Design of PSO-based Fuzzy Classification Systems

Chia-Chong Chen

Department of Electronics Engineering, Wufeng Institute of Technology, Chiayi, Taiwan 621, R.O.C.

Abstract

In this paper, a method based on the particle swarm optimization (PSO) is proposed for pattern classification to select a fuzzy classification system with an appropriate number of fuzzy rules so that the number of incorrectly classified patterns is minimized. In the PSO-based method, each individual in the population is considered to automatically generate a fuzzy classification system for an M-class classification problem. Subsequently, a fitness function is defined to guide the search procedure to select an appropriate fuzzy classification system such that the number of fuzzy rules and the number of incorrectly classified patterns are simultaneously minimized. Finally, two classification problems are utilized to illustrate the effectiveness of the proposed PSO-based approach.

Key Words: Particle Swarm Optimization, Fuzzy Classification System, Pattern Classification

1. Introduction

Fuzzy rule-based systems have been successfully applied to solve many classification problems. In many classification problems, fuzzy classification rules are derived from human experts as linguistic knowledge. Because it is not usually easy to derive fuzzy rules from human experts, many approaches have recently been proposed to generate fuzzy rules automatically from the training patterns of the considered classification problem [3−5,7,8,13]. In order to generate fuzzy rules from training patterns, fuzzy partition in the input space are generally considered to determine the premise part of a fuzzy classification system. The grid-type fuzzy partition of the input space [3,4,7,9,11,13] and the scatter-type fuzzy partition of the input data [8,10] have been often used to model fuzzy systems for training patterns. In [3], a heuristic method for generating fuzzy rules is applied to the grid-type fuzzy partitions, and a rule selection method, based on genetic algorithms, is then employed to select the relevant fuzzy rules from the generated fuzzy rules for classifying the training patterns in the considered classification problem. However, this approach to solving classification problems with high-dimensional pattern spaces has a significant shortcoming: the number of fuzzy rules becomes enormous as the number of dimensions increases and the learning time for genetic algorithms is too high. Furthermore, an adaptive grid partition in the input space is used to design the ANFIS-based fuzzy classifier in [4]. This approach takes the uniformly partitioned grid as the initial state. The grid evolves as the parameters in the premise membership functions are adjusted. However, the adaptive grid-partition scheme has two problems. First, the number of fuzzy sets for each input variable is predetermined in advance. Second, the learning complexity suffers from an exponential explosion as the number of inputs increases. In the scatter partition for constructing fuzzy classifiers, fuzzy min-max hyperbox classifiers are powerful tools for solving classification problems. Simpson [8] proposed a method for generating the hyperbox regions to construct a fuzzy classifier to the considered classification problem. In this approach to the classification problem, the learning parameter is very critical, since it directly affects the number and position of the resulting hyperboxes. Consequently, it influences the structure of the fuzzy classifier and the classification performance so that this approach suffers from a high sensitivity of the classification accu-

*Corresponding author. E-mail: ccchen@mail.wfc.edu.tw

racy with respect to the skill of the user to determine the appropriate learning parameter. Besides, the fixed learning parameter of the above approach to fuzzy min-max hyperbox classifier makes the same constraint on coverage resolution in the whole input space so that some small hyperboxes are generated even in the input space far from class boundaries. This reduces the generalization capability of the fuzzy min-max hyperbox classifier.

In this paper, a method based on the particle swarm optimization (PSO), called a PSO-based method, is implemented to select an appropriate fuzzy classification system such that a low number of training patterns are misclassified by the selected fuzzy classification system. In the PSO-based approach, each individual in the population is considered to represent a fuzzy classification system. Then, a fitness function is implemented to guide the search procedure to select an appropriate fuzzy classification system such that the number of fuzzy rules and the number of incorrectly classified patterns are simultaneously minimized.

The rest of this paper is organized as follows. Section 2 describes the structure of the fuzzy classification system. Section 3 proposes a PSO-based method to select an appropriate fuzzy classification system for the considered classification problem. Section 4 considers classification problems of a synthetic data set and the Iris data set to illustrate the learning ability and the generalization ability of the proposed approach, respectively. Finally, Section 5 draws conclusions about the proposed approach to solving the classification problem.

2. Structure of Fuzzy Classification Systems

When an M-class classification problem is considered, a rule base of a fuzzy classification system can be expressed as follows [3]:

$$\begin{aligned}&j-th\ \ rule:\\&\quad If\ \ x_1\ \ is\ \ A_{j1}\ \ and\ \ x_2\ \ is\ \ A_{j2}\ \ and\ \cdots\ and\ \ x_m\ \ is\ \ A_{jm},\\&\quad Then\ \ \underline{x}=(x_1,x_2,\cdots,x_m)\ \ belongs\ to\ Class\ \ H_j\ \ with\ \ CF=CF_j,\\&\quad j=1,2,\cdots,R,\end{aligned} \tag{1}$$

where R is the number of fuzzy rules in the rule base, A_{ji}, $i = 1,2,\cdots, m$, are the premise fuzzy sets of the *j-th* fuzzy rule, $H_j \in \{1,2,\cdots, M\}$, is the consequent class output of the *j-th* fuzzy rule, and $CF_j \in [0,1]$ is the grade of certainty of the *j-th* fuzzy rule. In this paper, the membership function of the fuzzy set A_{ji} is described by

$$\mu_{A_{ji}}(m_{(ji,1)}, m_{(ji,2)}, m_{(ji,3)}; x_i) = \begin{cases} \exp\left(-\left(\dfrac{x_i - m_{(ji,1)}}{m_{(ji,2)}}\right)^2\right), & if\ \ x_i \le m_{(ji,1)}, \\ \exp\left(-\left(\dfrac{x_i - m_{(ji,1)}}{m_{(ji,3)}}\right)^2\right), & if\ \ x_i > m_{(ji,1)}, \end{cases} \tag{2}$$

where $m_{(ji,1)}$, $m_{(ji,2)}$ and $m_{(ji,3)}$ determine the center position, the left and right width values of the membership function, respectively. Hence, the shape of the membership function is determined by a parameter vector $\underline{m}_{ji} = [m_{(ji,1)}\ m_{(ji,2)}\ m_{(ji,3)}]$. The j-th fuzzy rule in the rule base is determined by a parameter vector $\underline{r}_j = [\underline{m}_{j1}\ \underline{m}_{j2} \cdots \underline{m}_{jM}]$. Consequently, the set of parameters in the premise part of the rule base is defined as $\underline{r} = [\underline{r}_1\ \underline{r}_2 \cdots \underline{r}_R]$. According to (1), the set of parameters in the consequent part of the rule base is defined as $\underline{a} = [H_1\ CF_1\ H_2\ CF_2 \cdots H_R\ CF_R]$.

When the input $\underline{x} = (x_1, x_2,\cdots, x_m)$ is given, the firing strength of the premise of the *j-th* rule is calculated by $q_j(\underline{x}) = \prod_{i=1}^{M} \mu_{A_{ji}}(x_i)$. The class output of the fuzzy classification system with respect to the input $\underline{x}$ can be determined by

$$y = \arg\max_{j=1}^{R}\ q_j(\underline{x}) \cdot CF_j \tag{3}$$

According to the above description, each parameter set consisting of the premise and consequent parameters determines a fuzzy classification system. Thus, different parameter sets determine different fuzzy classification systems so that the generated fuzzy classification systems have different performances. The goal of this paper is to find an appropriate fuzzy classification system to minimize both of the number of incorrectly classified patterns and the number of fuzzy rules. In the next section, the PSO-based method is applied to select an appropriate fuzzy classification system with a low number of fuzzy rules such that the number of incorrectly classified patterns is minimized.

3. PSO-Based Fuzzy Classification Systems

The PSO is an evolutionary computation technique proposed by Kennedy and Eberhart [2,6]. Its development was based on observations of the social behavior of animals such as bird flocking, fish schooling, and swarm theory. Like the GA, the PSO is initialized with a population of random solutions. It also requires only the information about the fitness values of the individuals in the population. This differs from many optimization methods requiring the derivation information or the complete knowledge of the problem structure and parameter. Compared with the GA, the PSO has memory so that the information of good solutions is retained by all individuals. Furthermore, it has constructive cooperation between individuals, individuals in the population share information between them.

In the PSO-based method, each individual is represented to determine a fuzzy classification system. The individual is used to partition the input space so that the rule number and the premise part of the generated fuzzy classification system are determined. Subsequently, the consequent parameters of the corresponding fuzzy system are obtained by the premise fuzzy sets of the generated fuzzy classification system.

A set of L individuals, P, called population, is expressed in the following:

$$P = \begin{bmatrix} \underline{p}_1 \\ \underline{p}_2 \\ \vdots \\ \underline{p}_h \\ \vdots \\ \underline{p}_L \end{bmatrix} = \begin{bmatrix} \underline{r}_1 & \underline{g}_1 \\ \underline{r}_2 & \underline{g}_2 \\ \vdots & \vdots \\ \underline{r}_h & \underline{g}_h \\ \vdots & \vdots \\ \underline{r}_L & \underline{g}_L \end{bmatrix} \tag{4}$$

In order to evolutionarily determine the parameters of the fuzzy classification system, the individual $\underline{p}_h$ contains two parameter vectors: $\underline{r}_h$ and $\underline{g}_h$. That is, $\underline{p}_h = [\underline{r}_h\ \underline{g}_h]$. The parameter vector $\underline{r}_h = [\underline{r}_1^h\ \underline{r}_2^h \cdots \underline{r}_j^h \cdots \underline{r}_B^h]$ consists of the premise parameters of the candidate fuzzy rules, where B is a user-defined positive integer to decide the maximum number of fuzzy rules in the rule base generated by the individual $\underline{p}_h$. Here, $\underline{r}_j^h = [\underline{m}_{j1}^h\ \underline{m}_{j2}^h \cdots \underline{m}_{ji}^h \cdots \underline{m}_{jM}^h]$ is the parameter vector to determine the membership functions of the j-th fuzzy rule, where $\underline{m}_{ji}^h = [\underline{m}^h_{(ji,1)}\ \underline{m}^h_{(ji,2)}\ \underline{m}^h_{(ji,3)}]$ is the parameter vector to determine the membership function for the *i-th* input variable. The parameter vector $\underline{g}_h = [g_1^h\ g_2^h \cdots g_j^h \cdots g_B^h]$ is used to select the fuzzy rules from the candidate rules $\underline{r}_h = [\underline{r}_1^h\ \underline{r}_2^h \cdots \underline{r}_j^h \cdots \underline{r}_B^h]$ so that the fuzzy rule base is generated. $g_j^h \in [0,1]$ decides whether the *j-th* candidate rule $\underline{r}_j^h$ is added to the rule base of the generated fuzzy system or not. If $g_j^h \geq 0.5$, then the *j-th* candidate rule $\underline{r}_j^h$ is added to the rule base. Consequently, the total number of g_j^h $(j = 1,2,\cdots,B)$ whose value is greater than or equal to 0.5 is the number of fuzzy rules in the generated rule base. In order to generate the rule base, the index j of g_j^h $(j = 1,2,\cdots,B)$ whose value is greater than or equal to 0.5 is defined as $I_r^h \in \{1,2,\cdots,B\}$, $r = 1,2,\cdots,r_h$, where r_h represents the number of the fuzzy rules in the generated rule base. $\{\underline{r}^h_{I_1^h}, \underline{r}^h_{I_2^h}, \cdots, \underline{r}^h_{I_r^h}, \cdots, \underline{r}^h_{I_{r_h}^h}\}$ generates the premise part of the fuzzy rule base generated by the individual $\underline{p}_h = [\underline{r}_h\ \underline{g}_h]$. For example, assume that $\underline{r}_h$ and $\underline{g}_h$ are denoted as $[\underline{r}_1\ \underline{r}_2\ \underline{r}_3\ \underline{r}_4\ \underline{r}_5\ \underline{r}_6]$ and [0.12 0.72 0.82 0.35 0.29 0.81], respectively. According to g_h, the generated rule base has three fuzzy rules and $\{I_1^h, I_2^h, I_3^h\} = \{2,3,6\}$ so that $\{\underline{r}_2\ \underline{r}_3\ \underline{r}_6\}$ determines the premise part of the generated rule base. Consequently, the rule base of the generated fuzzy classification system is described as follows:

$$\begin{aligned} &r-th\ \ rule: \\ &\quad If\ \ x_1\ \ is\ \ A^h_{I_r^h 1}\ \ and\ \ x_2\ \ is\ \ A^h_{I_r^h 2}\ \ and\ \ \cdots\ \ and\ \ x_m\ \ is\ \ A^h_{I_r^h m}, \\ &\quad Then\ \ \underline{x} = (x_1, x_2, \cdots, x_m)\ \ belongs\ \ to\ \ Class\ \ H_r\ \ with\ \ CF = CF_r, \\ &\quad r = 1,2,\cdots,r_h, \end{aligned} \tag{5}$$

where $A^h_{I_r^h i}$, $i = 1,2,\cdots,m$, are the fuzzy sets of the generated *r-th* fuzzy rule. The membership function associated with the fuzzy set $A^h_{I_r^h i}$ is described as follows:

$$\mu_{A^h_{I_r^h i}}(m^h_{(I_r^h i,1)}, m^h_{(I_r^h i,2)}, m^h_{(I_r^h i,3)}; x_i) = \begin{cases} \exp\left(-\left(\dfrac{x_i - m^h_{(I_r^h i,1)}}{m^h_{(I_r^h i,2)}}\right)^2\right), & if\ \ x_i \leq m^h_{(I_r^h i,1)}, \\ \exp\left(-\left(\dfrac{x_i - m^h_{(I_r^h i,1)}}{m^h_{(I_r^h i,3)}}\right)^2\right), & if\ \ x_i > m^h_{(I_r^h i,1)}, \end{cases} \tag{6}$$

Consequently, the individual $\underline{p}_h$ determines the premise part of the generated fuzzy classification system. Assume that N training patterns $(\underline{x}_n, y_n)$, $n = 1,2,\cdots,N$, are gathered from the observation of the considered M-class classification problem, where $\underline{x}_n = (x_{n1}, x_{n2}, \cdots, x_{nm})$ is the input vector of the n-th training pattern and $y_n \in \{1,2,\cdots,M\}$ is the corresponding class output. In order to determine the consequent parameters H_r and CF_r of the *r-th* fuzzy rule, a procedure is described as follows [3]:

Step 1. Calculate θ_t, $t = 1,2,\cdots,M$ for the *r-th* fuzzy rule as follows:

$$\theta_t = \sum_{\underline{x}_p \in Class\ t} q_r(\underline{x}_p),\ t = 1,2,\cdots,M. \tag{7}$$

Step 2. Determine H_r for the *r-th* fuzzy rule by

$$H_r = \arg \max_{t=1}^{M} \theta_t. \tag{8}$$

Step 3. Determine the grade of certainty CF_r of the *r-th* fuzzy rule by

$$CF_r = \frac{\theta_{H_r} - \theta}{\sum_{t=1}^{M} \theta_t} \tag{9}$$

where

$$\theta = \sum_{\substack{t=1 \\ t \neq H_r}}^{M} \frac{\theta_t}{M-1}. \tag{10}$$

Thus, the consequent parameters of the generated fuzzy classification system are determined by the above procedure. According to the above description, each individual corresponds to a fuzzy classification system. In order to construct a fuzzy classification system which has an appropriate number of fuzzy rules and minimize incorrectly classified patterns simultaneously, the fitness function is defined as follows:

$$f_h = fit(\underline{p}_h) = g_1(\underline{p}_h) g_2(\underline{p}_h) \tag{11}$$

where f_h is the fitness value of the individual $\underline{p}_h$, $g_1(\underline{p}_h)$ and $g_2(\underline{p}_h)$ are defined respectively as follows:

$$g_1(\underline{p}_h) = \exp\left(-\frac{NICP(\underline{p}_h)}{\sigma_e} \right) \tag{12}$$

and

$$g_2(\underline{p}_h) = \exp\left(-\frac{r_h}{\sigma_r} \right). \tag{13}$$

Here, $NICP(\underline{p}_h)$ is the number of incorrectly classified patterns, r_h is the number of fuzzy rules in the rule base of the generated fuzzy classification system, and σ_e and σ_r are user-defined constants for the fitness function. Consequently, the fitness function is designed to deal with the tradeoff between the number of incorrectly classified patterns and the number of fuzzy rules. In this way, as the fitness function value increases as much as possible based on the guidance of the proposed fitness function, the fuzzy classification system corresponding to the individual will satisfy the desired objective as well as possible. That is, the selected fuzzy system has a low number of rules and a low number of incorrectly classified patterns simultaneously. Subsequently, a PSO-based method is proposed to find an appropriate individual so that the corresponding fuzzy classification system has the desired performance. The procedure is described as follows:

Step 1. Initialize the PSO-based method.

(a) Set the number of individuals (L), the maximum number of rules (B), the number of generations (K), the constants for the fitness function (σ_e and σ_r) and the constants for the PSO algorithm (ψ, d_1, d_2, c_1, and c_2).

(b) Generate randomly initial population P.
Each individual of the population is expressed as follows:

$$\underline{p}_h = [\underline{r}_h\ \underline{g}_h], \tag{14}$$

where $\underline{r}_h = [m^h_{(11,1)}\ m^h_{(11,2)}\ m^h_{(11,3)} \cdots\ m^h_{(1m,1)}\ m^h_{(1m,2)}\ m^h_{(1m,3)} \cdots\ m^h_{(B1,1)}\ m^h_{(B1,2)}\ m^h_{(B1,3)} \cdots m^h_{(Bm,1)}\ m^h_{(Bm,2)}\ m^h_{(Bm,3)}]$ and $\underline{g}_h = [g_1\ g_2 \cdots g_B]$. $m^h_{(ji,k)}$, $j \in \{1,2,\cdots,B\}$, $i \in \{1,2,\cdots,M\}$, $k \in \{1,2,3\}$, is randomly generated as follows:

$$m^h_{(ji,k)} = m^{\min}_{(ji,k)} + (m^{\max}_{(ji,k)} - m^{\min}_{(ji,k)}) \cdot rand(), \tag{15}$$

where the range of the parameter $m^h_{(ji,k)}$ is defined as $[m^{\min}_{(ji,k)}, m^{\max}_{(ji,k)}]$ and $rand()$ is a uniformly distributed random numbers in [0,1]. $g_j^h \in [0,1]$, $j \in \{1,2,\cdots,B\}$, is randomly generated as follows:

$$g_j^h = rand(). \tag{16}$$

(c) Generate randomly initial velocity vectors $\underline{v}_h$, $h = 1,2,\cdots,L$.
Each velocity vector is expressed as follows:

$$\underline{v}_h = [\underline{\alpha}_h\ \underline{\beta}_h], \quad (17)$$

where $\underline{\alpha}_h = [\alpha^h_{(11,1)}\ \alpha^h_{(11,2)}\ \alpha^h_{(11,3)} \cdots \alpha^h_{(1m,1)}\ \alpha^h_{(1m,2)}\ \alpha^h_{(1m,3)} \cdots \alpha^h_{(B1,1)}\ \alpha^h_{(B1,2)}\ \alpha^h_{(B1,3)} \cdots \alpha^h_{(Bm,1)}\ \alpha^h_{(Bm,2)}\ \alpha^h_{(Bm,3)}]$ and $\underline{\beta}_h = [\beta_1^h\ \beta_2^h \cdots \beta_B^h]$. $\alpha^h_{(ji,k)}$, $j \in \{1,2,\cdots,B\}$, $i \in \{1,2,\cdots,M\}$, $k \in \{1,2,3\}$, is randomly generated as follows:

$$\alpha^h_{(ji,k)} = \frac{m^{\max}_{(ji,k)} - m^{\min}_{(ji,k)}}{20} \cdot rand(). \quad (18)$$

β_j^h, $j \in \{1,2,\cdots,B\}$ is randomly generated as follows:

$$\beta_j^h = \frac{rand()}{20}. \quad (19)$$

Step 2. Calculate the fitness value of each individual and set initial p_h', f_h' for each individual and initial p^{best}, f^{best} for the initial population.

(a) $f_h = fit(\underline{p}_h)$, $h = 1,2,...,L$,

Set $f_h' = f_h$, $\underline{p}_h' = \underline{p}_h$, $= 1,2,...,L$.

(b) Find the index J of the individual with the best fitness $J = \arg\max_{h=1}^{L} f_h$.

Set $f^{best} = f_J$, $\underline{p}^{beat} = \underline{p}_J$.

(c) Set gen = 1.

Step 3. Update the vector $\underline{g}_h = [g_1^h\ g_2^h\ g_j^h \cdots g_B^h]$, $h \in \{1,2,\cdots,L\}$, as follows:

If $\psi \ge rand()$, then $g^h_{j^*} = 1 - g^h_{j^*}$, where $j^* = round(B\ rand() + 0.5)$. $round(B\ rand() + 0.5)$ rounds $B\ rand() + 0.5$ to the nearest integer.

Step 4. Update $\underline{p}'_h, f'_h$, and $\underline{p}^{best}, f^{best}$.

(a)Update $\underline{p}'_h, f'_h$ in the following:

Calculate $f_h = fit(\underline{p}_h)$, if $f_h > f_h'$, , then $f_h' = f_h$, $\underline{p}'_h = \underline{p}_h$, $h = 1,2,\cdots,L$,

(b)Update $\underline{p}^{best}, f^{best}$ in the following:

If $f_h' > f^{best}$, then $f^{best} = f_h'$, $\underline{p}^{best} = \underline{p}_h'$, $h = 1,2,\cdots,L$.

Step 5. Update the velocity vector $\underline{v}_h$ and the parameter vector $\underline{p}_h$.

(a) Update the velocity vectors in the following:

$$\begin{aligned}\underline{v}_h &= \underline{v}_h + c_1 \cdot rand() \cdot (\underline{p}^{best} - \underline{p}_h) \\ &\quad + c_2 \cdot rand() \cdot (\underline{p}_h' - \underline{p}_h),\ h = 1,2,\cdots,L.\end{aligned} \quad (20)$$

(b) Update the parameter vectors in the following:

$$\underline{p}_h = \underline{p}_h + \underline{v}_h,\ h = 1,2,\cdots,L. \quad (21)$$

Step 6. Decrease $\underline{v}_h$ and ψ by the constants $d_1 \in [0,1]$ and $d_2 \in [0,1]$, respectively.

$$\underline{v}_h = \underline{v}_h \cdot d_1,\ h = 1,2,\cdots,L, \quad (22)$$

$$\psi = \psi * d_2. \quad (23)$$

Step 7. gen = gen+1, if gen > K then go to Step 8; otherwise go to Step 3.

Step 8. Based on the individual $\underline{p}^{best} = [\underline{r}^{best}\ \underline{g}^{best}]$ with the best fitness f^{best}, the desired fuzzy classification system can be determined.

Each $\underline{g}_h$ is randomly updated by Step 3 so that the fuzzy classification system generated by the individual $\underline{p}_h$ has an appropriate number of fuzzy rules for the considered classification problem. The adjustment by Step 5 is conceptually similar to the mutation operator utilized by the genetic algorithm. Each individual $\underline{p}_h$ keeps track of its own best solution, which is associated with the best fitness f_h', it has achieved so far in a vector $\underline{p}'_h$. Furthermore, the best solution among all the individuals obtained so far in the population is kept track of as the vector $\underline{p}^{best}$ associated with the global best fitness f^{best}. According to Step 5, each parameter vector $\underline{p}_h$ is assigned with a randomized velocity vector $\underline{v}_h$ according to its own and its companions' flying experiences so that the vector $\underline{p}_h$ searches around its best vector $\underline{p}'_h$ and the global best vector $\underline{p}^{best}$. The adjustment toward $\underline{p}'_h$ and $\underline{p}^{best}$ by Step 5 is conceptually similar to the crossover operator utilized by the genetic algorithm. Consequently, the parameter vector $\underline{p}_h$ is updated by the PSO so that the fuzzy classification system generated by the individual $\underline{p}_h$ has an appropriate number of rules and a low number of incorrectly classified patterns simultaneously.

4. Simulation Results

In this section, a synthetic data set and the Iris data set are employed to examine the learning ability and the generalization ability of the proposed PSO-based fuzzy classification system, respectively.

Example 1. A synthetic data set

In this example, a synthetic data set containing of three clusters of various sizes, shapes and orientations

[12] is utilized to examine the learning ability of the proposed PSO-based fuzzy classification system. The total number of patterns in this data set is 579. Figure 1 shows the data set with a mixture of spherical and ellipsoidal clusters. Following the proposed method, simulation results of the proposed approach to classifying this data set is shown in Figure 2, where the initial conditions for the proposed method in Example 1 are given in the following: The number of individuals: $L = 100$, the maximum number of rules: $B = 20$, the number of generations: $K = 50$, the range of $m^h_{(j1,1)}, j \in \{1,2,\cdots,20\}$: [0,1], the range of $m^h_{(j1,2)}, j \in \{1,2,\cdots,20\}$: [0.05,0.5], the range of $m^h_{(j1,3)}, j \in \{1,2,\cdots,20\}$: [0.05,0.5], the range of $m^h_{(j2,1)}, j \in \{1,2\cdots,20\}$: [0,1], the range of $m^h_{(j2,2)}, j \in \{1,2,\cdots,20\}$: [0.05,0.5], the range of $m^h_{(j2,3)}, j \in \{1,2,\cdots,20\}$: [0.05, 0.5], the constants for the fitness function: $\{\sigma_e, \sigma_r\} = \{5,10\}$ and the constants for the PSO: $\{\psi, c_1,c_2,d_1,d_2\} = \{1,1,1,0.75,0.75\}$. The rule base of the selected fuzzy classification system has four rules such that the number of incorrectly classified patterns is zero. The parameters of the selected fuzzy classification system are shown in Table 1. From simulation results of this example, it is clear the proposed PSO-based method can select significant fuzzy rules to construct a fuzzy classification such that the number of incorrectly classified patterns are minimized.

Example 2. Iris data set

The Iris data set [1] contains 150 patterns with four features, that belong to three classes (Iris Setosa, Iris Versicolour and Iris Virginica). The four features are the sepal length in cm, the sepal width in cm, the petal length in cm and the petal width in cm. The data set contains three classes, each of 50 patterns; each class refers to a type of iris plant. One class is linearly separable from the other two; the latter are not linearly separable from each other. In this example, the two-fold cross validation (2CV) is employed to examine the generalization ability of the proposed approach to classifying the iris data. In the 2CV procedure, the Iris data are separated into two subsets of the same size. That is, each subset consists of

Table 1. Parameters of the selected fuzzy classification system by the proposed PSO-based method in Example 1

j	$m_{(j1,1)}$	$m_{(j1,2)}$	$m_{(j1,3)}$	$m_{(j2,1)}$	$m_{(j2,2)}$	$m_{(j2,3)}$	H_j	CF_j
1	0.7361	0.2883	0.4669	0.5187	0.2625	0.366	1	0.6762
2	0.1958	0.05	0.4141	0.8314	0.209	0.3121	3	0.9637
3	0	0.2181	0.2570	0.8126	0.455	0.5	3	0.9153
4	0.0154	0.3087	0.4406	0.1336	0.3009	0.3604	2	0.9343

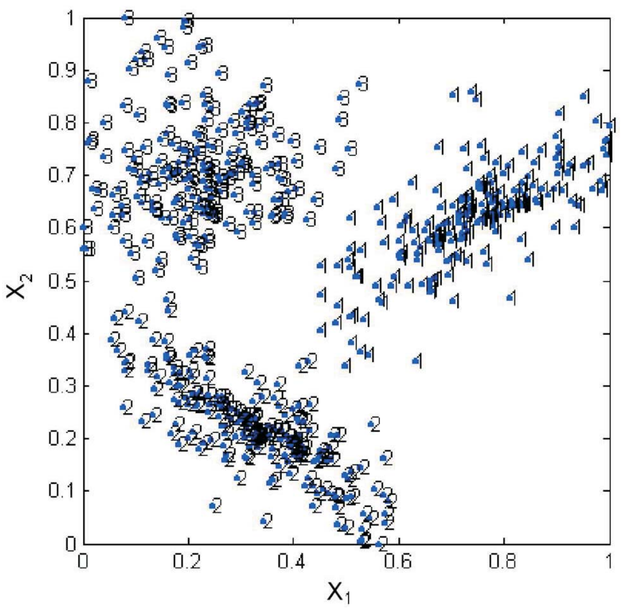

Figure 1. A synthetic data set with a mixture of spherical and ellipsoidal clusters in Example 1.

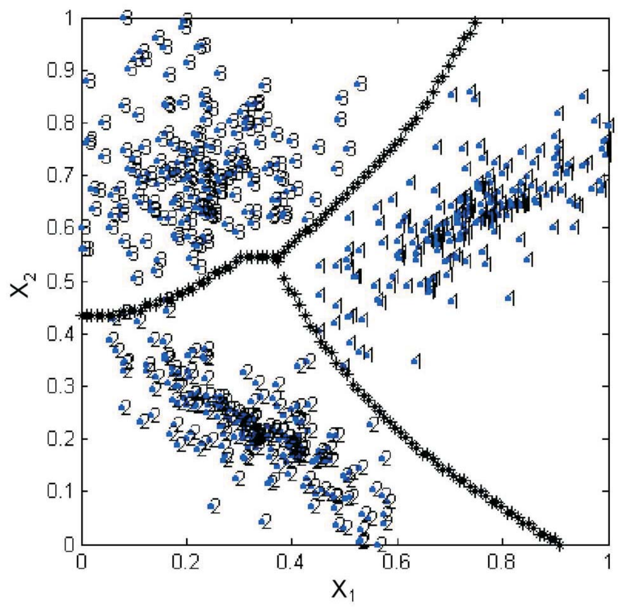

Figure 2. Simulation results of the proposed PSO-based method for the synthetic data set in Example 1.

Table 2. Simulation results of the proposed PSO-based method for the Iris data set using 2CV method in Example 2

Average classification rate on training patterns	Average classification rate on test patterns	Average number of fuzzy rules
98.87%	96.8%	4.725

Table 3. Generalization ability for test patterns in the Iris data classification using the same 2CV method

Method	Average classification rate on test patterns	Average number of fuzzy rules
Pruning	93.30 %	28.00
Multi-rule-table	94.30 %	597.75
GA-based	90.67 %	10.10
Proposed method	96.80 %	4.725

75 patterns. One subset is used as training patterns to construct a fuzzy classification system by the proposed approach. The other subset is used as test patterns to evaluate the generated fuzzy classification system. The same training-and-testing procedure is also followed after the roles of the subsets are exchanged with each other. The above procedure is iterated 20 times using different partitions of the 150 patterns into two subsets for calculating the average classification rates on these data. The initial conditions for the proposed method in Example 2 are given in the following: The number of individuals: $L = 100$, the maximum number of rules: $B = 100$, the number of generations: $K = 50$, the range of $m^h_{(j1,1)}$, $j \in \{1,2,\cdots,20\}$: [4.4,7.7], the range of $m^h_{(j1,2)}$, $j \in \{1,2,\cdots,20\}$: [0.1,1.65], the range of $m^h_{(j1,3)}$, $j \in \{1,2,\cdots,20\}$: [0.1,1.65], the range of $m^h_{(j2,1)}$, $j \in \{1,2,\cdots,20\}$: [2,4.4], the range of $m^h_{(j2,2)}$, $j \in \{1,2,\cdots,20\}$: [0.1,1.2], the range of $m^h_{(j2,3)}$, $j \in \{1,2,\cdots,20\}$: [0.1,1.2], the range of $m^h_{(j3,1)}$, $j \in \{1,2,\cdots,20\}$: [1.2,6.7], the range of $m^h_{(j3,2)}$, $j \in \{1,2,\cdots,20\}$: [0.1,2.75], the range of $m^h_{(j3,3)}$, $j \in \{1,2,\cdots,20\}$: [0.1,2.75], the range of $m^h_{(j4,1)}$, $j \in \{1,2,\cdots,20\}$: [0.1,2.5], the range of $m^h_{(j4,2)}$, $j \in \{1,2,\cdots,20\}$: [0.1,1.2], the range of $m^h_{(j4,3)}$, $j \in \{1,2,\cdots,20\}$: [0.1,1.2], the constants for the fitness function: $\{\sigma_e, \sigma_r\} = \{5,10\}$ and the constants for the PSO: $\{\psi, c_1, c_2, d_1, d_2\} = \{1,1,1,0.75,0.75\}$. The average number of fuzzy rules of the selected fuzzy classification systems and the average classification rates on training patterns and test patterns are summarized in Table 2. Table 3 compares our results with the results in [7] for the classification of the Iris data using the same 2CV method. From simulation results for the Iris data set, it is obvious that the proposed PSO-based fuzzy classification system has high generalization ability for the classification problem of the Iris data set.

5. Conclusions

A PSO-based method is proposed to design an appropriate fuzzy classification system for pattern classification. In the proposed approach, each individual $\underline{p}_h$ consists of two parameter vectors: $\underline{r}_h$ and $\underline{g}_h$ The parameter vector $\underline{g}_h$ is updated so that the generated fuzzy classification system has an appropriate number of fuzzy rules. The parameter vector $\underline{r}_h$ is updated so that the premise part of the generated fuzzy classification system has appropriate membership functions for the considered classification problem. Then, the obtained premise fuzzy sets are used to determine the consequent parameters of the corresponding fuzzy classification system. Consequently, each individual corresponds to a fuzzy classification system. Subsequently, a fitness function is defined to guide the searching procedure to select a fuzzy classification system with the desired performance. The simulation results show that the selected fuzzy classification system not only has an appropriate number of rules for the considered classification problem but also has a low number of incorrectly classified patterns.

Acknowledgement

This research was supported in part by the National Science Council of the Republic of China under contract NSC 93-2218-274-001

References

[1] Blake, C., Keogh, E. and Merz, C. J., UCI Repository of Machine Learning Database, Univ. California, Irvine (1998). http://www.ics.uci.edu/~mlearn/

[2] Eberhart, R. and Kennedy, J., "A New Optimizer Using Particle Swarm Theory," *Proc. Int. Sym. Micro Machine and Human Science*, Nagoya Japan, Oct., pp. 39–43 (1995).

[3] Ishibuchi, H., Nozaki, K., Yamamoto, N. and Tanaka, H., "Selecting Fuzzy If-Then Rules for Classification Problems Using Genetic Algorithms," *IEEE Trans. Fuzzy Systems*, Vol. 3, pp. 260–270 (1995).

[4] Jang, J. S., "ANFIS: Adaptive-Network-Based Fuzzy Inference systems," *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 23, pp. 665–685 (1993).

[5] Jang, J. S., Sun, C. T. and Mizutani, E., *Neuro-Fuzzy and Soft Computing*, Prentice Hall, New Jersey, U.S.A. (1997).

[6] Kennedy, J. and Eberhart, R., "Particle Swarm Optimization," *Proc. IEEE Int. Conf. Neural Networks*, Perth, Australia, pp. 1942–1948 (1995).

[7] Nozaki, K., Ishibuchi, H. and Tanaka, H., "Adaptive Fuzzy Rule-Based Classification Systems," *IEEE Trans. on Fuzzy Systems*, Vol. 4, No. 3, Aug., pp. 238–250 (1996).

[8] Simpson, P. K., "Fuzzy Min-Max Neural Networks-Part 1: Classification," *IEEE Trans. Neural Networks*, Vol. 3, Sep., pp. 776–786 (1992).

[9] Wang, L. X. and Mendel, J. M., "Generating Fuzzy Rules by Learning from Examples," *IEEE Trans. on Systems, Man and Cybernetics*, Vol. 22, pp. 1414–1427 (1992).

[10] Wong, C. C. and Chen, C. C., "A Hybrid Clustering and Gradient Descent Approach for Fuzzy Modeling," *IEEE Trans. on Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 29, pp. 686–693 (1999).

[11] Wong, C. C. and Chen, C. C., "A GA-Based Method for Constructing Fuzzy Systems Directly from Numerical Data," *IEEE Trans. on Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 30, pp. 904–911 (2000).

[12] Wong, C. C., Chen, C. C. and Su, M. C., "A Novel Algorithm for Data Clustering," *Pattern Recognition*, Vol. 34, pp. 425–442 (2001).

[13] Yager, R. R. and Filev, D. P. *Essentials of Fuzzy Modeling and Control*, John Wiley, New York, U.S.A. (1994).

Manuscript Received: Apr. 27, 2005
Accepted: Jul. 13, 2005