Multiobjective Formulations of Fuzzy Rule-Based Classification System Design

Hisao Ishibuchi and **Yusuke Nojima**
Graduate School of Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Sakai, Osaka 599-8531, Japan
{hisaoi, nojima}@cs.osakafu-u.ac.jp

Abstract

We examine several formulations of fuzzy rule selection for the design of fuzzy rule-based classification systems in our two-stage approach. The first stage is heuristic rule extraction where a large number of candidate rules are extracted. The second stage is evolutionary rule selection where fuzzy rule-based systems are constructed by choosing a small number of candidate rules. Rule selection is formulated as single-, two-, and three-objective optimization problems using an accuracy measure and two complexity measures.

Keywords: Fuzzy rule-based classification systems, Evolutionary multiobjective optimization, Accuracy-complexity tradeoff.

1 Introduction

Genetic algorithms have been used for the design of fuzzy rule-based systems in many studies [4]. While the accuracy maximization by genetic optimization has been mainly discussed in those studies, recently the existence of the accuracy-complexity tradeoff in the design of fuzzy rule-based systems has been realized by some researchers [1], [2]. The accuracy-complexity tradeoff can be handled in two different approaches. One approach is the use of the weighted sum of an accuracy measure and a complexity measure. The other approach uses a multiobjective formulation where an accuracy measure and a complexity measure are optimized by multiobjective optimization techniques. In this case, multiple non-dominated (i.e., Pareto optimal) fuzzy rule-based systems are obtained. Each fuzzy rule-based system corresponds to a different tradeoff solution of the multiobjective optimization problem.

For the design of fuzzy rule-based classification systems, a weighted sum-based approach to fuzzy rule selection was proposed in Ishibuchi et al. [12], [13] where the number of correctly classified training patterns was maximized and the number of fuzzy rules was minimized. These two objectives were optimized by a two-objective genetic algorithm in [8]. The two-objective approach was further extended in [10] to the case of three objectives using an additional objective: to minimize the total number of antecedent conditions. A large number of non-dominated fuzzy rule-based classification systems were obtained from the three-objective formulation. The generalization ability of non-dominated fuzzy systems was examined in [14], [15]. A three-objective memetic algorithm was used to efficiently search for non-dominated fuzzy systems in [16].

In this paper, we examine several formulations of fuzzy rule selection for the design of fuzzy rule-based classification systems. We compare single-, two-, and three-objective formulations with each other through computational experiments on data sets from the UCI Machine Learning Repository. Experimental results demonstrate advantages and disadvantages of each formulation.

2 Fuzzy Rule-Based Classification Systems

In this section, we briefly explain fuzzy rule-based classification. For details, see the textbook on fuzzy data mining by Ishibuchi et al. [11].

Let us assume that we have m training patterns $\mathbf{x}_p = (x_{p1}, x_{p2}, ..., x_{pn})$, $p = 1, 2, ..., m$ from M classes in the n-dimensional unit hyper-cube $[0, 1]^n$. That is, our pattern classification problem is an M-class problem with m training patterns in the n-

dimensional pattern space $[0, 1]^n$.

We use fuzzy if-then rules of the following type:

Rule R_q : If x_1 is A_{q1} and ... and x_n is A_{qn} then Class C_q with CF_q, (1)

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 for the i-th attribute, C_q is a consequent class, and CF_q is a certainty grade (i.e., rule weight). We denote the fuzzy rule in (1) as "$\mathbf{A}_q \Rightarrow C_q$". As the antecedent fuzzy set A_{qi}, we use one of the 14 fuzzy sets in Fig. 1. The antecedent fuzzy set A_{qi} can be also "*don't care*". Thus the total number of combinations of the n antecedent fuzzy sets in (1) is 15^n.

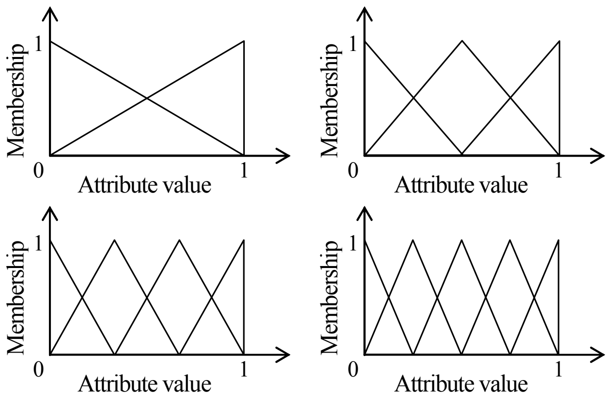

Figure 1: Antecedent fuzzy sets.

When the antecedent part of the fuzzy rule in (1) is given, the consequent class and the rule weight are determined in a heuristic manner from compatible training patterns. First we calculate the compatibility grade of each training pattern with the antecedent part using the product operation as

$$\mu_{\mathbf{A}_q}(\mathbf{x}_p) = \mu_{A_{q1}}(x_{p1}) \times \cdots \times \mu_{A_{qn}}(x_{pn}), \quad (2)$$

where $\mu_{A_{qi}}(\cdot)$ is the membership function of the antecedent fuzzy set A_{qi}. Then we calculate the confidence for each class as follows:

$$c(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{\sum_{p=1}^{m} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}, \quad h = 1, 2, ..., M. \quad (3)$$

The consequent C_q of the fuzzy rule "$\mathbf{A}_q \Rightarrow C_q$" in (1) is determined by finding the class with the maximum confidence for the antecedent $\mathbf{A}_q$ as

$$c(\mathbf{A}_q \Rightarrow C_q) = \max_{h=1,2,...,M} \{c(\mathbf{A}_q \Rightarrow \text{Class } h)\}. \quad (4)$$

We define the rule weight by the difference between the confidence of the consequent class and the sum of the confidences of the other classes as

$$CF_q = c(\mathbf{A}_q \Rightarrow C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^{M} c(\mathbf{A}_q \Rightarrow \text{Class } h). \quad (5)$$

See [9], [11], [18] for other specifications of rule weights and their effects on the accuracy of fuzzy rule-based classification systems.

In this manner, the consequent class and the rule weight of each fuzzy rule can be easily determined from compatible training patterns. Let us denote the set of generated fuzzy rules by *S*. The rule set *S* can be viewed as a fuzzy rule-based classification system. The classification of an input pattern $\mathbf{x}_p = (x_{p1}, x_{p2}, ..., x_{pn})$ is performed in the fuzzy rule-based classification system S by choosing a single winner rule R_w from *S* as follows:

$$\mu_{\mathbf{A}_w}(\mathbf{x}_p) \cdot CF_w = \max\{\mu_{\mathbf{A}_q}(\mathbf{x}_p) \cdot CF_q \,|\, R_q \in S\}. \quad (6)$$

The input pattern $\mathbf{x}_p$ is classified as Class C_w, which is the consequent class of the winner rule R_w.

3 Two-Stage Fuzzy Rule Selection Approach

Fuzzy rule selection in this paper is to find non-dominated rule sets from the 15^n fuzzy rules of the form in (1) with respect to accuracy and complexity. Since any subset of the 15^n fuzzy rules can be represented by a binary string of length 15^n, the size of the search space is 2^{15^n}. Except for the case of low-dimensional problems, it is very difficult to handle such a huge search space. Thus a two-stage fuzzy rule selection approach has been proposed in [11], [16]. The first stage is heuristic rule extraction where a tractable number of promising candidate rules are extracted from numerical data using a heuristic rule evaluation measure in the same manner as data mining. The second stage is evolutionary rule selection where evolutionary optimization algorithms are used to find non-dominated subsets of the extracted candidate rules with respect to accuracy and complexity.

A number of heuristic rule evaluation measures were examined in [17]. In this paper, we use the following rule evaluation measure:

$$f(R_q) = s(\mathbf{A}_q \Rightarrow C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^{M} s(\mathbf{A}_q \Rightarrow \text{Class } h), \quad (7)$$

where $s(\cdot)$ is the support of fuzzy rules, which is defined as

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

The heuristic rule evaluation measure in (7) is a modified version of a fitness function used in an iterative fuzzy GBML (genetics-based machine learning) algorithm called SLAVE [7].

Using the rule evaluation measure in (7), we generate a prespecified number of promising candidate rules for each class. In this heuristic rule extraction stage, we only examine short fuzzy rules with a few antecedent conditions. This is because we want to construct interpretable fuzzy rule-based classification systems (i.e., because it is very difficult for human users to intuitively understand long fuzzy rules with many antecedent conditions). More specifically, we choose 300 rules with the largest values of the rule evaluation measure in (7) for each class among short rules of length three or less in our computational experiments except for the sonar data set with 60 attributes. For the sonar data set, we only examine short rules of length two or less. The total number of candidate rules is $300M$ where M is the number of classes.

Let N be the total number of extracted candidate fuzzy rules (i.e., $N = 300M$ in our computational experiments). Any subset S of the candidate fuzzy rules can be represented by a binary substring of length N as $S = s_1 s_2 \cdots s_N$ where $s_j = 1$ and $s_j = 0$ mean that the j-th candidate rule is included in S and excluded from S, respectively. Such a binary coding is used in the second stage of our two-stage fuzzy rule selection approach.

In the second stage (i.e., evolutionary rule selection stage), evolutionary multiobjective optimization (EMO) algorithms are used to search for non-dominated rule sets (i.e., non-dominated binary strings of length N) with respect to accuracy and complexity. Formulations of fuzzy rule selection as multiobjective optimization problems are discussed in the next section. For EMO algorithms, see [3], [5]. We use the NSGA-II algorithm of Deb et al. [6].

4 Formulations of Rule Selection

We use an accuracy measure $f_1(S)$, which is the number of correctly classified training patterns by S, as in our former studies [8], [10]-[16]. We also use two complexity measures $f_2(S)$ and $f_3(S)$: $f_2(S)$ is the number of fuzzy rules in S, and $f_3(S)$ is the total number of antecedent conditions (i.e., $f_3(S)$ is the total rule length of fuzzy rules in S).

Several formulations are possible using these three measures in fuzzy rule selection. When we use all the three measures, we have the following three-objective optimization problem:

$$\text{Maximize } f_1(S) \text{ and Minimize } f_2(S),\ f_3(S). \quad (9)$$

We can formulate two two-objective optimization problems using one of the two complexity measures:

$$\text{Maximize } f_1(S) \text{ and Minimize } f_2(S), \quad (10)$$

$$\text{Maximize } f_1(S) \text{ and Minimize } f_3(S). \quad (11)$$

From each multiobjective optimization problem, we can formulate a single-objective maximization problem using a weighted sum fitness function as

$$fitness(S) = w_1 \cdot f_1(S) - w_2 \cdot f_2(S) - w_3 \cdot f_3(S), \quad (12)$$

$$fitness(S) = w_1 \cdot f_1(S) - w_2 \cdot f_2(S), \quad (13)$$

$$fitness(S) = w_1 \cdot f_1(S) - w_3 \cdot f_3(S). \quad (14)$$

5 Comparison among Six Formulations

Through computational experiments on six data sets in Table 1, we compared the six formulations of fuzzy rule selection in the previous section.

For the multiobjective optimization problems in (9)-(11), the NSGA-II algorithm [6] was executed using the following parameter specifications:

Population size: 200 strings,
Crossover probability: 0.8 (uniform crossover),
Biased mutation probabilities:
$$p_{\mathrm{m}}(0 \to 1) = 1/300M \quad \text{and} \quad p_{\mathrm{m}}(1 \to 0) = 0.1,$$
Stopping condition: 5000 generations.

We used a standard genetic algorithm with the same generation update scheme as the NSGA-II for the single-objective maximization problems in (12)-(14). The same parameter values were used in the NSGA-II algorithm and the standard genetic algorithm.

We mainly report experimental results on training patterns where we used all patterns in each data set as training patterns. Each formulation was examined by 20 independent runs on each data set. In Table 2, we show the average number of obtained non-dominated rule sets from each of the three multiobjective formulations. From this table, we can see that more non-dominated rule sets were obtained when $f_3(S)$ is used as a complexity measure. It should be noted that only a single rule set was always obtained from each of the single-objective formulations.

Table 1: Data sets in computational experiments.

Data set	Attributes	Patterns	Classes
Breast W	9	683*	2
Diabetes	8	768	2
Glass	9	214	6
Heart C	13	297*	5
Sonar	60	208	2
Wine	13	178	3

*Incomplete patterns with missing values are not included.

Table 2: Average number of non-dominated rule sets.

Data set	$\{f_1, f_2\}$	$\{f_1, f_3\}$	$\{f_1, f_2, f_3\}$
Breast W	9.50	11.30	**12.55**
Diabetes	8.25	11.85	**15.55**
Glass	16.95	27.30	**31.05**
Heart C	28.60	48.65	**49.60**
Sonar	8.75	11.65	**12.80**
Wine	5.90	9.55	**12.15**

As shown in Table 2, a number of non-dominated rule sets were obtained by a single run of the NSGA-II algorithm. Examples of obtained rule sets by a single run for the Glass data set are shown in Fig. 2 and Fig. 3. In Fig. 2, we show experimental results using the three-objective formulation and the two-objective formulation with $f_1(S)$ and $f_2(S)$. On the other hand, experimental results in Fig. 3 are from the three-objective formulation and the two-objective formulation with $f_1(S)$ and $f_3(S)$.

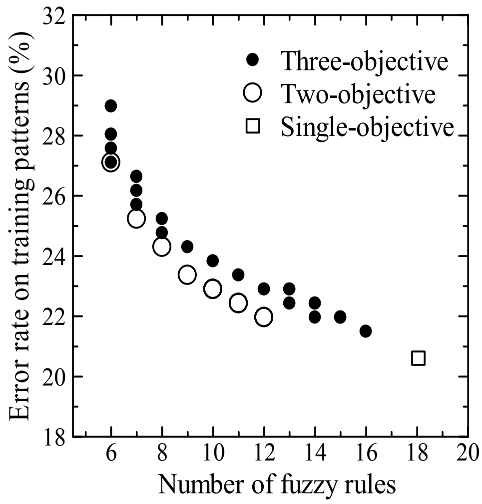

Figure 2: Non-dominated rule sets obtained by a single run for the three-objective formulation and the two-objective formulation with $f_1(S)$ and $f_2(S)$. For comparison, the average result by the single-objective formulation in (12) is also shown.

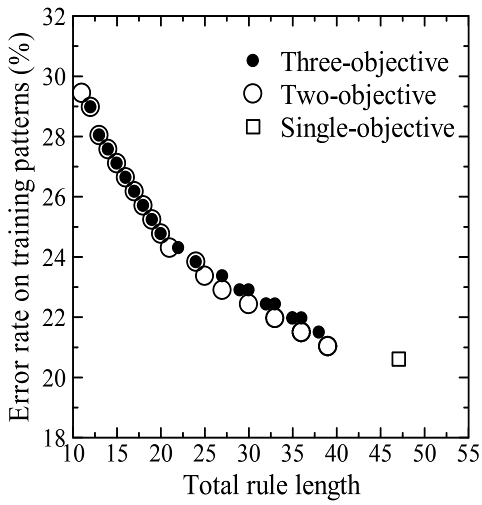

Figure 3: Non-dominated rule sets obtained by a single run for the three-objective formulation and the two-objective formulation with $f_1(S)$ and $f_3(S)$. For comparison, the average result by the single-objective formulation in (12) is also shown.

In Fig. 2 and Fig. 3, the same rule sets from the three-objective formulation are shown by small closed circles with different horizontal axes. From Fig. 2 and Fig. 3 (and from Table 2), we can see that more non-dominated rule sets are obtained from the three-objective formulation than the two-objective ones.

To examine the search ability of the NSGA-II algorithm, we examined the best rule set with respect to the accuracy among non-dominated rule sets obtained from each run. The average accuracy of the best rule set over 20 runs is shown in Table 3. For comparison, we also show the average accuracy over 20 runs for each of the three single-objective formulations in Table 4 and Table 5. Different weight values were used in each table. The average result in Table 4 for the Glass data set using the single-objective formulation in (12) is also shown in Fig. 2 and Fig. 3. From Tables 3-5 (and also from Fig. 2 and Fig. 3), we can see that the multiobjective formulation is inferior to the single-objective formulations in terms of accuracy for some data sets. This observation suggests the necessity of further improvement of the search ability of the NSGA-II algorithm.

Table 3: Average value of the best error rate among obtained non-dominated rule sets in each run.

Data set	$\{f_1, f_2\}$	$\{f_1, f_3\}$	$\{f_1, f_2, f_3\}$
Breast W	**1.76**	1.83	1.78
Diabetes	22.15	**22.08**	22.23
Glass	**21.36**	21.40	21.47
Heart C	28.40	28.40	**28.38**
Sonar	10.63	**10.53**	10.72
Wine	**0.00**	**0.00**	**0.00**

Table 4: Average error rate by the three single-objective formulations ($w_1 = 100$, $w_2 = w_3 = 1$).

Data set	$\{f_1, f_2\}$	$\{f_1, f_3\}$	$\{f_1, f_2, f_3\}$
Breast W	**1.79**	1.81	1.80
Diabetes	**21.96**	22.01	22.01
Glass	**20.44**	20.51	20.61
Heart C	**28.45**	28.60	28.52
Sonar	**9.59**	9.95	9.69
Wine	**0.00**	**0.00**	**0.00**

Table 5: Average error rate by the three single-objective formulations ($w_1 = 10$, $w_2 = w_3 = 1$).

Data set	$\{f_1, f_2\}$	$\{f_1, f_3\}$	$\{f_1, f_2, f_3\}$
Breast W	**1.77**	1.79	1.79
Diabetes	22.08	22.11	**22.02**
Glass	**20.37**	20.72	20.47
Heart C	28.64	**28.52**	**28.52**
Sonar	9.76	**9.71**	**9.71**
Wine	**0.00**	**0.00**	**0.00**

While we can not observe any advantages of the three-objective formulation over the two-objective ones in experimental results on training data (except for the increase in the number of obtained non-dominated rule sets), the use of $f_3(S)$ together with $f_1(S)$ and $f_2(S)$ has a positive effect on generalization ability to test patterns for some data sets. In Fig. 4, we compare error rates of obtained non-dominated rule sets from the three-objective formulation and the two-objective one with $f_1(S)$ and $f_2(S)$. Fig. 4 is experimental results of a single run of each formulation for the diabetes data set with 50% training patterns and 50% test patterns. For comparison, Fig. 4 also shows the corresponding result by the single-objective formulation in (12).

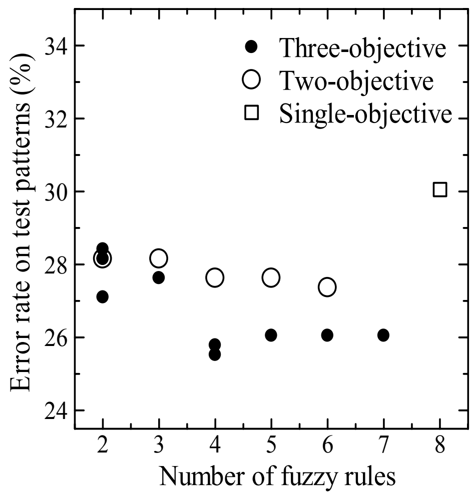

Figure 4: Generalization ability of non-dominated rule sets obtained by a single run for the diabetes data set with 50% training and 50% test patterns. The three-objective formulation is compared with the single-objective formulation in (12) and the two-objective formulation with $f_1(S)$ and $f_2(S)$.

6 Conclusions

In this paper, we compared six formulations of fuzzy rule selection. The main advantage of multiobjective formulations over single-objective ones is that multiple non-dominated rule sets are obtained from its single run, which visually show the accuracy-complexity tradeoff. Experimental results suggested that the increase in the number of objectives has a possibility to improve generalization ability of fuzzy rule-based classification systems while it makes the search for non-dominated rule sets difficult.

The authors would like to thank the financial support from the Okawa Foundation for Information and Telecommunications.

References

[1] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Interpretability Issues in Fuzzy Modeling*, Springer, 2003.
[2] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Accuracy Improvements in Linguistic Fuzzy Modeling*, Springer, 2003.
[3] C. A. Coello Coello, D. A. van Veldhuizen, and G. B. Lamont, *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer Academic Publishers, 2002.
[4] O. Cordon, F. Herrera, F. Hoffman, and L. Magdalena, *Genetic Fuzzy Systems*, World Scientific, 2001.
[5] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, 2001.
[6] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. on Evolutionary Computation*, vol. 6, no. 2, pp. 182-197, 2002.
[7] A. Gonzalez and R. Perez, "SLAVE: A genetic learning system based on an iterative approach," *IEEE Trans. on Fuzzy Systems*, vol. 7, no. 2, pp. 176-191, 1999.
[8] 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*, vol. 89, no. 2, pp. 135-150, 1997.
[9] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *IEEE Trans. on Fuzzy Systems*, vol. 9, no. 4, pp. 506-515, 2001.
[10] H. Ishibuchi, T. Nakashima, and T. Murata, "Three-objective genetics-based machine learning for linguistic rule extraction," *Information Sciences*, vol. 136, no. 1-4, pp. 109-133, 2001.
[11] H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and Modeling with Linguistic Information Granules*, Springer, 2004.
[12] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Construction of fuzzy classification systems with rectangular fuzzy rules using genetic algorithms," *Fuzzy Sets and Systems*, vol. 65, no. 2/3, pp. 237-253, 1994.
[13] H. Ishibuchi, K. Nozaki, N. Yamamoto, and H. Tanaka, "Selecting fuzzy if-then rules for classification problems using genetic algorithms," *IEEE Trans. on Fuzzy Systems*, vol. 3, no. 3, pp. 260-270, 1995.
[14] H. Ishibuchi and T. Yamamoto, "Effects of three-objective genetic rule selection on the generalization ability of fuzzy rule-based systems," *Lecture Notes in Computer Science 2632: EMO 2003*, pp. 608-622, Springer, 2003.
[15] H. Ishibuchi and T. Yamamoto, "Evolutionary multiobjective optimization for generating an ensemble of fuzzy rule-based classifiers," *Lecture Notes in Computer Science 2723: GECCO 2003*, pp. 1077-1088, Springer, 2003.
[16] 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*, vol. 141, no. 1, pp. 59-88, 2004.
[17] H. Ishibuchi and T. Yamamoto, "Comparison of heuristic criteria for fuzzy rule selection in classification problems," *Fuzzy Optimization and Decision Making*, vol. 3, no. 2, pp. 119-139, 2004.
[18] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. on Fuzzy Systems* (in press).