Journal of Advances in Computer Research
Quarterly ISSN: 2008-6148
Sari Branch, Islamic Azad University, Sari, I.R.Iran
(Vol. 3, No. 2, May 2012), Pages: 1-9
www.jacr.iausari.ac.ir

On Mining Fuzzy Classification Rules for Imbalanced Data

Mohsen Rahmanian[1]✉, Eghbal Mansoori[2], Mehdi Zareian Jahromi[3]

(1) Department of Computer Engineering, Jahrom University, Jahrom, Iran
(2) Department of Computer Science and Eng., School of Engineering, Shiraz University, Shiraz, Iran
(3) Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran

rahmanian@jahrom.ac.ir; mansoori@shirazu.ac.ir; m.jahromi@aut.ac.ir

Received: 2012/4/18; Accepted: 2012/5/8

Abstract

Fuzzy rule-based classification system (FRBCS) is a popular machine learning technique for classification purposes. One of the major issues when applying it on imbalanced data sets is its biased to the majority class, such that, it performs poorly in respect to the minority class. However many cases the minority classes are more important than the majority ones. In this paper, we have extended the basic FRBCS in order to decrease the side effects of imbalanced data by employing data-mining criteria such as confidence *and* support. *These measures are computed from information derived from data in the sub-spaces of each fuzzy rule. The experimental results show that the proposed method can improve the classification accuracy when applied on benchmark data sets.*

Keywords: *Imbalanced data-sets, Fuzzy rule based classification systems, Data-mining*

1. Introduction

In many real-world applications, the available data sets are imbalanced. The problem of learning from imbalanced data arises when one class (the majority class) contains many more samples than the other class (the minority class). Recently, machine learning community acknowledged that the current learning methods (e.g. SVM, C4.5, NN, k-NN, FRBCS) perform poorly in applications dealing with imbalanced data sets [1]. Traditionally, most classification algorithms do not make any special allowance concerning the class imbalance, assuming that in effect the training data is roughly balanced. By the way, in almost all cases of imbalanced data, the minority class is the class of interest and the classification accuracy on this class is more important, or, as it is usually stated in machine learning literature, it has a higher penalty errors [2][3]. Some application domains dealing with highly imbalanced data sets include [6][7]. helicopter gear-box fault monitoring, shuttle system failure, earthquakes and nuclear explosions, diagnoses of rare disease and rare genes mutations, text classification, oil spill detection, detecting computer security intrusions, all kinds of fraud detection (credit card, phone calls, insurance, etc.).

Several solutions have been introduced to tackle the problem of imbalanced data. We can divide these solutions into two main groups: First, the inner solutions which try to build a new algorithm, or make changes to the existing algorithms [4][6]. Second, the outer solutions that seek to re-sample the original data, either by over-sampling the minority class and/or under-sampling the majority class until the classes are approximately equally represented [17][19].

The main problem with the first solutions is that they are limited to one special approach, while the second solutions are independent of any classifier. On the other hand, making changes to the data is not always practical or is expensive.

A fuzzy rule-based classification system is a special case of fuzzy modelling, in which the output of the system is crisp and discrete (class label). The fuzzy rule-based classification system presents two main advantages: first, they permit to work with imprecise data, and also, provide a comfortable way to naturally represent the missing values, and second, the acquired knowledge with these models may be more human understandable[8]. Stressing the second preference, in this paper we have presented a new method for generating fuzzy classification rules from an imbalanced data set.

In the literature of imbalanced data classification, a few fuzzy algorithms for dealing with such data (without preprocessing steps) have been introduced [7]. The *E-algorithm* proposed by Le Xu *et al* [6]. has achieved satisfactory performance on some well-known data sets by extending the *support* and *confidence* measures used in Ishibuchi *et al's* algorithm (*I*-algorithm)[5]. It uses a coefficient for the weight measure that depends on the number of samples for each class rather than all data. Equations (1) and (2) show this values.

$$Conf'(A_j \Rightarrow Class\ T) = \frac{N}{N_T} Conf(A_j \Rightarrow Class\ T) \tag{1}$$

$$Supp'(A_j \Rightarrow Class\ T) = \frac{N^2}{N_T} Supp(A_j \Rightarrow Class\ T) \tag{2}$$

When the data set is not balanced, a larger coefficient is used for the smaller class. Thus, both extended measures for the minority classes are proportionally adjusted through the normalization to decrease the bias due to the imbalanced data. But, carefully in the formula provided by the authors (Eq. 1 and 2) , we noticed that the coefficients used for the support and confidence don't influence in choose the best condidate rules for each class, because $\frac{N}{N_T}$ and $\frac{N^2}{N_T}$ values for all rules of a class are equal.

In this paper, an effective modification and extension of the *I*-algorithm (*MI*-algorithm) is proposed to decrease the influence of the unbalancing in data. To show its effectiveness, we have applied *MI*-algorithm to some well-known data sets. The results shows an improvement over the *I*-algorithm and *E*-algorithm on imbalance data sets.

The rest of this paper is organized as follows. In section 2, the structure of basic fuzzy rule-based classification systems is explained. Our proposed method is discussed in section 3. In section 4, we present experimental results. Finally, Section 5 concludes the paper and expressed the limitations of our method.

2. General design of fuzzy rule-based classification system

Fuzzy if-then rules for a pattern classification problem with *n* attributes can be written as:

$$\textit{Rule } R_j\textit{: If } x_1 \textit{ is } A_{j1} \textit{ and ... and } x_n \textit{ is } A_{jn} \textit{ then class } C_j \textit{ with } CF_j,\ \textit{for } j=1,...,N \quad (3)$$

where X=[x_1 , x_2 , ..., x_n] is an n-dimensional pattern vector, A_{ji} (i=1,...,n) is an antecedent linguistic value of R_j , C_j is the consequent class, CF_j is the rule weight, and N is the number of fuzzy rules. Generally, for an M-class problem with m labeled patterns X_p=[x_{p1} , x_{p2} , ..., x_{pn}], p=1,... ,m , the task of designing the classifier is to generate a set of N fuzzy rules in the form (3).

For this purpose, first each attribute is rescaled to unit interval [0, 1] using a linear transformation that preserving the distribution of the training patterns. Then, the pattern space is partitioned into fuzzy subspaces each of which is identified by a fuzzy rule provided that patterns exist in that subspace. To do partitioning, usually K suitable membership functions are used to assign K linguistic values to each input attribute. Traditionally, triangular membership functions are used for this purpose, as they are less complex and more human understandable.

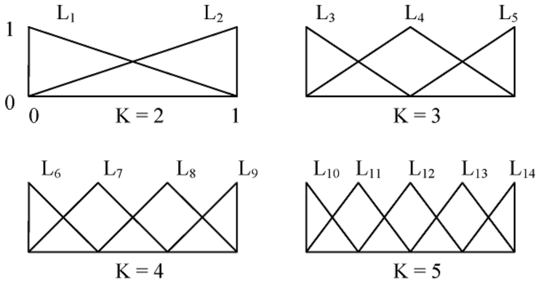

Figure 1. Different partitioning of each input attribute.

Figure 1 shows these membership functions for four different values of K, where the linguistic labels L_3 , L_4 and L_5 can be for example interpreted as the linguistic values *small, medium* and *large*, respectively.

Given an input partitioning of the pattern space, there are several approaches to generate fuzzy classification rules from the data [5][8]. The approach used in this paper, is the method proposed in [5], that is named as I-algorithm.

Fuzzy rules in I-algorithm are in the form of (3). The consequent class C_j of fuzzy rule R_j in (3) is determined by the training patterns lying in the corresponding fuzzy subspace. The compatibility grade of training pattern X_p is defined with the antecedent part $\mathbf{A}_j = A_{j1} \times A_{j2} \times \ ... \ \times A_{jn}$ of the rule R_j, using the product operator as:

$$\mu_j(X_p) = \prod_{i=1}^{n} \mu_{ji}(X_p) \quad (4)$$

where $\mu_{ji}(.)$ is the membership function of the antecedent fuzzy set A_{ji}. To select the consequent class of a rule, we have used the heuristic method employed in [5], which is based on the *confidence*. The confidence for rule $\mathbf{A}_j \Rightarrow$ *Class T* is defined as:

$$Conf(A_j \Rightarrow Class\ T) = \frac{\sum_{X_p \in Class T} \mu_j(X_p)}{\sum_{p=1}^{m} \mu_j(X_p)} \quad (5)$$

The confidence can be viewed as a measuring of the validity of $\mathbf{A}_j \Rightarrow Class\ T$. It can be also viewed as a numerical approximation of the conditional probability of *Class T* given $\mathbf{A}_j$ [14]. On the other hand, the support of $\mathbf{A}_j \Rightarrow Class\ T$ is written as follows:

$$Supp(A_j \Rightarrow Class\ T) = \frac{\sum_{X_p \in Class T} \mu_j(X_p)}{m} \quad (6)$$

where m is the number of given training patterns.

Using (5), the consequent class C_j of a fuzzy rule R_j is specified as the class with the maximum confidence. That is, the consequent class C_j is chosen as:

$$C_j = \arg\max_T \{Conf(A_j \Rightarrow Class\ T)\} \quad (7)$$

When the consequent C_j cannot be uniquely determined in (7), we do not generate any fuzzy rule with the antecedent $\mathbf{A}_j$.

For evaluating the condidate rules, befor selecting the best ones, some heuristic measures have been used in [5]. Their basic criterion, which is a fuzzy version of the difference between the number of true positives and false positives, is specified as:

$$f1(A_j \Rightarrow Class\ C_j) = \sum_{X_p \in Class\ C_j} \mu_j(X_p) - \sum_{X_P \notin Class\ C_j} \mu_j(X_P) \quad (8)$$

In this paper we use the product of the confidence and support as a criterion for selecting the best rules.

$$f2(A_j \Rightarrow Class\ C_j) = Supp(A_j \Rightarrow Class\ C_j) \times Conf(A_j \Rightarrow Class\ C_j) \quad (9)$$

The most popular reasoning method, in fuzzy rule-based classifiers, is the single winner rule [9]. This method is simple and understandable for human users. Using this method, a new pattern $\mathbf{X}_t = [x_{t1}, x_{t2}, \ldots, x_{tn}]$ is classified according to the consequent class of the winner rule R_w. Indeed, the winner rule has the maximum product value of compatibility grade with X_t and the rule weight. This can be stated as:

$$\mu_w(X_t) = \max_j \{\mu_j(X_t) \times CF_j\} \quad (10)$$

where $\mu_j(X_t)$ is the compatibility grade of rule R_j with pattern X_t using (4).

Since the main objective, in fuzzy rule-based classifier, is intrepretability of the system, the classification accuracy is not too high. However, it is possible to adjust the membership functions or to use the weighted fuzzy rules to achieve a higher accuracy. While modifying the membership functions of fuzzy sets will degrade the interpretability of the rules, assigning weights to the rules [5][10], or finding some suitable weighting functions [8] can increases the accuracy of the classifier, yet preserves the comprehensibility of the fuzzy rules.

In order to assign a weight to each fuzzy classification rule, some heuristic measures proposed in the past researches [5][12][13]. One of the more effective method [5] can be stated as:

$$CF_j = Conf(\mathbf{A}_j \Rightarrow Class\ C_j) - Conf_{\text{Sum}} \tag{11}$$

where, C_{Sum} is the sum of confidence over fuzzy association rules whose antecedent are $\mathbf{A}_j$ and whose consequents are not equal to T .

$$Conf_{\text{Sum}} = \sum_{\substack{T=1 \\ T \neq C_j}}^{M} Conf(\mathbf{A}_j \Rightarrow Class\ T) \tag{12}$$

Note that with the definition of rule weight (11), the weight of a rule can be negative. Possible negative weigths obtained by (11) is then set to zero.

3. Generating fuzzy classification rules for imbalanced data

When the data to be classified are imbalanced, the confidence and support can be significantly biased, because the sum of compatibility grades of the major class can be much larger than that of the minority class, even when the compatibility grades of the majority class are small. As an example, consider an imbalanced two-class data set, where the major class, C_-, has 98% of the total samples. Assume that each data sample of this class has a small compatibility grade of 0.02 with the antecedent part of the rule R_j, while each data sample of class C_+ (the minor class) has a large compatibility grade of 0.8. According to the definition of *Confidence* in (5) we obtain:

$$Conf(\mathbf{A}_j \Rightarrow C_-) = \frac{\sum_{X_p \in C_-} \mu_j(\mathbf{X}_p)}{\sum_{p=1}^{m} \mu_j(\mathbf{X}_p)} = \frac{0.02 \times 98\%}{3.56} = \frac{1.96}{3.56} = 0.55 \tag{13}$$

$$Conf(\mathbf{A}_j \Rightarrow C_+) = \frac{\sum_{X_p \in C_+} \mu_j(\mathbf{X}_p)}{\sum_{p=1}^{m} \mu_j(\mathbf{X}_p)} = \frac{0.8 \times 2\%}{3.56} = \frac{1.6}{3.56} = 0.45 \tag{14}$$

So, though the compatibility grade of samples in second class are large, but the value of confidence for class C_+ is smaller than for class C_-. This bias in confidence will affect the weight of rule as shown in (11), and consequently affect the single winner rule method whose selection criterion is the product of the compatibility grade and rule weight. So the rules of minority class will never win in this competition.

In order to decrease the negative effect of imbalanced data, in this paper we propose an effective coefficients for confidence and support.

***Creating an effective coefficients for confidence and support*:** In original equation of confidence (5) and support (6), we don't attention to number of samples of each class in the covering space of rule R_j than total samples of class. Although the number of minor class in the covering space R_j is low, but this number is significant regard to the total number of the minor class. This ratio can help the minor class that become stronger. For this purpose we introduce a coefficient for confidence and support as:

$$Conf'(\mathbf{A}_j \Rightarrow Class\ T) = \frac{N_{T_j}}{N_T} \times Conf(\mathbf{A}_j \Rightarrow Class\ T) \tag{15}$$

$$Supp'(\mathbf{A}_j \Rightarrow Class\ T) = \frac{N_{T_j}}{N_T} \times Supp(\mathbf{A}_j \Rightarrow Class\ T) \tag{16}$$

where N_T is the number of data belonging to class T whose N_{T_j} of them are in the subspace of the rule R_j. When the data set is not balanced, the coefficient of the minority class is larger than the coefficient of the majority class. Thus, with this coefficient, the minority classes are proportionally adjusted through methods illustrated in (15) and (16) to alleviate the bias due to the imbalanced data constitution.

4. Experimental results

In this section we present experimental results. It is organized as follows. In subsection 4.1 we introduced an evaluation method for imbalanced data. Our experimental results is discussed in sub-section 4.2.

4.1 Evaluation in imbalanced domains

A straightforward way to measure the accuracy of a classifier is to examine how many of the data are correctly classified, denoted by the correct classification rate (True Positive plus True Negative). However, it is usually insufficient to draw a conclusion of the performance of a classifier by simply observing classification rate, especially when the data is imbalanced. Thus, the geometric mean (g-mean) [15] often used as a measure to evaluate the performance of the classifier when the data is imbalanced.

The g-mean measure is derived from the confusion matrix shown in Table 1. The true positive rate ($Acc^+ = \frac{TP}{TP+FN}$) indicates the percentage of the positive data being correctly classified; true negative rate ($Acc^- = \frac{TN}{TN+FP}$) indicates how many instances of the negative class are correctly classified. The g-mean is calculated as:

$$g-mean = \sqrt{Acc^+ \times Acc^-} \tag{17}$$

The g-mean is large when both Acc^+ and Acc^- are large and the difference between them is small, i.e., the classification accuracies on both positive and negative classes are high and there is no wide disparity between them, representing a balanced performance [6].

Table 1. The confusion matrix

	Predicted Positive Class	Predicted Negative Class
Actual Positive Class	True Positive (TP)	False Negative(FN)
Actual Negative Class	False Positive(FP)	True Negative (TN)

4.2 Results

In this paper, we have considered some data-sets from UCI with different values for the *Imbalance Ratio* (IR), defined as the ratio of the number of instances of the majority class to the minority class [16]. We consider the first class as the positive (minority) class and the other class as negative (majority) class. Table 2 summarizes the data employed in this study and shows, for each data-set, the number of examples (No. Exp), number of attributes (No. Attrs.), name of each class, and the IR.

Table 2. Data sets descriptions

Data set	No. Exp.	No. Attrs.	Class (min.,maj.)	IR
Low IR:				
Wine2	178	13	(2, 1+3)	1.51
Iris2	150	4	(Iris-versicolor, others)	2.00
Iris3	150	4	(Iris-virginica, others)	2.00
Wine1	178	13	(1, 2+3)	2.02
Glass1	214	9	(build-wind-float, others)	3.06
Medium IR:				
Ecoli2	336	7	(im, others)	4.36
Ecoli3	336	7	(pp, others)	6.46
Glass7	214	9	(headlamps, others)	7.38
High IR:				
Glass3	214	9	(vehic-wind-float, others)	12.59
Glass5	214	9	(containers, others)	16.46
Ecoli5	336	7	(om, others)	16.80
Glass6	214	9	(tableware, others)	23.78

Table 3. Comparing the performance of the I-Algorithm, E-Algorithm and MI-Algorithm for low imbalance data-sets

Data set	I-algorithm	E-algorithm	MI-algorithm
Wine2	90.63	84.40	87.26
Iris2	93.47	70.71	80.62
Iris3	94.47	84.26	86.22
Wine1	94.38	90.28	90.75
Glass1	0.00	64.15	64.55
Avg.	74.59	78.76	81.85

Table 4. Comparing the performance of the I-Algorithm, E-Algorithm and MI-Algorithm for median imbalance data-sets

Data set	I-algorithm	E-algorithm	MI-algorithm
Ecoli2	0.00	88.30	86.84
Ecoli3	0.00	84.75	87.21
Glass7	74.28	84.87	87.69
Avg.	24.76	85.97	87.25

Table 5. Comparing the performance of the I-Algorithm, E-Algorithm and MI-Algorithm for high imbalance data-sets

Data set	I-algorithm	E-algorithm	MI-algorithm
Glass3	0.00	23.69	39.82
Glass5	0.00	89.39	85.99
Ecoli5	0.00	91.97	93.20
Glass6	0.00	0.0	54.70
Avg.	0.00	51.26	68.43

In order to develop our study we use a leave-one-out approach (LV1), that is, one instance for testing and $N-1$ instances for training. For each data-set we consider the average results of the N partitions. The g-means achieved by each method in this study are shown in Tables 3, 4 and 5, each one shows the result for data-sets with low, median and high imbalance, respectively. These tables show the results for FRBCSs obtained by *I-Algorithm*, *E-Algorithm* and the methods that study in this paper (*MI-Algorithm*). *I-Algorithm* and *E-Algorithm* use product T-norm, CF (11) for the rule weight and single winner method for choosing the winning rule, but in *MI-Algorithm* the rule weight is:

$$CF_j = Conf'(A_j \Rightarrow ClassT) - Conf'_{\text{sum}} \tag{18}$$

$$Conf'_{\text{sum}} = \sum_{\substack{h=1 \\ h\neq T}}^{M} Conf'(\mathbf{A}_j \Rightarrow Class\ h) \tag{19}$$

As you see, in Table 3 almost all of the result of the *MI-Algorithm* is better than *E-Algorithm*. Tables 3 and 4 clearly indicate that the *MI-Algorithm* has a significant dominance of g-means.

5. Conclusion

The fuzzy rule-based classification system proposed by Ishibuchi et al. is based on support and confidence. This algorithm has good performance on balanced data sets, but fail in the problems with imbalanced data. In this paper, we examined the performance of *MI-Algorithm* in extracting fuzzy IF-THEN rules from numerical data for classification of imbalanced data. We demonstrated that, unlike standard FRBCS, the calculation of confidence and support in *MI-Algorithm* is not blindly, and uses the ratio of the number of members of the minority and the majority classes took into consideration. This episode more equitable outcome is determined by rule. The only limitation of this method is when the imbalanced ratio (IR) is low. As we have shown in section 4.2, this makes the classification results even of the standard classifier become worse.

6. References

[1] Sofa Visa, Anca Ralescu, "Issues in Mining Imbalanced Data Sets - A Review Paper", *in Proceedings of the Sixteen Midwest Artificial Intelligence and Cognitive Science Conference,* pp. 67-73(2005).

[2] Sofia Visa, "Fuzzy Classifiers for Imbalanced Data Sets", *PhD thesis, Department of Electrical and Computer Engineering and Computer Science of the College of Engineering*, (2006).

[3] N. Chawla, N. Japkowicz, A. Kolcz, "Special issue on learning from imbalanced data sets", *SIGKDD Explorations 6(1)*, pp. 1-6(2004).

[4] R. Barandelaa, J. S. Sanchezb, V. Garcia, E. Rangel, "Strategies for learning in class imbalance problems", *Pattern Recognition Society,* pp. 849-851(2003).

[5] H. Ishibuchi, T. Yamamoto, "Rule Weight Specification in Fuzzy Rule-Based Classification Systems", *IEEE Transaction on fuzzy systems*, vol. 13, no. 4, (2005).

[6] L. Xu, M. Chow, L. S. Taylor, "Data Mining Based Fuzzy Classification Algorithm for Imbalanced Data", *IEEE: International Conference on Fuzzy Systems,*, (2006).

[7] Alberto Fernandeza, Salvador Garcaa, Mara Jos del Jesusb, Francisco Herreraa, "A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets", *Fuzzy Sets and Systems*, vol. 159, no. 18, pp. 2378-2398(2008).

[8] E. G. Mansoori, M. J. Zolghadri, and S. D. Katebi, "A Weighting Function for Improving Fuzzy Classification System Performance", *Fuzzy Sets and Systems*, vol. 158, no. 5, pp. 583-591(2007).

[9] H. Ishibuchi, T. Yamamoto, and T. Morisawa, "Voting in fuzzy rule-based systems for pattern classification problems", *Fuzzy Sets and Systems*, vol. 103, no. 2, pp. 223-238(1999).

[10] M. J. Zolghadri, and E. G. Mansoori, "Weighting Fuzzy classification rules using Receiver Operating Characteristics (ROC) analysis", *Information Sciences*, vol. 177, no. 11, pp. 2296-2307(2007).

[11] O. Cordon, M. J. del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems", *Int. J. Approx. Reason.*, vol. 20, no. 1, pp. 21-45(1999).

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

[13] H. Ishibuchi and T. Nakashima, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining", *Fuzzy Sets and Systems*, pp. 59-88(2004).

[14] J. van den Berg, U. Kaymak, and W.-M. van den Bergh, "Fuzzy classification using probability based rule weighting", *in Proc. 11th IEEE Int. Conf. Fuzzy Systems*, Honolulu, HI, pp. 991-996(2002).

[15] M. Kubat, R. Holte, and S. Matwin, "Machine Learning for the Detection of Oil Spills in Radar Images", *Machine Learning*, vol.30, pp.195-215(1998).

[16] A. Orriols-Puig, E. Bernado-Mansilla, K. Sastry, D.E. Goldberg, "Substructural surrogates for learning decomposable classification problems: implementation and first results", *in: GECCO '07: Proceedings of the 2007 GECCO Conference Companion on Genetic and Evolutionary Computation,* ACM Press, New York, NY, USA, pp. 2875-2882(2007).

[17] N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer, "Smote: synthetic minority over-sampling technique", *Journal Artiï¬ cial Intelligent Research*, vol. 16, pp. 321â€“357(2002).

[18] P. Hart, "The condensed nearest neighbor rule", *IEEE Trans. Inform. Theory*, vol. 14, pp. 515-516(1968).

[19] T. Fawcett, F.J. Provost, "Adaptive fraud detection", *Data Mining Knowledge Discovery* vol. 1, no. 3, pp. 291-316(1997).