A Cost-Based Fuzzy System for Pattern Classification with Class Importance

T. Nakashima, Y. Yokota, H. Ishibuchi
Graduate School of Engineering
Osaka Prefecture University
Gakuen-cho 1-1, Sakai, Osaka 599-8531, Japan
{nakashi, hisaoi}@cs.osakafu-u.ac.jp
yokota@ci.cs.osakafu-u.ac.jp

G. Schaefer
School of Engineering and Applied Science
Aston University
Birmingham B4 7ET, U.K.
g.schaefer@aston.ac.uk

Abstract

This paper proposes a cost-based fuzzy classification system for pattern classification problems with an order class importance. The task here is to minimize the misclassification of patterns from an important class. It is assumed in this paper that the classification importance is given for each class, not for each pattern. Another assumtion in this paper is that only the order of importance is given for given classes without any numerical measures of importance. We show the performance of the proposed cost-based fuzzy classification system for a real-world pattern classification problem.

1 Introduction

Fuzzy rule-based systems have been mainly applied to control problems [1]-[3] while more recently they have also been applied to pattern classification problems. One advantage of fuzzy rule-based systems is its interpretability. There are many approaches to the automatic generation of fuzzy if-then rules from numerical data for pattern classification problems [4]-[11].

There are several cases where misclassification of a particular input pattern causes extra costs. For example, in the medical diagnosis of cancer diagnosing malignant tumors as benign and hence mistaking a cancer patient as healthy could be penalized more than interpreting benign tumors as malignant. In [12] a pattern classification problem is re-formulated as a cost minimization problem. The concept of a weight is introduced for each training pattern in order to handle this situation. The weight of an input pattern can be viewed as the cost of misclassification of the pattern. Fuzzy if-then rules were generated by considering the weights as well as the compatibility of training patterns.

In many cases it is difficult to specify an exact value of the cost for misclassification. Contrary to this, the order of importance is easily available for a problem domain. For example, in the medical diagnosis of cancer it is reasonable to consider that the misclassfication of malignant tumors as benign incurs higher cost than the other misclassification even if we do not know how high the cost is for both types of misclassification (i.e., misclassifying malignant tumors as benign and benign ones as malignant).

In this paper we propose a method for constructing a fuzzy classification system that considers the order of class importance. The assumption here is that a set of training patterns and the order of class importance is given a priori. We compare the performance of the proposed fuzzy classification systems and the conventional ones for a real-world pattern classification problem.

2 Classification Problem

Without loss of generality, we assume that given training patterns are distributed over an n-dimensional pattern space $[0,1]^n$. We also assume that m training patterns $\mathbf{x}_p = (x_{p1}, x_{p2}, \ldots, x_{pn})$, $p = 1, 2, \ldots, m$, from C classes are given a priori.

In many real-world problems, the misclassification cost is different for different patterns. It is often the case that misclassification cost is different class by class. As an example let us consider the medical diagnosis of cancer where the task is to diagnose a tumor as benign (not cancerous) or malignant (cancerous). Although any misclassification should be avoided, more attention should be put into the correct diagnosis of malignant than that of benign tumors because the misdiagnosis of the cancer patient as not having the disease leads to much higher costs than the other misclassification. In this paper, we handle this situation by specifying different misclassification costs for different classes and put more emphasis on minimizing the misclassification costs rather than minimizing the number of misclassified input patterns.

3 Fuzzy Rule-Based Classification

The proposed method is based on a fuzzy rule-based classification system that was proposed by Ishibuchi et al.[6]. A fuzzy rule-based classification system is composed of a set of fuzzy if-then rules. The antecedent part of a fuzzy if-then rule specifies a fuzzy subarea in the pattern space while the consequent part describes the class and the degree of certainty for the specified fuzzy subarea. A fuzzy if-then rule is automatically generated from numerical data that are given as a set of training patterns. This section explains the generation of fuzzy if-then rules from given training patterns.

The following type of fuzzy if-then rules is used in the fuzzy rule-based classification system for an n-dimensional C-class pattern classification problem:

$$\begin{aligned}\text{Rule } R_j\text{: If } x_1 \text{ is } A_{j1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{jn}\\ \text{then Class } C_j \text{ with } CF_j, \quad j = 1, 2, \ldots, N,\end{aligned} \tag{1}$$

where R_j is the label of the j-th fuzzy if-then rule, $A_{j1}, \ldots, A_{jn}$ are antecedent fuzzy sets on the unit interval $[0, 1]$, C_j is a consequent class (i.e. one of the C given classes), CF_j is the grade of certainty of rule R_j, and N is the total number of fuzzy rules. We use triangular-type membership functions in Fig. 1 for antecedent fuzzy sets $A_{j1}, \ldots, A_{jn}$. Figure 1 shows four fuzzy partitions. We denote the number of fuzzy sets in the unit interval as L. The total number of generated fuzzy if-then rules N depends on the dimensionality of a pattern classification problem n and the number of fuzzy partitions L. For a ten-dimensional pattern classification problem for instance, the total number of generated fuzzy if-then rules is $N = 2^{10} = 1024$ when the number of fuzzy sets for each attribute is specified as $L = 2$. While we use a homogeneous fuzzy partition as in Fig. 1, it is also possible to use a heterogeneous fuzzy partition that reflects the distribution of training patterns. In this paper only homogeneous fuzzy partition is used because of the interpretability of the fuzzy rule-based classification systems.

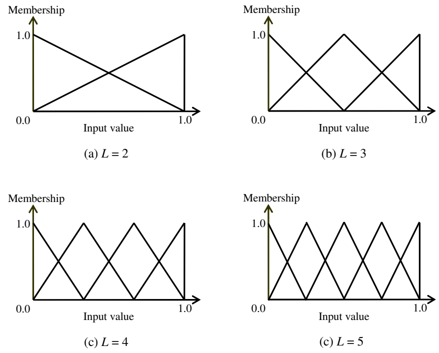

Figure 1: Triangular-type membership function used for antecedent fuzzy sets.

The generation procedure of fuzzy if-then rules consists of three steps: specification of the antecedent part, determination of the consequent class, and the calculation of the grade of certainty. Once the number of fuzzy sets L for each attribute is specified by an expert with the domain knowledge, the pattern space is divided into L^n fuzzy subareas where n is the dimensionality of a pattern classification problem at hand. Thus the specification of the antecedent part of a fuzzy if-then rule is already performed once a fuzzy subarea to generate a fuzzy if-then rule is fixed. The consequent part (i.e., the consequent class and the grade of certainty) is determined from the given training patterns [6]. In [15] it is shown that the use of the grade of certainty in fuzzy if-then rules allows us to generate comprehensible fuzzy rule-based classification systems with high classification performance. In the following we describe the remaining two steps in detail.

3.1 Determination of Consequent Class

The consequent class C_j for the fuzzy if-then rule R_j is determined from a set of the given training patterns $\mathbf{x}_p$, $p = 1, 2, \ldots, m$. In [13] first the sum of the compatibility is calculated for each class. Then the class with the largest value is taken as the consequent class of the fuzzy if-then rule. The procedure of determining the consequent class in the conventional fuzzy rule-based classification systems is summarized as follows:

[Conventional determination of C_j]

Step 1: Calculate $\beta_{\text{Class } h}(j)$ for Class h as

$$\beta_{\text{Class } h}(j) = \sum_{\mathbf{x}_p \in \text{Class } h} \mu_j(\mathbf{x}_p), \tag{2}$$

where

$$\mu_j(\mathbf{x}_p) = \mu_{j1}(x_{p1}) \cdot \ldots \cdot \mu_{jn}(x_{pn}), \quad (3)$$

$\mu_{j1}(\cdot),\ \mu_{j2}(\cdot), \ldots,\ \mu_j(\cdot)$ are the membership function of the fuzzy sets $A_{j1},\ A_{j2}, \ldots, A_{jn}$, respectively.

Step 2: Find Class $\hat{h}$ that has the maximum value of $\beta_{\text{Class } h}(j)$:

$$\beta_{\text{Class } \hat{h}}(j) = \max_{1 \le k \le C} \{\beta_{\text{Class } k}(j)\}. \quad (4)$$

Since the above procedure does not consider misclassification costs nor class importance, we propose a cost-based determination method. In this approach the consequent class of a fuzzy if-then rule is determined from its compatible training patterns. Let us define the number of compatible training patterns for Class k with the fuzzy if-then rule R_j as n_j^k. A pattern $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is compatible with the fuzzy if-then rule R_j if the following condition holds:

$$\mu_{j1}(x_1) \cdot \mu_{j2}(x_2) \cdot \ldots \cdot \mu_{jn}(x_n) > 0, \quad (5)$$

where $\mu_{j1}(\cdot), \mu_{j2}(\cdot), \ldots, \mu_{jn}(\cdot)$ are the membership functions of antecedent fuzzy sets $A_{j1}, A_{j2}, \ldots, A_{jn}$ of R_j. The consequent class of R_j is determined as the class with the maximum misclassification cost among the n_j^k training patterns. We use a multiplication operator to calculate the compatibility of an n-dimensional pattern with a fuzzy if-then rule in this paper.

The cost-based class determination procedure is performed as follows:

[Proposed determination of C_j]

Step 1: Calculate n_j^k for Class k, $k = 1, 2, \ldots, C$.

Step 2: Find Class $\hat{h}$ that is the most important (i.e., has the highest misclassification cost) among those classes with n_j^k as follows:

$$Cost_j^{\hat{h}} = \max_{\substack{1 \le k \le C \\ n_j^k > 0}} Cost_j^k. \quad (6)$$

3.2 Calculation of the Grade of Certainty

The procedure of calculating the grade of certainty in this paper is different from that in [13]. In [13] the grade of certainty CF_j of rule R_j is determined as follows:

$$CF_j = \frac{\beta_{\text{Class } \hat{h}}(j) - \bar{\beta}}{\sum_h \beta_{\text{Class } h}(j)}, \quad (7)$$

where

$$\bar{\beta} = \frac{\sum_{h \ne \hat{h}} \beta_{\text{Class } h}(j)}{C - 1}. \quad (8)$$

On the other hand, the proportion of $\beta_{\text{Class } \hat{h}}(j)$ over the sum of $\beta_{\text{Class } \hat{h}}(j),\ 1 \le h \le C$, is used in this paper. This is because the grade of certainty sometimes becomes an invalid value (e.g., negative) with the conventional determination.

3.3 Fuzzy Inference for Classification

Using the rule generation procedure outlined above we can generate N fuzzy if-then rules as in Equation (1). After both the consequent class C_j and the grade of certainty CF_j are determined for all N rules, a new pattern $\mathbf{x} = (x_1, \ldots, x_n)$ is classified by the following procedure:

Step 1: Calculate $\alpha_{\text{Class } h}(\mathbf{x})$ for Class h, $j = 1, \ldots, C$, as

$$\alpha_{\text{Class } h}(\mathbf{x}) = \max\{\mu_j(\mathbf{x}) \cdot CF_j | C_j = h\}. \quad (9)$$

Step 2: Find Class h' that has the maximum value of $\alpha_{\text{Class } h}(\mathbf{x})$:

$$\alpha_{\text{Class } h'}(\mathbf{x}) = \max_{1 \le k \le C} \{\alpha_{\text{Class } k}(\mathbf{x})\}. \quad (10)$$

If two or more classes take the maximal value, then the classification of $\mathbf{x}$ is rejected (i.e. $\mathbf{x}$ is left as an unclassifiable pattern), otherwise $\mathbf{x}$ is assigned to Class h'.

4 Computational Experiments

In the computational experiments performed to evaluate our proposed classifier we use Appendix data set. The appendix represents a seven-dimensional two-class pattern classification problem. There are 106 patterns in the data set, 21 patterns from Class 1 and 85 from Class 2. We assume in the computational experiments in this paper that Class 1 is more important than Class 2.

We apply both the cost-based fuzzy classification system and the conventional to the three data sets. First we examine the performance of both fuzzy classification systems on the training data. That is, the performance is measured on the whole data set that is used to generate fuzzy classification systems. We specified the number of fuzzy partitions $L = 3$ for each axis. We show the experimental results in Fig. 2.

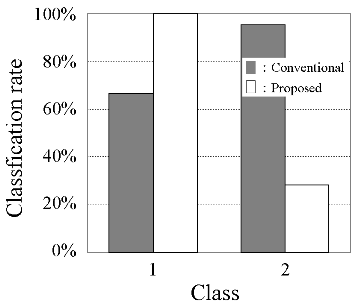

Figure 2: Results for training data.

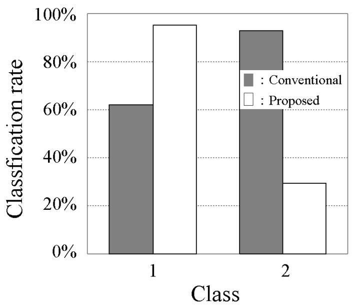

Figure 3: Results for test data.

5 Conclusions

In this paper we have proposed a cost-based fuzzy classification systems for pattern classification problems with an order of class importance. The task is to minimize the total misclassification cost. Experimental results demonstrated the effectiveness of the proposed method compared to a conventional fuzzy-rule classification system.

References

[1] C. C. Lee. 1990, "Fuzzy Logic in Control Systems: Fuzzy Logic Controller Part I and Part II", *IEEE Trans. Syst., Man, Cyberetics*, Vol. 20:404-435.

[2] C. T. Leondes. 1999, "Fuzzy theory Systems", *Techniques and Applications*. Academic Press, San Diego, Vol.1-4, 1999.

[3] M. Sugeno. 1985, "An Introductory Survey of Fuzzy Control", *Information Science*, Vol. 30, No. 1/2:59-83.

[4] K. Nozaki, H. Ishibuchi, H. Tanaka. 1996, "Adaptive fuzzy rule-based classification systems", *IEEE Trans. on Fuzzy Systems*, Vol. 4, No. 3:238-250.

[5] G.J. Klir, B. Yuan. 1995, *Fuzzy Sets and Fuzzy Logic*, Prentice-Hall.

[6] H. Ishibuchi, K. Nozaki, H. Tanaka. 1992, "Distributed representation of fuzzy rules and its application to pattern classification", *Fuzzy Sets and Systems*, Vol. 52, No. 1:21-32.

[7] M. Grabisch. 1996, "The representation of importance and interaction of features by fuzzy measures", *Pattern Recognition Letters*, Vol. 17:567-575.

[8] M. Grabisch, F. Dispot. 1992, "A comparison of some methods of fuzzy classification on real data", *Proc. of 2nd Int. Conf. on Fuzzy Logic and Neural Networks*, 659-662.

[9] M. Grabisch, and J.-M. Nicolas. 1994, "Classification by fuzzy integral: performance and tests", *Fuzzy Sets and Systems*, Vol. 65, No. 2/3:255-271.

[10] G.G. Yen, P. Meesad. 2001, "Constructing a fuzzy rule-based system using the ILFN network and genetic algorithm", *International Journal of Neural Systems*, Vol. 11, No. 5:427-443.

[11] Y. Chen, J.Z. Wang. 2003, "Support Vector Learning for Fuzzy Rule-Based Classification Systems", *IEEE Transactions on Fuzzy Systems*, Vol. 11, No. 6:716–728.

[12] T. Nakashima, Y. Yokota, H. Ishibuchi, A. Bargiela. 2005, "Constructing Fuzzy Classification Systems from Weighted Training Patterns," *Proc. of European Simulation Conference*, 2386-2391.

[13] H. Ishibuchi, K. Nozaki, N. Yamamoto, H. Tanaka. 1995, "Selecting fuzzy if-then rules for classification problems using genetic algorithms", *IEEE Trans. on Fuzzy Systems*, Vol. 3, No. 3:260-270.

[14] H. Ishibuchi, T. Nakashima. 1999b, "Improving the performance of fuzzy classifier systems for pattern classification problems with continuous attributes", *IEEE Trans. on Industrial Electronics*, Vol. 46, No. 6:1057- 1068.

[15] H. Ishibuchi, T. Nakashima. 2001, "Effect of rule weights in fuzzy rule-based classification systems", *IEEE Trans. on Fuzzy Systems*, Vol. 9, No. 4:506-515.