A Study on the Use of the Fuzzy Reasoning Method Based on the Winning Rule vs. Voting Procedure for Classification with Imbalanced Data Sets*

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

[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 Fuzzy Reasoning Methods for Fuzzy Rule Based Classification Systems in the framework of balanced and imbalanced data-sets with different degrees of imbalance. We analyze the behaviour of the Fuzzy Rule Based Classification Systems searching for the best type of Fuzzy Reasoning Method in each case, also studying the cooperation of some pre-processing methods of instances for imbalanced data-sets. To do so we use a fuzzy rule learning method that extends the well-known Wang and Mendel algorithm to classification problems.

The results obtained show the necessity to apply an instance pre-processing step and the differences for the most appropriate Fuzzy Reasoning Method in balanced and imbalanced data-sets, concluding that the choice of the Fuzzy Reasoning Method depends on the degree of imbalance, being the most adequate the use of the Winning Rule for high imbalanced data-sets and the Additive Combination method for the remaining data-sets.

Keywords: Fuzzy Rule Based Classification Systems, Instance Selection, Over-sampling, Imbalanced Data-sets, 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

* Supported by Spanish Projects TIN-2005-08386-C05-01 & TIC-2005-08386-C05-03.
** Corresponding author. Tel:+34-958-240598; Fax: +34-958-243317.

F. Sandoval et al. (Eds.): IWANN 2007, LNCS 4507, pp. 375–382, 2007.

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 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.

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

Our aim is to locate the best Fuzzy Reasoning Method (FRM) based on the class choice mechanism for each type of data-set, specifically we compare the Winning Rule mechanism versus a voting procedure based on Additive Combination. We use an FRBCSs constituted by a Rule Base generated by the technique that extends the Wang and Mendel algorithm [10] to fuzzy classification rules [4]. We employ triangular membership functions with five labels per variable.

In order to do that, this contribution is organized as follows. In Section 2 we introduce the FRBCS and the inductive learning algorithm used. Then in Section 3 we present the general model of FRM used in this work. In Section 4 we propose some preprocessing techniques for imbalanced data-sets. Section 5 introduces our experimentation framework while Section 6 shows the experimental study carried out with twenty-eight different data-sets. Finally, in Section 7 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 [10] 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.

Rule weights are used in FRBCSs in order to improve their performance [7] and different heuristic methods have been employed for rule weight specification [9]. In this contribution we will use the Certainty Factor defined in [5] as:

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

3 Fuzzy Reasoning Methods

An FRM is an inference procedure that derives conclusions from a set of fuzzy if-then rules and a pattern.

We will follow the general model of fuzzy reasoning presented in [5] for classification. Considering a new pattern $x_p = (x_{p1}, \ldots, x_{pn})$ the steps of the reasoning model are the following:

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 product or minimum T-norm.*

$$\mu_{A_j}(x_p) = T(\mu_{A_{j1}}(x_{p1}), \ldots, \mu_{A_{jn}}(x_{pn})), \qquad j = 1, \ldots, L. \tag{3}$$

2. *Association degree.* To compute the *association degree of the pattern* x_p *with the M classes according to each rule in the RB.* Function h is usually modeled as a T-norm. When using rules with the form of (1) this association degree only refers to the consequent class of the rule (i.e. $k = Class(R_j)$).

$$b_j^k = h(\mu_{A_j}(x_p), CF_j^k), \qquad k = 1, \ldots, M, \quad j = 1, \ldots, L. \tag{4}$$

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.

$$Y_k = f(b_j^k,\ j = 1, \ldots, L \text{ and } b_j^k > 0), \qquad k = 1, \ldots, M. \tag{5}$$

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.

$$F(Y_1, \ldots, Y_M) = l \qquad \text{such that} \qquad Y_l = \{max(Y_k), k = 1, \ldots, M\}. \tag{6}$$

At step 3 we may use different aggregation operators. In this contribution we study the performance of two FRMs for classifying new patterns by the rule set: the Winning Rule method (classic approach) and the Additive Combination method (voting approach).

1. *Winning Rule*: every new pattern is classified as the consequent class of a single winner rule which is determined as:

$$\mu_{A_w}(x_p) \cdot CF_w = max\{\mu_{A_j}(x_p) \cdot CF_j | j = 1, \ldots, L.\} \quad (7)$$

2. *Additive Combination*: each fuzzy rule casts a vote for its consequent class. The total strength of the vote for each class is computed as follows:

$$V_{Class_k}(x_p) = \sum_{j=1;\, Class_j=k}^{L} \mu_{A_j}(x_p) \cdot CF_j, k = 1, 2, \ldots, M. \quad (8)$$

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

4 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 preprocessing methods 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.
  - **"Smote 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.
- 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 [6] when using FRBCSs.

5 Experimental Framework: Data-Sets and Parameters

In this study we have considered twenty-eight data sets from UCI with different degrees of imbalance: from balanced data-sets to highly imbalanced data-sets. Tables 1 and 2 summarize the data employed in this study and shows, for each data set the number of examples (#Examples), number of attributes (#Atts.), number of classes (#Classes) in the case of balanced data-sets and class name of each class (majority and minority) and class attribute distribution in the case of imbalanced data-sets. Table 2 is ordered by the degree of imbalance, from low to high imbalance degree.

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.

Table 1. Data Sets summary descriptions for balanced datasets

Data set	**#Examples**	**#Atts.**	**#Classes**
Ecoli	336	7	8
Glass	214	9	7
Iris	150	4	3
New-Thyroid	215	5	3
Vehicle	846	18	4
Segment	2308	19	7
Yeast	1484	8	10

We consider the following parameters and functions for the Chi et al. algorithm[4]:

- 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.

For balanced data-sets we use the accuracy measure as performance metric. For the imbalanced data-sets a properly evaluation measure must be used. We employ the geometric mean metric (9), suggested in [1] where acc^+ is the accuracy classification on the positive instances, and acc^- the accuracy on the negative ones.

$$GM = \sqrt{acc^+ \cdot acc^-} \tag{9}$$

6 Results and Analysis

Our study is oriented to compare two different FRMs for FRBCSs in the framework of balanced and imbalanced data-sets. In this section we first present the

Table 2. Data Sets summary descriptions for imbalanced datasets

Data set	#Examples	#Atts.	Class (min., maj.)	%Class(min.,maj.)
Low Imbalanced Datasets (between 25 and 40%)				
EcoliCP-IM	220	7	(im, cp)	(35.00, 65.00)
Wisconsin	683	9	(malignant, benign)	(35.00, 65.00)
Pima	768	8	(tested-positive, tested-negative)	(34.84, 66.16)
Iris1	150	4	(Iris-Setosa, remainder)	(33.33, 66.67)
Yeast2	1484	8	(NUC, remainder)	(28.91, 71.09)
Vehicle3	846	18	(bus,remainder)	(28.37, 71.63)
Haberman	306	3	(Die, Survive)	(27.42, 73.58)
Medium Imbalanced Datasets (between 10 and 25%)				
GlassNW	214	9	(non-window glass, remainder)	(23.83, 76.17)
Vehicle1	846	18	(van,remainder)	(23.64, 76.36)
New-thyroid3	215	5	(hypo,remainder)	(16.89, 83.11)
New-thyroid2	215	5	(hyper,remainder)	(16.28, 83.72)
Segment1	2308	19	(brickface, remainder)	(14.26, 85.74)
Ecoli2	336	7	(iMU, remainder)	(10.88, 89.12)
Page-blocks	5472	10	(remainder, text)	(10.23, 89.77)
High Imbalanced Datasets (lower than 10%)				
Vowel0	988	13	(hid, remainder)	(9.01, 90.99)
Glass	214	9	(Ve-win-float-proc, remainder)	(8.78, 91.22)
EcoliMO	336	7	(MO, remainder)	(6.74, 93.26)
Abalone9-18	731	8	(18, 9)	(5.65, 94.25)
YeastCYT-POX	482	8	(POX,CYT)	(4.15, 95.85)
Yeast5	1484	8	(ME5, remainder)	(3.40, 96.60)
Abalone19	4174	8	(19, remainder)	(0.77, 99.23)

average results for the FRBCSs obtained by the Chi et al. method for each T-norm and FRM configuration. Then we analyze the best FRM found in each case and finally we demonstrate the necessity to apply a preprocessing step to transform the imbalanced data into a more balanced set.

Tables 3 to 6 show a comparative of the average results obtained with the FR-BCS method (Chi et al.) for each type of FRM in the balanced and imbalanced data-sets. In Table 3 we show the average accuracy for training and test for each FRM and T-norm for balanced data-sets. For the result tables of imbalanced data-sets (Tables 4 to 6), the following information is showed by columns:

– In the first column "FRM" we distinguish between each type of FRM, where WR stands for the Winning Rule method (classic approach) and AC stands for the Additive Combination method (voting procedure approach).
– Inside column "T-norm" we note if the results correspond to minimum or product T-norm.
– 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).

We focus our analysis on the generalization capacity via the test partition. In **bold** the best result for test are stressed and in underline we may observe the best results in columns, that is, for the different preprocessing method applied.

In the balanced data-sets the best accuracy is achieved when using the Additive Combination FRM together with the product T-norm but without significant differences with the FRM of the Winning Rule.

Table 3. Global comparison for the FRMs in balanced datasets

	Minimum T-norm		Product T-norm	
FRM	Acc_{Tr}	Acc_{Tst}	Acc_{Tr}	Acc_{Tst}
Winning Rule	85.98	<u>76.9</u>	87.25	76.62
Additive Combination	83.61	76.84	86.89	**77.48**

Table 4. Global comparison for the FRMs in low imbalanced datasets

		None		RandomOS		SMOTE		SMOTE-TL		SMOTE-ENN	
FRM	T-norm	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}
WR	Minimum	82.14	63.69	88.43	71.8	88.23	73.06	84.91	71.28	85.46	71.98
WR	Product	84.17	<u>66.37</u>	89.08	71.79	88.87	74.05	86.72	72.79	86.61	73.04
AC	Minimum	77.44	60.87	87.75	71.74	87.57	74.26	84.29	71.65	85.4	72.58
AC	Product	82.57	65.44	89.0	<u>72.8</u>	89.0	**74.73**	86.61	<u>73.08</u>	86.9	<u>73.58</u>

Table 5. Global comparison for the FRMs in medium imbalanced datasets

		None		RandomOS		SMOTE		SMOTE-TL		SMOTE-ENN	
FRM	T-norm	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}
WR	Minimum	87.96	78.05	95.07	87.87	95.2	89.16	94.81	89.17	94.75	88.96
WR	Product	90.32	<u>83.01</u>	96.25	88.97	96.01	89.73	95.34	89.29	95.19	89.16
AC	Minimum	84.43	76.36	93.95	87.24	93.81	88.39	93.68	88.34	93.91	88.48
AC	Product	89.07	81.65	96.78	<u>90.17</u>	96.64	<u>90.7</u>	96.07	**91.04**	95.94	<u>90.63</u>

Table 6. Global comparison for the FRMs in high imbalanced datasets

		None		RandomOS		SMOTE		SMOTE-TL		SMOTE-ENN	
FRM	T-norm	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}	GM_{Tr}	GM_{Tst}
WR	Minimum	51.03	40.21	84.25	<u>73.59</u>	83.52	74.4	83.06	**75.17**	83.01	75.05
WR	Product	53.96	<u>40.51</u>	85.74	73.1	84.56	<u>74.68</u>	84.17	75.07	84.09	**75.17**
AC	Minimum	46.39	36.8	74.54	60.99	81.4	66.31	81.16	68.18	81.44	67.95
AC	Product	51.25	40.06	81.01	64.84	83.77	70.72	83.43	70.93	83.76	70.22

The same occurs in the case of low and medium imbalanced data-bases in which the Additive Combination is the best choice for the FRM but it is still too similar to the Winning Rule one. For this type of data-sets we can see a good cooperation between the product T-norm and the FRM of Additive Combination, providing the best results independently of the preprocessing mechanism.

In the case of highly imbalanced data-sets the FRM of the Winning Rule outperforms the Additive Combination. We may observe clear differences between the two FRM regardless the preprocessing method used. It seems that when there are very few instances of the positive class it is more accurate to select the best rule in order to classify new examples. Also in this case there are no significant differences when using the minimum or product T-norm, although the latter achieves higher classification results.

Regarding the imbalanced data-sets, we can see in Tables 4 to 6 that there is a great difference in the results when we apply a preprocessing mechanism to balance the data comparing with the results without preprocessing (first column), which confirms that the preprocessing methods are useful and the necessity to transform the data-sets into a more balanced format.

7 Concluding Remarks

In this work we have analyzed the performance of the FRBCSs searching for the best FRM between the Winning Rule and voting procedure based on the Additive Combination focused in the framework of imbalanced data-sets. We have also studied the cooperation of some pre-processing methods of instances.

We have analyzed the differences for the most appropriate FRM in balanced and imbalanced data-sets, concluding that the choice of the FRM depends on the degree of imbalance, being the most adequate the use of the Winning Rule for the high imbalanced data-sets and the Additive Combination method for the remaining data-sets.

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. Symposium on Fuzzy Systems in Computer Science (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, Berlin (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. 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)