

# Generating Classification Rules from Numerical Data with Misclassification Cost

**Tomoharu Nakashima**

Osaka Prefecture University  
Naka-ku Gakuen-cho 1-1, Sakai  
Osaka 599-8531, Japan  
nakashi@cs.osakafu-u.ac.jp

**Yasuyuki Yokota**

Osaka Prefecture University  
Naka-ku Gakuen-cho 1-1, Sakai  
Osaka 599-8531, Japan  
yokota@ci.cs.osakafu-u.ac.jp

**Gerald Schaefer**

Nottingham Trent University  
Clifton Campus  
Nottingham NG11 8NS, United Kingdom  
Gerald.Schaefer@ntu.ac.uk

**Hisao Ishibuchi**

Osaka Prefecture University  
Naka-ku Gakuen-cho 1-1, Sakai  
Osaka 599-8531, Japan  
hisaoi@cs.osakafu-u.ac.jp

*Abstract*—This paper compares the performance of various rule-based classification systems. In the classification problems in this paper it is assumed that a misclassification cost is associated with each training pattern. Thus, the task of classification is to minimize the total sum of misclassification costs rather than to maximize the classification rate. If-then rules are being generated from a given set of training patterns. The differences between the classification systems used in this paper are (a) whether fuzzy sets or interval sets are used in the antecedent part of if-then rules, and (b) how the consequent part of the if-then rules is determined. In the determination of the consequent part of if-then rules we consider cost-based and compatibility-based determination. In cost-based determination, the consequent class of a rule is determined so that the misclassification costs is minimal over the covered training patterns by the antecedent part of the rule. On the other hand, in compatibility-based determination the consequent class of an if-then rule is determined from the compatibility of training patterns covered by the antecedent part of the rule. The grade of certainty of the rules in both determination types is calculated by using the compatibility of training patterns from each class. In a series of computational experiments, we examine the performance of the classification systems for three real-world pattern classification problems.

## I. INTRODUCTION

Fuzzy rule-based systems have been successfully applied mainly to control problems [1], [2], [3] with one their advantages being their 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 [4], [5], [6], [7], [8], [9]. Genetic algorithms can also be used to generate fuzzy classification systems. For example [10] proposed a genetic algorithm that selects a small number of fuzzy if-then rules from a large number of candidate fuzzy if-then rules. Ishibuchi *et al.* [11] proposed a Michigan-style fuzzy genetics-based machine learning method that generates fuzzy if-then rules from training patterns. Several heuristics for improving the performance of Michigan-style fuzzy genetics-based machine learning method were proposed in [12]. In [10] and [12] the antecedent fuzzy sets are determined by genetic algorithms while the consequent part is determined from given training

patterns that are covered by the antecedent part of the rule.

In some cases misclassification of a particular input pattern causes extra costs. For example, in medical diagnosis of cancer, diagnosing malignant tumors as benign and hence mistaking a cancer patients as healthy could be penalized more than interpreting benign tumors as malignant. In [13] 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/rejection of the pattern. Fuzzy if-then rules are generated by considering the weights as well as the compatibility of training patterns.

This paper compares the performance of various rule-based classification systems. For classification problems in this paper it is assumed that a misclassification cost is associated with each training pattern. Thus the task of the classification is to minimize the total sum of the misclassification costs rather than to maximize the classification rate. If-then rules are being generated from a given set of training patterns. The differences between the classification systems used in this paper are (a) whether fuzzy sets or interval sets are used in the antecedent part of rules, and (b) how the consequent part of the rules is determined. In the determination of the consequent part of a rule we consider cost-based and compatibility-based determination. In cost-based determination, the consequent class of a rule is determined so that the misclassification costs is minimum over the covered training patterns by the antecedent part of rule. On the other hand, in compatibility-based determination the consequent class of a rule is determined from the compatibility of training patterns covered by the antecedent part. The grade of certainty of the if-then rules in both determination methods is calculated by using the compatibility of training patterns from each class. In a series of computational experiments, we examine the performance of the classification systems for three real-world pattern classification problems.

## II. CLASSIFICATION PROBLEM

In this section we describe the pattern classification problems considered in this paper. 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  $\mathbf{x}_p = (x_{p1}, x_{p1}, \dots, x_{pn})$ ,  $p = 1, 2, \dots, m$ , from  $C$  classes are given a priori. In standard pattern classification problems the main focus is on how to construct classifiers that minimize the number of misclassified input patterns. In this case it is assumed that misclassification cost is specified equally for all classes. If  $c_k$  denotes the cost of misclassifying an pattern from class  $k$  then the cost for all classes are specified as follows in the conventional pattern classification problems:

$$c_k = 1, \quad k = 1, 2, \dots, C. \quad (1)$$

In many real-world problems however, the misclassification cost is different for different classes. 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 of benign tumors as the misdiagnosis of a cancer patient as not having the disease leads to a much higher costs than the other way around. In this paper, we handle this situation by specifying different misclassification costs for different classes and put more emphasis on minimizing the misclassification cost rather than minimizing the number of misclassified input patterns. Thus maximizing the classification rates is less important than minimizing the misclassification cost.

Let us consider a simple example where ten training patterns are given in a unit interval  $[0, 1]$  as shown in Figure 1. Six training patterns are of Class 1, and the other four of Class 2. The task here is to describe the class of the unit interval by only a single rule without any condition (i.e. to perform a base classification). In standard pattern classification the interval will be described as Class 1 as the number of Class 1 patterns is larger than that of Class 2 patterns. This is because the misclassification cost of both classes is assumed to be the same ( $c_1 = c_2$ ).

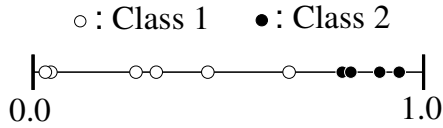


Fig. 1. Ten training patterns in a unit interval  $[0, 1]$ .

On the contrary, if we know that the misclassification costs are specified as  $C_1 = 0.001$  and  $C_2 = 1.0$  beforehand, the classification of the unit interval should be Class 2 as this is the optimal decision in terms of misclassification costs.

## III. CLASSIFICATION SYSTEM

Classification systems for an  $n$ -dimensional pattern classification problem considered in this paper consist of a set of

classification rules of the following type:

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

where  $R_j$  is a rule label,  $A_{ji}$  is an antecedent condition for the attribute value  $x_i$ ,  $i = 1, 2, \dots, n$ ,  $C_j$  is a consequent class,  $CF_j$  is the degree of certainty, and  $N$  is the number of classification rules in a classification system.

We categorize classification systems according to how classification rules in (2) are generated from given training patterns. Table I shows the categorization of classification systems. Fuzzy sets and intervals are used for antecedent conditions  $A_{ji}$ ,  $i = 1, 2, \dots, n$ . For determining the consequent class  $C_j$  and the degree of certainty  $CF_j$ , we employ two formulae. One is cost-based calculation, and the other is compatibility-based calculation. Detailed explanations on the antecedent conditions and the formulae for determining the consequent class and the degree of certainty are given in the following.

### A. Fuzzy and Interval Antecedent Part

The first step to generate a classification rule is the specification of the antecedent part (i.e.  $A_{ji}$ ,  $j = 1, 2, \dots, N$ , and  $i = 1, 2, \dots, n$ ). Fuzzy and interval antecedent parts are considered in this paper. Let us assume that we divide an input axis into  $L$  subsets. In the case of fuzzy antecedent part, we use triangular fuzzy sets in the antecedent part as shown in Figure 2 where four different partitions of an input axis are illustrated.

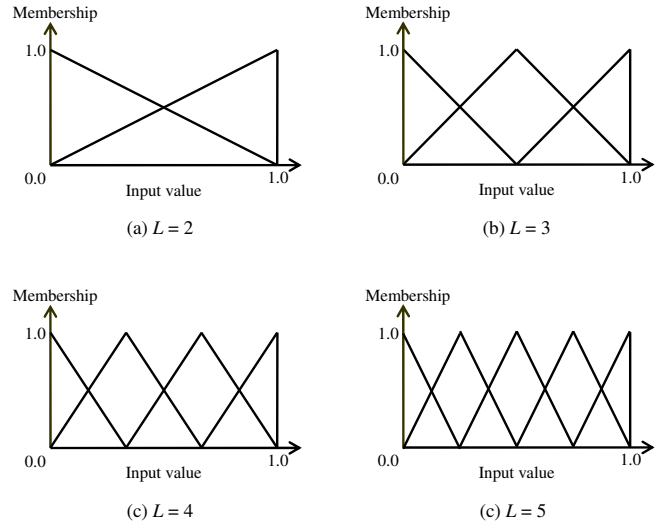


Fig. 2. Triangular fuzzy sets.

The compatibility of an  $n$ -dimensional input vector  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  with a fuzzy rule  $R_j$  is calculated as a product of the membership value for input attributes as follows:

$$\mu_j(\mathbf{x}) = \mu_{j1}(x_1) \cdot \mu_{j2}(x_2) \cdot \dots \cdot \mu_{jn}(x_n), \quad (3)$$

where  $\mu_j(\mathbf{x})$  is the compatibility of  $\mathbf{x}$  and  $\mu_{ji}(\cdot)$ ,  $j = 1, 2, \dots, N$  is the membership function of the antecedent fuzzy set  $A_{ji}$ .

TABLE I: Categorization of classification systems.

Name	Antecedent part	Determination of $C_j$
Fuzzy cost-based	Fuzzy	Cost-based
Fuzzy compatibility-based	Fuzzy	Compatibility-based
Interval cost-based	Interval	Cost-based

In the case of interval antecedent part, we divide an input axis based on the fuzzy partition. An interval set is generated based on the fuzzy partitions as in Figure 3 where the four different interval partitions corresponding to the fuzzy partitions of Figure 2 are given.

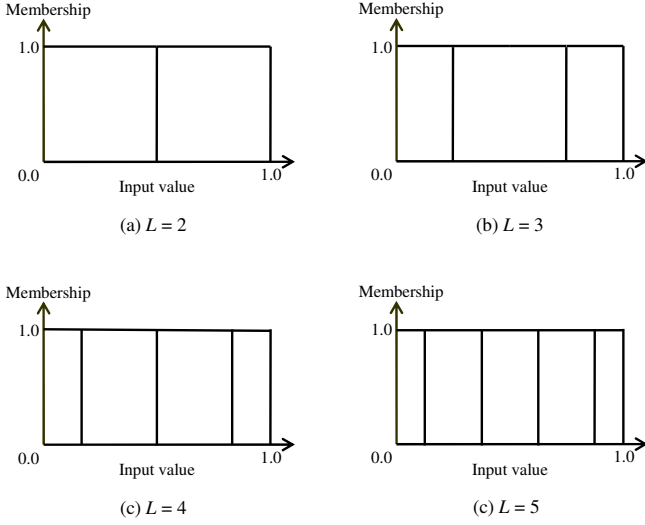


Fig. 3. Interval sets.

Once the number  $L$  of partitions for each input attribute is specified, the total number  $N$  of generated if-then rules in a fuzzy classification system is given as:

$$N = L^n, \quad (4)$$

where  $n$  is the dimensionality of the pattern classification problem. From Equation (4), we can see that the total number of generated if-then rules explodes when the dimensionality of a pattern classification problem is high. On the other hand, if the number of partitions is specified as a small value, the constructed fuzzy classification system does not suffer from the huge number of generated if-then rules although the classification performance often degrades.

### B. Determination of Consequent Class

The consequent class  $C_j$  for the if-then rule  $R_j$  is determined from the given training patterns  $\mathbf{x}_p$ ,  $p = 1, 2, \dots, m$ . In [10] first the sum of the compatibility is calculated for each class. Then the class with the largest value of the sum is taken as the consequent class of the rule. The procedure of determining the consequent class in the conventional fuzzy rule-based classification systems in [10] is summarized in the following procedure:

#### [Compatibility-based determination of $C_j$ ]

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

$$\beta_{\text{Class } h}(j) = \sum_{\mathbf{x}_p \in \text{Class } h} \mu_j(\mathbf{x}_p), \quad (5)$$

where

$$\mu_j(\mathbf{x}_p) = \mu_{j1}(x_{p1}) \cdot \dots \cdot \mu_{jn}(x_{pn}), \quad (6)$$

$\mu_{jn}(\cdot)$  is the membership function of the fuzzy set  $A_{jn}$ .

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

$$\beta_{\text{Class } \hat{h}}(j) = \max_{1 \leq k \leq C} \{\beta_{\text{Class } k}(j)\}. \quad (7)$$

Since the above procedure does not consider misclassification cost, we propose a cost-based determination method where 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 from Class  $k$  with the fuzzy rule  $R_j$  as  $n_j^k$ . A pattern  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  is compatible with rule  $R_j$  if the following condition holds:

$$\mu_{j1}(x_1) \cdot \mu_{j2}(x_2) \cdot \dots \cdot \mu_{jn}(x_n) > 0, \quad (8)$$

where  $\mu_{j1}(\cdot), \mu_{j2}(\cdot), \dots, \mu_{jn}(\cdot)$  are the membership functions of antecedent fuzzy sets  $A_{j1}, A_{j2}, \dots, A_{jn}$  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.

Cost-based class determination hence proceeds as follows:

#### [Cost-based determination of $C_j$ ]

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

Step 2: Calculate the misclassification cost  $Cost_j^k$  for Class  $k$  by  $R_j$  as

$$Cost_j^k = n_j^k \times c_k, \quad (9)$$

where  $c_k$  is the cost of misclassifying a pattern from Class  $k$ .

Step 3: Find Class  $\hat{h}$  that has the maximum value of  $cost_j^k$ :

$$Cost_j^{\hat{h}} = \max_{1 \leq k \leq C} Cost_j^k. \quad (10)$$

### C. Calculation of the Grade of Certainty

The grade of certainty can be viewed as a rule weight. Thus if a rule has a high grade of certainty, it can be used for classification even if it has only a low compatibility grade with

an input pattern. The procedure of calculating the grade of certainty is the same as the one proposed in [10]:

$$CF_j = \frac{\beta_{\text{Class } \hat{h}(j)} - \bar{\beta}}{\sum_h \beta_{\text{Class } h(j)}}, \quad (11)$$

where

$$\bar{\beta} = \frac{\sum_{h \neq \hat{h}} \beta_{\text{Class } h(j)}}{C - 1}. \quad (12)$$

TABLE II: Data sets used in experiments.

Data set	Attributes	Number of training patterns	
		Class 1	Class 2
Appendix	7	21	85
Haberman	3	225	81
Wisconsin	9	444	239

Note that the grade of certainty cannot be calculated when there are multiple classes that have the maximal value of  $\beta$  in Equation (7) or the maximal cost in Equation (10) depending on whether compatibility-based or cost-based determination of the consequent class is employed. In these cases the rule is not generated as the consequent class  $C_j$  cannot be determined uniquely.

In the case of interval antecedent sets (i.e. interval cost-based classification systems), the same values of the grade of certainty as the fuzzy version are used.

TABLE III: Specification of misclassification cost.

Data set	Misclassification cost	
	Class 1	Class 2
Appendix	4.05	0.25
Haberman	0.36	2.78
Wisconsin	0.54	1.86

#### D. Fuzzy Inference for Classification

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

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

$$\alpha_{\text{Class } h}(\mathbf{x}) = \max\{\mu_j(\mathbf{x}) \cdot CF_j | C_j = h\}. \quad (13)$$

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

$$\alpha_{\text{Class } h'}(\mathbf{x}) = \max_{1 \leq k \leq C} \{\alpha_{\text{Class } k}(\mathbf{x})\}. \quad (14)$$

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  $h'$ .

## IV. COMPUTATIONAL EXPERIMENTS

### A. Data Sets

For evaluating the different classifiers we use three real-world data sets: Appendix, Haberman, and Wisconsin breast cancer. The appendix data set was used in [15] to examine the performance of various pattern classification methods. The Wisconsin breast cancer and the Haberman data sets are available from UCI machine learning repository. All the three data sets are two-class pattern classification problems, Table II provides some details.

TABLE IV: Classification rates on training data for Appendix data set(%).

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	86.79	84.91	74.53
3	89.62	88.68	63.21
4	92.45	87.74	74.53
5	95.28	90.57	74.53

TABLE V: Misclassification costs on training data for Appendix data set.

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	11.18	9.43	5.4
3	7.24	6.37	8.0
4	7.06	6.43	2.6
5	5.00	8.12	9.2

TABLE VI: Classification rates on training data for Haberman data set(%).

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	73.86	26.47	26.80
3	74.18	30.07	30.07
4	77.12	42.81	35.29
5	78.76	47.39	40.20

### B. Specification of Misclassification Cost

Since misclassification cost for training patterns is not pre-specified in the original data sets, we assign a misclassification cost to each training pattern in order to generate a synthetic situation where misclassification costs are given a priori. We assume that the misclassification cost depends on the class from which a misclassified pattern comes. It is also assumed that the misclassification cost is determined based on the class distribution and determined the cost  $c_k$  of an input pattern from Class  $k$  as:

$$c_k = \frac{m - N_k}{N_k} = \frac{m}{N_k} - 1, \quad (15)$$

where  $N_k$  is the number of training patterns from Class  $k$  and  $m$  is the total number of given training patterns. It is hence assumed that misclassification cost is high for rare classes with a small number of training patterns, a situation that is typical in medical diagnosis (for example, the number of patients with malignant cancer is usually much smaller than that with benign cancer).

We show misclassification costs for each data set in Table III. For example, if an input pattern from Class 1 is misclassified as Class 2 for the Appendix data set, its misclassification cost is 4.05.

### C. Performance on Training Data

We applied the three classification systems to the three data sets. First, we examined the performance on training data. That is, the performance is measured on the whole data set that is used to generate the classification systems. We examined the performance with different numbers of fuzzy/interval partitions,  $L = 2, 3, 4$ , and 5. The results are given in Tables IV to IX.

TABLE VII: Misclassification costs on training data for Haberman data set.

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	222.22	88.25	80.64
3	214.61	108.47	79.46
4	198.61	77.51	78.53
5	170.88	89.39	82.8

TABLE VIII: Classification rates on training data for Breast cancer data set(%).

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	95.90	97.36	95.31
3	98.24	98.24	97.95
4	100	99.27	98.54
5	100	100	100

TABLE IX: Misclassification costs on training data for Breast cancer data set.

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	23.74	13.03	15.66
3	7.74	8.45	9.03
4	0	3.58	4.32
5	0	0	0

From these tables, we can see that the misclassification costs of the cost-based fuzzy classification systems are lower than that by the fuzzy compatibility-based one for all the three data sets. On the other hand, the classification rates by the fuzzy compatibility-based classification systems are higher than that by the fuzzy cost-based ones which implies the following: The cost-based fuzzy classification systems focuses on the correct classification of those patterns with a high misclassification cost and put less emphasis on patterns with a low misclassification cost.

We can also see that the classification rate of interval cost-based systems is similar to that of the fuzzy cost-based systems. The major difference is that the performance of fuzzy cost-based system improves as the number of fuzzy partitions increases while this is not the case for interval cost-based systems. This is because fuzzy classification systems can achieve

flexible classification boundaries with a large number of fuzzy partitions for each input attribute.

Since the results shown in Tables IV to IX were obtained by classifying the whole given training data and hence cannot be considered as the true performance indicator we will also examine their performance on unseen data.

### D. Performance on Unseen Data

Ten-fold cross-validation technique was used where a set of given training data is divided into ten disjoint subsets and each subset is used as unseen patterns while the other nine subsets are used to generate a classification system. The performance is measured by the average performance on the ten test subsets. We show the experimental results in Tables X to XV.

TABLE X: Classification rates on unseen data for Appendix data set(%).

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	87.00	82.27	65.27
3	86.91	85.09	61.37
4	86.00	78.55	66.09
5	82.37	72.82	60.27

TABLE XI: Misclassification costs on unseen data for Appendix data set.

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	1.12	1.06	0.69
3	0.84	0.85	1.00
4	1.03	1.08	0.88
5	0.96	1.40	1.48

TABLE XII: Classification rates on unseen data for Haberman data set(%).

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	73.21	24.18	26.14
3	72.20	31.02	28.74
4	71.87	40.79	33.69
5	71.23	44.45	35.97

TABLE XIII: Misclassification costs on unseen data for Haberman data set.

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	22.78	10.08	8.40
3	22.66	11.75	9.34
4	21.75	11.00	10.02
5	21.34	12.57	12.01

From these tables, we can see that the misclassification costs by the fuzzy/interval cost-based classification systems are lower than that by the fuzzy compatibility-based ones for the Haberman data set. For the other two data sets, the misclassification costs by the fuzzy/interval cost-based classification systems is low when the number of fuzzy partitions  $L$  for each axis is small such as in the cases of  $L = 2$  and  $L = 3$ . This shows that the performance of the fuzzy classification systems depends on the number of fuzzy partitions.

TABLE XIV: Classification rates on unseen data for Breast cancer data set (%)

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	95.32	96.50	94.01
3	96.34	89.47	88.73
4	90.50	77.62	76.74
5	81.71	66.34	66.34

TABLE XV: Misclassification costs on unseen data for Breast cancer data set.

$L$	Fuzzy compat.	Fuzzy cost	Interval cost
2	2.77	1.69	2.12
3	1.72	6.14	6.21
4	5.37	13.81	13.99
5	11.08	21.09	21.09

Although the fuzzy compatibility-based classification systems perform well when the number of fuzzy partitions is large, they suffer from the curse of dimensionality in terms of the number of generated fuzzy if-then rules. For example, when a fuzzy classification system is applied to a ten-dimensional pattern classification problem with the number of fuzzy partitions  $L = 5$ , the total number of generated fuzzy if-then rules is  $5^{10} = 9765625$ . On the other hand, the total number of fuzzy if-then rules is only  $2^{10} = 1024$  when the number of fuzzy partitions for each axis is two. Thus, there exists a tradeoff between the performance and the compactness of the rule-based classification systems.

## V. CONCLUSIONS

In this paper we examined the performance of various classification systems for pattern classification problems. The assumption in this paper is that a misclassification cost is assigned to each training pattern. The task of the classification systems is then to minimize the total misclassification cost rather than to maximize the classification rates.

We considered three classification systems. Two classifiers are based on fuzzy if-then rules while the third is based on interval if-then rules. Interval partitions are generated from fuzzy partitions by separating the input attributes on the crossing points of the triangular membership functions. We also considered two determination methods of the consequent classes: cost-based and compatibility-based determination. In cost-based determination the consequent class of a rule is determined as the class with the largest misclassification cost. The dominating class in terms of the sum of compatibility is determined as the class of a rule in the compatibility-based determination.

In a series of computer simulations we examined the performance of the three classification systems for three real-world data sets. The experimental results showed that the performance of the fuzzy/interval cost-based systems is better than that of fuzzy compatibility-based ones.

In future works we will derive a more detailed analysis of the performance of the classification systems. For example,

the sensitivity of the performance of the classification systems against the cost specification should be investigated. Improvement of the performance will also be addressed in near future research.

## REFERENCES

- [1] Lee, C. C. 1990. "Fuzzy Logic in Control Systems: Fuzzy Logic Controller Part I and Part II". *IEEE Trans. Syst., Man, Cybernetics*. Vol. 20:404-435.
- [2] Leondes, C. T. 1999. "Fuzzy theory Systems". *Techniques and Applications*. Academic Press, San Diego, Vol.1-4.
- [3] Sugeno, M. 1985. "An Introductory Survey of Fuzzy Control". *Information Science*. Vol. 30, No. 1/2:59-83.
- [4] Nozaki, K.; H. Ishibuchi; and H. Tanaka. 1996. "Adaptive fuzzy rule-based classification systems". *IEEE Trans. on Fuzzy Systems*. Vol. 4, No. 3:238-250.
- [5] Klir, G.J. and B. Yuan. 1995. *Fuzzy Sets and Fuzzy Logic*. Prentice-Hall.
- [6] Ishibuchi, H.; K. Nozaki; and 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] Grabisch, M. 1996. "The representation of importance and interaction of features by fuzzy measures". *Pattern Recognition Letters*. Vol. 17:567-575.
- [8] Grabisch, M. and F. Dispot. 1992. "A comparison of some methods of fuzzy classification on real data". *Proc. of 2nd Intl. Conf. on Fuzzy Logic and Neural Networks*. 659-662.
- [9] Grabisch, M. and J.-M. Nicolas. 1994. "Classification by fuzzy integral: performance and tests". *Fuzzy Sets and Systems*. Vol. 65, No. 2/3:255-271.
- [10] Ishibuchi, H.; 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.
- [11] Ishibuchi, H. and T. Nakashima. 1999. "Performance evaluation of fuzzy classifier systems for multi-dimensional pattern classification problems". *IEEE Trans. on Syst., Man, Cybernetics, Part B*. Vol. 29:601-618.
- [12] Ishibuchi, H. and T. Nakashima. 1999. "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.
- [13] Nakashima, T.; Y. Yokota; H. Ishibuchi; and A. Bargiela. 2005. "Constructing Fuzzy Classification Systems from Weighted Training Patterns". *Proc. of European Simulation Conference*. 91-96.
- [14] Ishibuchi, H. and T. Nakashima. 2001. "Effect of rule weights in fuzzy rule-based classification systems". *IEEE Trans. on Fuzzy Systems*. Vol. 9, No. 4:506-515.
- [15] Weiss, S. M. and C. A. Kulikowski. 1991. *Computer Systems That Learn*. Morgan Kaufmann Publishers, San Mateo, CA.