

CLASSIFICATION CONFIDENCE OF FUZZY RULE-BASED CLASSIFIERS

Tomoharu Nakashima
Department of Computer Science and Intelligent Systems
Osaka Prefecture University
Gakuen-cho 1-1, Naka-ku, Sakai, Osaka 599-8531, Japan
Email: nakashi@cs.osakafu-u.ac.jp

Ashish Ghosh
Machine Intelligence Unit
Indian Statistical Institute
203 B.T. Road, Kolkata 700108, India
Email: ash@isical.ac.in

KEYWORDS

Fuzzy if-then rule, pattern recognition, classification boundary, classification confidence, and cost-sensitive classification.

ABSTRACT

In this paper we first introduce the concept of classification confidence in fuzzy rule-based classification. Classification confidence shows the strength of classification for an unseen pattern. Low classification confidence for an unseen pattern means that the classification of that pattern is not very clear compared to that with high classification confidence. Then we focus on the minimum classification confidence for fuzzy rule-based classifiers using the classification confidence. The minimum classification confidence represents the worst classification among given training patterns. Some discussion on assigning a weight to training pattern is given to show that cost-sensitive fuzzy rule-based classifiers are advantageous for producing a large minimum-confidence classifiers. A series of experiments are done in order to show that reasonable classification boundaries can be obtained by cost-sensitive fuzzy rule-based classifiers if appropriate weights are assigned to training patterns.

INTRODUCTION

Fuzzy systems based on fuzzy if-then rules have been researched in various fields such as control (Lee, 1990), classification and modeling (Ishibuchi et al., 2004). A fuzzy rule-based classifier is composed of a set of fuzzy if-then rules. Fuzzy if-then rules are generated from a set of given training patterns. Advantages of fuzzy classifiers are mainly two-folds: First, the classification behavior can be easily understood by human users. This can be done by carefully checking the fuzzy if-then rules in the fuzzy classifier because fuzzy if-then rules are inherently expressed in linguistic forms. Another advantage is nonlinearity in classification. It is well known that non-fuzzy rule-based classifiers are difficult to perform non-linear classification because classification boundaries are always parallel to attribute axes in most cases. The non-linearity of fuzzy classification leads to high generalization ability of fuzzy rule-based classifiers while its classification behavior is linguistically understood.

While fuzzy rule-based classifiers have high generalization ability and linguistic interpretability, much research work has not been done on how well they assign input patterns with class labels. For example, high generalization ability of support vector machine (Abe, 2010) is discussed from mathematical view point based on the concept of margin. The margin is defined as the minimum distance between training patterns and the classification boundaries. However, it is difficult to find the margin for fuzzy rule-based classifiers because the classification boundaries are usually non-linear.

In this paper, we introduce the concept of confidence of classification for fuzzy rule-based classifiers. The concept of confidence is defined as the degree of how well an input pattern is assigned to a class label by a fuzzy rule-based classifier. We focus on the minimum confidence of classification among given training patterns as it represents the worst confidence of classification. It is expected that a fuzzy rule-based classifier with high generalization ability has a large minimum confidence of classification. We also discuss the relationship between the minimum confidence of classification and the classification boundaries. After a cost-sensitive version of fuzzy rule-based classifiers is introduced, we show that fuzzy rule-based classifiers with a large minimum confidence of classification can be obtained by appropriately assigning weights to training patterns.

FUZZY RULE-BASED CLASSIFIER

In this paper, a fuzzy rule-based classifier proposed in (Ishibuchi et al., 2004) is used. It should be noted that the idea of the classification confidence can be applied to any forms of fuzzy classifiers if they are rule-based systems. An overview of the system in (Ishibuchi et al., 2004) is given below.

Fuzzy If-Then Rule

In a pattern classification problem with n dimensional-ity and M classes, we suppose that m labeled patterns, $\mathbf{x}_p = \{x_{p1}, x_{p2}, \dots, x_{pn}\}$, $p = 1, 2, \dots, m$, are given as training patterns. We also assume that without loss of generality, each attribute of \mathbf{x}_p is normalized to a unit interval $[0, 1]$. From the training patterns we generate fuzzy

if-then rules of the following type:

$$R_q: \text{If } x_1 \text{ is } A_{q1} \text{ and } \dots \text{ and } x_n \text{ is } A_{qn} \\ \text{then Class } C_q \text{ with } CF_q, \quad (1) \\ q = 1, 2, \dots, N,$$

where R_q is the label of the q -th fuzzy if-then rule, $\mathbf{A}_q = (A_{q1}, \dots, A_{qn})$ represents a set of antecedent fuzzy sets, C_q a the consequent class, CF_q is the confidence of the rule R_q , and N is the total number of generated fuzzy if-then rules.

We use triangular membership functions as antecedent fuzzy sets. Figure 1 shows triangular membership functions which divide the attribute axis into five fuzzy sets. Suppose that an attribute axis is divided into L fuzzy sets. The membership function of the k -th fuzzy set is defined as follows:

$$\mu_k(x) = \max \left\{ 1 - \frac{|x - x_k|}{v}, 0 \right\}, k = 1, \dots, L, \quad (2)$$

where

$$x_k = \frac{k-1}{L-1}, k = 1, \dots, L, \quad (3)$$

and

$$v = \frac{1}{L-1}. \quad (4)$$

Let us denote the compatibility of a training pattern \mathbf{x}_p with a fuzzy if-then rule R_q as $\mu_{\mathbf{A}_q}(\mathbf{x}_p)$. The compatibility $\mu_{\mathbf{A}_q}(\mathbf{x}_p)$ is calculated as follows:

$$\mu_{\mathbf{A}_q}(\mathbf{x}_p) = \prod_{i=1}^n \mu_{A_{qi}}(x_{pi}), q = 1, 2, \dots, N, \quad (5)$$

where $\mu_{A_{qi}}(x_{pi})$ is the compatibility of x_{pi} with the fuzzy set A_{qi} and x_{pi} is the i -th attribute value of \mathbf{x}_p . Note that $\mu_{A_{qi}}(x_{pi})$ is calculated by (2).

The number of fuzzy rules to be generated is L^n . That is, the number of rules increases exponentially for the division number and the dimensionality.

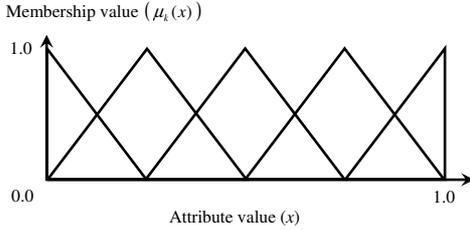


Figure 1: Triangular fuzzy sets.

Generating Fuzzy If-Then Rules

A fuzzy classification system consists of a set of fuzzy if-then rules. The fuzzy if-then rules are generated from the training patterns $\mathbf{x}_p, p = 1, 2, \dots, m$. The number of generated fuzzy if-then rules is determined by the number of fuzzy partitions for each axis (i.e., L in (2) ~ (4)). That is, the number of generated fuzzy if-then rules is

the number of combinations of fuzzy sets that are used for attribute axes. Although different numbers of fuzzy partitions can be used for different axes, in this paper we assume that it is the same for all axes. In this case, the number of fuzzy if-then rules is calculated as $N = L^n$ where n is the dimensionality of the pattern classification problem at hand. In this paper, it is supposed that all attributes are divided in the same way (i.e., the same fuzzy partition). An illustrative example is shown in Fig. 2. In Fig. 2, a two-dimensional pattern space is divided into $3^2 = 9$ fuzzy subspaces as each attribute is divided into three fuzzy sets. Each subspace is labeled with a rule label ($R_1 \sim R_9$). For example, the antecedent part of Rule R_6 has the fuzzy set A_3 for attribute x_1 and A_2 for attribute x_2 . In this way, the total number of generated fuzzy if-then rules and the antecedent part of each fuzzy if-then rule are automatically determined after the number of fuzzy sets for each attribute is determined.

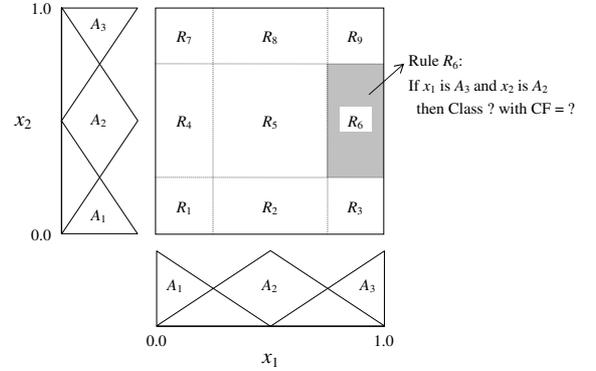


Figure 2: Two-dimensional illustrative example of specifying the antecedent part of a fuzzy if-then rule (three fuzzy sets for both the two attributes).

The consequent part of fuzzy if-then rules (i.e., C_q and CF_q in (1)) is determined from the given training patterns once the antecedent part is specified. The consequent class C_q of the fuzzy if-then rule R_q is determined as follows:

$$C_q = \arg \max_{h=1, \dots, M} \beta_h^q, \quad (6)$$

where

$$\beta_h^q = \sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p). \quad (7)$$

That is, the most matching class with the fuzzy if-then rule is selected considering the given training patterns. If there is not any training pattern that is covered by the fuzzy if-then rules, the consequent class is set as empty. Also, in the case where multiple classes have the maximum value in (6), the consequent class is set as empty. The confidence CF_q is determined as follows:

$$CF_q = \frac{\beta_{C_q} - \bar{\beta}}{\sum_{h=1}^m \beta_h^q}, \quad (8)$$

where

$$\bar{\beta} = \frac{1}{M-1} \sum_{h \neq C_q} \beta_h^q. \quad (9)$$

There are other formulations for determining the confidence. Interested readers are referred to Ishibuchi et al. (2004) for the discussion on the confidence calculation and the performance evaluation.

Classification of Unseen Patterns

Generated fuzzy if-then rules in the previous subsections are used to assign a class label to an unseen pattern which is not included in the set of training patterns. Let us denote an n -dimensional unseen pattern as $\mathbf{x} = (x_1, x_2, \dots, x_n)$. The fuzzy inference is employed to classify unseen patterns in the fuzzy classification system in this paper. The class of an unseen pattern \mathbf{x} is classified as Class C that is determined by the following equation:

$$C = \arg \max_{h=1, \dots, M} \{\alpha_h\}, \quad (10)$$

where

$$\alpha_h = \max_{\substack{q=1, \dots, N \\ C_q=h}} \{\mu_{\mathbf{A}_q(\mathbf{x})} \cdot CF_q\}. \quad (11)$$

In the above equations, M is the number of classes and N is the number of fuzzy if-then rules. In (10), if there are multiple classes that have the same maximum value of α_h , the classification of the unseen pattern is rejected.

CLASSIFICATION CONFIDENCE

Overview

In general, one of the most important things in constructing classification systems is to achieve high generalization ability. For the fuzzy rule-based classifiers, numerous computational experiments have been made so far in order to show its generalization ability. Machine learning community has tackled with this perspective and mathematically shown the limit of the generalization ability. One of the most famous discussion is the large margin classifiers (Bartlett et al., 2000). On the other hand, there have not been many discussions on the margin of fuzzy rule-based classifiers. This is because it is difficult to mathematically define the margin for rule-based classifiers. Instead of defining the margin, this paper proposes a concept of classification confidence for fuzzy rule-based classifiers.

Classification Confidence

Classification confidence represents how well an input pattern is classified by a fuzzy rule-based classifier. The fuzzy inference process assigns an input pattern to the class with the maximum product of the compatibility and the certainty factor (i.e., the maximum value of α_h in (11)). We can also see that the value of α_h represents the degree with which the input pattern belongs to Class h .

Let us consider the difference between the maximum and the second maximum of α_h for an input pattern \mathbf{x} .

We call this the classification confidence for \mathbf{x} and defined as follows:

$$\gamma(\mathbf{x}) = \alpha_C - \arg \max_{h \neq C} \{\alpha_h\}, \quad (12)$$

where C is the assigned class for \mathbf{x} that is determined by the fuzzy inference (10). As the value of α_h is always between 0 to 1, the value of γ ranges in the interval $[0, 1]$. If $\gamma \approx 1$, the fuzzy rule-based classifier has high confidence for the classification of the input pattern because the value of α for the assigned class is much larger than that for the second largest class. On the other hand, the classification is not certain if $\alpha \approx 0$ because the two α 's are close to each other.

Minimum Classification Confidence

We apply the confidence of the classification for given training patterns that are used to construct fuzzy rule-based classifiers. In this case, the minimum value of the confidence represents the worst classification among the given training patterns. Let us denote the minimum confidence of classification as γ^* .

$$\gamma^* = \min_{p=1, \dots, m} \{t_p \cdot \gamma(\mathbf{x}_p)\}, \quad (13)$$

where m is the number of given training patterns and

$$t_p = \begin{cases} 1, & \text{if } \mathbf{x}_p \text{ is correctly classified,} \\ -1, & \text{otherwise.} \end{cases} \quad (14)$$

The above equation is required in order to ensure that a misclassified training pattern has a negative value of confidence so that the misclassification with a large confidence of classification should be more punished than those with a low confidence of classification.

In the next subsection, we examine the relationship between the classification boundaries and the minimum classification confidence γ^* .

Preliminary Experiments

In this subsection, we evaluate the minimum classification confidence for the conventional fuzzy rule-based classifiers. For this purpose, a two-dimensional pattern classification problem shown in Fig. 3 is synthetically generated. This data set consists of 205 training patterns from two classes: 200 patterns from Class 1 and five patterns from Class 2. There is a large gap between the two classes.

We examined the minimum confidence of the conventional fuzzy rule-based classifiers for the data set in Fig. 3. For the fuzzy partition of each axis (i.e., L in (2)), two, three, five, and seven fuzzy sets are examined. The minimum confidence of fuzzy rule-based classifiers for the two-dimensional data set is shown in Table 1. From this table, we can see that the minimum confidence becomes larger as the number of fuzzy sets increases. When two fuzzy sets are used for each axis, all Class-2 patterns are misclassified because the consequent class of the four

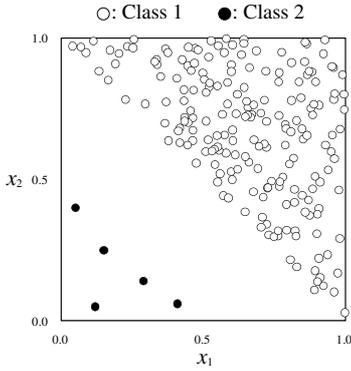


Figure 3: Two-dimensional synthetic data set.

Table 1: Minimum confidence of the conventional fuzzy rule-based classifiers.

# of fuzzy sets	MCC
2	-0.576943
3	-0.182721
5	0.256629
7	0.289514

MCC=Minimum classification confidence

(2×2) fuzzy if-then rules is Class 1. Thus, every training pattern was classified as Class 1.

We also investigated the classification boundaries produced by the conventional fuzzy rule-based classifiers for the test problem in Figs. 4 through 7. From these figures, we can see that the classification boundaries only comes in the middle of the two classes when the number of fuzzy sets is large (e.g., seven fuzzy sets in Fig. 7). We can also see that when the number of fuzzy sets is small (e.g., two and three fuzzy sets), classification boundaries are pushed toward the minor class (Class 2). This is because the conventional fuzzy rule-based classifier does not consider the imbalance in the number of training patterns for the two classes.

COST-SENSITIVE FUZZY RULE-BASED CLASSIFIER

The idea of cost-sensitive classifiers is based on an understanding that in certain cases misclassification of a particular input pattern will cause extra costs. For example in diagnosis of cancer, diagnosing people with cancer as not having the disease could be penalized more than diagnosing healthy individuals as cancer candidates. For this purpose, a cost-sensitive fuzzy rule-based classifier was proposed by (Nakashima et al., 2007b) and (Nakashima et al., 2004). The cost-sensitive fuzzy rule-based classifiers have been applied to image processing (Nakashima et al., 2007a), medical diagnosis (Schaefer et al., 2007), and gene expression analysis (Schaefer et al., 2009).

Extending the principle of fuzzy rule-based classifiers

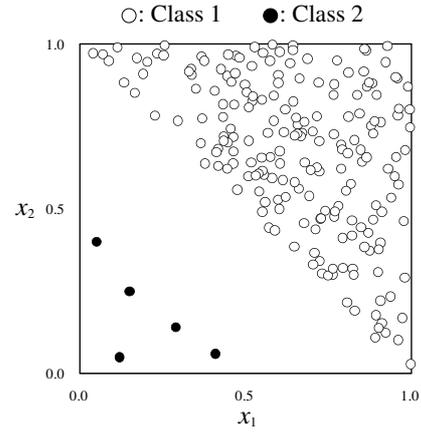


Figure 4: Classification boundary of the conventional fuzzy rule-based classifier (two fuzzy sets for each axis).

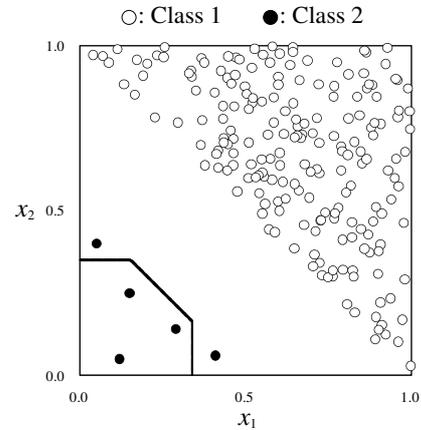


Figure 5: Classification boundary of the conventional fuzzy rule-based classifier (three fuzzy sets for each axis).

to accommodate weighted training patterns is straightforward. We reformulate the pattern classification problem as a cost minimization problem. The concept of 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 it. Fuzzy if-then rules are generated by considering the weights as well as the compatibility of training patterns.

In order to incorporate the concept of weight, the fuzzy rule generation process in (7) is modified as follows:

$$\beta_h^q = \sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p) \cdot \omega_p, \quad (15)$$

where ω_p is the weight associated with training pattern \mathbf{x}_p .

We note that this fuzzy rule generation procedure can also be applied to the standard pattern classification problem where there are no pattern weights. In this case, the class and the grade of certainty are determined from training patterns by specifying a pattern weight as $\omega_p =$

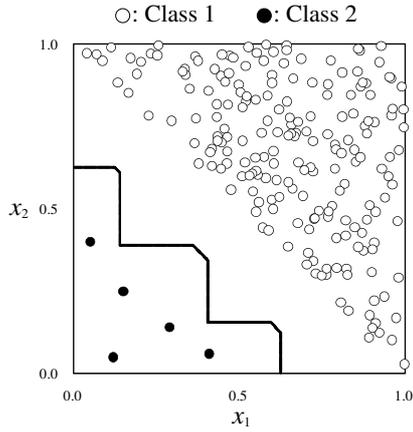


Figure 6: Classification boundary of the conventional fuzzy rule-based classifier (five fuzzy sets for each axis).

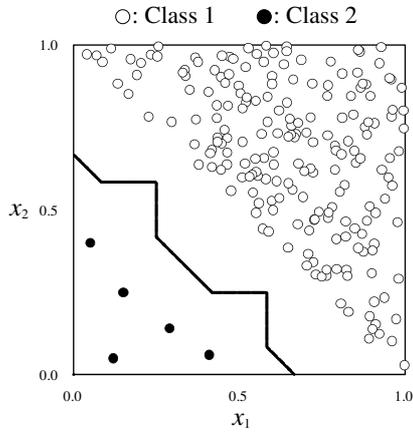


Figure 7: Classification boundary of the conventional fuzzy rule-based classifier (seven fuzzy sets for each axis).

1 for $p = 1, \dots, m$.

Evaluating Minimum Classification Confidence

As in the case of the conventional fuzzy rule-based classifiers, we investigated the minimum confidence of classification of the cost-sensitive fuzzy rule-based classifiers. The data set in Fig. 3 was used as well. The experimental settings are the same as in the conventional ones except that the weight of training patterns is assigned in the following manner: First, each class is given an equal amount of weights (in this case of two class problems, each class has 0.5). Then, it is further divided by the number of training patterns that belongs to each class. Thus, the weight for a Class-1 pattern is $0.5/200 = 0.0025$ while that for a Class-2 pattern is $0.5/5 = 0.1$. We obtained the minimum confidence and boundaries shown in Table 2 and Figs. 8 through 11.

From the above results, we can see that the minimum confidence is positive in all fuzzy partitions. This means that the cost-sensitive fuzzy rule-based classifiers

Table 2: Minimum confidence of the conventional fuzzy rule-based classifiers.

# of fuzzy sets	MCC
2	0.028547
3	0.093917
5	0.144178
7	0.289514

MCC=Minimum classification confidence

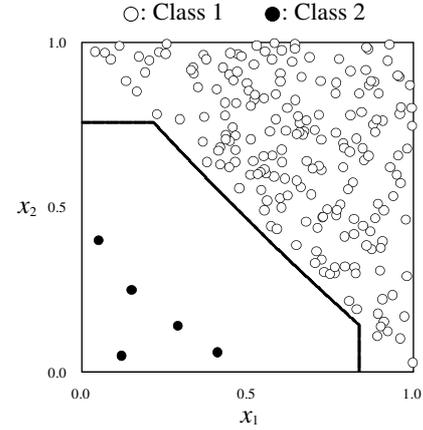


Figure 8: Classification boundary of the cost-sensitive fuzzy rule-based classifier (two fuzzy sets for each axis).

achieved perfect classification for training patterns. This is because the weight of Class-1 patterns is much smaller than that of Class 2 considering the bias in the number of patterns. Thus fuzzy if-then rules with the consequent class of Class 2 are generated by the cost-sensitive fuzzy rule-based classifiers. On the other hand, when the number of fuzzy sets are large (e.g., five and seven), The minimum confidence by the cost-sensitive fuzzy rule-based classifiers are not larger than the conventional ones. This is because when the number of fuzzy sets is large, the area covered by a fuzzy if-then rule becomes small. Thus the bias in the number of patterns for classes does not have much effect on the fuzzy rule-generation process. However, it should be noted that usually a large number of fuzzy sets is not favored in real-world application when human users want to linguistically understand the classification behavior of fuzzy systems. It is better for the number of fuzzy rules to be as small as possible.

CONCLUSIONS

In this paper, we proposed the concept of minimum-confidence for fuzzy rule-based classifiers. The minimum confidence is defined as how well an input pattern is classified. We also showed that the minimum confidence plays a similar role to the well-known idea of margin that was defined for linear discriminant-type classifiers. Cost-sensitive fuzzy rule-based classifiers have potential for producing a large minimum-confidence by assigning

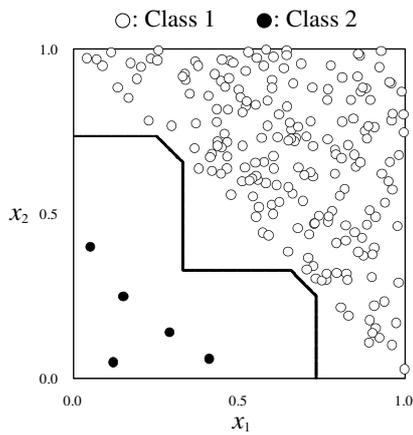


Figure 9: Classification boundary of the cost-sensitive fuzzy rule-based classifier (three fuzzy sets for each axis).

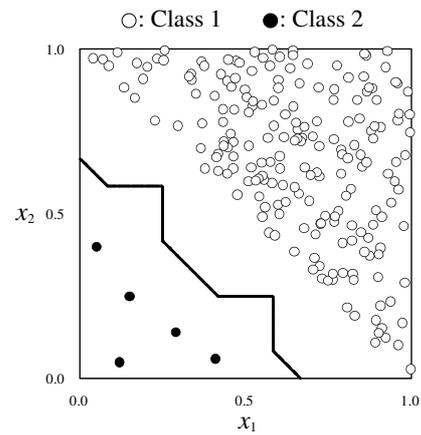


Figure 11: Classification boundary of the cost-sensitive fuzzy rule-based classifier (seven fuzzy sets for each axis).

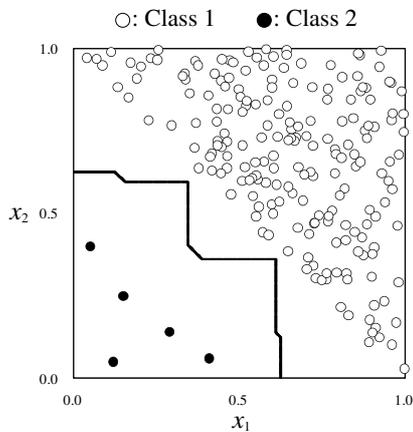


Figure 10: Classification boundary of the cost-sensitive fuzzy rule-based classifier (five fuzzy sets for each axis).

appropriate weights to training patterns.

Only a two-dimensional data set was used in the computational experiments because the main aim of this paper is to clearly introduce the idea of minimum confidence. Thus future works include the investigation of minimum confidence for real-world data sets with high-dimensionality.

REFERENCES

- Abe, S. (2010). *Support Vector Machines for Pattern Classification*. Springer, London, 2nd edition.
- Bartlett, P. J., Schölkopf, B., Schuurmans, D., and Smola, A. J. (2000). *Advances in Large-Margin Classifiers*. The MIT Press, 1st edition.
- Ishibuchi, H., Nakashima, T., and Nii, M. (2004). *Classification And Modeling With Linguistic Information Granules: Advanced Approaches To Linguistic Data Mining*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition.

- Lee, C. C. (1990). Fuzzy logic in control systems: Fuzzy logic controller – part I and II. *IEEE Transactions on Systems Man And Cybernetics*, 20(2):404–435.

- Nakashima, T., Ishibuchi, H., and Bargiela, A. (2004). Constructing fuzzy classification systems from weighted training patterns. In *SMC*, pages 2386–2391. IEEE.

- Nakashima, T., Schaefer, G., Yokota, Y., and Ishibuchi, H. (2007a). A weighted fuzzy classifier and its application to image processing tasks. *Fuzzy Sets and Systems*, 158(3):284–294.

- Nakashima, T., Yokota, Y., Ishibuchi, H., Schaefer, G., Drastich, A., and Zaviscek, M. (2007b). Constructing cost-sensitive fuzzy-rule-based systems for pattern classification problems. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 11(6):546–553.

- Schaefer, G., Nakashima, T., and Ishibuchi, H. (2009). Gene expression analysis by fuzzy and hybrid fuzzy classification. In Jin, Y. and Wang, L., editors, *Fuzzy Systems in Bioinformatics and Computational Biology*, volume 242 of *Studies in Fuzziness and Soft Computing*, pages 127–140. Springer.

- Schaefer, G., Nakashima, T., Yokota, Y., and Ishibuchi, H. (2007). Cost-sensitive fuzzy classification for medical diagnosis. In *CIBCB*, pages 312–316. IEEE.

AUTHOR BIOGRAPHIES

TOMOHARU NAKASHIMA is an associate professor at Osaka Prefecture University. His email is nakashi@cs.osakafu-u.ac.jp and his personal webpage at <http://www.cs.osakafu-u.ac.jp/~nakashi/>.

ASHISH GHOSH is a professor at Indian Statistical Institute. His email is ash@isical.ac.in and his personal webpage at <http://www.isical.ac.in/~ash/>.