3rd International Workshop on Genetic and Evolving Fuzzy Systems
Witten-Bommerholz, Germany, March 2008

Designing Fuzzy Rule-Based Classifiers That Can Visually Explain Their Classification Results to Human Users

Hisao Ishibuchi, *Member, IEEE,* Yutaka Kaisho, and Yusuke Nojima, *Member, IEEE*

Abstract - **In various application areas of fuzzy rule-based systems, human users want to know why a particular reasoning result is obtained. That is, fuzzy rule-based systems are required to have high explanation ability. In this paper, we propose an approach to the design of fuzzy rule-based classifiers that can visually explain their classification results to human users. That is, our fuzzy rule-based classifiers can explain to human users why an input pattern is classified as a particular class in an understandable manner. The proposed approach consists of a rule selection method and a visualization interface. Our idea is to design fuzzy rule-based classifiers using fuzzy rules with only two antecedent conditions. A genetic algorithm is employed to construct a compact fuzzy rule-based classifier by choosing only a small number of fuzzy rules. In the classification phase, we use a single winner rule-based method for classifying an input pattern. The classification result of the input pattern is visually explained in a two-dimensional space where the two antecedent conditions of the winner rule are defined. Our approach is compared with feature selection by computational experiments.**

I. Introduction

THE main advantage of fuzzy rule-based systems over other non-linear systems such as neural networks is their interpretability. Usually fuzzy rule-based systems are easily understood through linguistic interpretation of each fuzzy rule in a human understandable manner. In the design of fuzzy rule-based systems, emphasis has often been placed on both their accuracy and interpretability. Thus it is essential to find a good accuracy-interpretability tradeoff [1], [2].

A number of approaches have been proposed to improve both the accuracy and the interpretability of fuzzy rule-based systems [3]-[5]. In those approaches, fuzzy rule-based system design can be viewed as a single-objective optimization problem of the following integrated objective function of these two objectives:

$$f(S) = f(Accuracy(S), Interpretability(S)), \qquad (1)$$

where S is a fuzzy rule-based system to be optimized, $f(S)$ is an objective function, $Accuracy(S)$ is an accuracy measure, and $Interpretability(S)$ is an interpretability measure. Various techniques in genetic fuzzy systems [6]-[8] can be employed to optimize the integrated objective function in (1).

Multiobjective approaches have been proposed to search for non-dominated fuzzy rule-based systems along the accuracy-complexity tradeoff surface in the literature [9]-[12]. In those studies, a large number of non-dominated fuzzy rule-based systems are obtained by a single run of multiobjective optimizers such as multiobjective genetic algorithms.

The definition of the interpretability of fuzzy rule-based systems is an important research issue. There exist a number of aspects related to the interpretability of fuzzy rule-based systems: the type of fuzzy rules, the number of fuzzy rules, the number of antecedent conditions in each fuzzy rule, the number of input variables, the granularity of fuzzy partitions (i.e., the number of fuzzy sets for each input/output variable), the separability of neighboring fuzzy sets, and the shape of fuzzy sets [1], [2], [13]-[16].

Whereas the interpretability of fuzzy rule-based systems has been discussed from various points of view, the ability to explain their reasoning results has not been taken into account in many studies. This does not mean that the explanation ability of fuzzy rule-based systems is unnecessary. In many application areas involving human decision makers such as business and healthcare, it is important for fuzzy rule-based systems to explain their reasoning results to human users in an understandable manner. For example, human users want to know why a particular action is suggested for the current situation by their fuzzy rule-based decision support system. In classification tasks, human users want to know why an input pattern is classified as a particular class.

In this paper, we propose an approach to the design of fuzzy rule-based classifiers with high explanation ability. Our fuzzy rule-based classifiers can visually explain to human users why an input pattern is classified as a particular class. Our approach consists of a rule selection method and a visualization interface. The point of our approach is to use fuzzy rules with only two antecedent conditions. A small number of fuzzy rules are selected from a large number of candidate rules by a genetic algorithm to design a compact fuzzy rule-based classifier. In the classification phase of an input pattern by the designed classifier, a single winner rule is identified. The classification result, which is the consequent class of the winner rule, is explained in a two-dimensional pattern space where the two antecedent conditions of the winner rule are defined. Our visualization interface shows some related statistics such as the compatibility grade of the input pattern with the winner rule, the certainty grade of the winner rule, and other possible classes of the input pattern.

This work was partially supported by the ICOM Electronic Communication Engineering Promotion Foundation, Japan, and Japan Society for the Promotion of Science (JSPS) through Grand-in-Aid for Scientific Research (B): KAKENHI (17300075).

Hisao Ishibuchi, Yutaka Kaisho and Yusuke Nojima are with the Department of Computer Science and Intelligent Systems, Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan (hisaoi@cs.osakafu-u.ac.jp, kaisho@ci.cs.osakafu-u.ac.jp, nojima@cs.osakafu-u.ac.jp).

In this paper, we first explain fuzzy rule-based classifiers in Section II. Next we explain our approach to the design of fuzzy rule-based classifiers with high explanation ability in Section III. Then we discuss the visualization of classification results in Section IV. We also examine the performance of our fuzzy rule-based classifiers in comparison with a feature selection-based approach in Section V. Finally we conclude this paper in Section VI.

II. Fuzzy Rule-Based Classifiers

A. Fuzzy Rules for Classification Problems

Let us assume that we have m training (i.e., labeled) patterns $\mathbf{x}_p = (x_{p1}, ..., x_{pn}), p = 1, 2, ..., m$ from M classes in the n-dimensional continuous pattern space where x_{pi} is the attribute value of the p-th training pattern for the i-th attribute ($i = 1, 2, ..., n$). For the simplicity of explanation, we assume that all the attribute values have already been normalized into real numbers in the unit interval [0, 1]. Thus the pattern space of our pattern classification problem is an n-dimensional unit-hypercube $[0, 1]^n$.

For our n-dimensional pattern classification problem, we use fuzzy rules of the following type [17]:

$$\text{Rule } R_q : \text{If } x_1 \text{ is } A_{q1} \text{ and ... and } x_n \text{ is } A_{qn} \text{ then Class } C_q \text{ with } CF_q , \quad (2)$$

where R_q is the label of the q-th fuzzy rule, $\mathbf{x} = (x_1, ..., x_n)$ is an n-dimensional pattern vector, A_{qi} is an antecedent fuzzy set ($i = 1, 2, ..., n$), C_q is a class label, and CF_q is a rule weight (i.e., certainty factor). For other types of fuzzy rules for pattern classification problems, see [18]-[21].

Since we usually have no *a priori* information about an appropriate granularity of the fuzzy discretization for each attribute, we simultaneously use multiple fuzzy partitions with different granularities for fuzzy rule generation. In our computational experiments, we use four homogeneous fuzzy partitions with triangular fuzzy sets in Fig. 1. In addition to the 14 fuzzy sets in Fig. 1, we also use the domain interval [0, 1] as an antecedent fuzzy set in order to represent a *don't care* condition [22]. That is, we use the 15 antecedent fuzzy sets for each attribute in our computational experiments.

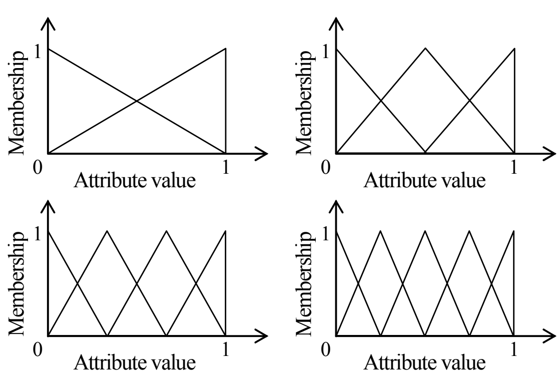

Fig. 1. Fuzzy partitions in our computational experiments.

B. Single Winner Rule-Based Fuzzy Reasoning Method

Let S be a set of fuzzy rules of the form in (2). That is, S is a fuzzy rule-based classifier. When an input pattern $\mathbf{x}_p$ is to be classified by S, first we calculate the compatibility grade of $\mathbf{x}_p$ with the antecedent part $\mathbf{A}_q$ of each fuzzy rule R_q in S using the product operation as

$$\mu_{\mathbf{A}_q}(\mathbf{x}_p) = \mu_{A_{q1}}(x_{p1}) \cdot \ ... \ \cdot \mu_{A_{qn}}(x_{pn}) , \quad (3)$$

where $\mu_{A_{qi}}(\cdot)$ is the membership function of the antecedent fuzzy set A_{qi}. Then a single winner rule R_w is identified from S as follows:

$$\mu_{\mathbf{A}_w}(\mathbf{x}_p) \cdot CF_w = \max\{\mu_{\mathbf{A}_q}(\mathbf{x}_p) \cdot CF_q \mid R_q \in S\} . \quad (4)$$

The input pattern $\mathbf{x}_p$ is classified as the consequent class C_w of the winner rule R_w. When multiple rules with different consequent classes have the same maximum value in (4), the classification of $\mathbf{x}_p$ is rejected. If there is no compatible fuzzy rule with $\mathbf{x}_p$, its classification is also rejected.

As shown in (4), the winner rule R_w has the maximum product of the compatibility grade and the rule weight. See [18]-[21] for other fuzzy reasoning methods for pattern classification problems. We use the single winner rule-based method since human users easily know which fuzzy rule is responsible for the classification of each input pattern.

C. Heuristic Rule Generation Method

Since we use the 14 antecedent fuzzy sets in Fig. 1 and an additional *don't care* fuzzy set [0, 1] for each attribute of our n-dimensional problem, the total number of possible fuzzy rules (i.e., the total number of combinations of antecedent fuzzy sets) is 15^n. Among these combinations, we use fuzzy rules with only two antecedent conditions as candidate rules. Thus the number of candidate fuzzy rules is ${}_nC_2 \cdot 14^2$ where ${}_nC_2$ is the total number of combinations of two attributes out of the n attributes. We generate all of these candidate rules. It should be noted that the number of candidate fuzzy rules with two antecedent conditions increases with the order of n^2.

The consequent class C_q and the rule weight CF_q of each fuzzy rule R_q are specified from compatible training patterns in the following heuristic manner [23]. First we calculate the confidence of each class for the antecedent part $\mathbf{A}_q$ of R_q:

$$c(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{\sum_{p=1}^{m} \mu_{\mathbf{A}_q}(\mathbf{x}_p)} . \quad (5)$$

Then the consequent class C_q is specified by identifying the class with the maximum confidence:

$$c(\mathbf{A}_q \Rightarrow \text{Class } C_q) = \max_{h=1,2,...,m} \{c(\mathbf{A}_q \Rightarrow \text{Class } h)\} . \quad (6)$$

The consequent class C_q can be viewed as the dominant class in the fuzzy subspace defined by the antecedent part $\mathbf{A}_q$. If there is no training pattern in $\mathbf{A}_q$, we do not generate any fuzzy rule with the antecedent part $\mathbf{A}_q$.

The rule weight CF_q of each fuzzy rule R_q has a large effect on the performance of fuzzy rule-based classifiers [24]. Its different specifications have been proposed and examined in the literature. We use the following specification because good results were reported in the literature [21], [23]:

$$CF_q = c(\mathbf{A}_q \Rightarrow \text{Class } C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^{M} c(\mathbf{A}_q \Rightarrow \text{Class } h). \quad (7)$$

III. DESIGN OF FUZZY RULE-BASED CLASSIFIERS

A. Outline of Our Approach

Using the heuristic rule generation method in the previous section, we can generate a number of candidate rules with two antecedent conditions. We use genetic algorithms to choose only a small number of candidate rules to construct a compact fuzzy rule-based classifier. When the number of candidate rules is large (e.g., tens of thousands), it is not easy for genetic algorithms to efficiently perform rule selection. Thus we use a prescreening procedure [11] to decrease the number of candidate rules using a heuristic rule evaluation measure [25]. Thus the outline of our rule selection-based approach to the design of fuzzy rule-based classifiers with high explanation ability can be written as follows:

[Outline of Our Approach to Classifier Design]

1. Candidate Rule Generation: Generate a large number of fuzzy rules with only two antecedent conditions using the heuristic rule generation method in Section II.
2. Candidate Rule Prescreening: Choose a tractable number of candidate rules from the generated fuzzy rules using heuristic rule evaluation criteria.
3. Genetic Rule Selection: Select a small number of candidate rules to construct a compact fuzzy rule-based classifier.

We explain candidate rule prescreening and genetic rule selection method in this section.

B. Candidate Rule Prescreening

In the field of data mining, two measures called *support* and *confidence* have been often used to evaluate association rules in the literature [26]. The confidence evaluates the accuracy of the association from the antecedent part to the consequent part. In Section II, we have already shown a fuzzy version of the confidence in (5). On the other hand, the support indicates the percentage of covered patterns by each rule. Its fuzzy version can be written as follows [21]:

$$s(R_q) = s(\mathbf{A}_q \Rightarrow \text{Class}\, C_q) = \frac{\sum_{\mathbf{x}_p \in \text{Class } C_q} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{m}. \quad (8)$$

Roughly speaking, large support values mean general rules that cover many training patterns. On the other hand, small support values mean specific rules that cover only a few training patterns. As candidate rules, we use fuzzy rules that have larger support values than a pre-specified minimum support level (0.01 in our computational experiments). This means that we exclude too specific fuzzy rules even when they have high confidence values. We also use a minimum confidence level (0.6 in our computational experiments). This means that we exclude too general rules with low accuracy.

Among candidate fuzzy rules satisfying the minimum support and confidence levels, we choose a prespecified number of good rules for each class. As a rule evaluation measure, we use the product of support and confidence. That is, we choose a prespecified number of the best candidate rules for each class in terms of the product $s(R_q)\cdot c(R_q)$. Various rule evaluation measures including the product of support and confidence have been examined in [25].

C. Genetic Rule Selection

Let us assume that we have N candidate fuzzy rules after the prescreening phase. Now our problem is to construct a fuzzy rule-based classifier S by choosing only a small number of fuzzy rules from the N candidate fuzzy rules. Since S is a subset of the N candidate fuzzy rules, S is represented by a binary string of length N as $S = s_1 s_2 \cdots s_N$. In this binary string, $s_j = 1$ and $s_j = 0$ mean that the j-th rule is included in S and excluded from S, respectively. Such a binary string is handled as an individual in genetic fuzzy rule selection.

Whereas we can perform multiobjective genetic rule selection [9]-[11], we use a single-objective formulation [3] in this paper for simplicity. That is, we use the following weighted sum fitness function in genetic fuzzy rule selection:

$$\mathit{fitness}(S) = w_1 \cdot \mathit{Accuracy}(S) - w_2 \cdot \mathit{Complexity}(S), \quad (9)$$

where w_1 and w_2 are non-negative weights, $\mathit{Accuracy}(S)$ is an accuracy measure, and $\mathit{Complexity}(S)$ is a complexity measure. As in [3], we use the number of correctly classified patterns and the number of fuzzy rules as $\mathit{Accuracy}(S)$ and $\mathit{Complexity}(S)$, respectively. The weights in (9) are specified as $w_1 = 5$ and $w_2 = 1$ in our computational experiments.

Since genetic rule selection in this paper is formulated as a single-objective problem with binary strings in (9), we can directly use standard genetic algorithms with no modification. In our computational experiments, we use a standard genetic algorithm with the $(\mu + \lambda)$-ES generation update scheme where $\mu = \lambda$. We do not use any heuristic tricks (e.g., biased mutation to decrease the number of fuzzy rules, and the deletion of unnecessary rules [3], [9]-[11]).

IV. EXPLANATION OF CLASSIFICATION RESULTS

As we have already explained, an input pattern $\mathbf{x}_p$ is classified by a single winner rule R_w. Since we use fuzzy rules with two antecedent conditions, each fuzzy rule (including the winner rule R_w) can be always depicted in a two-dimensional pattern space where its antecedent conditions are defined. We can also show the input pattern $\mathbf{x}_p$ together with the training patterns used for the design of our fuzzy rule-based classifier S. That is, we can visually show not only the winner rule responsible for the classification of

$\mathbf{x}_p$ but also the relation between $\mathbf{x}_p$ and the training patterns.

In order to illustrate the above-mentioned visualization method, we performed computational experiments on five data sets in Table 1 from the UCI Machine Learning Repository. Each data set was divided into 90% training patterns and 10% test patterns. First, a large number of fuzzy rules were generated from the training patterns by the heuristic rule generation method. Next, 100 candidate rules were selected from the generated fuzzy rules for each class by the heuristic rule prescreening method. Then, we applied our genetic algorithm to the candidate fuzzy rules for genetic rule selection using the following parameter specifications:

Population size: 200 ($\mu = \lambda = 200$),
Crossover probability: 0.9 (uniform crossover),
Mutation probability: $1/N$ (N: Number of candidate rules),
Termination condition: 1000 generations.

Table 1. Data sets in our computational experiments.

Data set	Attributes	Patterns	Classes
Breast W	9	683	2
Glass	9	214	6
Heart C	13	297	5
Iris	4	150	3
Wine	13	178	3

As an example, we show a winner rule for a test pattern from Iris data set in Fig. 2 where the 0.5-level set of its antecedent part is depicted together with training patterns. The test pattern is classified as Class 2 using the consequent class of the winner rule. We can see from Fig. 2 that the test pattern is located close to many Class 2 training patterns. We can also see that there are only a few training patterns from Class 3 close to the test pattern. From these observations, one may think that the classification of the test pattern as Class 2 has somewhat high reliability. Actually the winner rule has a large rule weight (i.e., certainty factor: 0.566) and a large compatibility grade with the test pattern (i.e., 0.678).

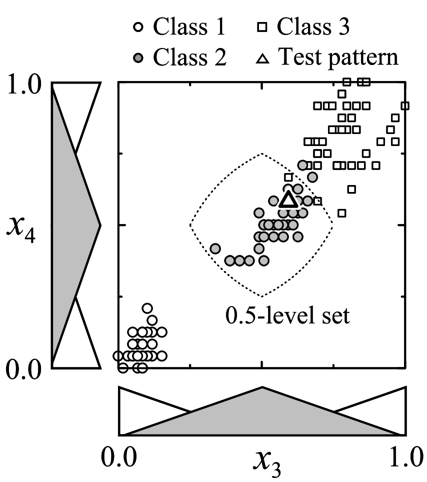

Fig. 2. A test pattern and its winner rule (Iris).

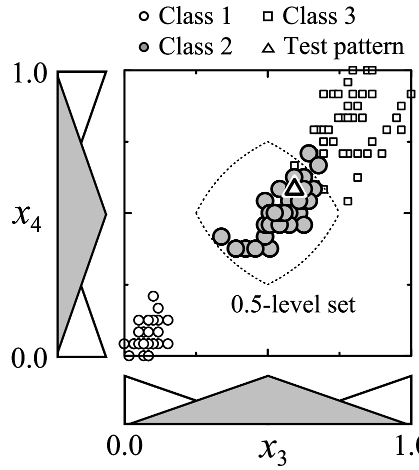

Fig. 3. Training patterns classified by the winner rule (Iris).

In Fig. 3, training patterns classified by the same winner rule are highlighted by larger closed circles. From this figure, we can see that all training patterns from Class 2 are classified by the same winner rule as the test pattern.

We can also examine the possibility of classifying the test pattern as another class (i.e., Class 1 or Class 3). This is easily performed by searching for a class-specific winner rule from fuzzy rules with each consequent class. For the test pattern in Fig. 2, a Class 3 winner rule was found. On the other hand, no fuzzy rules with Class 1 consequent were compatible with the test pattern. In Fig. 4, we show the Class 3 winner rule. It should be noted that the test pattern is not actually classified by the fuzzy rule in Fig. 4 (it is classified by the winner rule in Fig. 3). In Fig. 4, training patterns actually classified by the depicted fuzzy rule are shown by larger squares. We can see that many training patterns from Class 3 are classified by the same winner rule in Fig. 4. We can also see that the compatibility grade of the fuzzy rule with the test pattern is very small in Fig. 4 (i.e., 0.099). From such a small compatibility grade and the location of the test pattern in the two-dimensional pattern space in Fig. 4, we intuitively feel that the test pattern is not likely to be classified as Class 3.

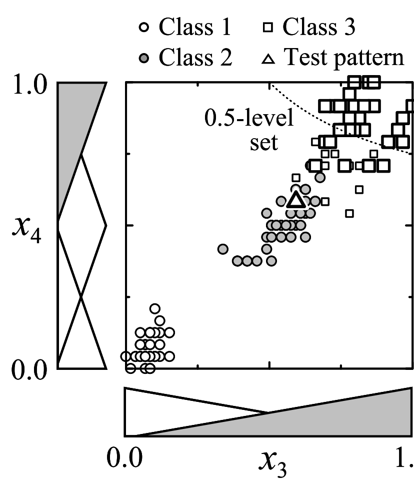

Fig. 4. The winner rule among Class 3 fuzzy rules.

It is well-known that x_3 and x_4 have dominant effects on the classification of Iris data set. These two attributes are also used in Figs. 2-4. Whereas this observation empirically supports the validity of our approach, our experimental results in Figs. 2-4 are not so interesting because we expected these two attributes to be used in fuzzy rule-based classifiers for Iris data set. So we show a different example for a data set with more attributes.

In Fig. 5, we show a winner rule for a test pattern from Wine data set with 13 attributes. This test pattern is classified as Class 2 by the winner rule. As in Fig. 3, training patterns classified by the same winner rule are highlighted by large closed circles in Fig. 5. The test pattern in Fig. 5 is surrounded by a number of Class 2 training patterns classified by the same winner rule. Thus we feel that the classification of the test pattern as Class 2 is reliable whereas some Class 1 training patterns are located close to the test pattern. It should be noted that those Class 1 training patterns are not classified by the winner rule in Fig. 5. We also examined the possibility of classifying the test pattern as another class (i.e., Class 1 or Class 3). We found that the test pattern was not covered by any fuzzy rules with Class 1 or Class 3 consequent. Thus there is no possibility that the test pattern is classified as another class by our fuzzy rule-based classifier.

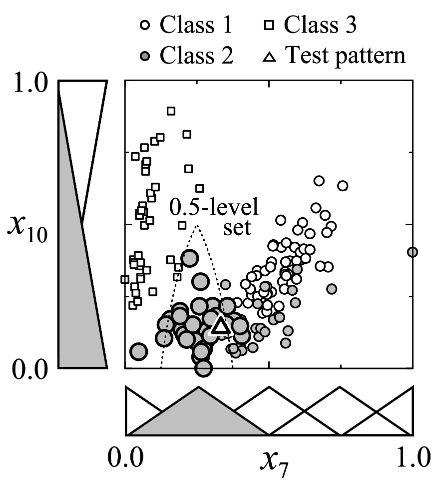

Fig. 5. A test pattern and its winner rule (Wine).

V. PERFORMANCE EVALUATION

The point of our rule selection-based approach is the use of fuzzy rules with only two antecedent conditions. Since each fuzzy rule can use an arbitrary pair of attributes, our fuzzy rule-based classifier may have more than two attributes. Thus our approach is different from feature selection where the same attributes are used in all fuzzy rules.

In this section, we examine how this difference between rule selection and feature selection affects the performance of fuzzy rule-based classifiers designed by each approach through computational experiments on the five UCI data sets in Table 1. In our computational experiments, we examined all combinations of two attributes out of the given n attributes (i.e., ${}_nC_2$ combinations for an n-dimensional problem). That is, we used an explicit enumeration method for selecting the best two features. For each combination of two attributes, we examined 16 fuzzy partitions of the two-dimensional pattern space. More specifically, we generated 4^2 fuzzy rule-based classifiers by assigning one of the four fuzzy partitions in Fig. 1 to each of the two attributes. Each of the generated fuzzy rule-based classifiers was evaluated by the classification rate on training patterns. As a result, we examined ${}_nC_2 \cdot 4^2$ fuzzy rule-based classifiers in total for each data set in a single run of our computational experiment. The best fuzzy rule-based classifier in terms of the classification rate on training patterns was chosen as the final result of feature selection.

The performance of such an exhaustive feature selection method was evaluated through five iterations of the ten-fold cross-validation procedure (i.e., 5 x 10CV). The performance of our rule selection-based approach was examined in the same manner. We examined two settings of the prespecified number of candidate rules for each class: 100 and 1000. The performance of fuzzy rule-based classifiers was examined before and after genetic fuzzy rule selection. Genetic fuzzy rule selection was performed using the same parameter values as the previous section. Experimental results are summarized in Tables 2-4 where H and G mean heuristic selection (i.e., before genetic selection) and genetic selection (i.e., after genetic selection), respectively. The subscripts 100 and 1000 denote the upper bound on the number of extracted candidate rules for each class by heuristic rule selection. On the other hand, FS_{2A} denotes feature selection of two attributes. In Table 2 and Table 3, the best average result for each data set is highlighted by bold face with underlines.

Table 2. Average classification rates on training patterns.

Data set	H_{100}	H_{1000}	GA_{100}	GA_{1000}	FS_{2A}
Breast W	95.31	95.83	97.54	**<u>98.45</u>**	96.35
Glass	64.73	64.86	74.19	**<u>75.49</u>**	64.69
Heart C	55.24	54.10	57.70	**<u>57.81</u>**	57.71
Iris	96.44	96.44	**<u>99.19</u>**	98.92	97.36
Wine	97.30	96.39	99.78	**<u>99.93</u>**	94.81

Table 3. Average classification rates on test patterns.

Data set	H_{100}	H_{1000}	GA_{100}	GA_{1000}	FS_{2A}
Breast W	95.11	**<u>95.55</u>**	95.49	95.29	95.08
Glass	59.21	59.10	**<u>62.40</u>**	62.04	58.48
Heart C	**<u>54.13</u>**	53.79	53.12	52.90	53.52
Iris	**<u>96.13</u>**	**<u>96.13</u>**	94.27	94.13	95.73
Wine	95.29	93.12	**<u>95.64</u>**	94.63	90.01

Table 4. Average number of fuzzy rules in each classifier.

Data set	H_{100}	H_{1000}	GA_{100}	GA_{1000}	FS_{2A}
Breast W	200.0	2000.0	8.3	17.5	16.9
Glass	286.5	795.5	13.4	15.8	13.2
Heart C	125.0	1025.0	5.7	6.1	10.4
Iris	300.0	742.4	5.9	6.1	9.5
Wine	300.0	3000.0	6.6	10.8	15.9

In Table 2, higher classification rates on training patterns were obtained for all the five data sets by genetic rule selection (G_{100} and G_{1000}) than feature selection of two attributes (FS_{2A}). Better results on test patterns were also obtained for three data sets by rule selection than feature selection in Table 3 on test patterns. The difference in test data accuracy between these two approaches is large for Glass and Wine. On the other hand, slightly better results on test patterns were obtained by feature selection for Heart C and Iris than genetic rule selection.

A small number of fuzzy rules were selected by our genetic algorithm from a large number of candidate rules in Table 4. We can see from Table 4 that the number of fuzzy rules obtained by feature selection was also small. This is because only two attributes were selected by feature selection for generating fuzzy rules.

In our computational experiments, we used a simple genetic algorithm with no heuristic tricks for genetic rule selection. Our experimental results by genetic rule selection in Tables 2-4 (especially Table 2 and Table 4) may be improved by using a more sophisticated genetic algorithm and/or some problem specific heuristic tricks [3], [9]-[11].

VI. Conclusion

In this paper, we proposed an approach to the design of fuzzy rule-based classifiers that can visually explain their classification results to human users in an understandable manner. That is, our fuzzy rule-based classifiers can explain to human users why each input pattern is classified as a particular class. The point of our approach is to design fuzzy rule-based classifiers using fuzzy rules with only two antecedent conditions. A small number of fuzzy rules are selected by a genetic algorithm to design a compact fuzzy rule-based classifier. Our approach is similar to but different from feature selection of two attributes. In our approach, fuzzy rule-based classifiers may have more than two attributes while each fuzzy rule has only two antecedent conditions. This is because each fuzzy rule can use an arbitrary pair of attributes. On the other hand, the same attributes are used by all fuzzy rules in feature selection. Rule selection and feature selection were compared with each other through computational experiments. Experimental results showed that more accurate fuzzy rule-based classifiers with less fuzzy rules were obtained by rule selection than feature selection for some data sets.

References

[1] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Interpretability Issues in Fuzzy Modeling*, Springer, Berlin, 2003.

[2] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Accuracy Improvements in Linguistic Fuzzy Modeling*, Springer, Berlin, 2003.

[3] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Selecting fuzzy if-then rules for classification problems using genetic algorithms," *IEEE Trans. on Fuzzy Systems* 3 (3), pp. 260-270, August 1995.

[4] Y. Jin, "Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement," *IEEE Trans. on Fuzzy Systems* 8 (2), pp. 212-221, April 2000.

[5] M. Setnes and H. Roubos, "GA-based modeling and classification: Complexity and performance," *IEEE Trans. on Fuzzy Systems* 8 (5), pp. 509-522, October 2000.

[6] O. Cordon, F. Herrera, F. Hoffmann, and L. Magdalena, *Genetic Fuzzy Systems*, World Scientific, Singapore, 2001.

[7] O. Cordon, F. Gomide, F. Herrera, F. Hoffmann, and L. Magdalena, "Ten years of genetic fuzzy systems: Current framework and new trends," *Fuzzy Sets and Systems* 141 (1), pp. 5-31, January 2004.

[8] F. Herrera, "Genetic fuzzy systems: Status, critical considerations and future directions," *International Journal of Computational Intelligence Research* 1 (1), pp. 59-67, December 2005.

[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 and Systems* 89 (2), pp. 135-150, July 1997.

[10] H. Ishibuchi, T. Nakashima, and T. Murata, "Three-objective genetics-based machine learning for linguistic rule extraction," *Information Sciences* 136 (1-4), pp. 109-133, August 2001.

[11] H. Ishibuchi and T. Yamamoto, "Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining," *Fuzzy Sets and Systems* 141 (1), pp. 59-88, January 2004.

[12] H. Ishibuchi and Y. Nojima, "Analysis of interpretability-accuracy tradeoff of fuzzy systems by multiobjective fuzzy genetics-based machine learning," *International Journal of Approximate Reasoning* 44 (1), pp. 4-31, January 2007.

[13] J. V. de Oliveira, "Semantic constraints for membership function optimization," *IEEE Trans. on Systems, Man, and Cybernetics: Part A - Systems and Humans* 29 (1), pp 128-138, January 1999.

[14] D. Nauck and R. Kruse, "Obtaining interpretable fuzzy classification rules from medical data," *Artificial Intelligence in Medicine* 16 (2), pp. 149-169, June 1999.

[15] S. Guillaume, "Designing fuzzy inference systems from data: An interpretability-oriented review," *IEEE Trans. on Fuzzy Systems* 9 (3), pp. 426-443, June 2001.

[16] R. Mikut, J. Jakel, and L. Groll, "Interpretability issues in data-based learning of fuzzy systems," *Fuzzy Sets and Systems* 150 (2), pp. 179-197, March 2005.

[17] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets and Systems* 52 (1), pp. 21-32, November 1992.

[18] O. Cordon, M. J. del Jesus, and F. Herrera, "A proposal on reasoning methods in fuzzy rule-based classification systems," *International Journal of Approximate Reasoning* 20 (1), pp. 21-45, January 1999.

[19] H. Ishibuchi, T. Nakashima, and T. Morisawa, "Voting in fuzzy rule-based systems for pattern classification problems," *Fuzzy Sets and Systems* 103 (2), pp. 223-238, April 1999.

[20] L. I. Kuncheva, *Fuzzy Classifier Design*, Physica-Verlag, Heidelberg, 2000.

[21] H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining*, Springer, Berlin, 2004.

[22] H. Ishibuchi, T. Nakashima, and T. Murata, "Performance evaluation of fuzzy classifier systems for multi-dimensional pattern classification problems," *IEEE Trans. on Systems, Man, and Cybernetics - Part B: Cybernetics* 29 (5), pp. 601-618, October 1999.

[23] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. on Fuzzy Systems* 13 (4), pp. 428-435, August 2005.

[24] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *IEEE Trans. on Fuzzy Systems* 9 (4), pp. 506-515, August 2001.

[25] H. Ishibuchi and T. Yamamoto, "Comparison of heuristic criteria for fuzzy rule selection in classification problems," *Fuzzy Optimization and Decision Making* 3 (2), pp. 119-139, June 2004.

[26] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo, "Fast discovery of association rules," in U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy (eds.), *Advances in Knowledge Discovery and Data Mining*, AAAI Press, Menlo Park, pp. 307-328, 1996.