# Adaptive Fuzzy Rule-Based Classification Systems

Ken Nozaki, Hisao Ishibuchi, *Member, IEEE*, and Hideo Tanaka, *Member, IEEE*

***Abstract*—This paper proposes an adaptive method to construct a fuzzy rule-based classification system with high performance for pattern classification problems. The proposed method consists of two procedures: an error correction-based learning procedure and an additional learning procedure. The error correction-based learning procedure adjusts the grade of certainty of each fuzzy rule by its classification performance. That is, when a pattern is misclassified by a particular fuzzy rule, the grade of certainty of that rule is decreased. On the contrary, when a pattern is correctly classified, the grade of certainty is increased. Because the error correction-based learning procedure is not meaningful after all the given patterns are correctly classified, we cannot adjust a classification boundary in such a case. To acquire a more intuitively acceptable boundary, we propose an additional learning procedure. We also propose a method for selecting significant fuzzy rules by pruning unnecessary fuzzy rules, which consists of the error correction-based learning procedure and the concept of forgetting. We can construct a compact fuzzy rule-based classification system with high performance. Finally, we test the performance of the proposed two methods on the well-known iris data.**

## I. Introduction

Fuzzy systems based on fuzzy rules have been successfully applied to various control problems [19], [28]. In many application tasks, fuzzy rules were usually derived from human experts as linguistic knowledge. Because it is not always easy to derive fuzzy rules from human experts, recently several methods have been proposed for automatically generating fuzzy rules from numerical data (for example, see [29], [31], and [32]). For pattern classification problems, several classification methods based on fuzzy set theory [34] have been proposed (for example, [12], [13], [7], [8], and [26]). In Grabisch *et al.* [7], [8], those fuzzy classification methods were classified into four categories. We summarize the four categories in Table I. Fuzzy rule-based classification methods were classified into the first category, i.e., the methods based on fuzzy relations.

Let us consider fuzzy rules in a two-dimensional pattern space $X_1 \times X_2$ such as "If $x_1$ is $A$ and $x_2$ is $B$ then $(x_1, x_2)$ belongs to Class 1 with a 0.9 degree of certainty" where $A$ and $B$ are fuzzy subsets on $X_1$ and $X_2$, respectively. Each fuzzy rule defines a fuzzy subspace of the pattern space. Fig. 1 illustrates how a fuzzy rule defines the corresponding fuzzy subspace. As we can see from Fig. 1, the fuzzy subspace is locally represented by the antecedent of the fuzzy rule. That is, the knowledge for classification problems is expressed by each fuzzy rule with local representation [27] or local approximation [17]. In general, local representation-based methods,

Manuscript received September 16, 1994; revised August 25, 1995.

The authors are with the Department of Industrial Engineering, Osaka Prefecture University, Gakuen-cho 1-1, Sakai, Osaka 593, Japan.

Publisher Item Identifier S 1063-6706(96)05631-7.

TABLE I
Fuzzy Classification Methods

| method based on | example |
|---|---|
| fuzzy relations | fuzzy-rule-based methods [12, 13, 30] |
| | linguistic recognition system [25] |
| fuzzy pattern matching | weighted fuzzy pattern matching [5] |
| | fuzzy integral [8] |
| fuzzy clustering | fuzzy c-means [2] |
| other methods | fuzzy k-nearest-neighbor [3, 18] |
| | fuzzy decision tree [4] |

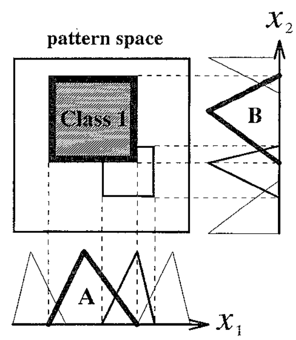



Fig. 1. Local representation by fuzzy rules.

including look-up tables, nearest-neighbor algorithms, and RBF networks [21], [22], have the advantage of requiring relatively small computation time and yield intelligible representation. One of the features of local representation by fuzzy rules is that the fuzzy rules partially overlap with one another as shown in Fig. 1. Therefore, the obtained classification boundary can be smooth because of the interpolation technique well known in fuzzy logic control. On the other hand, we cannot obtain a smooth classification boundary by using, for example, a look-up table based on the crisp partition of a pattern space. This is illustrated in Fig. 2.

Now let us focus our attention on fuzzy rule-based classification methods. Generation of fuzzy rules from numerical data for pattern classification problems consists of two phases: how to partition a pattern space into fuzzy subspaces and how to define a fuzzy rule for each fuzzy subspace. Ishibuchi *et al.* [12] have proposed a fuzzy rule-based classification method based on a fuzzy partition by a simple fuzzy grid (known, henceforth, as the simple fuzzy grid method). Fig. 3 illustrates an example of the fuzzy partition of a two-dimensional pattern space by the simple fuzzy grid method. The simple fuzzy grid method can be viewed as a method based on local representation. When we use the simple fuzzy grid method, the classification performance directly depends on the choice of a

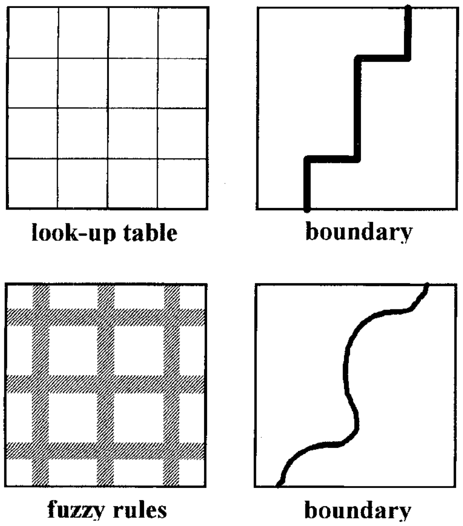



Fig. 2. Classification boundaries by a look-up table and a fuzzy rule-based classification system.

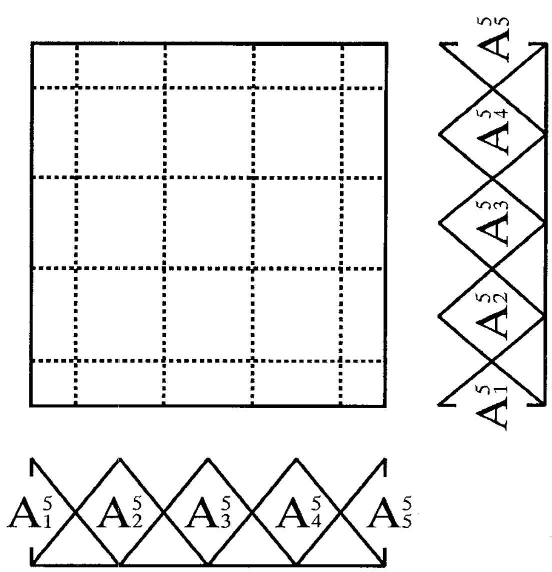


Fig. 3. Fuzzy partition by a simple fuzzy grid.

fuzzy partition. That is, if the fuzzy partition is too coarse, a number of patterns may be misclassified. On the contrary, if it is too fine, a lot of fuzzy rules cannot be generated because of a lack of training patterns in the corresponding fuzzy subspaces. Therefore, the choice of the fuzzy partition is very important but difficult.

To reduce the dependency of the classification performance on the choice of the fuzzy partition, Ishibuchi *et al.* [12] have proposed a fuzzy classification method that simultaneously employs several fuzzy partitions of different sizes in a single fuzzy rule-based classification system. The idea of this method is illustrated in Fig. 4 where $K$ is the number of fuzzy subsets on each axis of the pattern space. It can be seen from Fig. 4 that this method employs multiple fuzzy rule tables. Therefore, we call this method the multirule table method. The main drawback of the multirule table method is that

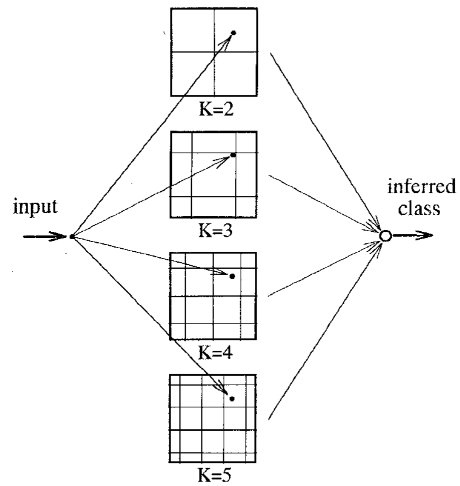



Fig. 4. Multirule table method.

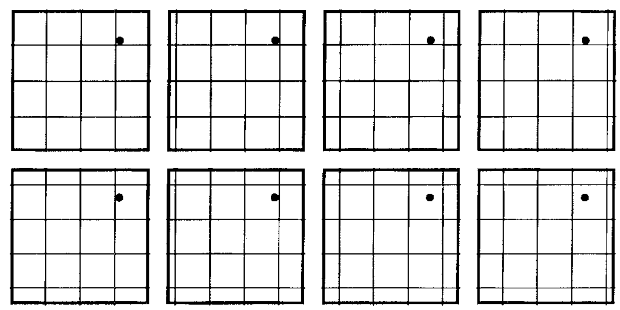


Fig. 5. CMAC.

the number of fuzzy rules would be enormous, especially for high-dimensional classification problems.

The idea of employing plural partitions of an input space has been used in the cerebellar model articulation controller (CMAC) [1] illustrated in Fig. 5 where several partitions have been generated by sliding the grid. As seen from Figs. 4 and 5, CMAC employs several partitions with the same size, whereas the multirule table method uses those with different sizes. In the multirule table method, we can acquire a generalized fuzzy rule table from a coarse fuzzy partition (e.g., see the top table with $K = 2$ in Fig. 4). On the contrary, we can get a specialized fuzzy rule table from a fine fuzzy partition (e.g., see the bottom table with $K = 5$ in Fig. 4). Therefore, the process of choosing a fuzzy partition in the simple fuzzy grid method can be viewed as finding the compromise between generalization and specialization in machine learning [20]. Such a compromise is not necessary in the multirule table method because generalized and specialized fuzzy rule tables are simultaneously employed.

In Ishibuchi *et al.* [12], the grade of certainty of each fuzzy rule is determined by a heuristic method without learning (i.e., one-pass algorithm). The advantage of such a heuristic method is that we require neither time-consuming iterative computation nor complicated learning procedures. However, it can be expected that the classification power may be improved by modifying the grade of certainty of each fuzzy rule. In this study, we will propose an adaptive fuzzy rule-based classification method (known, henceforth, as the adaptive method) that can adjust the grade of certainty of each fuzzy rule by a simple error correction-based learning procedure and an additional learning procedure. The advantages of the adaptive method are: i) we can obtain a fuzzy rule-based classification system with high performance by adjusting the grades of certainty of the fuzzy rules, and ii) we do not require a fine fuzzy partition of a pattern space for even complex classification problems. That is, we can get a high performance classification system with a coarse fuzzy partition. This means that the adaptive method can construct a compact classification system with a small number of fuzzy rules.

We also suggest a rule pruning method to remove unnecessary fuzzy rules by applying the concept of forgetting in a destructive method [16] to a fuzzy rule-based classification system. In the pruning method, we also employ the error correction-based learning procedure of the adaptive method.

This paper is organized as follows. In Section II, we give a brief description of the simple fuzzy grid method and the multirule table method. In Section III, we propose an adaptive method of a fuzzy rule-based classification system, which consists of an error correction-based learning procedure and an additional learning procedure. We examine various specifications of learning constants in the error correction-based learning procedure. We also demonstrate that the additional learning procedure can generate an intuitive acceptable classification boundary. In Section IV, we propose a pruning method that is based on the concept of forgetting for selecting significant fuzzy rules and removing unnecessary fuzzy rules to build a compact fuzzy rule-based classification system. Section V provides the performance evaluation of the proposed methods by applying them to the iris data [6]. We discuss the limitations of the proposed methods and suggest possible extensions in Section VI. Section VII gives the conclusion.

## II. Fuzzy Rule-Based Classification Systems

In this section, we briefly describe the fuzzy rule-based classification method proposed in Ishibuchi *et al.* [12]. The fuzzy rule-based classification method consists of two procedures: a fuzzy rule generation procedure and a classification procedure. We also give a brief description of the multirule table method.

### A. Fuzzy Rule Generation Procedure

Let us assume that a pattern space is the unit square [0, 1] × [0, 1] for enhancing graphical illustration. Suppose that $m$ patterns $\boldsymbol{x}_p = (x_{p1}, x_{p2})$, $p = 1, 2, \cdots, m$, are given as training patterns from $M$ classes: Class 1 ($C1$), Class 2 ($C2$), $\cdots$, Class M ($CM$). That is, the classification of each $\boldsymbol{x}_p (p = 1, 2, \cdots, m)$ is known as one of $M$ classes. The classification problem here is to generate fuzzy rules that divide the pattern space into $M$ disjoint decision areas. For this problem, we employ fuzzy rules of the following type

$$\begin{aligned} \text{Rule } R_{ij}^K\text{:}\quad & \text{If } x_{p1} \text{ is } A_i^K \text{ and } x_{p2} \text{ is } A_j^K \\ & \text{then } \boldsymbol{x}_p \text{ belongs to } C_{ij}^K \text{ with } CF = CF_{ij}^K, \\ & i = 1, 2, \cdots, K; \quad j = 1, 2, \cdots, K \end{aligned} \tag{1}$$

where $K$ is the number of fuzzy subsets on each axis of the pattern space, $R_{ij}^K$ is the label of the fuzzy rule, $A_i^K$ and $A_j^K$ are fuzzy subsets on the unit interval [0, 1], $C_{ij}^K$ is the consequent (i.e., one of $M$ classes), and $CF$ is the grade of certainty of the fuzzy rule.

In this paper, we employ symmetric triangle-shaped membership functions for $A_i^K$ and $A_j^K$ in the antecedent of (1) though we can use other types of membership functions e.g., trapezoid-shaped or bell-shaped. Let us assume that each axis of the pattern space is evenly partitioned into $K$ fuzzy subsets $A_1^K, A_2^K, \cdots, A_K^K$ where $A_i^K$ is defined by the symmetric triangle-shaped membership function

$$\mu_i^K(x) = \max\left\{1 - \frac{|x - a_i^K|}{b^K}, 0\right\}, \quad i = 1, 2, \cdots, K \tag{2}$$

where

$$a_i^K = \frac{i-1}{K-1}, \quad i = 1, 2, \cdots, K \tag{3}$$

$$b^K = \frac{1}{K-1}. \tag{4}$$

Fig. 3 shows the fuzzy partition and the triangle-shaped membership functions for $K = 5$. The simple fuzzy grid method in Ishibuchi *et al.* [12] uses the $K^2$ fuzzy rules in (1) generated by the $K \times K$ simple fuzzy grid such as Fig. 3.

The consequent $C_{ij}^K$ and the grade of certainty $CF_{ij}^K$ of the fuzzy rule $R_{ij}^K$ in (1) can be determined by the following procedure.

*Procedure 1—Generation of Fuzzy Rules:*

1) Calculate $\beta_{CT}$ for $T = 1, 2, \cdots, M$ as

$$\beta_{CT} = \sum_{p \in CT} \mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}). \tag{5}$$

2) Find Class X ($CX$) by

$$\beta_{CX} = \max\{\beta_{C1}, \beta_{C2}, \cdots, \beta_{CM}\}. \tag{6}$$

If multiple classes take the maximum value in (6), the consequent $C_{ij}^K$ of the fuzzy rule corresponding to the fuzzy subspace $A_i^K \times A_j^K$ cannot be determined uniquely. Thus, let $C_{ij}^K$ be $\phi$ and the procedure is terminated. Otherwise, $C_{ij}^K$ is determined as $CX$ in (6).

3) $CF_{ij}^K$ is determined as

$$CF_{ij}^K = \frac{\beta_{CX} - \beta}{\sum_{T=1}^{M} \beta_{CT}} \tag{7}$$

where

$$\beta = \sum_{CT \neq CX} \frac{\beta_{CT}}{M-1}. \tag{8}$$

In this procedure, the consequent $C_{ij}^K$ is determined as the class that has the largest sum of $\mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2})$. The fuzzy rules with the consequent of $\phi$ are called dummy rules that have no effect on fuzzy inference in the classification phase. Thus, when there is no pattern in the fuzzy subspace $A_i^K \times A_j^K$,

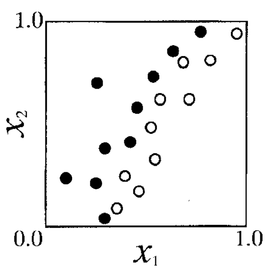



Fig. 6. Two-class classification problem.

this procedure generates a dummy rule as the fuzzy rule $R_{ij}^K$. By Procedure 1, $K^2$ fuzzy rules are generated from the given training patterns for the $K \times K$ simple fuzzy grid.

The compatibility grade of the training pattern $\boldsymbol{x}_p$ to the fuzzy rule $R_{ij}^K$ is defined by the product operator in (5) instead of the minimum operator which has been used in many fuzzy control systems. These two operators were compared in Ishibuchi *et al.* [14] where the superiority of the product operator in classification performance was demonstrated by computer simulations.

There is one possible way to extend the fuzzy rules in (1) when multiple classes take the maximum value of $\beta_{CT}$. Let us assume that Class 1 and Class 3 take the maximum value of $\beta_{CT}$. In this case, we can generate a fuzzy rule with the consequent "Class 1 or Class 3" instead of a dummy rule. While such a fuzzy rule with multiple classes in the consequent part can give useful information (i.e., possible classes in the corresponding fuzzy subspace), it is not used in this paper because its introduction may spoil the simplicity of the simple fuzzy grid method. Utilization of such a fuzzy rule is left for future work.

Let us consider a two-class classification problem shown in Fig. 6 where closed circles and open circles denote training patterns from Class 1 and Class 2, respectively. We applied Procedure 1 to the classification problem shown in Fig. 6 using various values of $K$ (i.e., $K = 2$–7). The generated fuzzy rules are shown in Fig. 7. In each figure, hatched area, dotted area, and painted area represent the following:

1) *Hatched area*: The consequent of the generated fuzzy rule in this area is Class 1 (closed circles).
2) *Dotted area*: The consequent of the generated fuzzy rule in this area is Class 2 (open circles).
3) *Painted area*: A dummy rule is generated in this area.

As seen from Fig. 7, dummy rules denoted by the painted areas were generated for $K = 4$–7 because there is no training pattern in the corresponding fuzzy subspaces. It should be noted that the grid lines in Fig. 7 are not crisp boundaries between fuzzy rules. As is shown in Fig. 3, the grid lines in the pattern space show the 0.5 levels of the fuzzy subsets on each axis. This explains why fuzzy rules are generated at some areas with no training patterns such as the top-left and bottom-right ones for the case of $K = 3$. For example, the top-left fuzzy rule generated for $K = 3$ seems to have no training patterns in Fig. 7, but its actual fuzzy subspace ranges from the top-left corner (0, 1) to the center point (0.5, 0.5) of the two-dimensional pattern space $[0, 1]^2$. Thus,

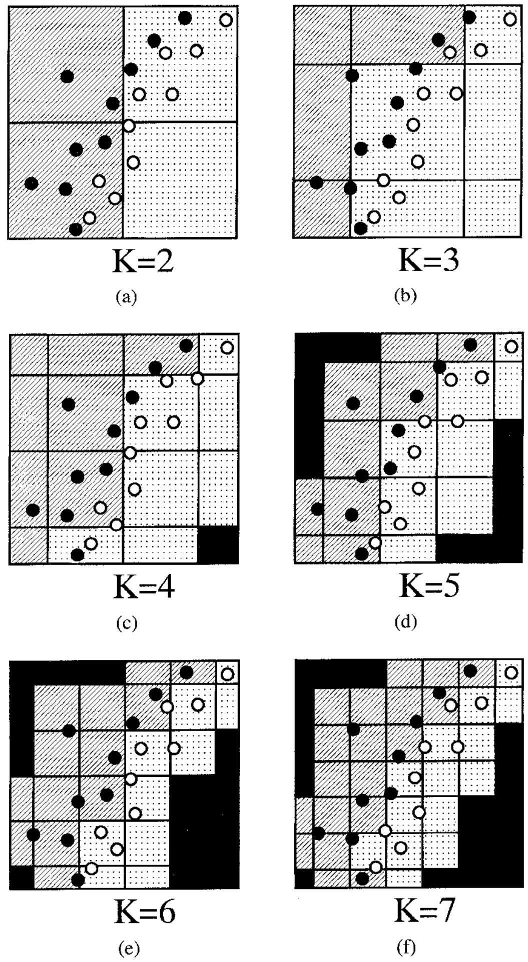



Fig. 7. Generated fuzzy rules.

that fuzzy rule includes two closed circles (i.e., Class 1 patterns).

## B. Classification Procedure

Let us assume that a rule set $S$ is given to form a fuzzy rule-based classification system. Using the fuzzy rules in $S$, we can classify an unknown pattern $\boldsymbol{x}_p = (x_{p1}, x_{p2})$ by the following procedure.

*Procedure 2—Classification of an Unknown Pattern* $\boldsymbol{x}_p = (x_{p1}, x_{p2})$:

1) Calculate $\alpha_{CT}$ for $T = 1, 2, \cdots, M$ as

$$\alpha_{CT} = \max\{\mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}) \cdot CF_{ij}^K | C_{ij}^K = CT; \quad R_{ij}^K \in S\}. \tag{9}$$

2) Find Class X ($CX$) maximizing $\alpha_{CT}$ by

$$\alpha_{CX} = \max\{\alpha_{C1}, \alpha_{C2}, \cdots, \alpha_{CM}\}. \tag{10}$$

When multiple classes take the maximum value in (10), $\boldsymbol{x}_p$ cannot be classified, i.e., $\boldsymbol{x}_p$ is considered as an unclassifiable pattern. Otherwise, $\boldsymbol{x}_p$ is classified as Class X ($CX$) determined by (10).

In this procedure, the result of the fuzzy inference is the consequent of the fuzzy rule with the maximum value of $\mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}) \cdot CF_{ij}^K$. Therefore, when there are no fuzzy rules with $\mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}) > 0$ and $CF_{ij}^K > 0$, the unknown pattern $\boldsymbol{x}_p$ cannot be classified. When multiple classes take the maximum value of $\alpha_{CT}$ in (10), the classification of the unknown pattern is rejected in this procedure. This procedure can easily be extended to indicate possible classes of the unknown pattern. For example, when Class 1 and Class 3 take the maximum value of $\alpha_{CT}$, these two classes can be suggested as the possible classes of the unknown pattern. While we do not allow such a classification strategy in computer simulations of this paper, information about possible classes of the unknown pattern may be useful in some application areas.

We here introduce a confidence level for the classification of an unknown pattern $\boldsymbol{x}_p$, which can be considered a membership value of that pattern to the classification result $CX$. Let $\sigma_p$ be a confidence level defined as

$$\sigma_p = \alpha_{CX} - \max_{CT \neq CX} \{\alpha_{C1}, \alpha_{C2}, \cdots, \alpha_{CM}\} \tag{11}$$

where $\alpha_{C1}, \alpha_{C2}, \cdots, \alpha_{CM}$ and $\alpha_{CX}$ are defined by (9) and (10), respectively.

We applied Procedure 2 to the classification problem in Fig. 6. Classification results by the fuzzy rules for $K$ = 2–7 are shown in Fig. 8 where (a)–(f) are corresponding to those in Fig. 7. In Fig. 8, the curves denote the classification boundaries between Class 1 and Class 2, and the painted areas describe unclassifiable regions. As seen from Fig. 8, when the fuzzy partition of the pattern space is too coarse ($K$ = 2, 3), we cannot correctly classify all the given training patterns. On the contrary, when it is too fine ($K$ = 6, 7), there are several unclassifiable regions because the dummy rules were generated in those fuzzy subspaces. In this example, we can correctly classify all the training patterns by the fuzzy rules for $K$ = 6, 7. In Fig. 7, we should note again that the grid lines in the pattern space are not crisp boundaries between fuzzy rules. This explains why the painted areas in Fig. 8 are much smaller than those in Fig. 7.

In general, when the fuzzy partition is too coarse, we cannot obtain a classification boundary without misclassification. On the contrary, when the fuzzy partition is too fine, we cannot generate a number of fuzzy rules because of a lack of training patterns in the corresponding fuzzy subspaces.

### C. Multirule Table Method

As one possible approach to cope with the above-mentioned issue, Ishibuchi *et al.* [12] have proposed the multirule table method, which simultaneously employs several fuzzy partitions of different sizes in a single fuzzy rule-based classification system as shown in Fig. 4. Fig. 9 shows the classification result of the multi-rule-table method using the fuzzy partitions with $K$ = 2–6, i.e., all the fuzzy rules corresponding to $K$ = 2–6. As seen from Fig. 9, all the training patterns are correctly classified without unclassifiable regions.

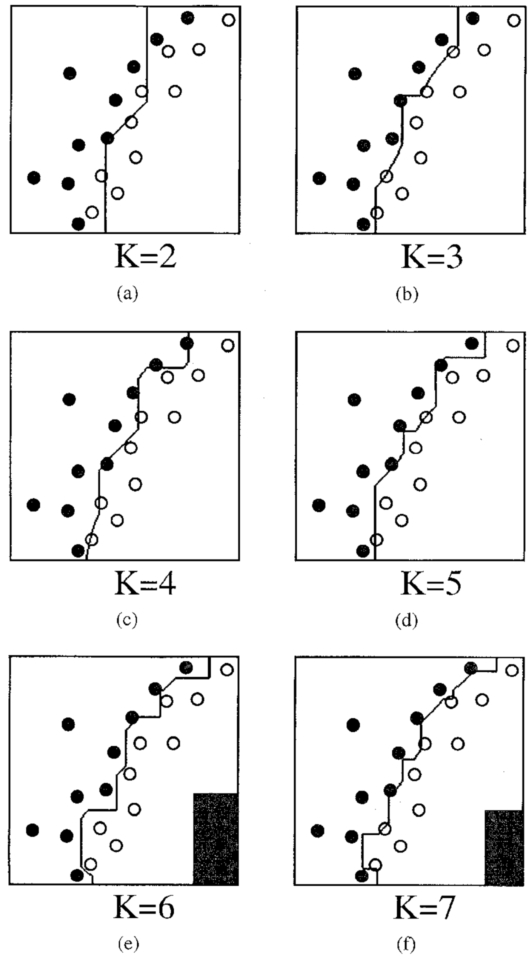



Fig. 8. Classification results by the simple fuzzy grid method.

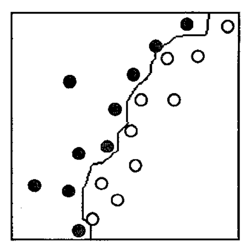


Fig. 9. Classification result by the multirule table method.

## III. Adaptive Fuzzy Rule-Based Classification Systems

In the fuzzy rule-based classification systems described previously, the classification boundary was determined by the two neighboring fuzzy rules with different consequent classes. To illustrate how a classification boundary is determined, let us consider the following fuzzy rules in a one-dimensional pattern space

$R_1$: If $x$ is $A_1$

then $x$ belongs to Class 1 with $CF = CF_1$, (12)

$R_2$: If $x$ is $A_2$

then $x$ belongs to Class 2 with $CF = CF_2$ (13)

where $A_1$ and $A_2$ are fuzzy subsets. Fig. 10 illustrates the classification boundary defined by these fuzzy rules. In Fig. 10, the dotted lines denote the membership functions of $A_1$ and $A_2$, and the solid lines describe the products of the membership functions and the corresponding grades of certainty. As seen from Fig. 10, the classification boundary is determined by the equation $\mu_1(x) \cdot CF_1 = \mu_2(x) \cdot CF_2$ where $\mu_1(x)$ and $\mu_2(x)$ are the membership functions of $A_1$ and $A_2$, respectively. That is, the classification boundary is defined by the ratio of $CF_1$ and $CF_2$.

In the simple fuzzy grid method, the grades of certainty of fuzzy rules are determined by Procedure 1, i.e., a heuristic method without learning capability. Therefore it is possible to improve the classification performance by adjusting the grades of certainty of the fuzzy rules generated by Procedure 1. To construct a fuzzy rule-based classification system with high performance, we here propose an adaptive fuzzy rule-based classification method. The adaptive method consists of two procedures: an error correction-based learning procedure to sequentially adjust the grades of certainty of the fuzzy rules and an additional learning procedure to acquire an intuitively acceptable classification boundary. First, we illustrate the error correction-based learning procedure, then we propose an adaptive fuzzy rule-based classification method with both the error correction-based learning procedure and the additional learning procedure.

### A. Learning Procedure

To adjust the grades of certainty of the fuzzy rules, we use the following error correction-based learning procedure (Procedure 3A).

*Procedure 3A—Error Correction-Based Learning Procedure:* For a training pattern $\boldsymbol{x}_p = (x_{p1}, x_{p2})$, find a fuzzy rule $R_{ij}^K$ satisfying the following equation:

$$\max\{\alpha_{C1}, \alpha_{C2}, \cdots, \alpha_{CM}\} = \mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}) \cdot CF_{ij}^K \tag{14}$$

where $\alpha_{CT}(T = 1, 2, \cdots, M)$ is calculated by (9). Adjust the grade of certainty of the fuzzy rule $R_{ij}^K$ obtained by (14) as follows:

1) when $\boldsymbol{x}_p$ is correctly classified by $R_{ij}^K$

$$CF_{ij}^K := CF_{ij}^K + \eta_1(1 - CF_{ij}^K) \tag{15}$$

and

2) when $\boldsymbol{x}_p$ is misclassified by $R_{ij}^K$

$$CF_{ij}^K := CF_{ij}^K - \eta_2 \cdot CF_{ij}^K \tag{16}$$

where $\eta_1$ and $\eta_2$ are learning constants.

In this procedure, first we find the fuzzy rule $R_{ij}^K$ having the maximum product of the compatibility and the certainty [i.e., $\mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}) \cdot CF_{ij}^K$] among all the fuzzy rules in the rule set $S$. That is, we choose the fuzzy rule that classifies the given training pattern $\boldsymbol{x}_p$. Next, when the selected fuzzy rule correctly classifies the training pattern, the grade of certainty of this rule is increased by (15). Otherwise, the grade of certainty is decreased by (16).

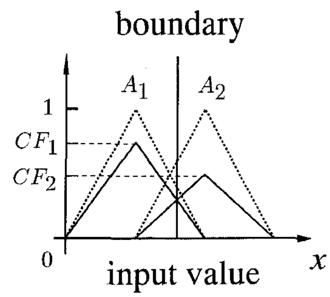



Fig. 10. Classification boundary determined by two fuzzy rules.

Generally, because the number of correctly classified patterns is much larger than that of misclassified patterns as shown in Fig. 8, the grade of certainty of each fuzzy rule tends to be increased to its upper limit (i.e., $CF_{ij}^K \approx 1$) by (15) if we assign the same value to the learning constants $\eta_1$ and $\eta_2$. To avoid this tendency, we should specify $\eta_1$ to be much smaller than $\eta_2$, i.e., $0 < \eta_1 \ll \eta_2 < 1$. Appropriate specifications of these two learning constants will be examined by computer simulations in this subsection later.

A learning method with Procedure 3A for generating a fuzzy rule-based classification system can be written as the following algorithm.

*Learning Method with Procedure 3A*

Step 1) *Initialization*:
1) The number of iterations of the algorithm is set as $J = 0$. Let $J_{\max}$ be the maximum number of iterations, and let $\varepsilon$ be the desirable classification rate.
2) Let $K$ be the number of the fuzzy subspaces on each axis of a pattern space.
3) Generate the fuzzy rules for the given $K$ by Procedure 1 mentioned in Section II. Then denote the generated fuzzy rules by the rule set $S$.

Step 2) *Classification*:
1) Classify all the training patterns by Procedure 2 mentioned in Section II with the fuzzy rules in the rule set $S$. Calculate the classification rate of the training patterns.
2) When $J = J_{\max}$, stop the algorithm.

Step 3) *Adjustment of the Grades of Certainty*:
1) When the classification rate of the training patterns is higher than or equal to the desirable classification rate $\varepsilon$, stop the algorithm.
2) Set $J := J + 1$.
3) Perform Procedure 3A for each $\boldsymbol{x}_p$, $p = 1, 2, \cdots, m$.
4) Go back to Step 2).

In this algorithm, the initial values of the grades of certainty of the fuzzy rules are determined by the heuristic method (i.e., Procedure 1), and the learning of the grades of certainty is performed pattern by pattern. Thus, the order in which the training patterns are used may have some influence on the result of the learning algorithm. This influence can be reduced by using small values of the learning constants. The learning method can be also implemented as a batch procedure (i.e., the grades of certainty are changed after all the training patterns are used).

When the fuzzy partition is too coarse to attain the desirable classification rate $\varepsilon$, the learning method is not terminated until the maximum iteration $J_{\max}$. In this case, the oscillation of classification boundaries may occur. Bad effect of this oscillation can be reduced by using small values of the learning constants. The introduction of the momentum term to the learning of the grades of certainty may also reduce the oscillation.

While we only use the maximum iteration number $J_{\max}$ and the desirable classification rate $\varepsilon$ as the stopping conditions of this algorithm for simplicity, other stopping conditions can be included in the learning method to avoid unnecessary learning iterations. For example, if the algorithm monitors the difference between the current status and the previous one, it can be automatically terminated when no improvement is expected from more iterations.

We illustrate this learning method with Procedure 3A by applying it to the two-class classification problem shown in Fig. 6. To investigate the relationship between the values of the learning constants (i.e., $\eta_1$ and $\eta_2$) and the learning speed, we performed nine numerical experiments where the learning constants $\eta_1$ and $\eta_2$ were chosen from 0.001, 0.01, and 0.1. In these experiments, the stopping conditions of this method were set as $J_{\max} = 1000$ and $\varepsilon$ = 100%. Table II shows the number of iterations required for correctly classifying all the training patterns under the various values of the learning constants. From Table II, we can summarize as follows:

1) When $\eta_1$ was larger than or equal to $\eta_2$, all the training patterns could not be correctly classified until 1000 iterations of the algorithm. That is, the learning constants should be specified as $0 < \eta_1 \ll \eta_2 < 1$.
2) When the learning constants were too small, the learning speed was slow. For example, the required number of iterations for correctly classifying all the training patterns in the second row (i.e., $\eta_1 = 0.001$ and $\eta_2 = 0.01$) was more than ten times as large as that in the third row (i.e., $\eta_1 = 0.001$ and $\eta_2 = 0.1$). Thus, if we use very small values of the learning constants such as $\eta_1 = 0.00001$ and $\eta_2 = 0.0001$, the learning is terribly slow while the learning constants satisfy the condition of $0 < \eta_1 \ll \eta_2 < 1$.

We employ the learning constants $\eta_1 = 0.001$ and $\eta_2 = 0.1$ in the later experiments, which gave the best results in Table II.

We applied the learning method with Procedure 3A using $K = 2$–$5$ to the classification problem shown in Fig. 6. The classification results are shown in Fig. 11. From the comparison of Fig. 11 with Fig. 8, we can see that the classification boundaries between two classes were moved by the learning in the direction of correctly classifying the training patterns as much as possible. Therefore, all the training patterns can be correctly classified by fewer fuzzy rules in the learning method with Procedure 3A than in the simple fuzzy grid method. For example, while the 36 fuzzy rules for $K = 6$ were required for correctly classifying all the training patterns in the simple fuzzy grid method [see Fig. 8(e)], the nine fuzzy rules for $K = 3$ can correctly classify them [see Fig. 11(b)]. This means that the learning method can improve the classification power by adjusting the grades of certainty of the fuzzy rules.

It should be noted that the learning method with Procedure 3A requires a relatively small number of learning iterations when the learning constants are appropriately specified (see Table II). That is, when the learning constants were specified as $\eta_1 = 0.001$ and $\eta_2 = 0.1$, all the training patterns were correctly classified after five iterations for $K = 3$, after three iterations for $K = 4$ and after two iterations for $K = 5$. It is because we use the grades of certainty determined by Procedure 1 as the initial values, which have been already well defined to classify the training patterns.

TABLE II
RELATIONSHIP BETWEEN THE VALUES OF THE LEARNING CONSTANTS AND THE NUMBER OF ITERATIONS

| $\eta_1$ | $\eta_2$ | Number of iterations | | | |
|---|---|---|---|---|---|
| | | $K=2$ | $K=3$ | $K=4$ | $K=5$ |
| 0.001 | 0.001 | * | * | * | * |
| 0.001 | 0.01 | * | 255 | 33 | 27 |
| 0.001 | 0.1 | * | 5 | 3 | 2 |
| 0.01 | 0.001 | * | * | * | * |
| 0.01 | 0.01 | * | * | * | * |
| 0.01 | 0.1 | * | 49 | 4 | 3 |
| 0.1 | 0.001 | * | * | * | * |
| 0.1 | 0.01 | * | * | * | * |
| 0.1 | 0.1 | * | * | * | * |

* :All the training patterns could not be correctly classified until 1000 iterations.

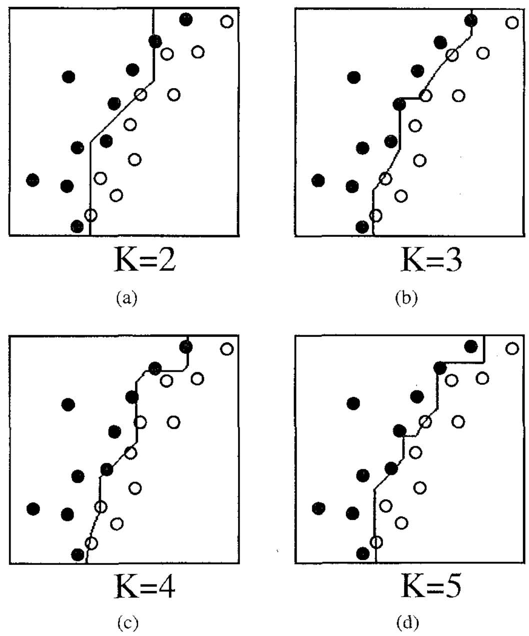


Fig. 11. Classification results by the learning method with Procedure 3A.

### B. Additional Learning

In the learning method with Procedure 3A, the algorithm is stopped when the classification rate for the training patterns becomes higher than or equal to the desirable classification rate $\varepsilon$. Therefore, when we specify $\varepsilon$ as $\varepsilon = 100\%$, the classification boundary obtained by that method is located in the neighborhood of the training pattern that was finally classified correctly. However, one may think that the classification boundary should lie in the center of training patterns belonging to different classes. Thus, we propose the following additional learning procedure (Procedure 3B) to adjust the location of the classification boundary in the direction of obtaining a more intuitively acceptable classification boundary after all the training patterns are correctly classified.

*Procedure 3B—Additional Learning Procedure:* Calculate the confidence level $\sigma_p$ of each training pattern $\boldsymbol{x}_p$, $p = 1, 2, \cdots, m$, by (11). Find the training pattern $\boldsymbol{x}_q$ having the minimum value of the confidence level as follows:

$$\sigma_q = \min_p \sigma_p. \tag{17}$$

When $\sigma_q$ is below the threshold $\theta_a$, adjust the grades of certainty of two fuzzy rules as follows:

1) Find the fuzzy rule $R_{ij}^K$ corresponding to the first term $\alpha_{CX}$ of (11) as

$$\alpha_{CX} = \mu_i^K(x_{q1}) \cdot \mu_j^K(x_{q2}) \cdot CF_{ij}^K. \tag{18}$$

Adjust the grade of certainty of the fuzzy rule satisfying (18) by

$$CF_{ij}^K := CF_{ij}^K + \eta_3(1 - CF_{ij}^K) \tag{19}$$

where $\eta_3$ is the learning constant such that $\eta_1 < \eta_3 < \eta_2$.

2) Find the fuzzy rule $R_{ij}^K$ corresponding to the second term of (11) as

$$\max_{CT \neq CX} \{\alpha_{C1}, \alpha_{C2}, \cdots, \alpha_{CM}\} = \mu_i^K(x_{q1}) \cdot \mu_j^K(x_{q2}) \cdot CF_{ij}^K. \tag{20}$$

Adjust the grade of certainty of the fuzzy rule satisfying (20) by

$$CF_{ij}^K := CF_{ij}^K - \eta_3 \cdot CF_{ij}^K. \tag{21}$$

From the definition in (11), we can see that the confidence level is equal to zero if two classes take the same maximum value of $\alpha_{CT}$. This means that it is equal to zero on classification boundaries. Therefore the confidence levels of training patterns lying near the classification boundaries are very small. In this learning procedure, first we find a training pattern with the minimum confidence level by (17) and choose the two fuzzy rules associated with the calculation of the confidence level by (18) and (20). Then we adjust the grades of certainty of those fuzzy rules in the direction of making the confidence level larger by (19) and (21).

We propose an adaptive method for generating a fuzzy rule-based classification system by both the error correction-based learning (i.e., Procedure 3A) and the additional learning (i.e., Procedure 3B). The adaptive method is the same as the learning method with Procedure 3A in the last subsection except for Step 3). Step 3) of the learning method is modified in the adaptive method as follows:

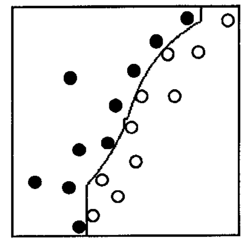


Fig. 12. Classification result by the adaptive method.

*Adaptive Method*

Step 3) *Adjustment of the Grades of Certainty*:

1) Set $J := J + 1$.
2) When the classification rate of the training patterns is below 100%, adjust the fuzzy rule by Procedure 3A for each $\boldsymbol{x}_p$, $p = 1, 2, \cdots, m$.
3) When the classification rate of the training patterns is 100%, adjust the fuzzy rules by Procedure 3B. When the minimum value of the confidence level is beyond the threshold value $\theta_a$, stop the algorithm.
4) Go back to Step 2).

In Step 3) of this adaptive method, when all the training patterns are not correctly classified, the grades of certainty of the fuzzy rules are adjusted by Procedure 3A in the direction of correctly classifying the misclassified training patterns. On the other hand, when all the training patterns are correctly classified, the grades of certainty are adjusted by Procedure 3B in the direction of making the minimum value of the confidence level larger. That is, the grades of certainty are adjusted in the direction of generating a classification boundary coinciding with our intuition.

For the illustration of the adaptive method, we applied it to the two-class classification problem shown in Fig. 6 using the parameter specifications $\eta_1 = 0.001$, $\eta_2 = 0.1$, $\eta_3 = 0.01$, and $\theta_a = 0.05$. Fig. 12 shows the classification result for $K = 3$. From the comparison of Fig. 12 with Fig. 11(b), we can see that the classification boundary by the adaptive method moved in the direction of acquiring an intuitively acceptable classification boundary by Procedure 3B, i.e., by the additional learning procedure.

## IV. Rule Pruning with Forgetting

Fuzzy rule-based classification methods, which generate fuzzy rules from numerical data, can be viewed as a knowledge acquisition tool for classification problems. Generally speaking, the fewer fuzzy rules a fuzzy rule-based classification system has, the more intelligible it is. Ishibuchi *et al.* [15] have proposed the genetic algorithm-based (GA-based) method to construct a compact fuzzy rule-based classification system by selecting significant fuzzy rules and removing unnecessary fuzzy rules by genetic operations.

We propose here a rule-pruning method based on the concept of forgetting [16] in a destructive-learning method to construct a compact fuzzy rule-based classification system. That is, we apply the concept of forgetting to the grades of certainty of fuzzy rules generated by Procedure 1 in Section II to prune unnecessary fuzzy rules. We propose the following method for selecting an appropriate rule set $S$ by introducing the concept of forgetting to the learning method in Section III-A.

*Pruning Method—Rule Pruning with Forgetting:*

Step 1) Set the number of iterations of the algorithm as $t := 1$. Specify the maximum number of iterations $t_{\max}$ as a stopping condition. Generate an initial rule set $S$.

Step 2) Perform Procedure 3A for each $\boldsymbol{x}_p(p = 1, 2, \cdots, m)$.

Step 3) Modify the grades of certainty of the fuzzy rules in the rule set $S$ by

$$CF_{ij}^K := CF_{ij}^K \cdot (1 - \gamma) \tag{22}$$

where $\gamma$ is a forgetting rate. We call this modification "forgetting" in this paper.

Step 4) When the grade of certainty $CF_{ij}^K$ falls below the pre-specified threshold $\theta$, remove the fuzzy rule $R_{ij}^K$ from the rule set $S$.

Step 5) When $t = t_{\max}$, stop the algorithm. Otherwise, let $t := t + 1$ and return to Step 2).

In Step 2) of the pruning method, we use Procedure 3A to adjust the grades of certainty of the fuzzy rules. Furthermore, in Step 3), the grades of certainty of the fuzzy rules in the rule set $S$ are decreased by (22). Therefore, the grades of certainty that are not increased at Step 2) are just decreased at Step 3). This means that the grades of certainty of fuzzy rules that are not used for the classification of the training patterns are reduced monotonically with the number of iterations of the algorithm. On the contrary, the grades of certainty of fuzzy rules that correctly classify the training patterns are increased at Step 2) and decreased at Step 3). When the grade of certainty of a fuzzy rule falls below the threshold $\theta$, that fuzzy rule is removed from the rule set $S$. In this manner, we can prune unnecessary fuzzy rules and obtain a compact rule set.

It should be noted that the basic idea of the pruning method is to remove fuzzy rules that are not used for the classification of the training patterns. For example, a fuzzy rule with a large value of the grade of certainty (e.g., 0.9) in the initial rule set will be removed after some iterations of the pruning method if that fuzzy rule is not used for the classification of any training pattern. On the contrary, a fuzzy rule with a small value of the grade of certainty (e.g., 0.3) will not be removed if that fuzzy rule correctly classifies some training patterns in every iteration of the pruning method.

To illustrate the pruning method, we applied it to the two-class classification problem shown in Fig. 6 with the parameter specifications $t_{\max} = 10\,000$, $\eta_1 = 0.001$, $\eta_2 = 0.1$, $\gamma = 0.001$, and $\theta = 0.1$. In the computer simulation, we used all the fuzzy rules corresponding to the fuzzy partitions for $K = 2$–$6$ shown in Fig. 7(a)–(e) as the initial rules in the rule set $S$. Our problem here is to prune unnecessary fuzzy rules in order to select a compact rule set $S$ from the initial rule set (i.e., all the fuzzy rules for $K = 2$–$6$). Fig. 13 shows the selected fuzzy rules in $S$ and the corresponding classification result. In Fig. 13, the selected fuzzy rules are denoted by the hatched and the dotted areas in the same manner as in Fig. 7. In the comparison of the classification result of the pruning method with those of the simple fuzzy grid method, the pruning method can correctly classify all the given training patterns by using fewer fuzzy rules than the simple fuzzy grid method. That is, in the pruning method, the selected 17 fuzzy rules correctly classified all the training patterns, whereas the 36 fuzzy rules did in the simple fuzzy grid method [see Fig. 8(e)].

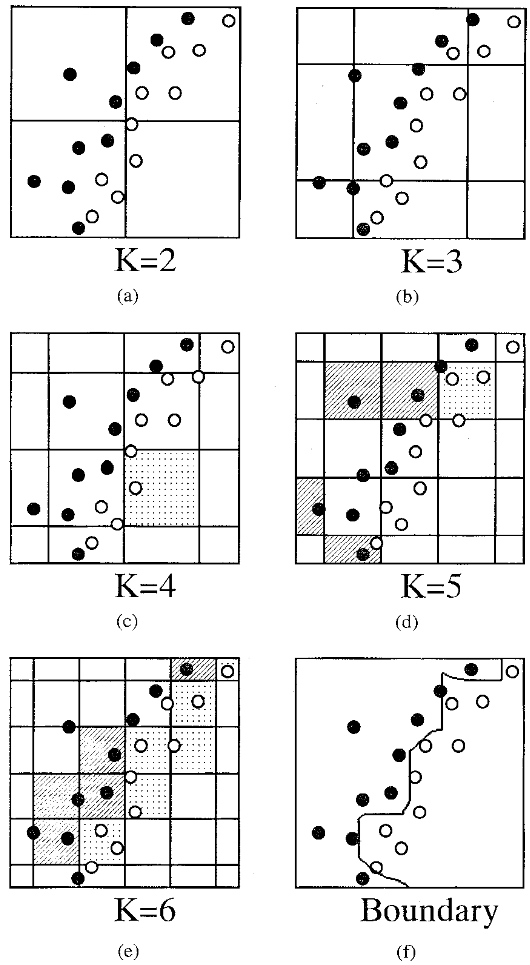



Fig. 13. Selected fuzzy rules and classification result by the pruning method.

There is no guarantee that the selected fuzzy rules by the pruning method cover the entire pattern space. That is, there may be unclassifiable regions in some classification problems. One possible approach to cope with this issue is to employ the bell-shaped fuzzy sets in the antecedents of the fuzzy rules instead of the triangle-shaped fuzzy sets. We applied the pruning method with the bell-shaped fuzzy sets to the classification problem in Fig. 6. The selected fuzzy rules and the corresponding classification result are shown

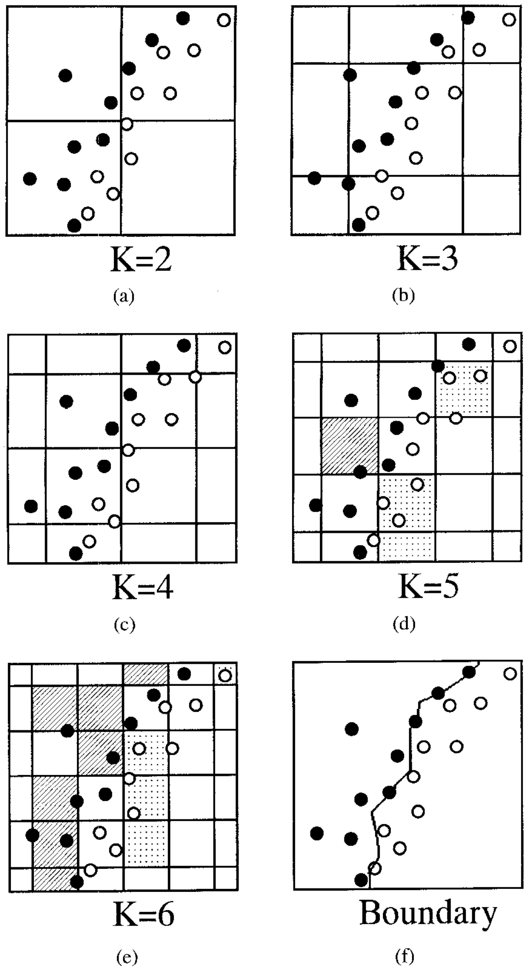



Fig. 14. Selected fuzzy rules and classification result by the pruning method with bell-shaped fuzzy sets.

in Fig. 14. Fig. 14 demonstrates that 15 fuzzy rules correctly classified all the training patterns and covered the entire pattern space.

In Fig. 13, we used all the fuzzy rules corresponding to $K = 2$–6 as the initial rule set. That is, the pruning was started from the rule set generated by the multirule table method. The classification result by the initial rule set was shown in Fig. 9 as the result of the multirule table method. From the comparison between Fig. 9 (before pruning) and Fig. 13 (after pruning), we can see that the classification boundary in Fig. 13(f) is a bit counter-intuitive while all the training patterns were correctly classified in both figures. This suggests that the pruning may slightly deteriorate the classification performance for unknown patterns that were not used for generating and pruning fuzzy rules. This issue will be addressed by computer simulations in the next section. The pruning method can be applied to the initial rule set generated by the simple fuzzy grid method as well as the multirule table method. In this case, fewer fuzzy rules may be selected. We used the multirule table method for generating the initial rule set of the pruning method in computer simulations in the next section as in this section to compare the pruning method with the GA-based method in Ishibuchi *et al.* [15].

TABLE III
IRIS DATA

| sepal length | sepal width | petal length | petal width | Class |
|---|---|---|---|---|
| 5.1 | 3.5 | 1.4 | 0.2 | C1 |
| 4.9 | 3.0 | 1.4 | 0.2 | C1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 7.0 | 3.2 | 4.7 | 1.4 | C2 |
| 6.4 | 3.2 | 4.5 | 1.5 | C2 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 6.3 | 3.3 | 6.0 | 2.5 | C3 |
| 5.8 | 2.7 | 5.1 | 1.9 | C3 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |

## V. SIMULATION RESULTS

We applied the proposed methods to the iris data [6] to verify the effectiveness of our methods. The classification problem of the iris data is to classify three species of iris (iris setosa: C1, iris versicolor: C2, and iris verginica: C3) by the four-dimensional pattern vectors consisting of sepal length ($x_1$), sepal width ($x_2$), petal length ($x_3$), and petal width ($x_4$). There are 50 samples of each class in this classification problem. Table III shows some of the 150 samples of the iris data. To employ the same membership functions for each axis of the pattern space, we normalized attribute values of each attribute as having the maximum value of one and the minimum value of zero. We examine the learning ability for training patterns used for learning and the generalization ability for testing patterns that are not used for learning. Table IV summarizes the proposed methods in this paper and our former methods [12], [15] together with the corresponding parameter specifications used in this section.

### *A. Learning Ability for Training Patterns*

To examine the learning ability for training patterns, we used all the 150 samples as training patterns and performed computer simulations with the parameter specifications listed in Table IV.

Table V shows the classification rates of the adaptive method (i.e., the combination of the error correction-based learning and the additional learning) and the simple fuzzy grid method. It should be noted in Table V that the number of fuzzy rules excludes dummy rules (i.e., fuzzy rules with $\phi$ in the consequent part). Table V indicates that the learning ability of the adaptive method is higher than that of the simple fuzzy grid method for all values of $K$. Especially for $K = 2$, the improvement of the classification rate was 26.67%. We can also see that the adaptive method can correctly classify all the training patterns with only 62 fuzzy rules for $K = 3$. On the contrary, the simple fuzzy grid method correctly classified 98.67% of the training patterns with even 295 fuzzy rules for $K = 6$. It should be noted that the results for the adaptive method in Table V can also be viewed as the results of the error correction-based learning method with no additional learning. This is because the additional learning procedure

TABLE IV
CLASSIFICATION METHODS AND PARAMETER SPECIFICATIONS

| Method | Parameter specifications | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $K$ | $J_{\max}$ | $\eta_1$ | $\eta_2$ | $\eta_3$ | $\theta_\alpha$ | $\gamma$ | $\theta$ |
| Adaptive | $2 \sim 6$ | 1000 | 0.001 | 0.1 | 0.01 | 0.05 | - | - |
| Pruning | $2 \sim 6$ | 10000 | 0.001 | 0.1 | - | - | 0.001 | 0.1 |
| Simple-fuzzy-grid [12] | $2 \sim 6$ | - | - | - | - | - | - | - |
| Multi-rule-table [12] | $2 \sim 6$ | - | - | - | - | - | - | - |
| GA-based [15] | $2 \sim 6$ | - | - | - | - | - | - | - |

TABLE V
LEARNING ABILITY FOR TRAINING PATTERNS (%)

| method | $K=2$ | $K=3$ | $K=4$ | $K=5$ | $K=6$ |
|---|---|---|---|---|---|
| Adaptive | 94.00 | 100.00 | 100.00 | 100.00 | 100.00 |
| Simple-fuzzy-grid | 67.33 | 94.00 | 92.67 | 96.00 | 98.67 |
| # of rules | 16 | 62 | 129 | 190 | 295 |

TABLE VI
LEARNING ABILITY FOR TRAINING PATTERNS

| Method | Correct(%) | Error(%) | # of rules |
|---|---|---|---|
| Pruning | 100.00 | 0.00 | 43 |
| Multi-rule-table | 95.33 | 4.67 | 692 |
| GA-based | 99.47 | 0.53 | 12.6 |

(i.e., Procedure 3B) was used only after a 100% classification rate was attained for the training patterns.

Table VI summarizes the classification performance of the pruning method, the multirule table method and the GA-based method. In those methods, all the fuzzy rules corresponding to the fuzzy partitions for $K$ = 2–6 were used as the initial rules in $S$, i.e., 692 fuzzy rules were used. It can be seen from Table VI that all the training patterns are correctly classified by the selected 43 fuzzy rules by the pruning method. We can also see that the pruning method outperforms the multirule table method from the viewpoint of the number of fuzzy rules as well as the classification rate. Therefore, we can conclude that the pruning method can construct a compact fuzzy rule-based classification system with high performance by removing unnecessary fuzzy rules from the initial rule set. From the comparison of the pruning method with the GA-based method, we can see that the pruning method has a little more fuzzy rules. This is because the number of fuzzy rules was used as a part of the fitness function in the GA-based method. That is, the number of fuzzy rules was directly minimized in the GA-based method while that is implicitly minimized by removing fuzzy rules with very small grades of certainty in the pruning method.

### B. Generalization Ability for Testing Patterns

To examine the generalization ability for testing patterns, we employed the two-fold cross validation (2CV) and the leaving-one-out (LV1) as follows (see [33] for the details of 2CV and LV1):

*Two-Fold Cross Validation (2CV):* Carry out the following procedure ten times and calculate the average classification rate for testing patterns.

TABLE VII
GENERALIZATION ABILITY FOR TESTING PATTERNS (%)

| method | | $K=2$ | $K=3$ | $K=4$ | $K=5$ | $K=6$ |
|---|---|---|---|---|---|---|
| Adaptive | 2CV | 91.73 | 94.80 | 94.53 | 94.80 | 95.37 |
| | LV1 | 92.00 | 95.33 | 98.00 | 94.67 | 96.67 |
| Simple-fuzzy-grid | 2CV | 69.27 | 92.43 | 90.03 | 95.27 | 95.57 |
| | LV1 | 67.33 | 93.33 | 89.33 | 95.33 | 96.67 |

TABLE VIII
GENERALIZATION ABILITY FOR TESTING PATTERNS

| Method | | Correct(%) | Error(%) | # of rules |
|---|---|---|---|---|
| Pruning | 2CV | 93.03 | 6.70 | 28.00 |
| | LV1 | 93.33 | 6.67 | 42.62 |
| Multi-rule-table | 2CV | 94.30 | 5.70 | 597.75 |
| | LV1 | 94.67 | 5.33 | 691.11 |
| GA-based | 2CV | 90.67 | 7.20 | 10.10 |
| | LV1 | 94.67 | 4.00 | 12.90 |

Step 1) Divide randomly 150 samples into two subsets having 75 samples in each subset. Use one subset as training patterns and the other subset as testing patterns.

Step 2) Generate and train the fuzzy rules by using the training patterns, and then classify the testing patterns.

Step 3) Exchange the training patterns for the testing patterns, and then repeat the same procedure as Step 2).

*Leaving One Out (LV1):* Select one testing pattern, and use the other 149 patterns as training patterns. Generate and train the fuzzy rules by using the training patterns, and classify the testing pattern. Repeat this procedure 150 times until each of all the samples is selected as a testing pattern, and calculate the average classification rate for testing patterns.

Tables VII and VIII show the average classification rates for testing patterns by 2CV and LV1. From the comparison of the classification rates of the adaptive method and the simple fuzzy grid method in Table VII, we can see that the generalization ability of the fuzzy rule-based classification system was improved by the adaptive method, especially when the fuzzy partition of the pattern space was coarse, i.e., $K$ = 2, 3, 4. For example, the improvement of the LV1 results was 24.67% for $K$ = 2. On the other hand, when the fuzzy partition is fine, the generalization ability of the adaptive method was slightly worse than that of the simple fuzzy grid method because of overfitting to training patterns.

From the comparison of the simulation results of the pruning method with those of the multirule table method in Table VIII, we can see that the pruning method substantially reduced the number of fuzzy rules at the cost of a slight deterioration of the performance.

## VI. Discussion

In the last section, we have demonstrated that the adjustment of the grade of certainty of each fuzzy rule can improve the classification performance of fuzzy rule-based classification systems generated by our previous methods (i.e., the simple fuzzy grid method and the multirule table method). The proposed methods, however, have the following limitations:

1) The fuzzy partition of the pattern space should be prespecified.
2) The membership function of each antecedent fuzzy set is not adjusted.
3) The number of fuzzy rules is not efficiently reduced by the pruning method in comparison with the GA-based rule selection method.
4) It is difficult to apply the proposed methods to pattern classification problems with many attributes because the number of fuzzy rules generated by fuzzy grids exponentially increases as the number of attributes.

To cope with the first two limitations, we can introduce the learning of the membership function of each antecedent fuzzy set to our adaptive method. The membership function can be adjusted by an error correction-learning method (see Nozaki *et al.* [24]). A genetic algorithm-based learning method similar to Nomura's algorithm [23] can be also used for the learning of the membership function of each antecedent fuzzy set (see Ishibuchi and Murata [9]). While the error-correction-learning algorithm in [24] does not change the number of fuzzy subsets on each axis, the GA-based learning methods in [9], [23] can adjust the number of fuzzy subsets as well as their membership functions.

The third limitation suggests that the GA-based rule selection method [15] rather than the proposed pruning method is appropriate when our aim is to minimize the number of fuzzy rules. Even in this case, the proposed adaptive method can be utilized for the learning of the selected fuzzy rules by the GA-based rule selection method (see Ishibuchi *et al.* [10]). The last limitation prevents us from applying the proposed methods to pattern classification problems with many attributes. For example, if we have five fuzzy subsets on each axis of a thirteen-dimensional pattern space, the number of fuzzy rules generated from a simple fuzzy grid is $5^{13}$ (more than one billion). Such a pattern classification problems involving many attributes can be handled by a fuzzy classifier system (see Ishibuchi *et al.* [11]) where each fuzzy rule is coded as an individual. The proposed adaptive method can also be used in the fuzzy classifier system for the learning of each fuzzy rule.

## VII. Conclusion

This paper proposed the adaptive fuzzy rule-based classification system that can automatically adjust the grades of certainty of fuzzy rules using numerical data. The adaptive method consists of two procedures: the error correction-based learning procedure and the additional learning procedure. We demonstrated by computer simulations that the adaptive method can improve the classification power of fuzzy rules by the error correction-based learning procedure and generate an intuitively acceptable classification boundary by the additional learning procedure. Furthermore, we proposed the pruning method which is based on the concept of forgetting, for constructing a compact fuzzy rule-based classification system with high performance.

In this paper, we just modified the grades of certainty of fuzzy rules to improve the classification performance. There are, however, many other approaches to more sophisticated fuzzy rule-based classification systems. Some approaches such as the learning of the membership functions and the combination of the error correction-based learning and genetic algorithms were suggested in this paper.

## References


[1] J. S. Albus, "A new approach to manipulator control: The cerebellar model articulation controller (CMAC)," *Trans. Amer. Soc. Mech. Eng.: J. Dyn. Syst., Meas. Contr.,* vol. 97, pp. 220–227, Sept. 1975.

[2] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms.* New York: Plenum, 1981.

[3] J. C. Bezdek, S. K. Chuah, and D. Leep, "Generalized $k$-nearest neighbor rules," *Fuzzy Sets Syst.,* vol. 18, no. 3, pp. 237–256, Apr. 1986.

[4] R. L. P. Chang and T. Pavlidis, "Fuzzy decision tree algorithms," *IEEE Trans. Syst., Man, Cybern.,* vol. 7, no. 1, pp. 28–35, Jan. 1977.

[5] D. Dubois, H. Prade, and C. Testermale, "Weighted fuzzy pattern matching," *Fuzzy Sets Syst.,* vol. 28, no. 3, pp. 313–331, Dec. 1988.

[6] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugen.,* vol. 7, pp. 179–188, 1936.

[7] M. Grabisch and F. Dispot, "A comparison of some methods of fuzzy classification on real data," in *Proc. 2nd Int. Conf. Fuzzy Logic Neural Networks,* Iizuka, Japan, July 1992, pp. 659–662.

[8] M. Grabisch and M. Sugeno, "Multi-attribute classification using fuzzy integral," in *Proc. FUZZ-IEEE,* San Diego, CA, Mar. 1992, pp. 47–54.

[9] H. Ishibuchi and T. Murata, "A genetic-algorithm-based fuzzy partition method for pattern classification problems," in *Genetic Algorithm and Soft Computing, Physica-Verlag,* F. Herrera and J. L. Verdegay, Eds. Heidelberg, Germany: Physica-Verlag, Sept. 1996, pp. 555–578.

[10] H. Ishibuchi, T. Muratam, and I. B. Turksen, "Single-objective and multi-objective genetic algorithms for selecting linguistic rules for pattern classification problems," *Fuzzy Sets Syst.,* to be published.

[11] H. Ishibuchi, T. Nakashima, and T. Muratam, "A fuzzy classifier system that generates fuzzy if-then rules for pattern classification problems," in *Proc. Int. Conf. Evolutionary Computat.,* Perth, Australia, Nov. 1995, vol. 2, pp. 759–764.

[12] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets Syst.,* vol. 52, pp. 21–32, Nov. 1992.

[13] ——, "Efficient fuzzy partition of pattern space for classification problems," *Fuzzy Sets Syst.,* vol. 59, pp. 295–304, Nov. 1993.

[14] ——, "Pattern classification by fuzzy rules," *J. Japan Soc. Fuzzy Theory Syst.,* vol. 5, no. 1, pp. 74–84, Feb. 1993 (in Japanese).

[15] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Selecting fuzzy if-then rules for classification problems using genetic algorithms," *IEEE Trans. Fuzzy Syst.,* vol. 3, pp. 260–270, Aug. 1995.

[16] M. Ishikawa, "A structural connectionist learning algorithm with forgetting," *J. Jap. Soc. Artificial Intell.,* vol. 5, no. 5, pp. 595–603, Sept. 1990 (in Japanese).

[17] R. A. Jacobs and M. I. Jordan, "A competitive modular connectionist architecture," in *Advances in Neural Information Processing System,* R. P. Lippmann, J. E. Moody, and D. S. Touretzky, Eds. San Mateo, CA: Morgan Kaufman, 1991, vol. 3, pp. 767–773.

[18] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy $k$-nearest neighbor algorithm," *IEEE Trans. Syst., Man, Cybern.,* vol. SMC-15, pp. 580–585, July/Aug. 1985.

[19] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller," *IEEE Trans. Syst., Man, Cybern.,* vol. 20, pp. 404–435, Mar./Apr. 1990.

[20] R. Michalski, J. G. Carbonell, and T. M. Mitchell, *Machine Learning, An Artificial Intelligence Approach.* Los Altos, CA: Morgan Kaufman, 1983, 1986, 1990, vols. 1–3.

[21] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Comput.,* vol. 1, pp. 281–294, 1989.

[22] M. T. Musavi, W. Ahmed, K. H. Chan, K. B. Faris, and D. M. Hummels, "On the training of radial basis function classifiers," *Neural Networks* vol. 5, pp. 595–603, July/Aug. 1992.

[23] H. Nomura, I. Hayashi, and N. Wakami, "A self-tuning method of fuzzy reasoning by genetic algorithm," in *Proc. '92 Int. Fuzzy Syst. Intell. Contr. Conf.,* Louisville, KY, Mar. 1992, pp. 236–245.

[24] K. Nozaki, T. Morisawa, and H. Ishibuchi, "Adjusting membership functions in fuzzy-rule-based classification systems," in *Proc. EUFIT '95,* Aachen, Germany, Aug. 1995, pp. 615–619.

[25] S. K. Pal and D. P. Mandal, "Linguistic recognition system based on approximate reasoning," *Informat. Sci.,* vol. 61, pp. 135–161, Apr. 1992.

[26] W. Pedrycz, "Fuzzy sets in pattern recognition: Methodology and methods," *Pattern Recogn.,* vol. 23, pp. 121–146, 1990.

[27] D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing.* Cambridge, MA: MIT Press, 1986, vol. 1.

[28] M. Sugeno, "An introductory survey of fuzzy control," *Informat. Sci.,* vol. 36, pp. 59–83, July 1985.

[29] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," *IEEE Trans. Fuzzy Syst.,* vol. 1, pp. 7–31, Feb. 1993.

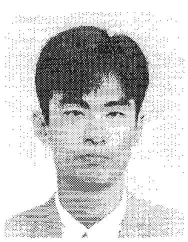
[30] C. T. Sun, "Rule-base structure identification in an adaptive-network-based fuzzy inference system," *IEEE Trans. Fuzzy Syst.,* vol. 2, no. 1, pp. 64–73, Feb. 1994.

[31] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *IEEE Trans. Syst., Man, Cybern.,* vol. SMC-15, pp. 116–132, Jan./Feb. 1985.

[32] L. X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. Syst., Man, Cybern.,* vol. 22, pp. 1414–1427, Nov./Dec. 1992.

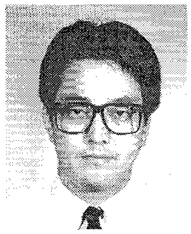
[33] S. M. Weiss and C. A. Kulikowski, *Computer Systems That Learn.* San Mateo, CA: Morgan Kaufmann, 1991.

[34] L. A. Zadeh, "Fuzzy sets," *Informat. Contr.,* vol. 8, pp. 338–353, 1965.

**Ken Nozaki** received the B.S., M.S., and Ph.D. degrees from the Department of Industrial Engineering, Osaka Prefecture University, Japan, in 1992, 1994, and 1995, respectively.

From 1993 to 1994, he was a Visiting Research Associate at the University of Toronto, Canada. He is currently working at Hitachi, Ltd., Systems Development Laboratory, Kawasaki, Japan. His research interests include fuzzy rule-based systems, genetic algorithms and autonomous decentralized systems.



**Hisao Ishibuchi** (M'93) received the B.S. and M.S. degrees in precision mechanics from Kyoto University, Japan, in 1985 and 1987, respectively, and the Ph.D. degree from Osaka Prefecture University, Japan, in 1992.

He is currently an Associate Professor in the Department of Industrial Engineering at Osaka Prefecture University, Japan. From 1994 to 1995, he was a Visiting Research Associate at the University of Toronto, Canada. His current research interests include fuzzy data analysis, fuzzy rule-based systems, fuzzy neural networks, genetic algorithms, and fuzzy scheduling.

Dr. Ishibuchi is a member of INNS and IFSA.

**Hideo Tanaka** (M'81) received the B.S. degree from the Faculty of Engineering at Kobe University, Japan, in 1962, and the M.S. and Ph.D. degrees from the Faculty of Engineering, Osaka City University, Japan, in 1966 and 1969, respectively.

He spent two years with the Central Research Laboratory of Daikin Corporation, Osaka, Japan. He is now with the Department of Industrial Engineering, Osaka Prefecture University, Japan where he has been a Professor since 1987. From 1972 to 1973, he was a Visiting Research Associate at University of California, Berkeley, from 1975 to 1977, an Alexander Von Humboldt Fellow at the Technical University of Aachen, Germany, and from 1981 to 1982 he was a Research Associate at Kansas State University. His research interests include possibility data analysis such as regression analysis and discriminant analysis based on possibility theory.

Dr. Tanaka is currently the President of SOFT (Japan Society for Fuzzy Theory and Systems).