Construct Interpretable Fuzzy Classification System based on Fuzzy Clustering Initialization

Xing Zong-yi[1], Zhang Yong[1], Jia Li-min[2], Hu Wei-li[1]

[1]Nanjing University of Science and Technology, Nanjing 210094, China
[2]Beijing Jiaotong University, Beijing, 100044, China

xingzongyi@tom.com

Abstract

An approach to construct interpretable fuzzy classification system based on fuzzy clustering initialization is proposed in this paper. First, the precision index is defined, and the necessary conditions of interpretability are analyzed. Second, the initial fuzzy classification system is identified using a fuzzy clustering algorithm, and the number of fuzzy rules is determined by cluster validity measure. Subsequently, the method of merging similar fuzzy sets is adopted to reduce the initial model in order to enhance its interpretability, and genetic algorithm is used to optimize the model in order to improve its precision. The proposed approach is applied to Iris benchmark classification problem, and the results show its validity.

Key words: fuzzy classification system; fuzzy clustering; interpretability; precision

I. Introduction

Fuzzy systems have been successfully applied to various classification problems due to its powerful capabilities of handling uncertainty and vagueness [1]. For a simple system, we can obtain a fuzzy classification system using the expert experience, however it is difficult to construct the fuzzy classification system for a complex system, where the expert experience is incomplete or inexistent. So how to model a fuzzy classification system from data has become research focus in recent years.

Different techniques including clustering methods [2], genetic algorithm [3] and neuro-fuzzy networks [4], have been employed for automatically determining and learning fuzzy classification system directly from data. For example, Abe *et al.* [2] applied the clustering technique to divide labeled data for each class into clusters and built fuzzy rules with an ellipsoidal region for each cluster. After fine-tuning, a fuzzy classifier with high recognition rate was obtained.

However all these technologies only focus on precision that simply fit data with highest possible accuracy, neglecting interpretability of the obtained fuzzy classifications, which is considered as a primary merit of fuzzy systems and is the most prominent feature that distinguishes fuzzy systems from many other models [5]. In order to improve interpretability of fuzzy classification system, some methods have been developed.

Jin [6] concludes some necessary conditions of interpretability. Ishibuchi *et al* [7] introduce a two-stage rule selection approach. In the first stage, a pre-specified number of candidate rules are extracted from numerical data using a data mining technique. In the second stage, an EMO algorithm is used for finding non-dominated rule sets with respect to three objectives: to maximize the number of correctly classified training patterns, to minimize the number of rules, and to minimize the total rule length. Abonyi *et al*. [8] initialize fuzzy classification system using decision tree, and then reduce the model in an iterative scheme by means of similarity-driven rule reduction. To improve classification performance of the reduced fuzzy classification system, a genetic algorithm with a multi-objective criterion searching for both redundancy and accuracy is applied. Chang *et al*. [9] proposed a new evolutionary approach for deriving a compact fuzzy classification system directly from data without any *a priori* knowledge of the distribution of the data. At the beginning, the fuzzy classification is empty with no rules in the rule base and no membership functions assigned to fuzzy variables. Then, rules and membership functions are automatically created and optimized in an evolutionary process.

This paper proposes an approach to construct interpretable fuzzy classification system based on fuzzy clustering initialization. Section 2 gives the used fuzzy classification system and details the concept of interpretability. Section 3 focuses on the method of constructing fuzzy classification system. The experimental evaluation is described in section 4, and section 5 concludes the paper.

II. Preliminary Issues

A. Fuzzy Classification System

Considering an n-dimensional classification problem for which N patterns $\mathrm{x} \in \boldsymbol{X} \subseteq R^n$, $\mathrm{x} = (x_1, x_2, \cdots, x_n)$ are given form M classes $\{C_1, C_2, \cdots, C_M\}$, A typical fuzzy rule of the classification system has the form:

$$\begin{aligned} R_i:\ & \text{If } x_1 \text{ is } A_{i1} \text{ and } x_2 \text{ is } A_{i2} \text{ and } \cdots \text{ and } x_n \text{ is } A_{in} \\ & \text{Then the pattern } (x_1, x_2, \cdots, x_n) \text{ belongs to} \\ & \text{class } C_i \text{ with } CF = CF_i \end{aligned} \tag{1}$$

where CF_i is a certainty factor that represents the desired impact of the rule, $A_{i1}, \cdots, A_{in}$ are fuzzy sets defined on the universe of discourse of the input variables.

In this paper, Gaussian membership functions are used to represent the fuzzy set:

$$A_{ij}(x_{jk}) = \exp(-\frac{1}{2}\frac{(x_{jk} - v_{ij})^2}{\sigma_{ij}^2}) \tag{2}$$

where v_{ij} and σ_{ij} represent center and variance of Gauss function respectively.

The output of the classification system is determined by winner take all strategy, i.e. the output of the class related to the consequent of the rule that gets the highest degrees of activation:

$$\mathrm{x}_k \in C_{i^*},\ i^* = \arg(\max(\beta_i(\mathrm{x}_k)))\ \ 1 \le i \le M \tag{3}$$

where β_i is the firing strength of *i*-th rule:

$$\beta_i(\mathrm{x}_k) = CF_i \prod_{j=1}^{n} A_{ij}(x_{jk}) \tag{4}$$

B. Precisions and Interpretability

The classification error for pattern x_k is defined as:

$$e_k = \begin{cases} 1 & \text{if } \mathrm{x_k} \text{ is classfied correctly} \\ 0 & \text{if } \mathrm{x_k} \text{ is classfied falsely} \end{cases} \tag{5}$$

Then the precision performance of fuzzy classification system is defined as:

$$J = \frac{1}{N}\sum_{k=1}^{N} e_k \tag{6}$$

Interpretability is a subjective property that depends on several factors, mainly the model structure, the number of feature variables, the number of fuzzy rules, the number of linguistic terms, and the shape of the fuzzy sets, etc. Although there is no formal definition for interpretability, several aspects are believed to be essential, which are described as follows [6].

1) The number of variables: a highly multi- dimensional fuzzy model is difficult to interpret. The model should use as fewer variables as possible.

2) The number of rules: a fuzzy model with a large rule base is less interpretable than a fuzzy system containing only few rules. Experientially, the number of fuzzy rules of an interpretable model is no more than ten, which is determined by intellect of human being.

3) Completeness, consistency and compactness of fuzzy rules: for each effective input variables combination, there must be at least one fuzzy rule being fired, i.e. fuzzy

rules cover the whole input space. The fuzzy rules in the rule base should be consistent. If there are rules that are contradictory to each other, it is hard to understand the fuzzy rules. There must be no rule whose antecedent is a subset of another rule, and no rule may appear more than once in the rule base.

4) Characteristics of membership functions: convexity and normality are two principle aspects which are satisfied naturally for most widely used membership functions, e.g., the Gauss function, triangle function. The fuzzy partition of all input variables should be complete to prevent unpredictable system outputs. Fuzzy sets should be distinguishing, thus meaningful to assign linguistic terms to fuzzy system. Usually, a minimum/maximum degree of overlapping between fuzzy sets muse be enforced.

III. Interpretable Fuzzy classification System

A. Constructing Fuzzy classification system based on fuzzy clustering method

It is generally acknowledged that fuzzy clustering is a well-recognized paradigm to construct fuzzy models. Numerous fuzzy clustering algorithms have been developed. Fuzzy C-means algorithm (FCM) is the base algorithm from the set of fuzzy clustering algorithm using the objective function and it has many modified versions. The Gustafson-Kessel [10] (GK) algorithm is an extension of FCM, whereas its clusters are ellipsoids and has different size in any dimension.

The objective function of GK algorithm is described following:

$$J(\mathbf{Z};\mathbf{U},\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m D_{ik}^2 \tag{7}$$

where $\mathbf{Z}$ is the set of data, $\mathbf{U} = [\mu_{ik}]$ is the fuzzy partition matrix, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_c]^T$ is the set of centers of the clusters, c is the number of clusters, N is the number of data, m is the fuzzy coefficient fuzziness, μ_{ik} is the membership degree between the ith cluster and kth data, which satisfy conditions:

$$\mu_{ik} \in [0,1];\ \sum\nolimits_{i=1}^{C}\mu_{ik} = 1;\ 0 < \sum\nolimits_{k=1}^{N}\mu_{ik} < N \tag{8}$$

The norm of distance between the ith cluster and kth data is:

$$D_{ik}^2 = \| \mathbf{z}_k - \mathbf{v}_i \|_{A_i}^2 = (\mathbf{z}_k - \mathbf{v}_i)^T \mathbf{A}_i (\mathbf{z}_k - \mathbf{v}_i) \tag{9}$$

where

$$\mathbf{A}_i = (\rho \det(\mathbf{F}_i))^{1/n} \mathbf{F}_i^{-1} \tag{10}$$

$$\rho = \det(\mathbf{A}_i) \tag{11}$$

F_i is the fuzzy covariance matrix of *i*-th cluster:

$$\mathbf{F}_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m (\mathbf{z}_k - \mathbf{v}_i)(\mathbf{z}_k - \mathbf{v}_i)^T}{\sum_{k=1}^{N} (\mu_{ik})^m} \tag{12}$$

Lagrange multiplier is used to optimize the objective function (5) and the minimum of (**U**, **V**) is calculated as follows:

$$\mu_{ik} = \frac{1}{\sum_{j=1}^{c} (D_{ik} / D_{jk})^{2/(m-1)}} \tag{13}$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^{N} (\mu_{ik})^m \mathbf{z}_k}{\sum_{k=1}^{N} (\mu_{ik})^m} \tag{14}$$

The variance of Gauss function is calculated as:

$$\sigma_{ij}^2 = \frac{\sum_{k=1}^{N} \mu_{ik} (x_{jk} - v_{jk})^2}{\sum_{k=1}^{N} \mu_{ik}} \tag{15}$$

In order to determine the label of rule, we define the function:

$$M_{ij} = \sum_{k=1}^{N} u_{ik} f_j(k) \Big/ \sum_{k=1}^{N} f_j(k) \tag{16}$$

where

$$f_j(k) = \begin{cases} 1 & \text{if } \mathrm{x}_k \in C_j \\ 0 & \text{if } \mathrm{x}_k \notin C_j \end{cases} \tag{17}$$

Then the class label of *i*-th rule is:

$$i^* = \arg(\max(M_{ij})) \quad j = 1, 2, \cdots, M \tag{18}$$

and the certainty factor of *i*-th rule is:

$$CF_i = \max_j (M_{ij}) \tag{19}$$

B. Cluster Validation Index

It is essential to determine the number of rules, i.e. the number of fuzzy clusters. This problem can be solved by validation analysis using cluster validity index.

There are two categories of fuzzy validity index. The first category uses only the membership values of a fuzzy partition of data. On the other hand the latter one involve both partition matrix and the data itself.

The partition coefficient (*PC*) and the partition entropy coefficient (*PE*) [11] are the typical cluster validity index of the first category:

$$PC(c) = \frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^{2} \tag{20}$$

$$PE(c) = -\frac{1}{n}\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}\log_a \mu_{ik} \tag{21}$$

where $PC(c) \in [1/c, 1]$, $PE(c) \in [0, \log_a c]$. With increase of *c*, the values of *PC* and *PE* are decrease/ increase respectively. The number corresponding to significant knee is selected as the optimal number of rules.

The above mentioned cluster validity index are sensitive to fuzzy coefficient *m*. when $m \to 1$, the index give the same values for all *c*. when $m \to \infty$, both *PC* and *PE* exhibit significant knee at *c*=2.

The compactness and separation validity function proposed by Xie and Beni (*XB*) [12] is a representation of the second category. The smallest value of *XB* indicates the optimal clusters.

$$XB(c) = \frac{\sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^{m}\left\|x_k - v_i\right\|^2}{n \cdot \min_{i,k}\left\|v_i - v_k\right\|^2} \tag{22}$$

Individual cluster validity index cannot determine the number of rules absolutely, so we always employ several cluster validity index, including Non-Fuzzy Index (*NFI*) [13], Min Hard Tendency (*MinHT*) [14], Fuzzy Hyper Volume (*FHV*) and Average Partition Density (*PA*) and Partition Density (*PD*) [15], Possibility Partition (*PP*) [16]. The maximum of *NFI*、*MinHT*、*MeHT*、*DPA* and *PD* correspond to the optimal number, while minimum of *XB*、 *FHV* and *PP* indicate adoptive number.

We employ ten cluster validity indices to determine the number of rules. If obtained results of there indices are different, we adopt the most appeared number as final result. Of course, the maximum number can be adopted, and then rule reduction is utilized to selected optimal rules. Setnes et al [17] and Castillo et al [18] detailed rule reduction method.

C. Similar Fuzzy Sets Merging

Fuzzy classification system obtained above may contain redundant information in the form of similarity between fuzzy sets. The similarity of fuzzy sets makes the fuzzy model uninterpretable, for it is difficult to assign qualitatively meaningful labels to similar fuzzy sets. In order to acquire an effective and interpretable fuzzy model, elimination of redundancy and making the fuzzy model as simple as possible are necessary.

There are three types of redundant or similar fuzzy sets in fuzzy model: 1) a fuzzy set similar to the universal set, 2) a fuzzy set similar to the singleton set, and 3) the fuzzy set *A* is similar to the fuzzy set *B*.

If a fuzzy set is similar to universal set or singleton set, it should be removed from the corresponding fuzzy rule antecedent without deteriorating precision. As for two similar fuzzy sets, a similarity measure is unutilized to determine if the fuzzy sets should be combined.

For the fuzzy sets *A* and *B*, a set-theoretic operation based similarity measure [19] is defined as

$$S(A,B)=\frac{|A\bigcap B|}{|A\bigcup B|} \tag{23}$$

where $|\cdot|$ denotes the cardinality of a set, the $\cap$ and $\cup$ operators represent the intersection and union respectively. For discrete universe $X=\{x_j \mid j=1,2,\cdots,m\}$, this can be rewritten as

$$S(A,B)=\frac{\sum_{k=1}^{N}[\mu_A(x_k)\wedge\mu_B(x_k)]}{\sum_{k=1}^{N}[\mu_A(x_k)\vee\mu_B(x_k)]} \tag{24}$$

where $\wedge$ and $\vee$ are the minimum and maximum operators respectively. *S* is a similarity measure in [0,1]. *S*=1 means the compared fuzzy sets are equal, while *S*=0 indicates that there is no overlapping between fuzzy sets.

If similarity measure $S(A,B)>\tau$, i.e. the fuzzy sets are very similar, then the two fuzzy sets *A* and *B* should be merged to create a new fuzzy set *C*, where λ is a predefined threshold. It should be pointed out that threshold τ influences the model performance significantly. Small threshold generates low accurate and high interpretable fuzzy model. In a general way, $\tau=[0.4-0.7]$ is a good choice.

For Gauss type of fuzzy sets used in this paper, the parameters of new merged fuzzy set *C* from *A* and *B* are defined as:

$$\begin{cases} v_c=(v_A+v_B)/2 \\ \sigma_c=\sqrt{\sigma_A^2+\sigma_B^2}/2 \end{cases} \tag{25}$$

Fuzzy sets merging process is carried out iteratively. For each iteration, the similarity measures between all pairs of fuzzy sets for each variable are calculated. The pair of highly similar fuzzy sets with $S>\lambda$ is merged to create a new fuzzy set, thus, the rule base of fuzzy model is updated. This process continues until there are no fuzzy sets for which $S > \tau$. Then the fuzzy sets that have similarity to the universal set or singleton set are removed.

D. Genetic Algorithm Optimization

AS a gradient free and parallel optimization algorithm, genetic algorithm (GA) has applied to various domains successfully [20]. After similar fuzzy sets merging process, the precision of obtained fuzzy classification system reduced, so genetic algorithm is adopted to improve its precision.

Chromosome Representation

Genetic algorithm with binary code is less efficient to cope with multidimensional or continuous problems. For complex system, the bit strings becomes very long and the search space blows up. In real-coded genetic algorithm, the variables appear directly in chromosome simply, and computation burden is relieved. So real-coded GA is adopted in this paper.

The obtained fuzzy classification system is encoded in a chromosome as a sequence of elements describing the fuzzy sets in the rule antecedents followed by the certainty factors:

$$H_1 = (v_{11}, \cdots, v_{cn}, \sigma_{11}, \cdots, \sigma_{cn}, CF_1, \cdots, CF_c) \tag{26}$$

Given search space $[H^{\min}, H^{\max}]$:

$$H^{\min} = (v_{11}^{\min}, \cdots, v_{cn}^{\min}, \sigma_{11}^{\min}, \cdots, \sigma_{cn}^{\min}, 0, \cdots, 0) \tag{27}$$

$$H^{\max} = (v_{11}^{\max}, \cdots, v_{cn}^{\max}, \sigma_{11}^{\max}, \cdots, \sigma_{cn}^{\max}, 1, \cdots, 1) \tag{28}$$

where $v_{ij}^{\max}$ 、 $v_{ij}^{\min}$ 、 $\sigma_{ij}^{\min}$ 、 $\sigma_{ij}^{\max}$ are maximum and minimum values of corresponding membership functions.

The initial population is created by random variation (uniform distribution) around H_1 with the search space $[H^{\min}, H^{\max}]$.

Fitness Function

The precision performance is adopted as fitness function directly:

$$Fit = \frac{1}{N} \sum_{k=1}^{N} e_k \tag{29}$$

Genetic Operators

There are three genetic operators in GA: selection, crossover and mutation. In order to hold variety of chromosomes, several randomly selected methods for each operator are adopted in this paper.

1) Selection:

The *roulette wheel* selection method is used to select chromosomes to operate. For chromosome H_p with fitness value f_p, the selected probability is:

$$p_p = f_p \Big/ \sum_{p=1}^{L} f_p \tag{30}$$

In order to prevent optimal chromosomes are ignored, *elitist* selection are used at the same time, i.e., the best chromosome is always preserved in population

2) Crossover:

$H_r^t = (r_1, \cdots, r_l)$ and $H_s^t = (s_1, \cdots, s_l)$ are selected chromosome for crossover in t generation. The following crossover operators are adopted randomly.

<u>Simple arithmetic crossover:</u> k is randomly selected position of chromosome. The result offspring are:

$$H_r^{t+1} = (r_1, \cdots, r_k, s_{k+1}, \cdots, s_l) \tag{31}$$

$$H_s^{t+1} = (s_1, \cdots, s_k, r_{k+1}, \cdots, r_l) \tag{32}$$

<u>Whole arithmetic crossover:</u> $\lambda \in [0,1]$ is a uniform distributed random number. The result offspring are:

$$H_r^{t+1} = \lambda(H_r^t) + (1-\lambda)H_s^t \tag{33}$$

$$H_s^{t+1} = \lambda(H_s^t) + (1-\lambda)H_r^t \tag{34}$$

3) Mutation:

$H_r^t = (r_1, \cdots, r_l)$ and $H_s^t = (s_1, \cdots, s_l)$ are selected chromosome for crossover in t generation. T is total number of generations. The following mutation operators are adopted randomly.

Uniform mutation: r_k is randomly selected element of chromosome. $\hat{r}_k \in [r_k^{\min}, r_k^{\max}]$ is random number where $[r_k^{\min}, r_k^{\max}]$ is search space of r_k. The result offspring is:

$$H_r^{t+1} = (r_1, \cdots, \hat{r}_k, \cdots, r_l) \quad (35)$$

Gauss mutation: x_k is a Gauss distributed random number with zero mean and adaptive variance σ_k:

$$\sigma_k = ((T-t)/T)((r_k^{\max} - r_k^{\min})/3) \quad (36)$$

The corresponding offspring is:

$$H_r^{t+1} = (\hat{r}_1, \cdots, \hat{r}_k, \cdots, \hat{r}_l) \quad (37)$$

where

$$\hat{r}_k = r_k + x_k \quad (38)$$

Procedure of GA

Given the number of generation T, select the population size L, crossover probability P_c, and mutation probability P_m. The procedure of GA is described as follows:

(1). Transform obtained fuzzy classification system into chromosome, and create initial population in search space.

(2). Calculate fitness values of chromosomes.

(3). Select $L/4$ pairs of chromosomes for crossover, and form $L/2$ new chromosomes, denote as $P_t^{'}$.

(4). Select $L/2$ chromosomes for mutation, and form $L/2$ new chromosomes, denote as $P_t^{''}$.

(5). Select L chromosomes from $(P_t, P_t^{'}, P_t^{''})$, preserve the best chromosome.

(6). $t = t + 1$, if $t > T$, then stop and transform the best chromosome into fuzzy classification system, otherwise go to (2).

E. Flow Chart

Fig. 1 gives flow chart of proposed approach.

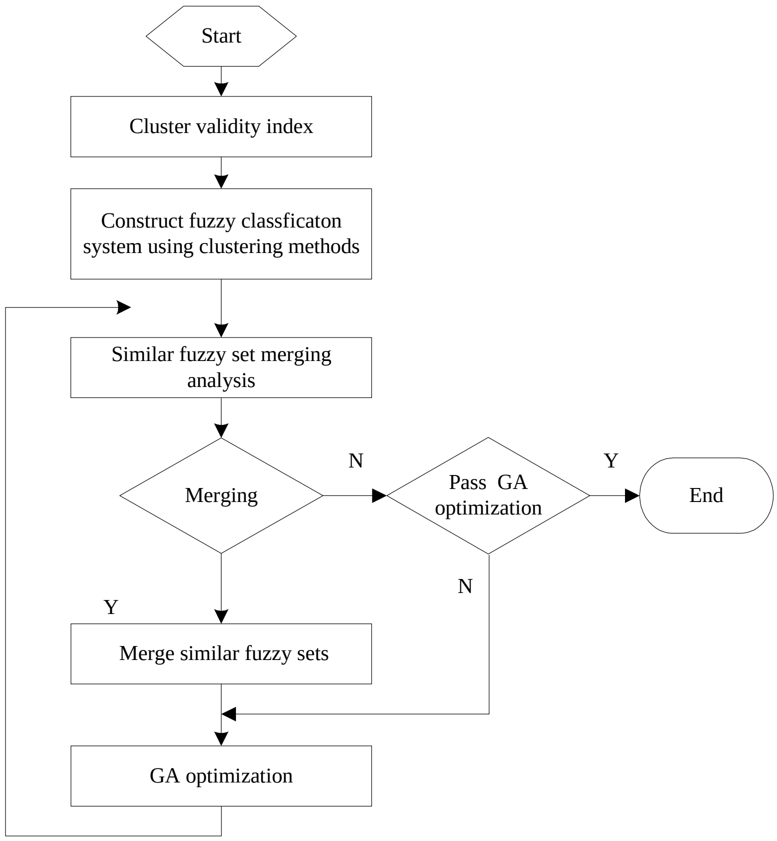


Fig.1. Flow chart of proposed approach

IV. Example

The Iris data is a common benchmark problem in classification and pattern recognition studies. It contains 50 measurements of four features (*sepal length*, *sepal width*, *pental length*, *pental width*) from each of three species (*setosa*, *versicolor*, *virinica*). The first class is separate from others clearly, while the second and third class are overlap slightly. Fig. 2 shows the two-dimension (*sepal length*, *sepal width*) measurement, where "*" denotes data of *setosa* class, "o" denotes data of *versicolor* class, "+" denotes data of *virinica* class.

Cluster validity indices are used to determine the number of fuzzy rules. Tab. 1 diagrams the results, and Fig.3 illustrates *PC* and *PE* intuitionally. All cluster validity indices except *XB* indicate that the optimal number of rules is three. *XB* results that

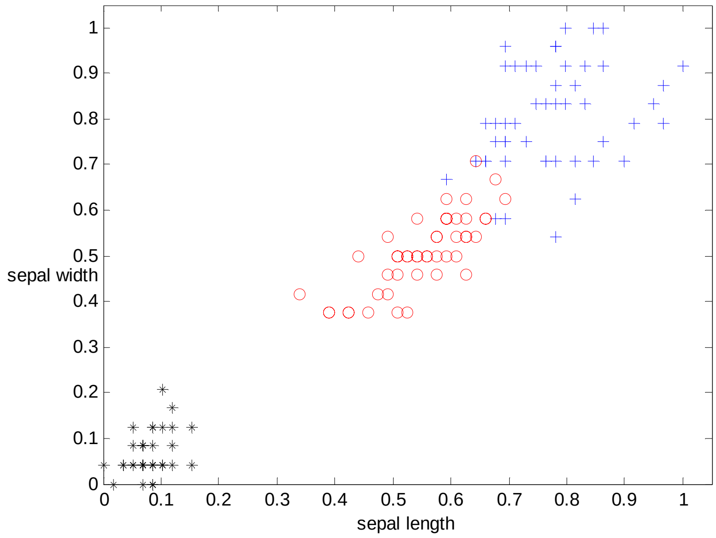


Fig.2. Iris data: setosa (*), versicolor (o), virinica (+)

Tab.1 Cluster validity indices

c	PC	PE	NFI	MinHT	MeHT	XB	FHV	DPA	PD	FP
2	0.7403	0.4074	0.4807	0.8468	0.6476	**0.1008**	0.0464	0.4236	0.4356	0.0002
3	**0.7363**	**0.4529**	**0.6044**	**1.4058**	**0.7710**	0.2278	**0.0414**	**0.8695**	**0.5999**	**-0.0009**
4	0.6204	0.6984	0.4938	1.0839	0.5859	0.4248	0.0678	0.7043	0.3312	0.0212
5	0.5351	0.8987	0.4189	0.8731	0.4275	0.8546	0.0875	0.6833	0.2960	0.0350
6	0.4959	0.9676	0.3951	0.4770	0.4172	0.6892	0.0751	0.6279	0.3669	0.0034
7	0.4479	1.1201	0.3559	0.4557	0.3669	1.3662	0.1131	0.4814	0.2557	0.0120
8	0.4396	1.1707	0.3595	0.5300	0.3686	1.9045	0.1546	0.3811	0.1638	0.0171
9	0.4282	1.2449	0.3567	0.6608	0.4162	0.6574	0.1275	0.6284	0.3004	0.0212
10	0.4068	1.2977	0.3409	0.6881	0.3954	0.6688	0.1178	0.5301	0.3380	0.0047

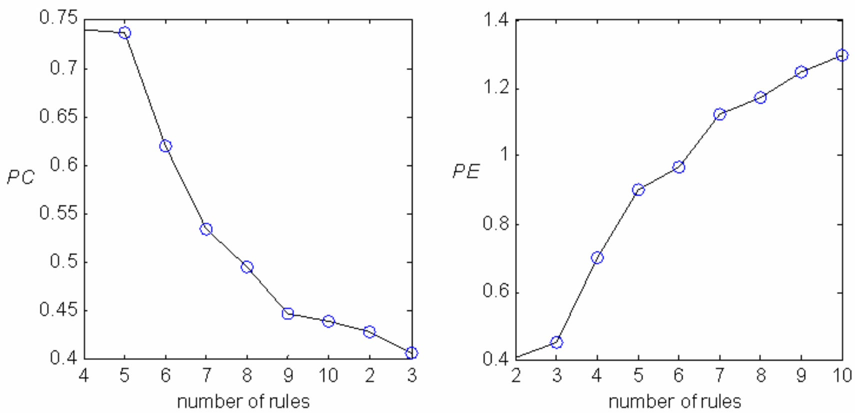


Fig.3. Cluster validity indices (PC and PE)

the optimal number is two, and suboptimal number is three. In this paper, the adopted number of rules is three.

The initial fuzzy classification system is constructed based on clustering method. Fig.4 shows the membership functions of the model with 88.67% precision. Obviously, fuzzy sets of membership functions are overlap highly.

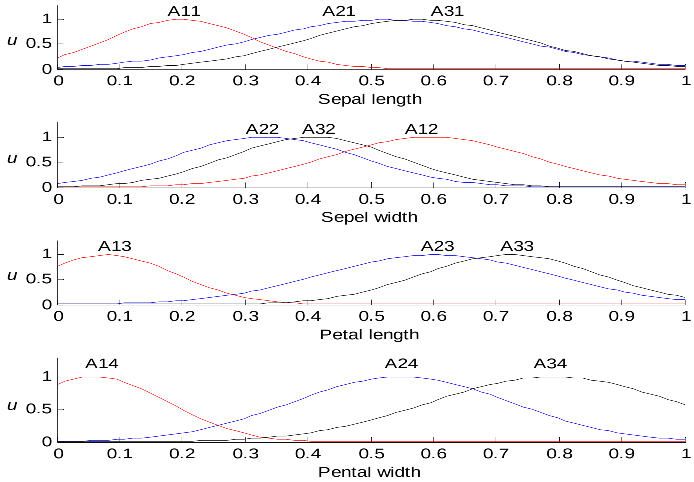


Fig.4. Membership functions of initial fuzzy classification system

Fuzzy sets merging process is carried out for reduced fuzzy model sequentially. The similarity measure between fuzzy sets of *flow rate* is 0.3420, and the measure between fuzzy sets of *PH* is 0.0943. Without considering precision and reality, the fuzzy sets of *flow rate* can be merged to a new fuzzy set, so the flow rate variable can also be deleted for containing only one fuzzy set. After fuzzy sets merging, the training error and validation errors of the model are 1.6290 and 1.7547 respectively.

Fuzzy sets merging process is carried out for initial fuzzy classification system sequentially with merging threshold 0.4. Similarity measure of A_{21} and A_{31} is 0.7909; Similarity measure of A_{22} and A_{32} is 0.6587; Similarity measure of A_{23} and A_{33} is 0. 3825; Similarity measure of A_{24} and A_{34} is 0. 3007. Similarity measure of (A_{21}, A_{31}) and (A_{22}, A_{32}) are bigger that threshold, and these fuzzy sets are merged. As result, the similarity measure of fuzzy sets of *Sepal length* is 0.1491, and the similarity measure of *Sepal width* is 0.2825.

GA is used to optimize the obtained fuzzy classification system to improve its precision with $L = 40$, $Pc = 0.8$, $Pm = 0.05$ and $T = 100$. The precision performance is 92%.

The process of similar fuzzy set merging and GA optimization are continue iteratively until it satisfy the following stop criteria: all similar measures of fuzzy sets are less than predetermined threshold, or there are similar measures bigger than threshold, however the new increased misclassified patterns is exceed 2 after merging and GA optimization.

During the process of merging and optimization, the membership functions of *Sepal length* and *sepal width* are merged to only one fuzzy set. The precision of the model is un-deteriorated if the two fuzzy sets are omitted, so feature of Sepal *length* and *sepal width* are deleted in final fuzzy classification system:

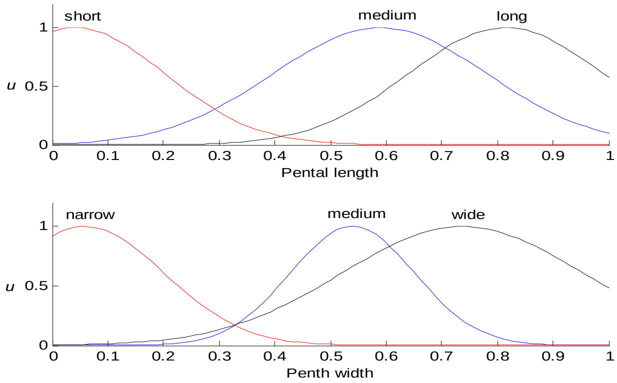


Fig.5. Membership functions of finial fuzzy classification system

$$
\begin{aligned}
&R^1: \text{ If } \textit{pental length} \text{ is medium and } \textit{pental width} \text{ is medium} \\
&\quad \text{Then pattern belongs to Class 1 with } CF=0.9197 \\
&R^2: \text{ If } \textit{pental length} \text{ is short and } \textit{pental width} \text{ is narrow} \\
&\quad \text{Then pattern belongs to Class 2 with } CF=0.3594 \\
&R^3: \text{ If } \textit{pental length} \text{ is long and } \textit{pental width} \text{ is wide} \\
&\quad \text{Then pattern belongs to Class 3 with } CF=0.4657
\end{aligned} \tag{39}
$$

Fig. 5 shows the membership functions of the final fuzzy classifications system. Obviously, it is easy to appoint semantic terms to fuzzy sets. The number of misclassified patterns is 3, and precision performance is 98%. The final fuzzy classification is both accurate and interpretable.

In order to illustrate performance of proposed approach, Tab.2 gives the comparison of other studies. The results show that the developed method can build accurate fuzzy classification system with few fuzzy rules and fuzzy sets. Russo [26] classifies all patterns correctly, however it is difficult to interpret the model for containing too many fuzzy rules and fuzzy sets, and its generalization capacity is unknown.

Tab.2. Comparison of different methods

	Number offuzzy sets	Number of fuzzy rules	Precision (%)
Wang [21]	11	3	97.5
Wu [22]	9	3	96.2
Shi [23]	12	4	98
Ishibuchi [24]	7	5	98
Tong [25]	12	3	98
This paper	6	3	98
Russo [26]	18	5	100

To estimate the performance of the proposed method on unseen data, the five-fold cross-validation experiment is performed on the Iris data. In the five-fold cross-validation test, Iris data are divided into five disjoint groups containing 30 different patterns each, with ten patterns belonging to each class. Then we derive fuzzy classification system via proposed approach on all data outside one group and test the obtained fuzzy classification system on data of that group. The result is 96.67%. Tong *et al*. [25] gives the result of 95.53%.

V. Future Work

This paper proposes an approach to construct interpretable fuzzy classification system based on fuzzy clustering initialization. However, for high-dimensional classification problems, there are always too many features, so feature selection and rule reduction are needed to yield more interpretable and accurate fuzzy classification system. This is the future work of this paper.

Acknowledgement

The paper is supported by National Natural Science Foundation (60332020) and scientific research foundation of Nanjing University of Science and Technology (2005)

Reference

[1] L. I. Kuncheva. "*Fuzzy classifier design* (*Studies in Fuzziness and Soft Computing*)". New York: Heidelberg, 2000.
[2] S. Abe, R. Thawonmas. "A fuzzy classifier with ellipsoidal regions". IEEE Trans. Fuzzy Systems, 1997, 5: 358–368.
[3] Y. Shi, R. Eberhart, Y. Chen. "Implementation of evolutionary fuzzy system". IEEE Trans. Fuzzy Systems, 1999, 7:109–119.
[4] G. Castellano, A. M. Fanelli1. "Modeling fuzzy classification systems with compact rule base". 1999 International Conference on Computational Intelligence for Modeling, Control and Automation. Vienna, Austria, 1999
[5] R. Babuska, H. Bersini, D. A. Linkens. "Future Prospects for Fuzzy Systems and Technology". http://www.erudit.de/erudit/newsletters/news_61/page5.htm.
[6] Y. Jin. "*Advanced fuzzy systems design and applications*". New York: Physical-Verl, 2003.
[7] H. Ishibuchi, T Yamamoto. "Effects of three-objective genetic rule selection on the generalization ability of fuzzy rule-based systems". Proc. 2nd Int. Conf. Evolutionary Multi-Criterion Optimization, Faro, Portugal, 2003.4.608–622.
[8] J. Abonyi, H. Roubos, F. Szeifert. "Data-Driven Generation of Compact, Accurate, and Linguistically Sound Fuzzy Classifiers Based on a Decision Tree Initialization". International Journal of Approximate Reasoning, 2003, 32: 1-21.
[9] X. Chang, J. H. Lilly. "Evolutionary Design of a Fuzzy Classifier from Data".

IEEE Trans System Man Cybernetic Part B. 2004, 34(4): 1894-906.

[10] D. Gustafson, W. Kessel. “Fuzzy clustering with a fuzzy covariance matrix”. Proc of IEEE conf on decision and control. San Diego, USA, 1979: 761-766.

[11] J. C. Bezdek. “*Pattern Recognition with fuzzy objective algorithm*”. New York: Plenum Press, 1981.

[12] X. L. Xie, G. Beni. “A Validity Measure for Fuzzy Clustering”. IEEE Trans on Pattern Analysis and Machine Intelligence, 1991, 13(8): 841-847.

[13] M. Roubens. “Pattern Classification Problems and Fuzzy Sets”. Fuzzy Sets and Systems. 1978, 1:239-253.

[14] F. F. Rivera, E. L. Zapata, J M Carazo. “Cluster validity based on the hard tendency of the fuzzy classification”. Pattern Recognition Letters. 1990, 11:7-12.

[15] I. Gath, A. B. Geva. “Fuzzy clustering for the estimation of the parameters of the components of mixtures of normal distributions”. Pattern Recognition Letters1989, 9:77-86.

[16] X. B. Gao. “*Fuzzy cluster analysis and its applications*”. Xi’an: Xidian University Publisher, 2004.

[17] M. Setnes, H. Hellendoorn. “Orthogonal transforms for ordering and reduction of fuzzy rules”. Proc. of the 9th IEEE International Conference on Fuzzy Systems, San Antonio, 2000, 700-705.

[18] L. Castillo. A. Gonzalez, R. perez. “Including a simplicity criterion in the selection of the best rule in a genetic fuzzy learning algorithm”. Fuzzy Sets and System. 2001, 120:309-321.

[19] M. Setnes, R. Babuska, U. Kaymak, H. R. N. Lemke. “Similarity Measures in Fuzzy Rule Base Simplification”. IEEE Trans on Systems Man and Cybernetics Part B. 1998, 28(3): 376-386.

[20] M. Gen, R. Cheng. “*Genetic algorithm and engineering design*”. New York: Wiley, 1999.

[21] J. S. Wang, G. C. S. Lee. “Self-adaptive neuro-fuzzy inference system for classification application”. IEEE Trans. Fuzzy System. 2002, 10: 790–802.

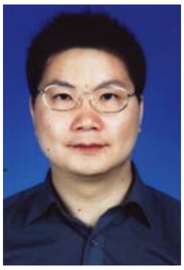
[22] T. P. Wu, S. M. Chen. “A new method for constructing membership functions and fuzzy rules from training examples”. IEEE Trans System Man Cybernetic Part B.1999, 29: 25–40.

[23] Y. Shi, R. Eberhart, Y. Chen. “Implementation of evolutionary fuzzy system”. IEEE Trans. Fuzzy System. 1999, 7: 109–119.

[24] H. Ishibuchi, T. Nakashima, T. Murata. “Three-objective genetic-based machine learning for linguistic rule extraction”. Information Science.2001, 136: 109–133.

[25] S. H. Tong, Y. Shen, Z .Y. Liu. “Approach to construct fuzzy classification system with clustering”. Control and decision. 2001, 16:737-740.

[26] M. Russo. “Genetic fuzzy learning. IEEE Trans. Evolutionary Computation”. 2000, 4: 259–273.



Xing Zong-Yi, Postdoctor at Nanjing University of Science and Technology. His research interests include fuzzy modeling and intelligent control of industry process.

Zhang Yong, Doctor candidate at Nanjing University of Science and Technology. His research interests are intelligent control and system, nonlinear control theory and application.

Jia Li-Min, Professor at Beijing Jiaotong University. His research interests include fuzzy set and system, intelligent control and application, railway intelligent transportation system.

Hu Wei-li was born in Jiangsu, China in January, 1941. He received his Bachelor degree in Automation Department in 1965 from Tsinghua University, and his Master's degree in Automatic Control in 1981 from Nanjing University of Science & Technology. He became a teaching assistant in the Department of Automatic Control in NUST in 1981. Since 1983, he was a Lecturer. From 1988 to 1992, he was a Associate Professor. He is a Professor since 1992, still in the Automation Department in NUST. His research interest is in Robust Adaptive control and nonlinear control, AC servo systems, networked control systems. He is the author of 80 journal and conference papers, and received 9 prizes of Science and Technology Progressive Award from a Ministry and Jiangsu Province of China.