Pruning and Learning Fuzzy Rule-Based Classifier

Do Wan Kim*, Jin Bae Park*, and Young Hoon Joo**

*Department of Electrical and Electronic Engineering, Yonsei University, Seodaemun-gu, Seoul, 120-749, Korea,
(e-mail: dwkim@control.yonsei.ac.kr; jbpark@control.yonsei.ac.kr)

**School of Electronic and Information Engineering, Kunsan National University, Kunsan, Chonbuk, 573-701, Korea,
(e-mail: yhjoo@kunsan.ac.kr).

Abstract: This paper presents new pruning and learning methods for the fuzzy rule-based classifier. The structure of the proposed classifier is framed from the fuzzy sets in the premise part of the rule and the Bayesian classifier in the consequent part. For the simplicity of the model structure, the unnecessary features for each fuzzy rule are eliminated through the iterative pruning algorithm. The quality of the feature is measured by the proposed correctness method, which is defined as the ratio of the fuzzy values for a set of the feature values on the decision region to one for all feature values. For the improvement of the classification performance, the parameters of the proposed classifier are finely adjusted by using the gradient descent method so that the misclassified feature vectors are correctly re-categorized. The cost function is determined as the squared-error between the classifier output for the correct class and the sum of the maximum output for the rest and a positive scalar. Then, the learning rules are derived from forming the gradient. Finally, the fuzzy rule-based classifier is tested on two data sets and is found to demonstrate an excellent performance.

Keywords: Fuzzy rule-based classifier, Bayesian classifier, pruning, learning, gradient descent method.

1. Introduction

The fuzzy rule-based classifier is a popular counter part of a fuzzy control system and a fuzzy modeling [1, 2], which carries out the pattern classification by using the membership grades of the feature variables. The point to be observed in the fuzzy rule-based classifier is that it takes the shape of the linguistic form and the discriminant function [3]. In numerous researches[4–11, 13–15], the excellent capabilities to the pattern classification of the fuzzy classifier have been shown.

In the design of the fuzzy rule-based classifier, there are two main issues which involve the model complexity and the classification performance. If too many free parameters are used, there is a danger of overfitting; conversely, if too few parameters are used, the training set may not be learned. Thus, the design process can be divided into two strategies: the feature selection and the learning. For the simplicity of the model, one possibility is to reduce the dimensionality by selecting an appropriate subset of the existing features. The key point in the feature selection is the measure of the quality of a set of the features, which concerns some measure of the predictive power of the features. An attractive approach [4–6] is to measure the similarity of the overlapping degree between two fuzzy sets. For the improvement of the classification performance, the learning in the fuzzy rule-based classifier can be formulated as minimizing a cost function. In [7, 8], the cost function was selected the squared-error between the classifier output and the desired value, and then the learning rules are derived from forming the gradient. The classifier outputs for the correct class and the rest are the upper bound and the lower bound, respectively, as the cost function approaches zero.

Despite of the existence of the excellent previous researches [4–8], there are still critical issues. In [4–6], the overlapping degree does not become the accurate criterion in respect of measuring the risk of the classification error because not all data on the overlapping region can be judged as the classification error. And also, a way to measure some degree between two classes may not be efficient in the multicategory case. In [7,8], the desired values for the correct class and the rest are defined as the upper bound and the lower bound of the classifier outputs, respectively. However, these desired values are very strict conditions for the learning objective, which is that the classifier output for the correct class gets bigger than one for the rest as the cost function approaches zero. These strict conditions make the learning inefficient. In addition, this approach does not used in the case of the classifier with the unbounded outputs.

This paper aims at developing the fuzzy rule-based classifier for the model complexity and the classification capability. The main contributions of this paper are to measure the qualities of the given feature variables by using a new analysis technique of the fuzzy sets, *correctness method* and to derive the learning rules for the classifier without the upper bound. Specifically, the proposed design procedure consists of three steps. The first step is to construct the initial fuzzy rule-based classifier. In the premise part, the fuzzy set is characterized by the Gaussian membership function, where the premise parameters are initially identified by the mean and the standard deviation of the training set. In the consequent part, the Bayesian classifier is applied. Thus, the proposed classifier can be viewed as a combining classifier applied in the fuzzy classifier and the Bayesian classifier. In the next step, an appropriate subset of the existing features for each fuzzy rule is selected through the iterative pruning algorithm. To measure the quality of the features, the correctness degree is defined as the ratio of the fuzzy values for a set of the feature values on the decision region to one for all feature values. In the final step, the premise parameters

This paper is supported by my funding agencies.

are finely tuned so that the misclassified feature vectors are correctly re-categorized. For this purpose, the cost function is determined as the squared-error between the classifier output for the correct class and the sum of the maximum output for the rest and a positive scalar. Then, the learning rules are derived from forming the gradient. Finally, to show the feasibility of the proposed algorithms, computer simulations are provided.

The layout of this paper is organized as follows: Section 2. reviews the fuzzy classifier and the Bayesian classifier. In Section 3, the structure of the fuzzy rule-based classifier is developed, the iterative pruning algorithms is proposed, and the parameters of the fuzzy sets are adjusted by the gradient descent method. Section 4performs pattern classification for the Iris data and the glass data to show the effectiveness of the proposed method. Section 5concludes this paper.

2. Preliminaries

Let $\mathcal{W} = \{w_1, w_2, \dots, w_m\}$ be a set of class labels and $\mathbf{x} = [x_1, x_2, \dots, x_n]^T \in \Re^n$ be a feature vector. A classifier is any mapping

$$D : \Re^n \longrightarrow \mathcal{W} \tag{1}$$

There are many different ways to represent pattern classifiers. One of the most useful is in terms of a set of m discriminant functions $d_i(\mathbf{x})$, $i \in \mathcal{I}_m = [1, 2, \dots, m]$, expressing the support for the respective classes. The classifier is said to assign a feature vector $\mathbf{x}$ to the class i if

$$d_i(\mathbf{x}) > d_j(\mathbf{x}), \qquad \forall j \neq i, j \in \mathcal{I}_m \tag{2}$$

Both a fuzzy classifier and a Bayesian classifier are easily represented in this way.

There are various fuzzy models to do pattern classification, but the most general [4, 9–11] is

$$\begin{aligned} R_i :\ & \text{IF } x_1 \text{ is } A_{i1} \text{ and } \dots \text{ and } x_n \text{ is } A_{in}, \\ & \text{THEN the class is } i, \end{aligned} \tag{3}$$

where $R_i, i \in \mathcal{I}_m$, denotes the ith fuzzy rule, $x_h, h \in \mathcal{I}_n = [1, 2, \dots, n]$, is the hth feature, and $A_{ih}, (i, h) \in \mathcal{I}_m \times \mathcal{I}_n$, is the fuzzy set. The conjunction rule to transform the fuzzy sets into a discriminant function is

$$d_i^F(\mathbf{x}) = \mathcal{A}\{A_{i1}(x_1), A_{i2}(x_2), \dots, A_{in}(x_n)\} \tag{4}$$

where $\mathcal{A}$ means a certain combination with inference engines and fuzzifiers. The fuzzy approaches would lie in guiding the steps by which one takes knowledge in a linguistic form and casts it into discriminant functions[3].

Using the prior probabilities $P(\mathbf{x})$ and the conditional densities $P(\mathbf{x}|w_i)$, especially, the multivariate Gaussian model, the Bayesian classifier is designed by the following discriminant function:

$$d_i^B(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m}_i)^T {\Sigma_i}^{-1} (\mathbf{x} - \mathbf{m}_i) \right) P(w_i) \tag{5}$$

where $\mathbf{m}_i = [m_{i1}^B, m_{i2}^B, \dots, m_{in}^B]^T$ is the n-component mean vector, Σ_i is the $n \times n$ covariance matrix, and $|\Sigma_i|$ and ${\Sigma_i}^{-1}$ are its determinant and inverse, respectively. For example, $d_i^B(\mathbf{x}), i \in \mathcal{I}_2$, divides the feature space into two decision regions $\mathcal{R}_1^B$ and $\mathcal{R}_2^B$. Due to the robust performance and simple implementation, the Bayesian classifier has become a popular classification tool in recent year [18].

3. Design Process of Fuzzy Rule-Based Classifier

This section describes all the phases of the proposed design process, which contains new methods to prune the unnecessary features and to tune the parameters of the fuzzy model-based classifier.

3.1. Initial Fuzzy Rule-Based Classifier

For a given feature vector $\mathbf{x}$, the proposed fuzzy rule-based classifier is formulated in the following form.

$$\begin{aligned} R_i :\ & \text{IF } x_1 \text{ is } A_{i1} \text{ and } \dots \text{ and } x_n \text{ is } A_{in}, \\ & \text{THEN } y_i = f_i(\mathbf{x}) \end{aligned} \tag{6}$$

where y_i is the output vector of R_i, and $f_i(\mathbf{x})$ is the discriminant function. Here, we equate the number of the classes to the one of the fuzzy rules. The premise parts of (6) divide the feature space into the number of fuzzy regions by the fuzzy sets, while the consequent parts describe the outputs of the classifier in these regions.

In the premise part of (6), the fuzzy set A_{ih} is characterized by the following Gaussian membership function.

$$A_{ih}(x_h) = \exp\left(-\frac{1}{2} \left(\frac{x_h - m_{ih}^F}{\sigma_{ih}^F} \right)^2 \right) \tag{7}$$

In the consequent part, the Bayesian classifier (5) is selected as $f_i(\mathbf{x})$. For simplicity, we assume that the covariance matrix Σ_i is the following diagonal matrix which contains the variance $(\sigma_{ih}^B)^2$.

$$\Sigma_i = \begin{bmatrix} (\sigma_{i1}^B)^2 & 0 & \dots & 0 \\ 0 & (\sigma_{i2}^B)^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & (\sigma_{in}^B)^2 \end{bmatrix} \tag{8}$$

Then, the Bayesian classifier (5) can be described by

$$d_i^B(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{h=1}^n \sigma_{ih}^B} \exp\left(-\frac{1}{2} \sum_{h=1}^n \left(\frac{x_h - m_{ih}^B}{\sigma_{ih}^B} \right)^2 \right) \tag{9}$$

The unknown parameters m_{ih}^F, m_{ih}^B, σ_{ih}^F, and σ_{ih}^B are initially identified by using the arithmetic average and the standard deviation.

The ith final output of (6) is inferred as follows:

$$\widehat{y}_i(\mathbf{x}) = d_i^F(\mathbf{x}) d_i^B(\mathbf{x}) \tag{10}$$

where $d_i^F(\mathbf{x}) = \max_{h \in \mathcal{I}_n} A_{ih}(x_h)$ by using the maximum inference engine and the singleton fuzzifier.

Remark 1: From the final output (10), the proposed fuzzy rule-based classifier can be viewed as the combining classifier, which is constructed by using the class-conscious fusion operator [12], the product.

From (10), the proposed classifier is said to assign a feature vector $\mathbf{x}$ to the class i if

$$\widehat{y}_i(\mathbf{x}) > \widehat{y}_j(\mathbf{x}), \quad \forall j \neq i \tag{11}$$

3.2. Pruning the Unnecessary Features for Each Fuzzy Rule

We suggest a pruning algorithm based on the analysis of the fuzzy sets for eliminating the unnecessary feature variables for each fuzzy rule. The difficult point of selecting the unnecessary feature is that pruning the feature has direct impact on the classification performance. The main reason is that the decision regions are changed by eliminating the labelled feature. Thus, in the analysis of the fuzzy set, the class label and the decision region are considerable matters. In addition, for the general application, the analysis technique is efficiently usable in the multicategory case. Dealing such issues are formulated as follows:

Problem 1: **(The analysis of the fuzzy set for pruning the feature)** If the analysis technique of the fuzzy set is used for checking to whether the feature is unnecessary for the supervised learning, it is sufficiently satisfied with following conditions:

- (i) The analysis tool must consider the class label and the decision region, which are important concepts in the supervised learning.
- (ii) The analysis technique of the fuzzy set must be simply applicable in the multicategory case.

To resolve Problem 1, the correctness degree of the fuzzy set A_{ih} for the hth feature x_h labelled as w_i is measured by using the cardinality, and it is defined as follows.

Definition 1: The correctness degree of A_{ih} is defined as

$$C(A_{ih}) = \frac{|A_{ih}|_{(x_h \in w_i) \in \mathcal{R}_{ih}}}{|A_{ih}|_{x_h \in w_i}} \tag{12}$$

where $|\cdot|$ denotes the cardinality of a set, and $\mathcal{R}_{ih}$ is one-dimensional fuzzy region according to $d_i^F(x_h) > d_j^F(x_h)$ for all $j \neq i$.

Remark 2: Compared with similarity method [6], the proposed correctness method (12) has the following advantages:

- (i) While, in the similarity method, the feature variable is only considered, in the correctness method, the class label as well as the feature variable are considered.
- (ii) While, in the similarity method, the overlapping of the fuzzy set is analyzed, in the correctness method, both the overlapping and the categorizing can be simultaneously analyzed.
- (iii) Compared with the similarity method, especially, in the multicategory case, the correctness method is easy to implement.

By applying the definition of the cardinality, (12) becomes

$$C(A_{ih}) = \frac{\sum_{(x_h \in w_i) \in \mathcal{R}_{ih}} A_{ih}(x_h)}{\sum_{x_h \in w_i} A_{ih}(x_h)} \tag{13}$$

Specifically, if A_{ih} is not overlapped with the others and/or all $x_h \in w_i$ fall on $\mathcal{R}_{ih}$, then $C(A_{ih}) = 1$. Conversely, if A_{ih} is completely overlapped with the others and/or no $x_h \in w_i$ falls on $\mathcal{R}_{ih}$, $C(A_{ih}) = 0$.

By using the correctness degree (12), we set the criterions of selecting the fuzzy rule and the feature, which are applied in the proposed pruning algorithm. To select the fuzzy rule, the following average correctness degree is employed.

$$\bar{C}(A_{ih}) = \frac{1}{n} \sum_{h=1}^{n} \frac{\sum_{(x_h \in w_i) \in \mathcal{R}_{ih}} A_{ih}(x_h)}{\sum_{x_h \in w_i} A_{ih}(x_h)} \tag{14}$$

The more large the value of (14) for any fuzzy rule is, the more separable the class described by the rule may be from the others. In this aspect, we know that the fuzzy rule corresponding to the large average correctness degree can describe the class with relatively small number of the feature variables. Thus, the fuzzy rule is selected in the order of large value of (14). On the other hand, the feature is selected in the order of small value of (12) because the feature corresponding to the small value of (12) has the bad impact on improving the recognition rate.

According to the above discussion, the proposed pruning algorithm becomes as follows.

Step 1 Select the fuzzy rule in the order of large value of (14).

Step 2 Prune any feature of the selected fuzzy rule that result in improving the recognition rate, where the feature is selected in the order of small value of (12).

Step 3 If no feature of the selected fuzzy rule is pruned, stop the algorithm; otherwise, update (12) and then repeat by going to Step 1.

Remark 3: In the current implementation of the algorithm after the initial model formulation, the unnecessary feature variables are eliminated by continuously checking the correctness degrees and the decrease of the performance of the classifier.

3.3. Learning Using Gradient Descent Method

Our goal is to develop a learning technique, which can be formulated as minimizing a cost function $\mathcal{J}$, for the parameters of (6) so that the misclassified feature vectors until the preceding step are correctly re-categorized. To this end, we formulate the following problem of adjusting the parameters for the misclassified feature vectors.

Problem 2: **(Learning parameters for the misclassified feature vectors)** To correctly categorize the misclassified feature vector $\mathbf{x}$ labelled as w_i, the learning method should be sufficiently satisfied with the following condition: The parameters σ_{ih}^F, σ_{ih}^B, m_{ih}^F, and m_{ih}^B of (6) should be finely adjusted so as to satisfy $\widehat{y}_i(\mathbf{x}) > \widehat{y}_j(\mathbf{x})$ for all $j \neq i$.

The difficult point for constructing the cost function $\mathcal{J}$ is to find the desired value of $\widehat{y}_i(\mathbf{x})$. The main reasons are as follows: First, we must deal with the inequality condition such as the Problem 2. Furthermore, any fixed upper bound of $\widehat{y}_i(\mathbf{x})$ is not existed since the upper bound $\frac{1}{(2\pi)^{\frac{n}{2}} \prod_{h=1}^{n} \sigma_{ih}^B}$ will vary as σ_{ih}^B is adjusted. To resolve these difficulty, define

Table 1. Classification results on glass data

Design procedure	Avg. number of features for each fuzzy rule	Avg. training recognition rate	Avg. testing recognition rate
Initial	9	47.44%	39.48%
Pruning	5.61	64.11%	53.23%
Learning	5.61	68.90%	58.28%

the following desired value of $\widehat{y}_i(\mathbf{x})$.

$$\max_{j\in\mathcal{I}_m, j\neq i} \widehat{y}_j(\mathbf{x}) + \epsilon \tag{15}$$

where ϵ is hopefully a small positive scalar. Thus, we can seek the parameters m_{ih}^F, σ_{ih}^F, m_{ih}^B, and σ_{ih}^B that minimize some function of the error between $\widehat{y}_i(\mathbf{x})$ and $\max_{j\in\mathcal{I}_m, j\neq i} \widehat{y}_j(\mathbf{x}) + \epsilon$. If error e is given by

$$e = \max_{j\in\mathcal{I}_m, j\neq i} \widehat{y}_j(\mathbf{x}) + \epsilon - \widehat{y}_i(\mathbf{x}) \tag{16}$$

then one approach is to try to minimize the squared-error. This is equivalent to minimizing the following cost function.

$$\mathcal{J} = \frac{(\max_{j\in\mathcal{I}_m, j\neq i} \widehat{y}_j(\mathbf{x}) + \epsilon - \widehat{y}_i(\mathbf{x}))^2}{2} \tag{17}$$

It is noticed that, because of $\epsilon > 0$, $\widehat{y}_i(\mathbf{x})$ obviously gets bigger than $\widehat{y}_j(\mathbf{x})$ for all $j \neq i$ as $\mathcal{J}$ approaches zero.
The problem of minimizing the squared-error can be numerically solved by a gradient descent method. The following theorems suggest the gradient descent method to finely tune the parameters σ_{ih}^F, σ_{ih}^B, m_{ih}^F, and m_{ih}^B of (6).

Theorem 1: Given the misclassified feature vector $\mathbf{x}$ labelled as w_i, the parameters σ_{ih}^F, σ_{ih}^B, m_{ih}^F, and m_{ih}^B of (6) can be precisely adjusted by the following learning rules, respectively: For the class i

$$\begin{aligned}\Delta m_{ih}^F &= \alpha_1(\max_{j\in\mathcal{I}_m, j\neq i} d_j^F(\mathbf{x})d_j^B(\mathbf{x}) - d_i^F(\mathbf{x})d_i^B(\mathbf{x}) + \epsilon) \\ &\quad\times d_i^B(\mathbf{x})\frac{\partial d_i^F(\mathbf{x})}{\partial m_{ih}^F} \end{aligned}\tag{18}$$

$$\begin{aligned}\Delta \sigma_{ih}^F &= \alpha_2(\max_{j\in\mathcal{I}_m, j\neq i} d_j^F(\mathbf{x})d_j^B(\mathbf{x}) - d_i^F(\mathbf{x})d_i^B(\mathbf{x}) + \epsilon) \\ &\quad\times d_i^B(\mathbf{x})\frac{\partial d_i^F(\mathbf{x})}{\partial \sigma_{ih}^F} \end{aligned}\tag{19}$$

and, for the class j

$$\begin{aligned}\Delta m_{jh}^F &= -\beta_1(\max_{j\in\mathcal{I}_m, j\neq i} d_j^F(\mathbf{x})d_j^B(\mathbf{x}) - d_i^F(\mathbf{x})d_i^B(\mathbf{x}) + \epsilon) \\ &\quad\times d_j^B(\mathbf{x})\frac{\partial d_j^F(\mathbf{x})}{\partial m_{jh}^F} \end{aligned}\tag{20}$$

$$\begin{aligned}\Delta \sigma_{jh}^F &= -\beta_2(\max_{j\in\mathcal{I}_m, j\neq i} d_j^F(\mathbf{x})d_j^B(\mathbf{x}) - d_i^F(\mathbf{x})d_i^B(\mathbf{x}) + \epsilon) \\ &\quad\times d_j^B(\mathbf{x})\frac{\partial d_j^F(\mathbf{x})}{\partial \sigma_{jh}^F} \end{aligned}\tag{21}$$

where α_1, α_2, β_1, and β_2 are the learning rates for m_{ih}^F, σ_{ih}^F, m_{jh}^F, and σ_{jh}^F, respectively.

Theorem 2: The consequent parameters of the fuzzy model can be finely adjusted by the following learning rules: for the class i

$$\Delta m_{ih}^B = \gamma_1(\max_{j\in\mathcal{I}_m, j\neq i} d_j^F(\mathbf{x})d_j^B(\mathbf{x}) - d_i^F(\mathbf{x})d_i^B(\mathbf{x}) + \epsilon)d_i^F\frac{\partial d_i^B(\mathbf{x})}{\partial m_{ih}^B} \tag{22}$$

$$\Delta \sigma_{ih}^B = \gamma_2(\max_{j\in\mathcal{I}_m, j\neq i} d_j^F(\mathbf{x})d_j^B(\mathbf{x}) - d_i^F(\mathbf{x})d_i^B(\mathbf{x}) + \epsilon)d_i^F\frac{\partial d_i^B(\mathbf{x})}{\partial \sigma_{ih}^B} \tag{23}$$

for the class j

$$\begin{aligned}\Delta m_{jh}^B &= -\delta_1(\max_{j\in\mathcal{I}_m, j\neq i} d_j^F d_j^F(\mathbf{x})d_j^B(\mathbf{x}) - d_i^F(\mathbf{x})d_i^B(\mathbf{x}) + \epsilon) \\ &\quad\times d_j^F\frac{\partial d_j^B(\mathbf{x})}{\partial m_{jh}^B} \end{aligned}\tag{24}$$

$$\begin{aligned}\Delta \sigma_{jh}^B &= -\delta_2(\max_{j\in\mathcal{I}_m, j\neq i} d_j^F d_j^F(\mathbf{x})d_j^B(\mathbf{x}) - d_i^F(\mathbf{x})d_i^B(\mathbf{x}) + \epsilon) \\ &\quad\times d_j^F(\mathbf{x})\frac{\partial d_j^B(\mathbf{x})}{\partial \sigma_{jh}^B} \end{aligned}\tag{25}$$

where γ_1, γ_2, δ_1, and δ_2 are the learning rates for m_{ih}^B, σ_{ih}^B, m_{jh}^B, and σ_{jh}^B, respectively.
Remark 4: Through the learning algorithm after the initial model formulation and the elimination of the unnecessary feature variables, the final classifier is identified by adjusting the premise and consequent parameters for the misclassified feature vectors.

4. Computer Simulations

4.1. Glass Data

The glass data set [17] is based on the chemical analysis of glass splinters. Nine feature variables are used to classify six types of glass: building windows float processed, building windows non float processed, vehicle windows float processed, containers, tableware, and headlamps. The feature variables are refractive index, sodium, magnesium, aluminum, silicon, potassium, calcium, barium, and iron. The unit of measurement of all features but refractive index is weight percent in corresponding oxide.
We attempt to perform the 25 times pattern classification. One half of 214 feature vectors are randomly selected as the training data and the other half are used as the testing data. Table 1 contains the simulation results of the proposed classifier for each step of the design procedure. Although the average number of the feature variables reduces from 9 to 5.61, the average recognition rates of the training and the testing set increase from 47.44% to 68.90% and from 39.48% to 58.28%, respectively. That definitely shows that the proposed design algorithm effectively provides the robustness for the overfitting and the decline of the dimensionality. Moreover, the classification performance of the proposed classifier is better than other classifiers as shown in Table 2.

Table 2. Comparison of classification results on glass data

Ref.	Avg. testing recognition rate
[19]	50.95%
[20]	52.70%
Ours	58.28%

5. Conclusions

In this paper, a novel design approach to the fuzzy rule-based classifier has been proposed for the model complexity and the classification performance. Unlike other pruning methods based on the similarity analysis between two fuzzy sets, the proposed method utilizes the correctness degree, which is the major factor that improves the simplicity of the model. In addition, the problem of learning the premise parameters is formulated as minimizing the cost function, which is determined as squared-error between the classifier output for the correct class and the sum of the maximum output for the rest and a positive scalar. Finally, the computer simulations for the Iris data and the glass data are given. The results show that the proposed fuzzy rule-based classifier has the low complexity, the very accurate classification ability, and the robustness for the overfitting in comparison with the conventional classifier. It indicates the great potential for reliable application of the pattern recognition.

References

[1] Y. H. Joo, H. S. Hwang, K. B. Kim, and K. B. Woo, "Linguistic model identification for fuzzy system," *Electron. Letter*, vol. 31, pp. 330–331, 1995.

[2] Y. H. Joo, H. S. Hwang, K. B. Kim, and K. B. Woo, "Fuzzy system modeling by fuzzy partition and GA hybrid schemes," *Fuzzy Set and Syst.*, vol. 86, pp. 279–288, 1997.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification,* A wiley-interscience publishing company, inc., 2001.

[4] T. P. Wu and S. M. Chen, "A new method for constructing membership functions and fuzzy rules from training examples," *IEEE Trans. Syst., Man, Cybern. B.*, vol. 29, pp. 25–40, 1999.

[5] H. Roubos and M. Setnes, "Compact transparent fuzzy models and classifiers through iterative complexity reduction," *IEEE Trans. Fuzzy Systems*, vol. 9, no. 4, pp. 516–524, 2001.

[6] M. Setnes and H. Roubos, "GA-fuzzy modeling and classification: complexity and performance," *IEEE Trans. Fuzzy Systems*, vol. 8, pp. 509–522, 2000.

[7] N. R. Pal and S. Chakraborty,"Fuzzy rule extraction from ID3-type decision trees for real data," *IEEE Trans. Syst., Man, Cybern. B.*, vol. 31, pp. 745–754, 2001.

[8] S. Paul and S. Kumar, "Subsethood based adaptive linguistic networks for pattern classification," *IEEE Trans. Syst., Man, Cybern. C.*, vol. 33, pp. 248–258, 2003.

[9] H. Ishibuchi, T. Murata, and I. B. Turksen, "Single-objective and two-objective genetic algorithms for selecting linguistic rules for pattern classification problems," *Fuzzy Sets Syst.*, vol. 89, pp. 135–149, 1997.

[10] S. Abe and R. Thawonmas, "A fuzzy classifier with ellipsoidal regions," *IEEE Trans. Fuzzy Systems*, vol. 5, pp. 358–368, 1997.

[11] R. Thawonmas and S. Abe, "A novel approach to feature selection based on analysis of class regions," *IEEE Trans. Syst., Man, Cybern. B.*, vol. 27, pp. 196–207, 1997.

[12] L. I. Kuncheva, J. C. Bezdek, and R. P. W. Duin, "Decision templates for multiple classifier fusion: an experimental comparison," *Pattern Recognition*, vol. 34, pp. 299–314, 2001.

[13] Y. Shi, R. Eberhart, and Y. Chen, "Implementation of evolutionary fuzzy systems," *IEEE Trans. Fuzzy Systems*, vol. 7, pp. 109–119, 1999.

[14] R. Li, M. Mukaidono, and I. B. Turksen, "A fuzzy neural network for pattern classification and feature selection," *Fuzzy Sets Syst.*, vol. 130, pp. 101–108, 2002.

[15] T. P. Hong and J. B. chen, "Processing individual fuzzy attributes for fuzzy rule induction," *Fuzzy Sets Syst.*, vol. 112, pp. 127–140, 2000.

[16] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics.*, vol. 7, no. 2, pp. 179−188, 1936.

[17] C. J. Merz and P. M. Murphy, "UCI repository of machine learning databases," http://www.ics.uci.edu/ mlearn/MLRepository.html, Irvine, Dept. of Information and Computer Science, Univ. of California, Irvine, 1996.

[18] H. Huang and C. Hsu, "Bayesian classification for data from the same unknown class," *IEEE Trans. Syst., Man, Cybern. B.*, vol. 32, pp. 137–145, 2002.

[19] G. Castellano, A. M. Fanelli, and C. Mencar, "An empirical risk functional to improve learning in a neuro-fuzzy Classifier," *IEEE Trans. Syst., Man, Cybern. B.*, vol. 34, pp. 725–731, 2004.

[20] N. R. Pal and S. Chakraborty, "Fuzzy rule extraction from ID3-type decision trees for real data," *IEEE Trans. Syst., Man, Cybern. B.*, vol. 31, pp. 745–754, 2001.