mendel, J.M.: Generating fuzzy rules by learning from examples. IEEE Transactions on Systems, Man, and Cybernetics 25(2), 353–361 (1992)