Selecting Fuzzy Rules with Forgetting in Fuzzy Classification Systems

Ken NOZAKI*, Hisao ISHIBUCHI and Hideo TANAKA
Department of Industrial Engineering, University of Osaka Prefecture
1-1 Gakuen-cho, Sakai, Osaka 593, JAPAN

Abstract

This paper proposes a rule selection method with the destructive learning algorithm to construct a compact fuzzy classification system with high performance. In this paper, first we construct a fuzzy classification system by generating fuzzy rules from numerical data, and consider the fuzzy classification system based on fuzzy rules a network. Then we select significant fuzzy rules from the rule set by the proposed method which can remove unnecessary fuzzy rules. We demonstrate the effectiveness of the proposed method by applying it to the classification problem of the iris data of Fisher.

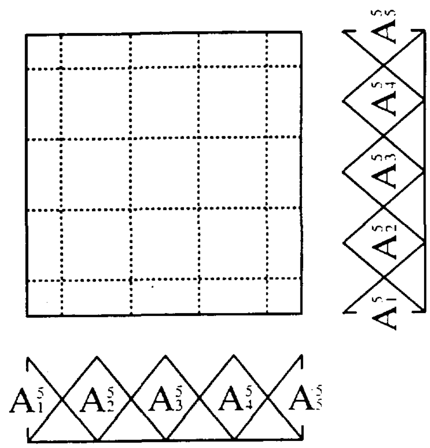

Figure 1: Simple fuzzy grid

1 Introduction

There have been many approaches to the generation of fuzzy rules from numerical data to construct fuzzy control systems based on fuzzy rules [1][7]. We have already proposed a rule generation method from numerical data for fuzzy classification problems [2][6]. In our former works, we have used fuzzy rules based on fuzzy partitions by simple fuzzy grids. Fig.1 illustrates an example of fuzzy partitions by simple fuzzy grids in a two-dimensional pattern space. In this case, a pattern space is divided into the fuzzy subspaces by the simple fuzzy grids and then a fuzzy rules corresponding to each fuzzy subspace is generated. When we use a fuzzy classification system based on simple fuzzy grids, the classification performance directly depends on fuzzy partitions, i.e., how the pattern space is divided into the fuzzy subspaces by a simple fuzzy grid.

To reduce the influence of fuzzy partitions on the classification performance, we also have proposed a fuzzy classification method based on distributed fuzzy rules [2] which simultaneously employs several fuzzy partitions of the different sizes in a fuzzy classification system. However, this method has a drawback such that it may have many fuzzy rules when applied to high dimensional classification problems. When a structure of neural networks is constructed, the above-mentioned problem has also existed [4]. That is, when the network structure is too small, it cannot be sufficiently learned. On the contrary, when the network structure is too large, the generalization power may be poor because it would be over-fitting to training patterns.

There has been the destructive learning algorithm which can find an appropriate network structure by sequentially removing unnecessary units and links from an initial network. In this paper, we consider a fuzzy classification system based on fuzzy rules a network, and propose a rule selection method with forgetting [4] for constructing a fuzzy classification system with an appropriate network structure. The proposed method can construct a compact fuzzy classification system with high performance by forgetting unnecessary fuzzy rules that have no effect on the classification of the given patterns.

*This work was done while the author was with I.B.Turksen's Expert Systems Lab. MRCO at University of Toronto, Toronto, Ontario, M5S 1A4, Canada.

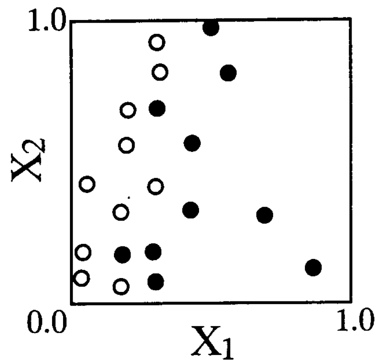

Figure 2: Two-class classification problem

2 Fuzzy Classification Systems

In this section, we briefly describe a rule generation method from numerical data proposed in Ishibuchi *et al.* [2][6]. We also explain the fuzzy classification system based on distributed fuzzy rules [2].

Let us consider a classification problem in the two dimensional pattern space, for enhancing graphical illustration. It is assumed that m patterns $\boldsymbol{x}_p = (x_{p1}, x_{p2}), p = 1, 2, \cdots, m$ are given as training data from M classes ($C1$: Class 1,$C2$: Class 2,$\cdots$,CM: Class M). Fig.2 shows an example of the classification problem of this kind (In Fig.2, $M = 2$ and $m = 20$). Our aim here is to generate the classification rules from the given training data to divide the pattern space into M disjoint subspaces.

Let R_{ij}^K be the label of the fuzzy rule corresponding to the fuzzy subspace $A_i^K \times A_j^K$. As in our former works [2][6], the fuzzy rule R_{ij}^K for two-dimensional classification problems can be written as follows.

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

where K is the number of the fuzzy subsets on each axis of the pattern space (see Fig.1), the consequent C_{ij}^K is one of M classes and CF_{ij}^K is the certainty of the fuzzy rule R_{ij}^K.

The consequent C_{ij}^K and the certainty CF_{ij}^K of each rule can be determined by the following procedure.

Procedure 1: Generation of fuzzy rules .

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

$$\beta_{CT} = \sum_{\boldsymbol{x}_p \in CT} \mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}), \tag{2}$$

where $\mu_i^K(\cdot)$ and $\mu_j^K(\cdot)$ are the membership functions of A_i^K and A_j^K, respectively.

(ii) Find Class $X(CX)$ such that

$$\beta_{CX} = \max\{\beta_{C1}, \beta_{C2}, \cdots, \beta_{CM}\}. \tag{3}$$

When plural classes take the maximum value in (3), let C_{ij}^K be ϕ to denote that R_{ij}^K is a dummy rule which has no effect on the classification of $\boldsymbol{x}_p$ and the procedure is completed. Otherwise, C_{ij}^K is determined as CX in (3).

(iii) CF_{ij}^K is determined as

$$CF_{ij}^K = \frac{\beta_{CX} - \sum_{CT \neq CX} \beta_{CT}/(M-1)}{\sum_{T=1}^{M} \beta_{CT}}. \tag{4}$$

Let us consider the two-class classification problem shown in Fig.2 where closed circles and open circles denote the patterns in Class 1 and Class 2, respectively. Fig.3 shows the generated fuzzy rules by Procedure 1 with $K = 2 \sim 7$. In Fig.3, hatched areas and dotted areas describe the consequent Class 1 and Class 2 of the generated fuzzy rules, respectively.

Let us denote the set of the generated fuzzy rules by S_R:

$$S_R = \{R_{ij}^K | i = 1, 2, \cdots, K; j = 1, 2, \cdots, K\}. \tag{5}$$

When using the fuzzy rules in S_R, we can classify a new pattern $\boldsymbol{x}_p = (x_{p1}, x_{p2})$ by the following procedure.

Procedure 2: Classification of a new pattern

.

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

$$\alpha_{CT} = \max\{\mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}) \cdot CF_{ij}^K \,|\, C_{ij}^K = CT; R_{ij}^K \in S_R\}. \tag{6}$$

(ii) Find Class X(CX) maximizing α_{CX} as

$$\alpha_{CX} = \max\{\alpha_{C1}, \alpha_{C2}, \cdots, \alpha_{CM}\}. \tag{7}$$

When several classes take the maximum value in (7), the classification of $\boldsymbol{x}_p$ is rejected. Otherwise, assign $\boldsymbol{x}_p$ to Class X determined by (7).

Fig.4 shows the classification results by the fuzzy rules in Fig.3. In Fig.4, the curves denote the boundaries between Class 1 and Class 2. It can be seen from Fig.4 that all the training patterns are correctly classified by the fuzzy rules for $K = 6$ and $K = 7$.

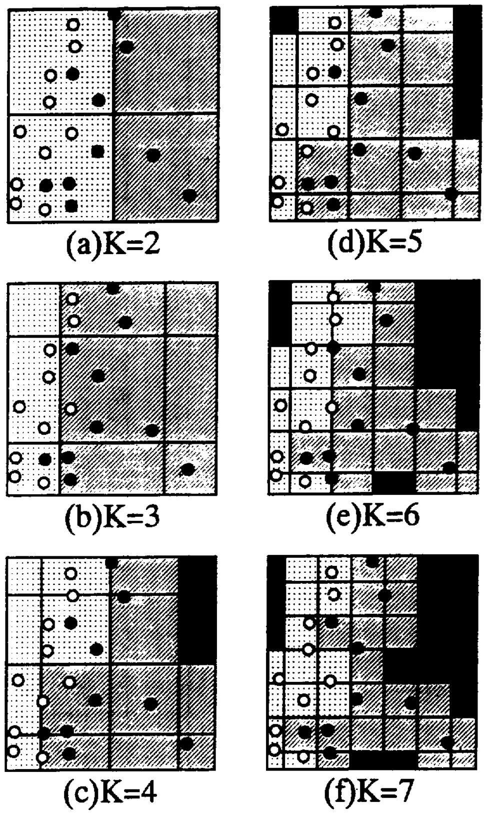

Figure 3: Generated fuzzy rules

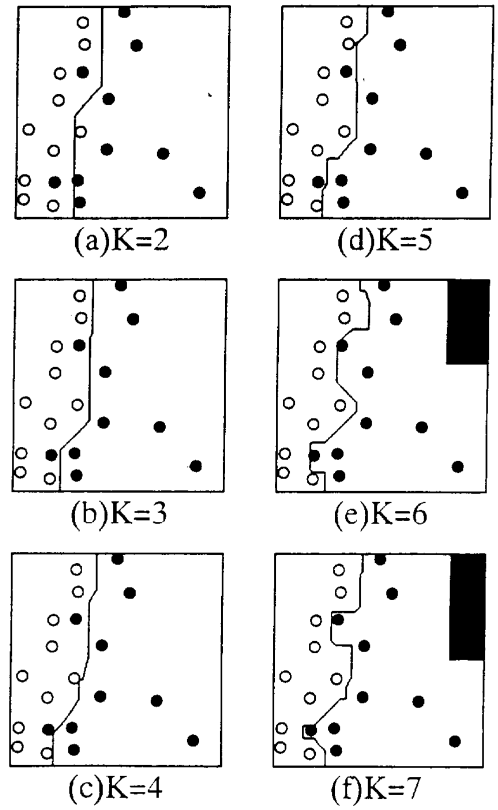

Figure 4: Classification results

We can see from Fig.4 that it is difficult to determine an appropriate fuzzy partition when fuzzy rules based on simple fuzzy grids are employed. To cope with this difficulty, we have proposed the fuzzy classification system based on distributed fuzzy rules [2] where simultaneously employs fuzzy rules corresponding to several fuzzy partitions of the different sizes. Fig.5 shows the classification result by using all the fuzzy rules corresponding to the fuzzy partitions for $K = 2 \sim 7$. In Fig.5, we can get the sufficient classification result such that there is no unclassifiable area. However, this approach has a drawback such that there may be many fuzzy rules in the fuzzy classification system, especially in high-dimensional classification problems.

3 Rule Selection with Forgetting

3.1 Forgetting algorithm

As mentioned in Section 1, the destructive learning method [4] has been proposed for constructing an appropriate neural network structure. We propose a rule selection method based on the concept of forgetting [4] in the destructive learning method to construct a compact fuzzy classification system. In the destructive learning method of neural networks, a weight of each link of the network is iteratively reduced after each learning phase of the network. That is, the weights which are not increased during learning phase are just reduced at each learning phase. Therefore we can construct an appropriate network structure by such an operation that significant links are survived and unnecessary links are forgotten, i.e., removed.

Fig.6 illustrates a fuzzy classification system mentioned in Section 2 as a network structure. The fuzzy

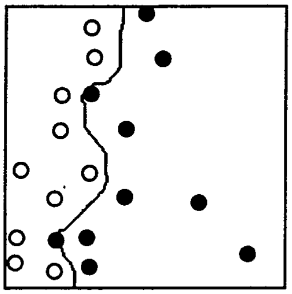

Figure 5: Classification result by distributed fuzzy rules

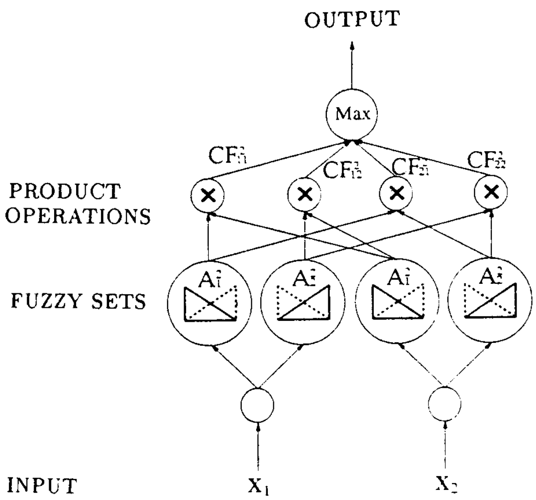

Figure 6: Network structure of fuzzy classification system

classification system in Fig.6 has four fuzzy rules corresponding to the fuzzy partition for $K = 2$. It can be seen from Fig.6 that we are able to use certainties of fuzzy rules instead of weights of neural networks in the destructive learning algorithm. Therefore we apply the concept of forgetting to certainties of fuzzy rules generated by Procedure 1 in Section 2 to select an appropriate rule set S from all the generated fuzzy rules. We propose the following procedure for selecting an appropriate fuzzy set S.

Rule selection procedure with forgetting :

(i) Set the number of iterations of the procedure at $t := 1$. Specify the maximum number of iterations $t_{\max}$ as a stopping condition and an initial rule set S.

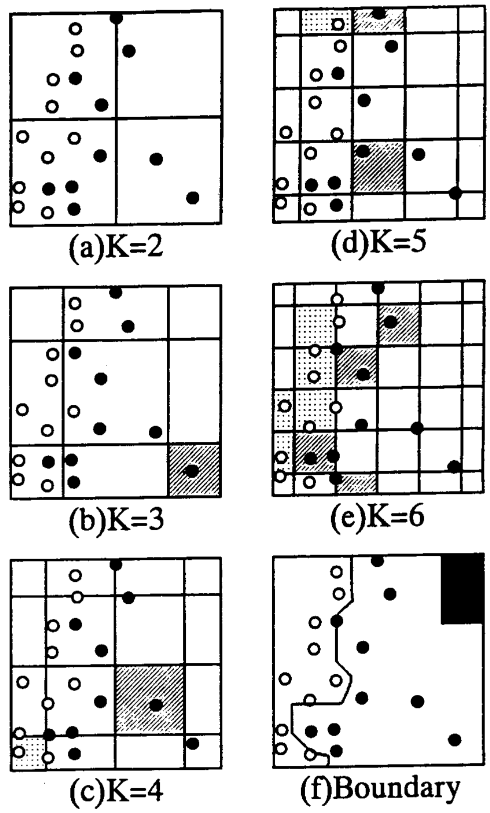

Figure 7: Classification result of the proposed method

(ii) For $\boldsymbol{x}_p$ ($p = 1, 2, \cdots, m$), find the fuzzy rule R_{ij}^K satisfying the following equation:

$$\max\{\alpha_{C1}, \alpha_{C2}, \cdots, \alpha_{CM}\} = \mu_i^K(x_{p1}) \cdot \mu_j^K(x_{p2}) \cdot CF_{ij}^K, \tag{8}$$

where α_{CT} ($T = 1, 2, \cdots, M$) is calculated by (6). When $\boldsymbol{x}_p$ is correctly classified by R_{ij}^K, adjust the certainty CF_{ij}^K of R_{ij}^K by:

$$CF_{ij}^K := CF_{ij}^K + \mu_1 \cdot (1 - CF_{ij}^K). \tag{9}$$

Conversely, when $\boldsymbol{x}_p$ is not correctly classified by R_{ij}^K, adjust the certainty CF_{ij}^K of R_{ij}^K by:

$$CF_{ij}^K := CF_{ij}^K - \mu_2 \cdot CF_{ij}^K, \tag{10}$$

where μ_1 and μ_2 are the learning constants such that $0 < \mu_1 \ll \mu_2 < 1$.

(iii) Modify the certainties of the fuzzy rules in the rule set S by:

$$CF_{ij}^K := CF_{ij}^K \cdot (1 - \gamma) \tag{11}$$

Table I: Simulation results of proposed method

	Learning	Generalization
Classification rate (%)	100.00	93.03
Error rate (%)	0.00	6.70
Number of rules	43	28.0

where γ is a forgetting rate. We call this modification 'forgetting' in this paper.

(iv) When the certainty CF_{ij}^K is less than the pre-specified threshold θ, remove the fuzzy rule R_{ij}^K from the rule set S.

(v) When $t = t_{\max}$, stop the procedure. Otherwise, let $t := t + 1$ and return to (ii).

To illustrate the proposed method, we applied it to the two-class classification problem shown in Fig.2 with the following parameter specifications: $t_{\max} = 10000$, $\eta_1 = 0.001$, $\eta_2 = 0.1$, $\gamma = 0.001$ and $\theta = 0.1$. In this computer simulations, we used all the fuzzy rules corresponding to the fuzzy partitions for $K = 2 \sim 6$ shown in Fig.3(a)$\sim$(e) as the initial rules in S. Our problem here is to select an appropriate rule set S from the initial rule set, i.e., all the fuzzy rules for $K = 2 \sim 6$. Fig.7 shows the selected fuzzy rules in S and the classification result. In the comparison of the classification result of the proposed method with those of our former works described in Section 2, the proposed method can correctly classify all the given training patterns by using fewer fuzzy rules than our former works (see Figs.3 and 4).

4 Simulation Results

To illustrate the effectiveness of the proposed method, we applied it to the iris data of Fisher [5]. In the computer simulations, we investigated the learning ability and the generalization ability of the proposed method. We also applied the fuzzy classification method of our former works [2][6] described in Section 2 and the rule selection method with genetic algorithm [3] to compare with the proposed method.

4.1 Learning ability to training patterns

To investigate the learning ability to training patterns, we used all the sample as the training patterns and performed the computer simulations with the same parameter specifications in Section 3. All the fuzzy rules corresponding to the fuzzy partitions for $K = 2 \sim 6$ were used the initial rules in S. The simulation result of the proposed method is shown in the left column of Table I. To compare with the proposed method, we also show the simulation results of our former works in Table II. It should be noted that the number of the fuzzy rules in the rule set S denotes the number of the fuzzy rules in S excluding the fuzzy rules having ϕ in the consequent. It can be seen from the left column of Table I that all the training patterns are correctly classified by the selected 43 fuzzy rules. From the comparison of Table I with Table II, we can see that the proposed method outperforms our former works. Moreover, in the comparison of the simulation result of the proposed method with that of the fuzzy classification system based on distributed fuzzy rules, we can conclude that the proposed method constructs the compact classification system with high performance by removing unnecessary fuzzy rules from the initial rule set.

4.2 Generalization ability to test patterns

To examine the generalization ability to test patterns, we applied 2-fold cross validation technique [8] in the computer simulations with the same parameter specifications in the previous subsection, and summarized the simulation results which are the average of classification rates, error rates and the numbers of fuzzy rules over 10 iterations.

The simulation results of the proposed method and our former works are shown in the right column of Table I and Table III, respectively. From the comparison of the simulation result of the proposed method with that of distributed fuzzy rules, we can see that the proposed method substantially reduced the number of fuzzy rules at the cost of the slight deterioration of the performance. Moreover, in the comparison of the proposed method with GA, we can say that the proposed method has a little more fuzzy rules but a much higher classification performance than GA.

5 Conclusion

In this paper, we performed all the computer simulations with $t_{\max} = 10000$ as the number of iterations of learning to suppose that we had sufficient learning time. In fact, we could get almost same classification performance at around $t = 3000$ in the computer

Table II: Learning ability of our former works

	Simple fuzzy grid					Dist.*	GA
	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=2\sim 6$	
Classification rate (%)	67.33	94.00	92.67	96.00	98.67	95.33	99.47
Error rate (%)	32.67	6.00	7.33	4.00	1.33	4.67	0.53
Number of rules	16	62	129	190	295	692	12.6

*: Fuzzy classification system based on distributed fuzzy rules

Table III: Generalization ability of our former works

	Simple fuzzy grid					Dist.*	GA
	$K=2$	$K=3$	$K=4$	$K=5$	$K=6$	$K=2\sim 6$	
Classification rate (%)	69.27	92.43	90.03	95.27	95.57	94.30	90.67
Error rate (%)	30.73	7.57	9.97	4.67	4.27	5.70	7.20
Number of rules	16.00	58.33	115.45	165.12	242.85	597.75	10.10

*: Fuzzy classification system based on distributed fuzzy rules

simulations. We believe that the classification performance of the proposed method will be improved by investigating the parameter specifications. It will be left for the future work.

References

[1] D.G.Burkhardt and P.P.Bonissone: "Automated fuzzy knowledge base generation and tuning," Proc. of FUZZ-IEEE'92, pp.179-188 (1992)

[2] H.Ishibuchi, K.Nozaki and H.Tanaka: "Distributed representation of fuzzy rules and its application to pattern classification," Fuzzy Sets and Systems 52, pp.21-32 (1992)

[3] H.Ishibuchi, K.Nozaki, N.Yamamoto and H.Tanaka: "Genetic Operations for Rule Selection in Fuzzy Classification Systems," Proc. IFSA'93, Seuol, Korea, pp.15-18, (July 4-9, 1993)

[4] M.Ishikawa: "A Structural Connectionist Learning Algorithm with Forgetting," Journal of Japanese Society for Artificial Intelligence, Vol.5, No.5, pp.595-603 (1991) (in Japanese)

[5] R.A.Fisher: "The use of multiple measurements in taxonomic problems," Annals of Eugenics 7, pp.170-188 (1936)

[6] K.Nozaki, H.Ishibuchi and H.Tanaka: "Performance Evaluation of Fuzzy-Rule-Based Classification Methods," Proc. IFSA'93, Seuol, Korea, pp.167-170, (July 4-9, 1993)

[7] L.X.Wang and J.M.Mendel: "Generating fuzzy rules by learning from example," IEEE Trans. on Systems, Man and Cybernetics, Vol.22, No.6, pp.1414-1427 (1992)

[8] S.M.Weiss and C.A.Kulikowski: "Computer System that Learn," Morgan Kaufmann (1991)