Constructing Fuzzy Classification Systems from Weighted Training Patterns *

Tomoharu Nakashima, Hisao Ishibuchi
College of Engineering
Osaka Prefecture University
Gakuen-cho 1-1, Sakai, Osaka 599-8531, Japan
{nakashi, hisaoi}@ie.osakafu-u.ac.jp

Andrzej Bargiela
Department of Computing
The Nottingham Trent University
Burton Street, Nottingham NG1 4BU, U.K.
andre@doc.ntu.ac.uk

Abstract - *In this paper, we examine the effect of weighting training patterns on the performance of fuzzy rule-based classification systems. A weight is assigned to each given pattern based on the class distribution of its neighboring given patterns. The values of weights are determined proportionally by the number of neighboring patterns from the same class. Large values are assigned to given patterns with many patterns from the same class. Patterns with small weights are not considered in the generation of fuzzy rule-based classification systems. That is, fuzzy if-then rules are generated from only patterns with large weights. These procedures can be viewed as preprocessing in pattern classification. The effect of weighting is examined for an artificial data set and several real-world data sets.*

Keywords: Fuzzy rule-based systems, pattern classification, data mining, cost minimization

1 Introduction

Fuzzy rule-based systems have been applied mainly to control problems [11, 12, 14]. One advantage of the fuzzy rule-based system is its interpretability. Recently fuzzy rule-based systems have also been applied to pattern classification problems. There are many approaches to the automatic generation of fuzzy if-then rules from numerical data for pattern classification problems. Genetic algorithms have also been used for generating fuzzy if-then rules for pattern classification [5, 6, 16].

In this paper, we examine the effect of weighting training patterns on the performance of fuzzy rule-based classification systems. A weight is assigned to each given pattern based on the class distribution of its neighboring given patterns. The values of weights are determined proportionally by the number of neighboring patterns from the same class. Large values are assigned to given patterns with many patterns from the same class. Patterns with small weights are not considered in the generation of fuzzy rule-based classification systems. That is, fuzzy if-then rules are generated from only patterns with large

weights. These procedures can be viewed as preprocessing in pattern classification. The effect of weighting is examined for an artificial data set and several real-world data sets. The results of the computer simulations show that the generalizing ability of fuzzy rule-based classification systems improved by constructing them from the weighted training patterns.

2 Fuzzy Rule-Based Classification

2.1 Pattern Classification Problems

Various methods have been proposed for fuzzy classification [10]. Let us assume that our pattern classification problem is an n-dimensional problem with C classes. We also assume that we have m given training patterns $\mathbf{x}_p = (x_{p1}, x_{p2}, ..., x_{pn})$, $p = 1, 2, ..., m$. Without loss of generality, each attribute of the given training patterns is normalized into a unit interval $[0,1]$. That is, the pattern space is n-dimensional unit hypercube $[0,1]^n$ in our pattern classification problems.

In this study, we use fuzzy if-then rules of the following type in our fuzzy rule-based classification systems:

Rule R_j: If x_1 is A_{j1} and ... and x_n is A_{jn}
　　then Class C_j with CF_j, $j = 1, 2, ..., N$, 　　(1)

where R_j is the label of the j-th fuzzy if-then rule, $A_{j1}, ..., A_{jn}$ are antecedent fuzzy sets on the unit interval $[0,1]$, C_j is the consequent class (i.e., one of the given C classes), CF_j is the grade of certainty of the fuzzy if-then rule R_j, and N is the total number of fuzzy if-then rules. As antecedent fuzzy sets, we use triangular fuzzy sets as in Fig. 1 where we show various partitions of a unit interval into a number of fuzzy sets.
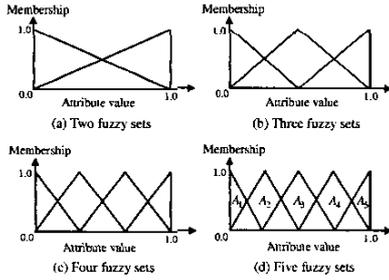
Fig. 1 An example of antecedent fuzzy sets.

2.2 Generating Fuzzy If-Then Rules

In our fuzzy rule-based classification systems, we specify the consequent class and the grade of certainty of each fuzzy if-then rule from the given training patterns [7, 8, 9]. In [7], 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.

The consequent class C_j and the grade of certainty CF_j of fuzzy if-then rule are determined in the following manner:

[Generation Procedure of Fuzzy If-Then Rule]

Step 1: Calculate $\beta_{\text{Class }h}(R_j)$ for Class h
($h = 1, \ldots, C$) as

$$\beta_{\text{Class }h}(R_j) = \sum_{x_p \in \text{Class }h} \mu_{j1}(x_{p1}) \cdot \ldots \cdot \mu_{jn}(x_{pn}),$$

$$h = 1, 2, \ldots, C. \tag{2}$$

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

$$\beta_{\text{Class }\hat{h}}(R_j) = \max\{\beta_{\text{Class }1}(R_j), \ldots, \beta_{\text{Class }C}(R_j)\}. \tag{3}$$

If two or more classes take the maximum value, the consequent class C_j of the rule R_j can not be determined uniquely. In this case, specify C_j as $C_j = \phi$. If a single class takes the maximum value, let C_j be Class $\hat{h}$. If a single class takes the maximum value of $\beta_{\text{Class }h}(R_j)$, the grade of certainty CF_j is determined as

$$CF_j = \frac{\beta_{\text{Class }\hat{h}}(R_j) - \bar{\beta}}{\sum \beta_{\text{Class }h}(R_j)}, \tag{4}$$

where

$$\bar{\beta} = \frac{\sum_{h \neq \hat{h}} \beta_{\text{Class }h}(R_j)}{c - 1}. \tag{5}$$

The number of fuzzy if-then rules in a fuzzy rule-based classification system is dependent on how each attribute is partitioned into fuzzy subsets. For example, when we divide each attribute into three fuzzy subsets in a ten-dimensional pattern classification problem, the total number of fuzzy if-then rules is $3^{10} = 59049$. This is what is called *the curse of dimensionality*. The grade of certainty CF_j can be adjusted by a learning algorithm [13].

2.3 Fuzzy Reasoning

By the rule generation procedure in subsection 2.2, we can generate N fuzzy if-then rules in (1). After both the consequent class C_j and the grade of certainty CF_j are determined for all the N fuzzy if-then rules, a new pattern x is classified by the following procedure [8]:

[Fuzzy reasoning procedure for classification]

Step 1: Calculate $\alpha_{\text{Class }h}(x)$ for Class h, $j = 1, 2, \ldots, C$ as

$$\alpha_{\text{Class }h}(x) = \max\{\alpha_{\text{Class }h}(x) \cdot CF_j \mid C_j = \text{Class }h\},$$

$$h = 1, 2, \ldots, C, \tag{6}$$

where

$$\mu_j(x) = \mu_{j1}(x_1) \cdot \ldots \cdot \mu_{jn}(x_n). \tag{7}$$

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

$$\alpha_{\text{Class }h'}(x) = \max\{\alpha_{\text{Class }1}(x), \ldots, \alpha_{\text{Class }C}(x)\}. \tag{8}$$

If two or more classes take the maximum value, then the classification of x is rejected (i.e., x is left as an unclassifiable pattern), otherwise assign x to Class h'.

3 Assigning Weights

The main aim of assigning weights is extract only necessary patterns for improving the performance of fuzzy rule-based classification systems. Generalization ability in specific is our main focus.

Let us consider a two-dimensional two-class pattern problem in Fig. 2. All of given patterns are shown in Fig.2. 250 patterns were generated from each of two normal distributions: a mean $(0,0)$ and a variance 0.3^2 for Class 1, and $(1,1)$ and 0.3^2 for Class 2. Both distributions do not have any correlation between two attributes. We can see that the two classes overlap with each other.
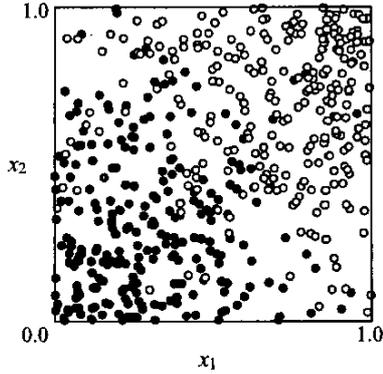

Fig. 2 A two-dimensional pattern classification problem.

In Fig. 3, we show classification boundaries that are generated by fuzzy rule-based classification systems with two, three, four, and five fuzzy sets for each attribute (see Fig. 1). From these figures, we can see that the classification boundaries are not diagonal when the number of fuzzy sets for each attribute is large (e.g., (d) in Fig. 3).

In order to determine the weights of given training patterns, we count the number of patterns from the same class in their neighborhood. Let us denote the neighborhood size as N_{size}. We examine N_{size} nearest patterns from each of given training patterns for determining the value of the weight. We use the following equation to determine the weight of the p-th given pattern w_p:

$$w_p = \frac{N_{same}}{N_{size}}, \qquad (9)$$

where N_{same} is the number of given patterns from the same class as the p-th given pattern. The weight w_p of the p-th given pattern can be viewed as a measure of overlaps. That is, if the value of w_p is large, that means are from the same class as p-th training pattern. On the other hand, the p-th given pattern is possibly an outlier if the value of w_p is low. Only given patterns that have higher weights than a prespecified threshold value are used as training patterns. In this paper, we denote the threshold as θ.

(a) Two fuzzy sets. (b) Three fuzzy sets.

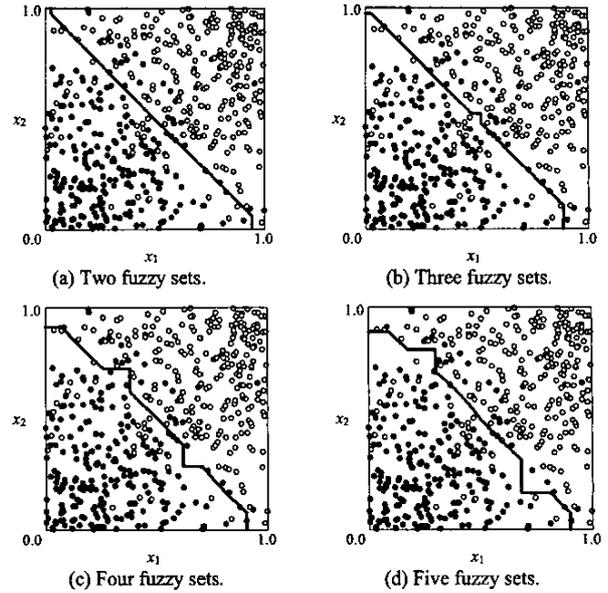(c) Four fuzzy sets. (d) Five fuzzy sets.

Fig. 3 Classification bounraries by fuzzy rule-based classification systems.

In Fig. 4, we show the result of the weight assignment to the two-dimensional patterns in Fig. 2. We specified the value of N_{size} as $N_{size} = 400$ and the value of the threshold as $\theta = 0.5$. From Fig. 4, we can see that overlapping areas are smaller than that in Fig. 2. In Fig. 5, we also show the classification boundaries in the same way as in Fig. 2 but by using the patterns in Fig. 4. From Fig. 5, we can see that the classification boundaries are more diagonal than those in Fig. 5 when the number of fuzzy sets for each attribute is large. Simple shape of classification boundaries such as those in Fig. 5 can lead to high generalization ability. We examine the performance of fuzzy rule-based classification systems on unseen data in the next section.
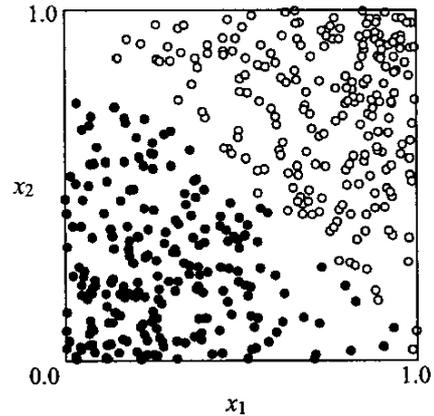

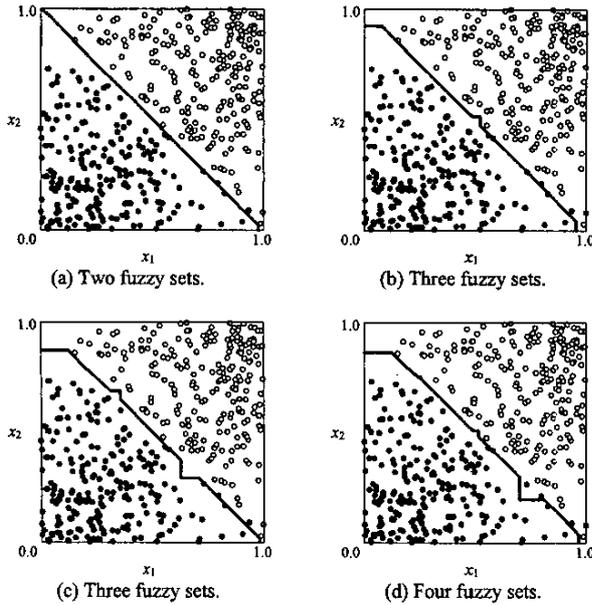Fig. 4 Result of assigning weights $(N_{size}, \theta) = (400, 0.5)$.

(a) Two fuzzy sets.

(b) Three fuzzy sets.

(c) Three fuzzy sets.

(d) Four fuzzy sets.

Fig. 5 Classification boundaries $(N_{size}, \theta) = (400, 0.5)$.

4 Computer Simulations

In this section, we examine the performance of our proposed ensemble method. First we explain data sets which are used in our computer simulations to examine the effect of assigning weights on the performance of fuzzy rule-based classification systems. In this section, we only show classification results on unseen data set because our focus is only on whether the generalization ability can be improved by the weight assignment in Section 3.

4.1 Performance on a Two-Dimensional Data Set

We use the two-dimensional two-class pattern classification problem in Section 3. The data set consists of 250 given training patterns from each class. We generate test patterns for evaluating the performance of fuzzy rule-based classification systems after the weight assignment to training pattern. From each class, we generate 500 patterns as test data by using the same class distribution as described in Section 3. That is, first we assign a weight to each of training patterns by using a prespecified neighborhood size N_{size}. Then, we generate fuzzy if-then rules from those training patterns that have higher value of weight than a prespecified threshold value θ. Test data are used to examine the performance of the generated fuzzy rule-based classification system. Table 1 shows the classification results of fuzzy rule-based classification system. We also show the performance of the conventional method (i.e., no weight is considered and all the given patterns are used to generate fuzzy if-then rules) in the last row of the table. From this table, we can

see that generalization ability is not improved very well by the weight assignment.

Table 1 Classification results for the two-dimensional data set.

N_{size}	The number of fuzzy sets			
	2	3	4	5
1	96.3%	96.4%	96.5%	96.6%
5	96.3%	96.3%	96.5%	96.4%
10	96.3%	96.3%	96.5%	96.5%
20	96.3%	96.3%	96.5%	96.4%
50	96.4%	96.4%	96.4%	96.4%
100	96.4%	96.4%	96.4%	96.5%
200	96.3%	96.4%	96.6%	96.4%
300	96.3%	96.4%	96.5%	96.4%
400	96.3%	96.3%	96.4%	96.4%
No Weight	96.3%	96.5%	96.5%	96.6%

4.2 Performance on Iris Data

Iris data set is a four-dimensional three-class problem with 150 given training patterns [1]. There are 50 training patterns from each class. This data set is one of the most well-known pattern classification problems. Many researchers have applied their classification methods to the iris data set. For example, Weiss and Kulikowski [15] examined the performance of various classification methods such as neural networks and nearest neighbor classifier for this data set. Grabisch and Dispot [3] has also examined the performance of various fuzzy classification methods such as fuzzy integrals and fuzzy k-nearest neighbor for the iris data set.

Table 2 Classification results for the iris data set.

N_{size}	The number of fuzzy sets			
	2	3	4	5
1	68.7%	93.3%	90.0%	95.3%
5	68.7%	93.3%	89.3%	96.0%
10	67.3%	93.3%	88.7%	96.0%
20	68.0%	93.3%	89.3%	95.3%
50	69.3%	90.7%	90.7%	96.0%
No Weight	67.3%	93.3%	89.3%	95.3%

We examined the performance of fuzzy rule-based classification systems on unseen data by using leaving-one-out method. In the leaving-one-out method, a single pattern is used as an unseen data and the other patterns are used to generate fuzzy if-then rules by the procedures in Section 2. Table 2 shows the classification results for the iris data. From this table, we can see that the

generalization ability of fuzzy rule-based classification systems for the two-dimensional two-class pattern classification system is improved by using our weight assignment.

4.3 Performance on Cancer Data

The cancer data set is a nine-dimensional two-class pattern classification problem. In Grabisch's works [2, 3, 4], various fuzzy classification methods have been applied to cancer data set in order to compare each of those fuzzy classification methods. In the same manner as for the iris data set in the last subsection, we examined the performance of the proposed ensemble method for the cancer data set. The performance of our ensemble method on appendicitis data set is shown in Table 3. We can see from this table that by assigning weights to training patterns the performance of the fuzzy rule-based classification system is improved in its classification ability. However, the performance of the fuzzy rule-based classification systems degrades if the neighborhood size N_{size} is large (e.g., $N_{size} = 200$). This is because the number of Class 1 patterns is 85 and the larger neighborhood size than that value does not make a sense.

Table 3 Classification results for the cancer data set.

N_{size}	The number of fuzzy sets			
	2	3	4	5
1	70.3%	64.3%	61.2%	51.4%
5	75.2%	68.5%	66.1%	58.0%
10	75.2%	69.2%	67.8%	59.1%
20	73.8%	70.3%	67.5%	59.8%
50	72.7%	68.5%	66.4%	59.8%
100	69.6%	65.7%	64.0%	58.0%
200	69.6%	65.7%	64.0%	58.0%
No Weight	71.3%	67.8%	65.4%	57.7%

5 Conclusions

In this paper, we examined the effect of weight assignment on the performance of fuzzy rule-based classification systems. A weight for each given pattern is determined by the proportion of patterns from the same class in its neighborhood. Only those given patterns that have larger weights than a threshold value are considered in the generation of fuzzy rule-based classification systems. Thus, the behavior of our weight assignment is determined by two factors: the neighborhood size and the threshold value.

We showed the performance of fuzzy rule-based classification systems with the weight assignment by comparing the performance without it. Through the

computer simulations, we showed that the generalization ability is improved by increasing the neighborhood size. However, the performance of fuzzy rule-based classification systems can be degraded when the neighborhood size is unnecessarily large.

We only used a simplest version of fuzzy rule-based classification systems in this paper. Our future works include the use of more sophisticated versions to show that generalization ability can be efficiently improved by using several techniques in the field of data mining, evolutionary computation, and so on.

Reference

[1] R.A. Fisher, "The use of multiple measurements in taxonomic problems," Ann. of Eugenics, Vol. 7:179-188, 1936.

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

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

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

[5] H. Ishibuchi, T. Nakashima, "Performance evaluation of fuzzy classifier systems for multi-dimensional pattern classification problems", IEEE Trans. on Syst., Man, Cybern. Part B Vol. 29:601-618, 1999.

[6] H. Ishibuchi, T. Nakashima, "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, 1999.

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

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

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

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

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

[12] C.T. Leondes (Ed.), "Fuzzy theory Systems", Techniques and Applications. Academic Press, San Diego, Vol.1-4, 1999.

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

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

[15] S.M. Weiss, C.A. Kulikowski, "Computer systems that learn", Morgan Kaufmann, San Mateo, 1991.

[16] Y. Yuan, H. Zhang, "A genetic algorithms for generating fuzzy classification rules", Fuzzy Sets and Systems, Vol. 84, No. 1:1-19, 1996.