Design of Fuzzy Classification System Using Genetic Algorithms

Ching-Chang Wong, Chia-Chong Chen, and Bo-Chen Lin

Department of Electrical Engineering, Tamkang University,
Tamsui, Taipei Hsien, Taiwan, Republic of China

E-mail: wong@ee.tku.edu.tw

Abstract-**This paper proposes a GA-based method to construct an appropriate fuzzy classification system to maximize the number of correctly classified patterns and minimize the number of fuzzy rules. In this method, a fuzzy classification system is coded as an individual in the GA. A fitness function is defined such that it can guide the search procedure to select an appropriate fuzzy classification system to maximize the number of correctly classified patterns and minimize the number of fuzzy rules. Finally, a two-class classification problem is utilized to illustrate the efficiency of the proposed method.**

Keywords: Fuzzy classification, genetic algorithms

I. Introduction

For a classification problem, a fuzzy classification system partitions the input space to learn the behavior of training patterns [3,4]. Therefore, the performance of a fuzzy classification system depends on the choice of a fuzzy partition. If a fuzzy partition is too coarse, many patterns may be misclassified to result in a low performance. If a fuzzy partition is too fine, many fuzzy rules cannot be generated because of the lack of training patterns in the corresponding fuzzy subspaces. The conventional approaches for the fuzzy classification system design are that the number of fuzzy sets of each input variable must be defined in advance. Therefore, they may have redundant fuzzy sets such that a large number of rules are generated for the conventional approaches. In this paper, a GA-based method is proposed to construct an appropriate fuzzy classification system. An individual in the GA is used to determine a fuzzy partition such that the rough fuzzy sets of each input variable are obtained, then the training method [3,4] is applied to determine the parameters of the consequent part. Therefore, a fitness function can guide the search process to select an appropriate fuzzy classification system to maximize the number of correctly classified patterns and minimize the number of fuzzy rules.

This paper is organized as the following: In Section II, the structure of the considered fuzzy classification system is described. In Section III, a GA-based method is proposed to select an appropriate fuzzy classification system. In Section IV, a two-class classification problem is utilized to illustrate the efficiency of the proposed approach. Finally, Section V concludes the paper.

II. Structure Of Fuzzy Classification System

Consider an M-class classification problem with n training patterns $\{\bar{x}_t, y_t\}$, $t = 1,2,\cdots,n$, where $\bar{x}_t = (x_{t1}, x_{t2}, \cdots, x_{tm}) \in \Re^m$ is the input vector corresponding to the t-th training pattern, and $y_t \in \{1,2,\cdots,M\}$ represents that the t-th pattern belongs to the y_t-th class. In order to deal with the M-class classification, the fuzzy classification method with fuzzy rules proposed in [3,4] is applied. A rule base of a fuzzy system can be expressed as follows:

$$
\begin{aligned}
&R_{(j_1,j_2,\ldots,j_m)}: \\
&\text{If } x_1 \text{ is } A_{(1,j_1)} \text{ and } x_2 \text{ is } A_{(2,j_2)} \text{ and } \cdots \text{ and } x_m \text{ is } A_{(m,j_m)}, \\
&\text{then } \bar{x} \text{ belongs to Class } y_{(j_1,j_2,\cdots,j_m)} \text{ with } CF = CF_{(j_1,j_2,\ldots,j_m)}, \\
&j_i \in \{1,2\ldots,d_i\},\ i \in \{1,2,\ldots,m\},
\end{aligned}
\tag{1}
$$

where $\bar{x}$ denotes the input vector $(x_1, x_2, \cdots, x_m)$, d_i denotes the number of fuzzy sets for the input variable x_i, $A_{(i,j_i)}$ is the j_i-th fuzzy set of the i-th input variable x_i in the premise part, and $y_{(j_1,j_2,\cdots,j_m)} \in \{1,2,\cdots,M\}$ determines that $\bar{x}$ belongs to the $y_{(j_1,j_2,\cdots,j_m)}$-th class, and $CF_{(j_1,j_2,\ldots,j_m)}$ denotes the grade of certainty of the corresponding fuzzy rule $R_{(j_1,j_2,\ldots,j_m)}$. In this paper, the membership function of the fuzzy set $A_{(i,j_i)}$ is described by

$$
A_{(i,j_i)}(c_{(i,j_i)}, w^l_{(i,j_i)}, w^r_{(i,j_i)}; x_i) =
\begin{cases}
\exp\left(-\left(\dfrac{x_i - c_{(i,j_i)}}{w^l_{(i,j_i)}}\right)^2\right) & \text{if } x_i \le c_{(i,j_i)}, \\
\exp\left(-\left(\dfrac{x_i - c_{(i,j_i)}}{w^r_{(i,j_i)}}\right)^2\right) & \text{if } x_i > c_{(i,j_i)}.
\end{cases}
\tag{2}
$$

where $c_{(i,j_i)}$ is the center of the membership function, $w^l_{(i,j_i)}$ is the left width value of the membership function, and $w^r_{(i,j_i)}$ is the right width value of the membership function. Therefore, the shape of the membership function corresponding to the fuzzy set $A_{(i,j_i)}$ in the premise part is determined by the parameters $(c_{(i,j_i)}, w^l_{(i,j_i)}, w^r_{(i,j_i)})$. When the t-th input $(x_{t1}, x_{t2}, \cdots, x_{tm})$ is given, the firing strength of the $(j_1, j_2, \cdots, j_m)$-th rule is calculated by

$$\mu^t_{(j_1,j_2,\cdots,j_m)} = \prod_{i=1}^{m} A_{(i,j_i)}(x_{ti}). \tag{3}$$

If there are d_i fuzzy sets corresponding to the input variable x_i, the complete fuzzy system will have $\prod_{i=1}^{m} d_i$ rules. On the other hand, the parameters $y_{(j_1,j_2,\cdots,j_m)}$ and $CF_{(j_1,j_2,\ldots,j_m)}$ in the consequent part of the corresponding fuzzy rule $R_{(j_1,j_2,\ldots,j_m)}$ are determined in the following procedure.

Procedure 1- Generation of fuzzy rules

Step 1. Calculate β_{CT} of each class $T \in \{1,2,\cdots,M\}$ for the fuzzy rule $R_{(j_1,j_2,\ldots,j_m)}$ as follows:

$$\beta_{CT} = \sum_{\bar{x}_t \in CT} \prod_{i=1}^{m} A_{(i,j_i)}(x_{ti}) = \sum_{\bar{x}_t \in CT} \mu^t_{(j_1,j_2,\cdots,j_m)}. \tag{4}$$

Step 2. Determine $y_{(j_1,j_2,\cdots,j_m)}$ for the fuzzy rule $R_{(j_1,j_2,\ldots,j_m)}$ such that

$$\beta_{Cy_{(j_1,j_2,\cdots,j_m)}} = \max\{\beta_{C1}, \beta_{C2}, \cdots, \beta_{CM}\}. \tag{5}$$

Step 3. Determine the grade of certainty $CF_{(j_1,j_2,\ldots,j_m)}$ of the fuzzy rule $R_{(j_1,j_2,\ldots,j_m)}$ by

$$CF_{(j_1,j_2,\ldots,j_m)} = \left(\beta_{Cy_{(j_1,j_2,\cdots,j_m)}} - \beta\right) \Big/ \sum_{T=1}^{M} \beta_{CT}, \tag{6}$$

where

$$\beta = \sum_{\substack{T=1 \\ T \neq y_{(j_1,j_2,\cdots,j_m)}}}^{M} \beta_{CT} \Big/ (M-1). \tag{7}$$

In this procedure, the consequent parameter $y_{(j_1,j_2,\cdots,j_m)}$ is determined by the largest sum of $\mu^t_{(j_1,j_2,\cdots,j_m)}$ over the M classes. The grade of certainty $CF_{(j_1,j_2,\ldots,j_m)}$ of the fuzzy rule $R_{(j_1,j_2,\ldots,j_m)}$ has two intuitively acceptable properties:

1. If all the patterns in the fuzzy subspace $A_{(1,j_1)} \times A_{(2,j_2)} \times \cdots \times A_{(m,j_m)}$ corresponding to the fuzzy rule $R_{(j_1,j_2,\ldots,j_m)}$ belong to the $y_{(j_1,j_2,\cdots,j_m)}$-th class, then the value of the parameter $CF_{(j_1,j_2,\ldots,j_m)}$ approximates to 1. In this case, it is certain that any pattern in $A_{(1,j_1)} \times A_{(2,j_2)} \times \cdots \times A_{(m,j_m)}$ belongs to the $y_{(j_1,j_2,\cdots,j_m)}$-th class.

2. If all values of β_{CT}'s are not so different from each other, then the value of the parameter $CF_{(j_1,j_2,\ldots,j_m)}$ approximates to 0. In this case, it is uncertain that a pattern in $A_{(1,j_1)} \times A_{(2,j_2)} \times \cdots \times A_{(m,j_m)}$ belongs to the $y_{(j_1,j_2,\cdots,j_m)}$-th class.

By the above training procedure, a rule set S is determined as

$$S = \{R_{(j_1,j_2,\ldots,j_m)} \mid j_i = 1,2,\ldots,d_i, i = 1,2,\ldots,m\}. \tag{8}$$

When a rule set S is given, a new pattern $\bar{x}_t = (x_{t1}, x_{t2}, \cdots, x_{tm})$ is classified by the following procedure based on the fuzzy rules in S.

Procedure 2- Classification of a new pattern

Step 1. Calculate α_{CT} of each class $T \in \{1,2,\cdots,M\}$ as follows:

$$\alpha_{CT} = \max\{\mu^t_{(j_1,j_2,\cdots,j_m)} \cdot CF_{(j_1,j_2,\ldots,j_m)} \mid y_{(j_1,j_2,\cdots,j_m)} = T \text{ and } R_{(j_1,j_2,\ldots,j_m)} \in S\}. \tag{9}$$

Step 2. Find Class Y such that

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

where the value of Y represents the new pattern belongs to the Y-th class.

According to the above description, the parameter set containing of the premise parameters determines a fuzzy classification system. Thus, different premise parameter sets determine different fuzzy classification systems resulting in different performances. In the next section, each individual in GA is represented as a fuzzy classification system and a method based on GA is proposed to select an appropriate fuzzy classification system to maximize the number of correctly classified patterns and minimize the number of fuzzy rules.

III. Selection Of Fuzzy Classification System Using Genetic Algorithm

The genetic algorithm (GA) is a method to obtain an optimal solution by applying a theory of biological evolution [1,2]. The GA employs the Darwinian survival-of-the-fittest theory to yield the best of the characters among the population and perform a random information exchange to produce superior individual. The GA requires only the information about the fitness value determined by each parameter set. This differs from many optimization methods requiring the derivation information or the complete knowledge of the problem structure and parameter. In this paper, the GA is applied to choose an appropriate fuzzy system for data classification. In the GA approach, each individual is represented to determine a fuzzy system. The individual is used to partition the input space so that the fuzzy sets in the premise part are determined, and then the number of rules is determined from the complete concept. Furthermore, the training procedure in the preceding section is applied to determine the parameters in the consequent part corresponding to the individual. Here, a set of L individuals, P, called population, is expressed in the following:

$$P=\{p^1,p^2,\cdots,p^L\}. \tag{11}$$

Each individual is viewed as a solution candidate and expressed by the following string expression:

$$p^k=\{p^k_{11},p^k_{12},\cdots,p^k_{1b_1},p^k_{21},p^k_{22},\cdots,p^k_{2b_2},\cdots,p^k_{m1},p^k_{m2},\cdots,p^k_{mb_m}\},\quad k\in\{1,2,\cdots,L\}, \tag{12}$$

where $p^k_{ij}, i\in\{1,2,\cdots,m\}, j\in\{1,2,\cdots,b_i\}$, is a bit variable taking of either "1" or "0". We use the substring, $\{p^k_{i1},p^k_{i2},\cdots,p^k_{ib_i}\}$, to partition the i-th input space and determine the shapes of membership functions for the i-th input variable. The bit with "1" is used to describe the center position of one membership function. If the number of bits with "1" is d_i, d_i is viewed as the number of membership functions for the i-th input variable. Therefore, the rule number of the fuzzy system described by this individual p^k is $\prod_{i=1}^{m} d_i$ in the sense of the complete concept.

By means of the input vectors $\bar{x}_t=(x_{t1},x_{t2},\cdots,x_{tm})$, $t=1,2,\cdots,n$, the range $[x_i^{min},x_i^{max}]$ of the i-th input variable is determined, where $x_i^{min}=\min_{t\in\{1,2,\cdots,n\}} x_{ti}$ and $x_i^{max}=\max_{t\in\{1,2,\cdots,n\}} x_{ti}$. Therefore, the center position of each membership function with respect to the i-th variable can be determined by:

$$c_{(i,j_i)}=x_i^{min}+(I_{(i,j_i)}-1)\cdot s_i,\ j_i\in\{1,2,\cdots,d_i\}, \tag{13}$$

where $s_i=(x_i^{max}-x_i^{min})/(b_i-1)$. The notation, $I_{(i,j_i)}\in\{1,2,\cdots,b_i\}$, is defined to represent the index of the substring, $\{p^k_{i1},p^k_{i2},\cdots,p^k_{ib_i}\}$, taking a value "1" with respect to the i-th input space. In order to determine the left and right width values of the membership function of the fuzzy set $A_{(i,j_i)}$, $i\in\{1,2,\cdots,m\}$, $j_i\in\{1,2,\cdots,d_i\}$, we choose a constant $\alpha\in[0,1]$ so that the membership function values of $A_{(i,j_i)}(c_{(i,j_i-1)})$ and $A_{(i,j_i)}(c_{(i,j_i+1)})$ are the value of α. That is, from (2),

$$A_{(i,j_i)}(c_{(i,j_i-1)})=\exp\left(-\left(\frac{c_{(i,j_i-1)}-c_{(i,j_i)}}{w^l_{(i,j_i)}}\right)^2\right)=\alpha \tag{14}$$

and

$$A_{(i,j_i)}(c_{(i,j_i+1)})=\exp\left(-\left(\frac{c_{(i,j_i+1)}-c_{(i,j_i)}}{w^r_{(i,j_i)}}\right)^2\right)=\alpha \tag{15}$$

Therefore, the left and right width values of each membership function with respect to the i-th input variable can be respectively calculated by

$$w^l_{(i,j_i)}=\sqrt{\frac{(c_{(i,j_i-1)}-c_{(i,j_i)})^2}{-\ln(\alpha)}},\quad j_i\in\{1,2,\cdots,d_i\} \tag{16}$$

and

$$w^r_{(i,j_i)}=\sqrt{\frac{(c_{(i,j_i+1)}-c_{(i,j_i)})^2}{-\ln(\alpha)}},\quad j_i\in\{1,2,\cdots,d_i\}, \tag{17}$$

where $c_{(i,0)}=x_i^{min}$, $c_{(i,d_i+1)}=x_i^{max}$ and $\alpha\in[0,1]$ is a constant to control the overlapping of two adjacent membership functions. By means of the preceding process, the number and shapes of membership functions in the premise part are constructed to partition the input space, and the number of fuzzy rules is determined.

An individual determines a fuzzy classification system.

In order to construct an appropriate fuzzy classification system to maximize the number of the correctly classified patterns and minimize the number of fuzzy rules, the fitness function is defined as follows [3,4]:

$$F(p^k) = W_{NCP} \cdot NCP(S(p^k)) - W_S \cdot |S(p^k)|, \qquad (19)$$

where $S(p^k)$ is a rule set corresponding to the individual p^k, $NCP(S(p^k))$ denotes the number of correctly classified patterns by $S(p^k)$, $|S(p^k)|$ denotes the number of fuzzy rules by $S(p^k)$, W_{NCP} is the weight of $NCP(S(p^k))$, and W_S is the weight of $|S(p^k)|$. In general, the classification power of a classification system is more important than its compactness. Therefore, the weights should be specified as $0 < W_S << W_{NCP}$. In this way, the fitness function will guide the individual to find an appropriate fuzzy classification system to satisfy the desired objective. That is, the final selected fuzzy classification system using GA can simultaneously achieve the combinatorial objectives for maximizing the number of correctly classified patterns and minimize the number of fuzzy rules.

Based on the above description, the procedure can be described as follows:

Step 1: Set the number of individuals (L), the length of bits for each input variable (b_i), the crossover probability (p_c), the mutation probability (p_m), the overlapping control constant (α), the number of generations (N), the weight W_{NCP} for $NCP(S(p^k))$, and the weight W_S for $|S(p^k)|$.

Step 2: Generate the initial individual for $G = 0$.

Step 3: Let $k = 1$.

Step 4: Determine the premise part of the fuzzy classification system from the k-th individual and determine the consequent part of the fuzzy classification system for the corresponding individual by the training procedure in Section II.

Step 5: Calculate the fitness value for the k-th individual by (19).

Step 6: If $k = L$, then go to Step 7; otherwise let $k = k + 1$ and go to Step 4.

Step 7: If $G = N$, then go to Step 9; otherwise let $G = G + 1$ and go to Step 8.

Step 8: Generate new individuals by the reproduction, crossover and mutation operators, and then go to Step 3.

Step 9: Based on the individual with the best fitness value of the final result, the desired fuzzy classification system can be determined.

IV. Illustrated Example

In order to illustrate the proposed method, we consider a two-class classification problem shown in Fig. 1, where closed circles and open circles denote the patterns belonging to Class 1 and Class 2, respectively. For the classification problem, a fine fuzzy partition is required for the left half of the pattern space but a coarse fuzzy partition is appropriate for the right half. The parameter specifications for the proposed method are given in the following. The number of individuals: $L = 50$, the length of bits for the input variable x_1: $b_1 = 15$, the length of bits for the input variable x_2: $b_2 = 15$, the length of individual: $b_1 + b_2 = 30$, the crossover probability: $p_c = 0.9$, the mutation probability: $p_m = 0.1$, the overlapping control constant: $\alpha = 0.15$, the number of generations: $N = 50$, the weight $W_{NCP} = 10$, and the weight $W_S = 1$. Following the proposed method, the selected significant fuzzy sets of each input variable are shown in Fig. 2, and the selected fuzzy classification system has ten fuzzy rules shown in Table. 1. The classification result corresponding to the selected fuzzy classification system is shown in Fig. 3. From the above results, it is obvious that the proposed method can extract significant fuzzy sets of each input variable such that the number of fuzzy rules is minimized and the data patterns can be correctly classified.

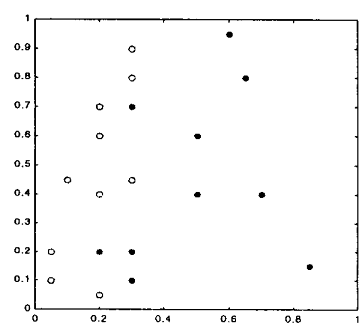

Fig. 1 A two-class classification problem

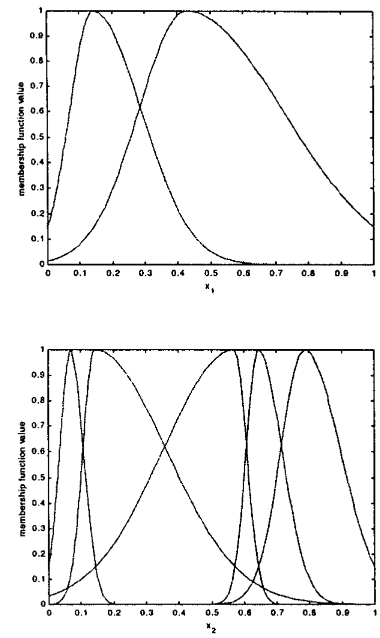

Fig. 2 The selected membership functions for each input variable

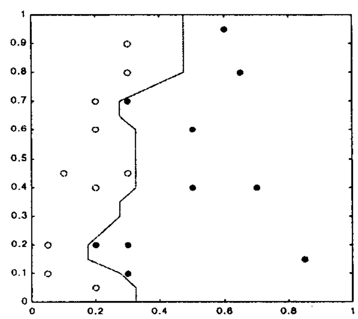

Fig. 3 The boundary corresponding to the selected fuzzy classification system

V. CONCLUSIONS

In this paper, a GA-based method was proposed to efficiently construct a fuzzy system for data classification. An individual in the GA is organized to represent a fuzzy system. A method based on GA is proposed such that the selected fuzzy system can maximize the number of correctly classified patterns and minimize the rule number of the selected fuzzy system. In the simulation, it was shown that the proposed method has the desired objective.

ACKNOWLEDGMENT

This research was supported in part by the National Science Council of the Republic of China under contract NSC-88-2213-E-032-024.

REFERENCES

[1] L. Davis (Ed.), Handbook of Genetic Algorithms, (Van Nostrand Reinhold, New York, 1991).

[2] D.E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning. (Addison-Wesley, Reading, MA, 1989).

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

[4] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Selecting fuzzy if-then rules for classification problems using genetic algorithms," IEEE Trans. on Fuzzy Systems, Vol. 3, No.3, pp. 260-270, 1995.

TABLE I
FUZZY RULES OF THE FUZZY CLASSIFICATION SYSTEM

Rule	Input variable x_1			Input variable x_2			Class	Grade of certainty
	Center position	Left width	Right width	Center position	Left width	Right width		
1	0.1429	0.1037	0.2074	0.0714	0.0519	0.0519	1	0.4532
2	0.1429	0.1037	0.2074	0.1429	0.0519	0.3112	1	0.0190
3	0.1429	0.1037	0.2074	0.5714	0.3112	0.0519	1	0.6994
4	0.1429	0.1037	0.2074	0.6429	0.0519	0.1037	1	0.4649
5	0.1429	0.1037	0.2074	0.7857	0.1037	0.1556	1	0.6537
6	0.4286	0.2074	0.4149	0.0714	0.0519	0.0519	2	0.3228
7	0.4286	0.2074	0.4149	0.1429	0.0519	0.3112	2	0.6480
8	0.4286	0.2074	0.4149	0.5714	0.3112	0.0519	2	0.3390
9	0.4286	0.2074	0.4149	0.6429	0.0519	0.1037	2	0.4175
10	0.4286	0.2074	0.4149	0.7857	0.1037	0.1556	2	0.0648