Fuzzy Sets and Systems 52 (1992) 21–32
North-Holland

Distributed representation of fuzzy rules and its application to pattern classification

Hisao Ishibuchi, Ken Nozaki and Hideo Tanaka
Department of Industrial Engineering, University of Osaka Prefecture, Osaka, Japan

Received September 1991
Revised December 1991

Abstract: This paper introduces the concept of distributed representation of fuzzy rules and applies it to classification problems. Distributed representation is implemented by superimposing many fuzzy rules corresponding to different fuzzy partitions of a pattern space. This means that we simultaneously employ many fuzzy rule tables corresponding to different fuzzy partitions in fuzzy inference. In order to apply distributed representation of fuzzy rules to pattern classification problems, we first propose an algorithm to generate fuzzy rules from numerical data. Next we propose a fuzzy inference method using the generated fuzzy rules. The classification power of distributed representation is compared with that of ordinary fuzzy rules which can be viewed as local representation.

Keywords: Distributed fuzzy rules; distributed representation; fuzzy inference; generation of fuzzy rules; pattern classification.

1. Introduction

Ever since fuzzy control research was started by Mamdani's work [5], various fuzzy control systems have been developed [6]. In most of those systems, fuzzy rules are generally derived from human experts using linguistic information. Recently several methods have been proposed to derive fuzzy rules from numerical data [7, 8].

Construction of fuzzy rules from numerical data for pattern classification problems consists of two phases: fuzzy partition of a pattern space and identification of a fuzzy rule for each fuzzy subspace. If the fuzzy partition is too coarse in the sense that the number of fuzzy subspaces is too small, the classification power of the generated fuzzy rules may be low. On the other hand, if the fuzzy partition is too fine in the sense that the number of fuzzy subspaces is too large, there may be a case where some fuzzy rules can not be constructed because of the lack of data points at the corresponding fuzzy subspaces. In general, it is difficult to construct the fuzzy rules whose fuzzy subspaces have no data point.

One approach to remedy the above difficulty is to partition a pattern space depending on the density of data points. That is, some area in a pattern space with high density is partitioned into fine fuzzy subspaces and other areas with low density into coarse fuzzy subspaces. This approach is natural and worth to be investigated. Another approach, which is proposed in this paper, is to simultaneously utilize the fuzzy rules corresponding to coarse fuzzy partitions in addition to those corresponding to fine fuzzy partitions. Even if some fuzzy rules corresponding to fine fuzzy partitions can not be constructed, coarse fuzzy rules may be generated from the given data. In the proposed method, the fuzzy rules corresponding to various fuzzy partitions are simultaneously utilized in fuzzy inference. In other words, various fuzzy rule tables with different fuzzy partitions are superimposed. We call a set of fuzzy rules from more than one fuzzy rule table 'distributed fuzzy rules' because the knowledge derived from numerical data is distributed over different fuzzy rule tables.

Correspondence to: Dr. H. Ishibuchi, Department of Industrial Engineering, University of Osaka Prefecture, 1-1 Gakuencho, Sakai, Osaka 593, Japan.

A similar idea of using many rule tables has already been proposed in CMAC by Albus [1]. In CMAC, not fuzzy but crisp partitions of an input space are employed and each subspace has basically the same size. For the discussions about distributed representation, see Hinton et al. [4].

In this paper, we first introduce the concept of distributed representation of fuzzy rules. Next we propose a method to derive fuzzy rules from numerical data for two-group classification problems. The proposed method is extended to the case of multi-group classification problems. A fuzzy inference method for classifying unknown samples using the derived fuzzy rules is also proposed. The classification power of distributed fuzzy rules is demonstrated by computer simulations for numerical examples.

2. Distributed representation of fuzzy rules

When we derive fuzzy rules from numerical data, it is important to determine an appropriate fuzzy partition since the performance of the derived fuzzy rules is affected by the fuzzy partition. The determination of fuzzy partition, however, is not easy since there is no established general method. In this paper, in order to obtain robust performance of fuzzy rules with respect to the fuzzy partition, we propose a simple method based on the concept of distributed representation of fuzzy rules.

Let us assume that the pattern space is the unit square $[0, 1] \times [0, 1]$. We also assume that the same fuzzy partition $\{A_1^L, A_2^L, \ldots, A_L^L\}$ is used for each of the two axes in the pattern space where L is the number of fuzzy subspaces in each axis. In this case, ordinary fuzzy rules can be represented as follows:

$$\text{If } x_1 \text{ is } A_i^L \text{ and } x_2 \text{ is } A_j^L \text{ then } \ldots, \quad i, j = 1, 2, \ldots, L. \tag{1}$$

The total number of fuzzy rules in (1) is L^2. In Figure 1, we show structures of fuzzy rule tables corresponding to $L = 2$ and $L = 3$.

On the other hand, we simultaneously use $L - 1$ fuzzy rule tables in the proposed method. That is, in the case of distributed fuzzy rules, we use fuzzy rules corresponding to $L - 1$ fuzzy partitions: $\{A_1^2, A_2^2\}$, $\{A_1^3, A_2^3, A_3^3\}$, $\ldots$, $\{A_1^L, A_2^L, \ldots, A_L^L\}$. In the case of $L = 3$, the two fuzzy rule tables shown in Figure 1 are simultaneously utilized. A set of distributed fuzzy rules can be represented as

$$\text{If } x_1 \text{ is } A_i^K \text{ and } x_2 \text{ is } A_j^K \text{ then } \ldots, \quad i, j = 1, 2, \ldots, K;\ K = 2, 3, \ldots, L \tag{2}$$

where K is the number of fuzzy subspaces in each axis in each fuzzy partition. The total number of the distributed fuzzy rules in (2) is $2^2 + 3^2 + \cdots + L^2$, i.e., $\frac{1}{6}L(L+1)(2L+1) - 1$. All the fuzzy rules corresponding to the fuzzy partitions with $K = 2, 3, \ldots, L$ are simultaneously used in fuzzy inference. The total number of the distributed fuzzy rules can be reduced by removing some unnecessary rules from (2). Since such selection of fuzzy rules may require a complicated procedure, we do not discuss it here in order to maintain the simplicity of our approach.

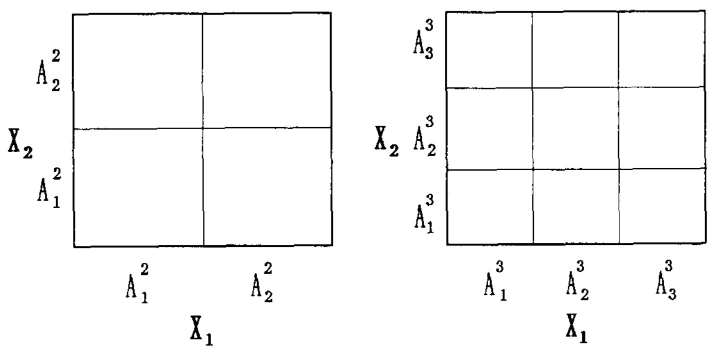

Fig. 1. Structures of fuzzy rule tables corresponding to $L = 2$ and $L = 3$.

3. Pattern classification using fuzzy rules

3.1. Derivation of fuzzy rules

Suppose that m patterns $\boldsymbol{x}_p = (x_{1p}, x_{2p})$, $p = 1, 2, \ldots, m$, are given as the training patterns in the pattern space $[0, 1] \times [0, 1]$ from two classes: $G1$ and $G2$. That is, $\boldsymbol{x}_p$ belongs to either $G1$ or $G2$. Our problem is to derive fuzzy rules from these data.

Let us assume that each axis of the pattern space is partitioned into K fuzzy sets $\{A_1^K, A_2^K, \ldots, A_K^K\}$ where A_i^K is defined by the symmetric triangular membership function

$$\mu_i^K(x) = \max\{1 - |x - a_i^K|/b^K, 0\}, \quad i = 1, 2, \ldots, K, \tag{3}$$

where

$$a_i^K = (i-1)/(K-1), \quad i = 1, 2, \ldots, K, \tag{4}$$

$$b^K = 1/(K-1). \tag{5}$$

In this membership function, a_i^K is the center where the grade of membership is equal to 1 and b^K is the spread of the membership function. Therefore the grade of membership is positive in the open interval $(a_i^K - b^K, a_i^K + b^K)$.

The following trapezoid fuzzy sets are also employed for fuzzy partitions in this paper:

$$\mu_i^K(x) = \max\{\min\{2 - 2\,|x - a_i^K|/b^K, 1\}, 0\}, \quad i = 1, 2, \ldots, K, \tag{6}$$

where a_i^K and b^K are defined by (4) and (5), respectively. The grade of membership defined by (6) is equal to 1 in the closed interval $[a_i^K - \frac{1}{2}b^K, a_i^K + \frac{1}{2}b^K]$ and positive in the open interval $(a_i^K - b^K, a_i^K + b^K)$. In Section 4, we compare the performance of the triangular fuzzy sets in (3) with that of the trapezoid fuzzy sets in (6).

As a fuzzy rule for two-group classification problems, we use the following rule with the grade of certainty:

$$\text{If } x_{1p} \text{ is } A_i^K \text{ and } x_{2p} \text{ is } A_j^K \text{ then } \boldsymbol{x}_p \text{ belongs to } G_{ij}^K \text{ with } \mathrm{CF} = \mathrm{CF}_{ij}^K, \tag{7}$$

where G_{ij}^K is either $G1$ or $G2$ and CF is the grade of certainty of the rule. The effect of introducing the grade of certainty is examined by computer simulations in Section 4.

As a method to derive the consequent G_{ij}^K and the grade of certainty CF_{ij}^K from the given patterns $\boldsymbol{x}_p = (x_{1p}, x_{2p})$, $p = 1, 2, \ldots, m$, we propose the following procedure.

Procedure. Derivation of fuzzy rules.

(i) Calculate the sum of compatibilities of the given patterns in each class to the premise, i.e., the if-part of the fuzzy rule (7), as follows:

$$\beta_{G1} = \sum_{p \in G1} \mu_i^K(x_{1p}) \cdot \mu_j^K(x_{2p}), \tag{8}$$

$$\beta_{G2} = \sum_{p \in G2} \mu_i^K(x_{1p}) \cdot \mu_j^K(x_{2p}). \tag{9}$$

(ii) If $\beta_{G1} = \beta_{G2}$ then we do not generate the fuzzy rule corresponding to the fuzzy subspace (A_i^K, A_j^K); else G_{ij}^K is determined as follows:

$$\text{If } \beta_{G1} > \beta_{G2} \text{ then } G_{ij}^K = G1. \tag{10}$$

$$\text{If } \beta_{G1} < \beta_{G2} \text{ then } G_{ij}^K = G2. \tag{11}$$

(iii) For the case of $\beta_{G1} \neq \beta_{G2}$, CF_{ij}^K is determined as follows:

$$\mathrm{CF}_{ij}^K = |\beta_{G1} - \beta_{G2}|/(\beta_{G1} + \beta_{G2}). \tag{12}$$

In this procedure, the consequent G_{ij}^K is determined as the class which has the larger sum of compatibilities to the premise of the fuzzy rule. As is known from (12), the grade of certainty CF_{ij}^K takes a value in the interval (0, 1]. In (i) of this procedure, we use the product operator to calculate the compatibilities. The min operator can be also used for calculating the compatibilities as follows:

$$\beta_{G1} = \sum_{p \in G1} \mu_i^K(x_{1p}) \wedge \mu_j^K(x_{2p}), \tag{13}$$

$$\beta_{G2} = \sum_{p \in G2} \mu_i^K(x_{1p}) \wedge \mu_j^K(x_{2p}), \tag{14}$$

where $\wedge$ is the min operator, i.e., $a \wedge b$ represents the smaller value of a and b. By computer simulations in Section 4, we compare the performances of the product operator and the min operator.

It should be noted that the fuzzy rule can not be generated by this procedure if there is no pattern in the fuzzy subspace (A_i^K, A_j^K). That is, if $\mu_i^K(x_{1p}) \cdot \mu_j^K(x_{2p}) = 0$ for all $\boldsymbol{x}_p$, the fuzzy rule corresponding to (A_i^K, A_j^K) can not be constructed from the given patterns. Therefore, when the fuzzy partition is too fine (i.e., K is too large), there may be many fuzzy rules which can not be generated because of the lack of given patterns in the corresponding fuzzy subspaces.

3.2. *Fuzzy inference for pattern classification*

By setting $K = 2, 3, \ldots, L$ in the proposed procedure in Subsection 3.1, the following distributed fuzzy rules can be generated from the given numerical data.

$$\text{If } x_{1p} \text{ is } A_i^K \text{ and } x_{2p} \text{ is } A_j^K \text{ then } \boldsymbol{x}_p \text{ belongs to } G_{ij}^K \text{ with CF} = \text{CF}_{ij}^K,$$
$$i, j = 1, 2, \ldots, K;\ K = 2, 3, \ldots, L. \tag{15}$$

We propose the following procedure to classify an unknown pattern $\boldsymbol{x}_p = (x_{1p}, x_{2p})$ using the distributed fuzzy rules (15).

Procedure. Fuzzy inference.

(i) Calculate α_{G1} and α_{G2} as follows:

$$\alpha_{G1} = \max\{\mu_i^K(x_{1p}) \cdot \mu_j^K(x_{2p}) \cdot \text{CF}_{ij}^K \mid G_{ij}^K = G1;\ i, j = 1, 2, \ldots, K;\ K = 2, 3, \ldots, L\}, \tag{16}$$

$$\alpha_{G2} = \max\{\mu_i^K(x_{1p}) \cdot \mu_j^K(x_{2p}) \cdot \text{CF}_{ij}^K \mid G_{ij}^K = G2;\ i, j = 1, 2, \ldots, K;\ K = 2, 3, \ldots, L\}. \tag{17}$$

(ii) If $\alpha_{G1} > \alpha_{G2}$ then infer $\boldsymbol{x}_p \in G1$ with the grade of support $\alpha_{G1} - \alpha_{G2}$. If $\alpha_{G1} < \alpha_{G2}$ then infer $\boldsymbol{x}_p \in G2$ with the grade of support $\alpha_{G2} - \alpha_{G1}$. If $\alpha_{G1} = \alpha_{G2}$ then $\boldsymbol{x}_p$ can not be classified.

In this procedure, the result of the fuzzy inference is the consequent of the fuzzy rule which has the maximum product of the compatibility and the certainty. The grade of support to the result of fuzzy inference is given as a value in the interval (0, 1]. This grade can be viewed as a kind of membership function which defines the belonging of an unknown pattern $\boldsymbol{x}_p$ to the inferred class. The use of the concept of membership function in the field of pattern classification was discussed by Gu and Dubuisson [3].

In the case where the min operator is employed for the calculation of compatibilities, α_{G1} and α_{G2} are calculated in the following manner instead of (16) and (17):

$$\alpha_{G1} = \max\{[\mu_i^K(x_{1p}) \wedge \mu_j^K(x_{2p})] \cdot \text{CF}_{ij}^K \mid G_{ij}^K = G1;\ i, j = 1, 2, \ldots, K;\ K = 2, 3, \ldots, L\}, \tag{18}$$

$$\alpha_{G2} = \max\{[\mu_i^K(x_{1p}) \wedge \mu_j^K(x_{2p})] \cdot \text{CF}_{ij}^K \mid G_{ij}^K = G2;\ i, j = 1, 2, \ldots, K;\ K = 2, 3, \ldots, L\}. \tag{19}$$

It should be noted that the unknown pattern $\boldsymbol{x}_p$ can not be classified if there is no fuzzy rule which has a positive compatibility to $\boldsymbol{x}_p$. That is, if $\mu_i^K(x_{1p}) \cdot \mu_j^K(x_{2p}) = 0$ for all the fuzzy rules, the unknown pattern $\boldsymbol{x}_p$ can not be classified.

3.3. Determination of fuzzy partition

If an appropriate value of L is given, we can derive fuzzy rules from the given data using the derivation procedure in Subsection 3.1 and classify unknown samples using the fuzzy inference procedure in Subsection 3.2. In order to determine the value of L, we propose the following simple procedure.

Procedure. Determination of L.

(i) Let $L := 2$. Set the values of L_{max} and ε where L_{max} and ε are the upper bound of L and the desirable rate of correctly classified patterns.

(ii) Derive the fuzzy rules corresponding to the current value of L using the procedure in Subsection 3.1.

(iii) Classify all the given patterns using the procedure in Subsection 3.2 with the generated fuzzy rules in (ii). If the rate of correctly classified patterns is equal to or greater than the desirable rate ε then stop the procedure.

(iv) If $L = L_{max}$ then stop the procedure else let $L := L + 1$ and go to (ii).

In this procedure, L_{max} should be determined from the viewpoint of available computer memory and computation time. It is expected that a high value of ε leads to a large value of L. In computer simulations of this paper, we set $\varepsilon = 100\%$ and $L_{max} = 20$. $\varepsilon = 100\%$ requires that all the given patterns should be correctly classified.

As is shown by computer simulations in Section 4, the classification power of the derived fuzzy rules is not sensitive to the value of L in the case of the distributed fuzzy rules. Therefore the determination of L is not so critical in our approach. On the other hand, the classification power of ordinary fuzzy rules is sensitive as is also shown by computer simulations in Section 4.

4. Results of computer simulations

4.1. Test problems

As test problems, we pick up seven classification problems. In each problem, the pattern space $[0, 1] \times [0, 1]$ is divided into two classes according to the value of the following function $f(\boldsymbol{x})$, i.e., if $f(\boldsymbol{x}) \geqslant 0$ then $\boldsymbol{x}$ belongs to $G1$ else $\boldsymbol{x}$ belongs to $G2$.

Problem 1: $$f(\boldsymbol{x}) = -\tfrac{1}{4}\sin(2\pi x_1) + x_2 - 0.5. \qquad (20)$$

Problem 2: $$f(\boldsymbol{x}) = -\tfrac{1}{3}\sin(2\pi x_1) + x_2 - 0.5. \qquad (21)$$

Problem 3: $$f(\boldsymbol{x}) = -\tfrac{1}{3}\sin(2\pi x_1 - \tfrac{1}{2}\pi) + x_2 - 0.5. \qquad (22)$$

Problem 4: $$f(\boldsymbol{x}) = -|-2x_1 + 1| + x_2. \qquad (23)$$

Problem 5: $$f(\boldsymbol{x}) = (x_1 + x_2 - 1)(-x_1 + x_2). \qquad (24)$$

Problem 6: $$f(\boldsymbol{x}) = -(x_1 - 0.5)^2/0.4^2 - (x_2 - 0.5)^2/0.3^2 + 1. \qquad (25)$$

Problem 7: $$f(\boldsymbol{x}) = -(x_1 - 0.5)^2/0.15^2 + (x_2 - 0.5)^2/0.2^2 + 1. \qquad (26)$$

For each classification problem, we randomly generate 20 problem instances where each of 10 problem instances has 20 patterns in each class as the given patterns (i.e., as the training patterns) and each of the other 10 problem instances has 50 patterns in each class. In Figure 2, we show examples of generated problem instances for Problem 1. Closed circles and open circles in Figure 2 represent given patterns belonging to $G1$ and $G2$, respectively. These patterns are used for deriving fuzzy rules in computer simulations.

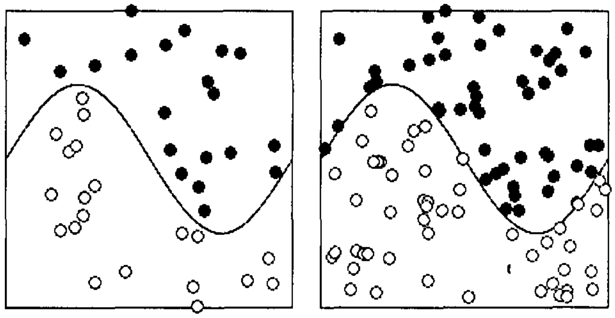

Fig. 2. Examples of problem instances of Problem 1. The numbers of given patterns are 40 and 100, respectively.

4.2. Simulation results

First we investigate the classification power of fuzzy rules under various conditions using ordinary, i.e., non-distributed fuzzy rules. More precisely, we examine the dependency of the classification power on the shape of fuzzy set and the type of operator. We also examine the effect of introducing the grade of certainty to fuzzy rules. This effect is examined by comparing the performances of the following two sets of fuzzy rules:

$$\text{If } x_{1p} \text{ is } A_i^L \text{ and } x_{2p} \text{ is } A_i^L \text{ then } \boldsymbol{x}_p \text{ belongs to } G_{ij}^L \text{ with CF} = \text{CF}_{ij}^L,\ i, j = 1, 2, \ldots, L. \tag{27}$$

$$\text{If } x_{1p} \text{ is } A_i^L \text{ and } x_{2p} \text{ is } A_j^L \text{ then } \boldsymbol{x}_p \text{ belongs to } G_{ij}^L,\ i, j = 1, 2, \ldots, L. \tag{28}$$

The fuzzy rules in (27) are ordinary fuzzy rules with the grade of certainty CF. Those in (28) are ordinary fuzzy rules without CF.

The dependency of the classification power on the shape of fuzzy set and the type of operator is also examined using the fuzzy rules in (27) and (28).

Computer simulations are performed by applying the following procedure to each of the generated 140 problem instances.

(i) Determine the value of L by the procedure in Subsection 3.3. Generate the corresponding fuzzy rules by the procedure in Subsection 3.1.

(ii) Randomly generate 100 patterns as the test patterns in the pattern space $[0, 1] \times [0, 1]$.

(iii) Classify the generated 100 patterns by the procedure in Subsection 3.2.

In computer simulations, we modify the three procedures proposed in Section 3 in order to accommodate them to the fuzzy rules in (27) or (28). For example, only L^2 fuzzy rules in (27) or (28) are used in fuzzy inference and the term CF_{ij}^K is removed from each procedure in the case of the fuzzy rules in (28).

The simulation results are summarized in Table 1. From the comparison between the first and the second rows or between the third and fourth rows, we can see that the shape of fuzzy set (triangular or trapezoid) has a slight effect on the classification power. From the comparison between the first and the fifth rows or between the third and sixth rows, we can see that the product operator slightly outperformed the min operator. On the other hand, the effect of introducing the grade of certainty CF is substantial. From the comparison between the first and the third rows, between the second and fourth rows or between the last two rows, we can see that the higher average rates of correctly classified patterns and the lower average rates of unclassifiable patterns are obtained from fuzzy rules with the grade of certainty. This means that the introduction of the grade of certainty improves the classification power of fuzzy rules.

Next we apply the following procedure to each of the generated 140 problem instances using seven fixed values of L ($L = 2, 4, 6, 8, 10, 12, 14$) in order to investigate the dependency of the performance on the choice of L.

(i) Derive fuzzy rules from the given data by the procedure in Subsection 3.1.

(ii) Randomly generate 100 patterns as the test patterns in the pattern space $[0, 1] \times [0, 1]$.

(iii) Classify the generated 100 patterns by the procedure in Subsection 3.2.

Table 1. Average rates of correctly classified patterns and unclassified patterns over 70 problem instances

Shape of fuzzy set	Type of operator	Grade of certainty	The number of patterns = 40		The number of patterns = 100	
			Correct	Unclass.	Correct	Unclass.
Triangular	Product	With CF	90.0	0.7	92.8	1.3
Trapezoid	Product	With CF	89.1	1.2	92.6	1.4
Triangular	Product	Without CF	84.8	4.1	88.9	4.3
Trapezoid	Product	Without CF	84.7	4.8	89.0	4.6
Triangular	Min	With CF	88.3	1.8	92.7	1.3
Triangular	Min	Without CF	84.2	5.0	88.8	4.7

This procedure is performed for two cases, i.e., ordinary fuzzy rules in (27) and distributed fuzzy rules in (15) to compare the performance of these two types of fuzzy rules. Simulation results are shown in Table 2 and Table 3. We also show the simulation results where L is determined by the procedure in Subsection 3.3 with $L_{\max} = 20$ and $\varepsilon = 100\%$ (see the row of $L = L^*$ in each table).

From Table 2, we can see that the performance of ordinary fuzzy rules is sensitive to the choice of L in the case of 40 patterns (e.g., the left figure in Figure 2) because the rate of unclassifiable patterns rapidly increases as the value of L increases. The increase of unclassifiable patterns is due to the increase of those fuzzy rules which can not be generated from the given data. The large value of L implies a fine fuzzy partition of the pattern space, i.e., a small fuzzy subspace for each fuzzy rule. Therefore many fuzzy rules can not be generated because of the lack of given patterns in the corresponding fuzzy subspaces as we mentioned in Subsection 3.1. As a result, when the value of L is large, some unknown patterns can not be classified because there is no fuzzy rule which is compatible to the unknown patterns. Since the number of non-generated fuzzy rules decreases as the number of given patterns increases, the average rate of unclassifiable patterns is low in the case of 100 training patterns (e.g., the right figure in Figure 2). Therefore the performance of ordinary fuzzy rules is not sensitive to L when a large number of patterns are given as the training patterns.

On the other hand, we can see from Table 3 that the performance of distributed fuzzy rules is not sensitive to the choice of L even if the number of given patterns (i.e., the number of training patterns) are small. This is because the fuzzy rules corresponding to coarse fuzzy partitions are simultaneously utilized in addition to those corresponding to fine fuzzy partitions. The robustness of the performance with respect to the fuzzy partition is one advantage of distributed fuzzy rules. Since the performance is not sensitive to L, the determination of L is not serious in the case of distributed fuzzy rules. From the

Table 2. Average rates of correctly classified patterns and unclassifiable patterns by ordinary fuzzy rules. L^* is determined by the proposed procedure in Subsection 3.3

L	The number of given patterns = 40		The number of given patterns = 100	
	Correct	Unclass.	Correct	Unclass.
2	69.6	0.0	68.4	0.0
4	87.2	0.0	90.2	0.0
6	89.8	0.1	92.6	0.0
8	89.5	0.6	94.0	0.0
10	87.4	2.9	93.6	0.0
12	83.0	7.7	93.1	0.4
14	77.5	14.7	92.4	1.2
L^*	90.0	0.7	92.8	1.3

Table 3. Average rates of correctly classified patterns and unclassifiable patterns by distributed fuzzy rules. L^* is determined by the proposed procedure in Subsection 3.3

	The number of given patterns = 40		The number of given patterns = 100	
L	Correct	Unclass.	Correct	Unclass.
2	69.6	0.0	68.4	0.0
4	88.6	0.0	91.2	0.0
6	90.3	0.0	93.3	0.0
8	90.8	0.0	94.6	0.0
10	90.5	0.0	94.4	0.0
12	90.3	0.0	94.2	0.0
14	90.2	0.0	94.2	0.0
L^*	91.0	0.0	94.5	0.0

comparison between Table 2 and Table 3, we can see that distributed fuzzy rules outperform ordinary fuzzy rules. The high performance and its robustness with respect to the value of L are the main advantages of distributed fuzzy rules over ordinary fuzzy rules.

4.3. Sensitivity to outliers

By the computer simulations in Subsection 4.2, we have already demonstrated the two advantages of distributed fuzzy rules: high performance and its robustness with respect to the choice of L. In addition to these two advantages, distributed fuzzy rules have a characteristic feature that the effect of existence of outliers is small. In this subsection, we demonstrate this feature of distributed fuzzy rules using a simple example.

Suppose that the patterns with an outlier are given as shown in Figure 3 where closed circles and open circles represent given patterns in $G1$ and $G2$, respectively. We derive ordinary fuzzy rules and distributed fuzzy rules from the given patterns. In this computer simulation, we set $L = 10$ and use the trapezoid fuzzy set and the product operator. In Figure 4 and Figure 5, we show the result of fuzzy inference using ordinary fuzzy rules and distributed fuzzy rules, respectively. The vertical axis of each figure represents the value of $\alpha_{G1} - \alpha_{G2}$ which is the grade of support to the result of fuzzy inference. The positive value of the vertical axis shows that $G1$ is inferred with the grade of support $\alpha_{G1} - \alpha_{G2}$. On the other hand, the negative value of the vertical axis shows that $G2$ is inferred with the grade of support $\alpha_{G2} - \alpha_{G1}$, i.e., the absolute value of the vertical axis. The two extreme values 1 and -1 represent that $G1$ and $G2$ are inferred with the full grade of support, respectively. Large absolute values of the vertical axis mean the strong support to the results of fuzzy inference. On the other hand, small absolute values mean a weak support.

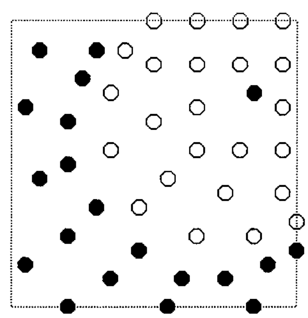

Fig. 3. Example of a classification problem with an outlier.

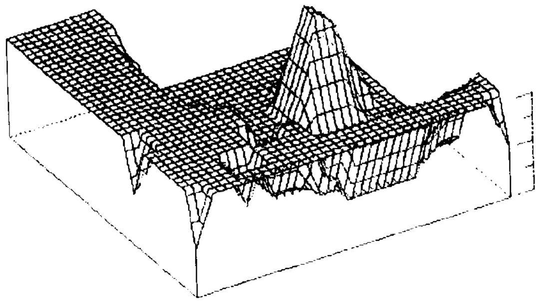

Fig. 4. Classification result using ordinary fuzzy rules.

From Figure 4, we can see that the ordinary fuzzy rules derived from the patterns in Figure 3 infer $G1$ with high grades of support around the outlier since the height of the peak in Figure 4 is about 1. On the other hand, from Figure 5, we can see that the distributed fuzzy rules do not infer $G1$ with high grades of support around the outlier since the height of the peak in Figure 5 is about 0. From these two figures, we can see that the result of fuzzy inference with distributed fuzzy rules is less sensitive to the existence of outliers than that with ordinary fuzzy rules.

5. Extension to multi-group classification problem

Our approach can be easily extended to multi-group classification problems with slight modifications. Let us assume that m patterns $\boldsymbol{x}_p, p = 1, 2, \ldots, m,$ are given as the training patterns in the pattern space $[0, 1] \times [0, 1]$ from M classes: $G1, G2, \ldots, GM$. That is, $\boldsymbol{x}_p$ belongs to one of the M classes.

The procedure for deriving fuzzy rules proposed in Subsection 3.1 is modified as follows:

Procedure. Derivation of fuzzy rules.

(i) Calculate the sum of compatibilities of the given patterns in each class to the premise, i.e., the if-part of the fuzzy rule (7), as follows:

$$\beta_{Gt} = \sum_{p \in Gt} \mu_i^K(x_{1p}) \cdot \mu_j^K(x_{2p}), \quad t = 1, 2, \ldots, M. \tag{29}$$

(ii) Find the class GX which satisfies the following condition.

$$\beta_{GX} = \max\{\beta_{G1}, \beta_{G2}, \ldots, \beta_{GM}\}. \tag{30}$$

If two or more classes take the maximum value in (30), then we do not generate the fuzzy rule corresponding to the fuzzy subspace (A_i^K, A_j^K); else G_{ij}^K is determined as GX in (30).

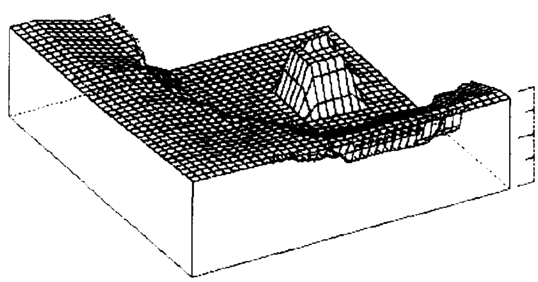

Fig. 5. Classification result using distributed fuzzy rules.

(iii) For the case where a unique class takes the maximum value in (30), CF_{ij}^K is determined as follows.

$$\mathrm{CF}_{ij}^K = |\beta_{GX} - \beta| \Big/ \sum_{t=1}^{M} \beta_{Gt} \tag{31}$$

where

$$\beta = \sum_{Gt \neq GX} \beta_{Gt}/(M-1). \tag{32}$$

In this procedure, the consequent G_{ij}^K is determined as the class which has the largest sum of compatibilities to the premise of the fuzzy rule. The grade of certainty CF_{ij}^K takes a value in the interval (0, 1]. It should be noted that this procedure coincides with the procedure in Subsection 3.1 in the case of $M = 2$.

The procedure for classifying an unknown pattern $\boldsymbol{x}_p$ proposed in Subsection 3.2 is modified as follows:

Procedure. Fuzzy inference.

(i) Calculate for $t = 1, 2, \ldots, M$,

$$\alpha_{Gt} = \max\{\mu_i^K(x_{1p}) \cdot \mu_j^K(x_{2p}) \cdot \mathrm{CF}_{ij}^K \mid G_{ij}^K = Gt;\ i, j = 1, 2, \ldots, K;\ K = 2, 3, \ldots, L\}. \tag{33}$$

(ii) Find the class GX which satisfies

$$\alpha_{GX} = \max\{\alpha_{G1}, \alpha_{G2}, \ldots, \alpha_{GM}\}. \tag{34}$$

If two or more classes take the maximum value in (34) then $\boldsymbol{x}_p$ can not be classified; else infer $\boldsymbol{x}_p \in GX$ with the following grade of support:

$$\text{Grade} = |\alpha_{GX} - \alpha| \Big/ \sum_{t=1}^{M} \alpha_{Gt} \tag{35}$$

where

$$\alpha = \sum_{Gt \neq GX} \alpha_{Gt}/(M-1). \tag{36}$$

In this procedure, the result of the fuzzy inference is the consequent of the fuzzy rule which has the maximum product of the compatibility and the certainty. The grade of support to the result of the fuzzy inference is given as a value in the interval (0, 1]. It should be noted that this procedure coincides with the procedure in Subsection 3.2 in the case of $M = 2$.

The procedure for determining the value of L in Subsection 3.3 can be applied to multi-group classification problems by using the two procedures in this section instead of those in Section 3.

We applied the extended procedures in this section to the three-group classification problem in Fisher [2]. There, four attribute values (sepal length, sepal width, petal length, petal width) are given to each of 150 samples from the three classes: 50 samples from Iris setosa, 50 samples from Iris versicolor and 50 samples from Iris virginica. Since each sample has four attribute values, this problem is a three-group classification problem in a four-dimensional pattern space. We apply the proposed procedures to this problem where our procedures are slightly modified to deal with the four-dimensional pattern space.

The following procedure is applied to Fisher's problem for each value of N: $N = 9$, 15, 21, 30, 60, 90, where N is the number of samples used for deriving fuzzy rules, i.e., N is the number of training patterns.

(i) Randomly pick up $\frac{1}{3}N$ samples from each of the three classes.
(ii) Determine the value of L using the N samples picked up in (i).
(iii) Derive fuzzy rules corresponding to the value of L in (ii) from the N samples.
(iv) Classify the remaining $150 - N$ samples using the derived fuzzy rules in (iii).

Table 4. Average rates of correctly classified samples and unclassifiable samples over 20 iterations

The number of samples used for deriving rules	Ordinary fuzzy rules		Distributed fuzzy rules	
	Correct	Unclass.	Correct	Unclass.
9	88.8	0.5	90.7	0.0
15	90.9	0.8	92.3	0.0
21	91.5	0.4	92.4	0.0
30	92.4	1.8	93.8	0.0
60	93.2	1.6	95.0	0.1
90	94.8	0.9	95.7	0.0

This procedure is iterated 20 times for each value of N. The simulation results are summarized in Table 4. Table 4 shows the average rates of correctly classified samples and unclassifiable samples in (iv) over 20 iterations. It is shown from this table that distributed fuzzy rules outperform ordinary fuzzy rules. It should be noted that the high average rates of correctly classified samples are obtained even when the number of samples used for deriving fuzzy rules is very small (e.g., 90.7% in the case of the distributed fuzzy rules derived from only 9 samples). This table also shows that the rate of correctly classified samples increases as the number of samples used for deriving fuzzy rules increases.

6. Concluding remarks

In this paper, we introduced the concept of distributed representation of fuzzy rules and applied it to classification problems. For two-group classification problems, we proposed three procedures to derive fuzzy rules from the given data, to classify unknown patterns by the derived fuzzy rules and to determine an appropriate fuzzy partition. These three procedures were extended to multi-group classification problems. One advantage of our approach is its simplicity, that is, it requires neither time-consuming iterative computations nor complicated procedures.

The classification power of distributed fuzzy rules was demonstrated by computer simulations, i.e., we showed that distributed fuzzy rules outperformed ordinary fuzzy rules. The robustness of the performance of distributed fuzzy rules with respect to the choice of fuzzy partition was also shown by computer simulations. In addition to the high classification power and its robustness with respect to the fuzzy partition, we demonstrated that the result of fuzzy inference with distributed fuzzy rules is less sensitive to the existence of outliers than that with ordinary fuzzy rules.

Distributed representation of fuzzy rules is a general concept which is not restricted within the field of pattern classification. Therefore it can be applied to various fuzzy systems based on fuzzy rules. The proposed method for distributed representation is very simple in the sense that all fuzzy rules in all rule tables corresponding to $K = 2, 3, \ldots, L$ are employed. Therefore our method can be sophisticated by removing unnecessary rules. Such improvement of our method is left for future study.

References

[1] J.S. Albus, A new approach to manipulator control: the cerebellar model articulation controller (CMAC), *Trans. ASME Ser. G, J. Dynamic Systems, Measurement and Control* **97** (1975) 220–227.

[2] R.A. Fisher, The use of multiple measurements in taxonomic problems, *Ann. of Eugenics* **7** (1936) 179–188.

[3] T. Gu and B. Dubuisson, Similarity of classes and fuzzy clustering, *Fuzzy Sets and Systems* **34** (1990) 213–221.

[4] G.E. Hinton, J.L. McClelland and D.E. Rumelhart, Distributed representations, in: D.E. Rumelhart, J.L. McLelland and the PDP Research Group, *Parallel Distributed Processing 1* (MIT Press, Cambridge, 1986).

[5] E.H. Mamdani, Applications of fuzzy algorithms for control of simple dynamic plant, *Proc. IEE* **121** (1974) 1585–1588.
[6] M. Sugeno, An introductory survey of fuzzy control, *Inform. Sci.* **36** (1985) 59–83.
[7] H. Takagi and I. Hayashi, NN-driven fuzzy reasoning, *Internat. J. Approximate Reasoning* **5** (1991) 191–212.
[8] T. Takagi and M. Sugeno, Fuzzy identification of systems and its application to modelling and control, *IEEE Trans. Systems Man Cybernet.* **15** (1985) 116–132.