2006 IEEE International Conference on Fuzzy Systems
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

# Fuzzy Data Mining by Heuristic Rule Extraction and Multiobjective Genetic Rule Selection

Hisao Ishibuchi, *Member*, *IEEE*, Yusuke Nojima, *Member*, *IEEE*, and Isao Kuwajima, *Student Member*, *IEEE*

*Abstract* — **In this paper, we demonstrate that multiobjective genetic rule selection can significantly improve the accuracy-complexity tradeoff curve of fuzzy rule-based classification systems generated by a heuristic rule extraction procedure for classification problems with many continuous attributes. First a prespecified number of fuzzy rules are extracted in a heuristic manner based on a rule evaluation criterion. This step can be viewed as fuzzy data mining. Then multiobjective genetic rule selection is applied to the extracted rules to find a number of non-dominated rule sets with respect to accuracy maximization and complexity minimization. This step can be viewed as a postprocessing procedure in fuzzy data mining. Experimental results show that multiobjective genetic rule selection finds a number of smaller rule sets with higher classification accuracy than heuristically extracted rule sets. That is, the accuracy-complexity tradeoff curve of heuristically extracted rule sets in fuzzy data mining is improved by multiobjective genetic rule selection. This observation suggests that multiobjective genetic rule selection plays an important role in fuzzy data mining as a postprocessing procedure.**

## I. Introduction

Evolutionary multiobjective optimization (EMO) is one of the most active research areas in the field of evolutionary computation [1]-[3]. Recently EMO algorithms have been employed in some studies on modeling and classification. For example, Kupinski & Anastasio [4] used an EMO algorithm to generate non-dominated neural networks on a receiver operating characteristic curve. Gonzalez et al. [5] generated non-dominated radial basis function networks of different sizes. Llora & Goldberg [6] used an EMO algorithm in Pittsburgh-style learning classifier systems. Abbass [7] used a memetic EMO algorithm (i.e., a hybrid EMO algorithm with local search) to speed up the back-propagation algorithm where multiple neural networks of different sizes were evolved to find an appropriate network structure. Non-dominated neural networks were combined into a single ensemble classifier in [8]-[10]. The use of EMO algorithms to design ensemble classifiers was also proposed in Ishibuchi & Yamamoto [11] where multiple fuzzy rule-based classifiers of different sizes were generated. In some studies on fuzzy rule-based systems, EMO algorithms were used to analyze the tradeoff structure between accuracy and interpretability [12]-[20]. For more recent studies, see [21].

In this paper, we use an EMO algorithm for fuzzy rule selection to examine the usefulness of multiobjective genetic rule selection as a postprocessing procedure in fuzzy data mining for pattern classification problems. A number of non-dominated rule sets (i.e., non-dominated fuzzy rule-based classification systems) with respect to their accuracy and complexity are found by EMO-based rule selection from fuzzy rules extracted by a heuristic rule extraction procedure in fuzzy data mining.

Genetic fuzzy rule selection for classification problems was first formulated as a single-objective combinatorial 0/1 optimization problem in Ishibuchi et al. [22], [23] where a fitness function of each rule set was defined as a weighted sum of its accuracy (i.e., the number of correctly classified training patterns) and its complexity (i.e., the number of fuzzy rules). This single-objective formulation was extended in [12] as a two-objective problem where non-dominated rule sets were found by an EMO algorithm. Then fuzzy rule selection was formulated as a three-objective problem in [13] where the total rule length (i.e., the total number of antecedent conditions over fuzzy rules in each rule set) was used as an additional complexity measure. This three-objective formulation was also handled by a memetic EMO algorithm in [16] and a multiobjective fuzzy genetics-based machine learning algorithm in [20].

This paper is organized as follows. First we briefly explain multiobjective optimization and fuzzy rule selection in Section II. We use two objectives in fuzzy rule selection: the minimization of the error rate on training patterns and the minimization of the number of fuzzy rules. Next we explain a heuristic procedure for extracting fuzzy classification rules in Section III. Several rule evaluation criteria used in the heuristic rule extraction procedure are compared in Section IV through computational experiments on some benchmark data sets in the UC Irvine machine learning repository. We also examine the accuracy-complexity tradeoff curve of extracted rule sets using various specifications of the number of fuzzy rules to be extracted. Then we demonstrate the usefulness of multiobjective genetic rule selection as a postprocessing procedure in fuzzy data mining in Section V. It is clearly shown that the accuracy-complexity tradeoff curve of heuristically extracted rule sets is improved by multiobjective genetic rule selection. That is, the accuracy of heuristically extracted rule sets is improved while their complexity is decreased by multiobjective genetic fuzzy rule selection. Finally Section VI concludes this paper.

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

Hisao Ishibuchi, Yusuke Nojima and Isao Kuwajima 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, nojima@cs.osakafu-u.ac.jp, kuwajima@ci.cs.osakafu-u.ac.jp).

0-7803-9489-5/06/$20.00/©2006 IEEE

## II. Multiobjective Fuzzy Rule Selection

In this section, we explain multiobjective optimization and multiobjective genetic fuzzy rule selection.

### A. Multiobjective Optimization

Let us consider the following $k$-objective minimization problem:

$$\text{Minimize } \mathbf{z} = (f_1(\mathbf{y}), f_2(\mathbf{y}), ..., f_k(\mathbf{y})), \quad (1)$$

$$\text{subject to } \mathbf{y} \in \mathbf{Y}, \quad (2)$$

where $\mathbf{z}$ is the objective vector, $f_i(\mathbf{y})$ is the $i$-th objective to be minimized, $\mathbf{y}$ is the decision vector, and $\mathbf{Y}$ is the feasible region in the decision space.

Let $\mathbf{a}$ and $\mathbf{b}$ be two feasible solutions of the $k$-objective minimization problem in (1)-(2). If the following condition holds, $\mathbf{a}$ can be viewed as being better than $\mathbf{b}$:

$$\forall i, f_i(\mathbf{a}) \le f_i(\mathbf{b}) \text{ and } \exists j, f_j(\mathbf{a}) < f_j(\mathbf{b}). \quad (3)$$

In this case, we say that $\mathbf{a}$ dominates $\mathbf{b}$ (equivalently $\mathbf{b}$ is dominated by $\mathbf{a}$).

When $\mathbf{b}$ is not dominated by any other feasible solutions (i.e., when there exists no feasible solution $\mathbf{a}$ that dominates $\mathbf{b}$), the solution $\mathbf{b}$ is referred to as a Pareto-optimal solution of the $k$-objective minimization problem in (1)-(2). The set of all Pareto-optimal solutions forms the tradeoff surface in the objective space. This tradeoff surface is referred to as the Pareto front. Various EMO algorithms have been proposed to efficiently search for Pareto-optimal solutions [1]-[3].

### B. Multiobjective Genetic Fuzzy Rule Selection

Let us assume that we have $N$ fuzzy rules extracted for a pattern classification problem by a heuristic rule extraction procedure. Multiobjective genetic fuzzy rule selection is used to find Pareto-optimal rule sets from these $N$ fuzzy rules with respect to the two goals of knowledge extraction: accuracy maximization and complexity minimization.

Let $S$ be a subset of the extracted $N$ fuzzy rules. The accuracy of the rule set $S$ is measured by the error rate when all the training patterns are classified by $S$. We use a single winner rule-based method to classify each training pattern by $S$. That is, each pattern is classified by the single winner rule in $S$ that has the maximum product of the rule weight and the compatibility grade with that pattern as explained in the next section. We include the rejection rate into the error rate (i.e., training patterns with no compatible fuzzy rules in $S$ are counted among errors in this paper).

On the other hand, we measure the complexity of the rule set $S$ by the number of fuzzy rules in $S$. Thus our fuzzy rule selection problem is formulated as follows:

$$\text{Minimize } f_1(S) \text{ and } f_2(S), \quad (4)$$

where $f_1(S)$ is the error rate on training patterns by the rule set $S$ and $f_2(S)$ is the number of fuzzy rules in $S$.

Any subset $S$ of the $N$ fuzzy rules can be represented by a binary string of length $N$ as

$$S = s_1 s_2 \cdots s_N, \quad (5)$$

where $s_j = 1$ and $s_j = 0$ mean that the $j$-th fuzzy rule is included in $S$ and excluded from $S$, respectively. Such a binary string is handled as an individual in multiobjective genetic fuzzy rule selection.

Since feasible solutions (i.e., any subsets of the $N$ fuzzy rules) are represented by binary strings in (5), we can apply almost all EMO algorithms with standard genetic operations to our multiobjective fuzzy rule selection problem in (4). In this paper, we use the NSGA-II algorithm [24] because it is a well-known high-performance EMO algorithm.

Let $P$ be the current population in NSGA-II. The outline of NSGA-II can be written as follows:

Step 1: $P$ := Initialize ($P$)
Step 2: while a termination condition is not satisfied, do
Step 3: $P'$ := Selection ($P$)
Step 4: $P''$ := Genetic Operations ($P'$)
Step 5: $P$ := Replace ($P \cup P''$)
Step 6: end while
Step 7: return (non-dominated solutions ($P$))

First an initial population is generated in Step 1 in the same manner as in single-objective genetic algorithms. Genetic operations in Step 4 are also the same as those in single-objective genetic algorithms. Parent selection in Step 3 and generation update in Step 5 of NSGA-II are different from single-objective genetic algorithms. Pareto ranking and a crowding measure are used to evaluate each solution in Step 3 for parent selection and in Step 5 for generation update. For details of NSGA-II, see Deb [1] and Deb et al. [24].

In the application of NSGA-II to multiobjective genetic fuzzy rule selection, we use two problem-specific heuristic tricks to efficiently find small rule sets with high accuracy. One trick is biased mutation where a larger probability is assigned to the mutation from 1 to 0 than that from 0 to 1. The other trick is the removal of unnecessary rules, which is a kind of local search. Since we use the single winner rule-based method for the classification of each pattern by the rule set $S$, some rules in $S$ may be chosen as winner rules for no training patterns. By removing these rules from $S$, we can improve the second objective (i.e., the number of fuzzy rules in $S$) without degrading the first objective (i.e., the error rate on training patterns). The removal of unnecessary rules is performed after the first objective is calculated and before the second objective is calculated.

## III. Heuristic Fuzzy Rule Extraction

In this section, we explain heuristic fuzzy rule extraction using rule evaluation criteria in data mining.

### A. Pattern Classification Problem

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]. That

is, $x_{pi} \in [0, 1]$ for $p = 1, 2, ..., m$ and $i = 1, 2, ..., n$. Thus the pattern space of our pattern classification problem is the $n$-dimensional unit-hypercube $[0, 1]^n$.

### B. Fuzzy Rules for Pattern Classification Problem

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

$$\text{Rule } R_q: \text{If } x_1 \text{ is } A_{q1} \text{ and } ... \text{ and } x_n \text{ is } A_{qn} \text{ then Class } C_q \text{ with } CF_q, \quad (6)$$

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 grade). For other types of fuzzy rules for pattern classification problems, see [17], [25], [26].

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 in fuzzy rule extraction. 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. That is, we use the 15 antecedent fuzzy sets for each attribute in our computational experiments.

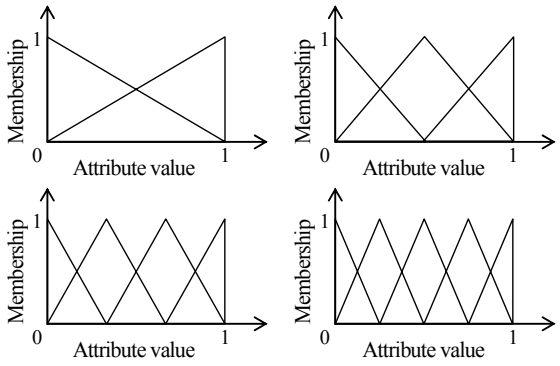


Fig. 1. Four fuzzy partitions used in our computational experiments.

### C. Fuzzy Rule Generation

Since we use the 15 antecedent fuzzy sets for each attribute of our $n$-dimensional pattern classification problem, the total number of combinations of the antecedent fuzzy sets is $15^n$. Each combination is used in the antecedent part of the fuzzy rule in (6). Thus the total number of possible fuzzy rules is also $15^n$. The consequent class $C_q$ and the rule weight $CF_q$ of each fuzzy rule $R_q$ are specified from the given training patterns in the following heuristic manner.

First we calculate the compatibility grade of each pattern $\mathbf{x}_p$ with the antecedent part $\mathbf{A}_q$ of the fuzzy rule $R_q$ using the product operation as

$$\mu_{\mathbf{A}_q}(\mathbf{x}_p) = \mu_{A_{q1}}(x_{p1}) \cdot ... \cdot \mu_{A_{qn}}(x_{pn}), \quad (7)$$

where $\mu_{A_{qi}}(\cdot)$ is the membership function of $A_{qi}$.

Next the confidence of the fuzzy rule "$\mathbf{A}_q \Rightarrow$ Class $h$" is calculated for each class ($h = 1, 2, ..., M$) as follows [17]:

$$c(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{\sum_{p=1}^{m} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}. \quad (8)$$

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

$$c(\mathbf{A}_q \Rightarrow \text{Class } C_q) = \max_{h=1,2,...,m} \{c(\mathbf{A}_q \Rightarrow \text{Class } h)\}. \quad (9)$$

The consequent class $C_q$ can be viewed as the dominant class in the fuzzy subspace defined by the antecedent part $\mathbf{A}_q$. When there is no pattern in the fuzzy subspace defined by $\mathbf{A}_q$, we do not generate any fuzzy rules with $\mathbf{A}_q$ in the antecedent part. This specification method of the consequent class of fuzzy rules has been used in many studies since [27].

The rule weight $CF_q$ of each fuzzy rule $R_q$ has a large effect on the performance of fuzzy rule-based classification systems [28]. Different specifications of the rule weight have been proposed and examined in the literature. We use the following specification because good results were reported by this specification in the literature [17], [29]:

$$CF_q = c(\mathbf{A}_q \Rightarrow \text{Class } C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^{M} c(\mathbf{A}_q \Rightarrow \text{Class } h). \quad (10)$$

Let $S$ be a set of fuzzy rules of the form in (6). A new pattern $\mathbf{x}_p$ is classified by a single winner rule $R_w$, which is chosen from the rule set $S$ as follows:

$$\mu_{\mathbf{A}_w}(\mathbf{x}_p) \cdot CF_w = \max\{\mu_{\mathbf{A}_q}(\mathbf{x}_p) \cdot CF_q \mid R_q \in S\}. \quad (11)$$

As shown in (11), the winner rule $R_w$ has the maximum product of the compatibility grade and the rule weight in $S$. For other fuzzy reasoning methods for pattern classification problems, see Cordon et al. [25] and Ishibuchi et al. [17], [26]. It should be noted that the choice of an appropriate rule weight specification depends on the type of fuzzy reasoning (i.e., single winner rule-based fuzzy reasoning) used in fuzzy rule-based classification systems [17], [29].

### D. Rule Evaluation Criteria

Using the above-mentioned procedure, we can generate a large number of fuzzy rules by specifying the consequent class and the rule weight for each of the $15^n$ combinations of the antecedent fuzzy sets. It is, however, very difficult for human users to handle such a large number of generated fuzzy rules. It is also very difficult to intuitively understand long fuzzy rules with many antecedent conditions. Thus we only generate short fuzzy rules with only a small number of antecedent conditions. It should be noted that *don't care* conditions with the special antecedent fuzzy set [0, 1] can be omitted from fuzzy rules. The rule length means the number of antecedent conditions excluding *don't care* conditions. We examine only short fuzzy rules of length $L_{\max}$ or less (e.g., $L_{\max} = 3$). This restriction is to find a small number of short (i.e., simple) fuzzy rules with high interpretability.

Among short fuzzy rules, we choose a prespecified number of good rules by a heuristic rule evaluation criterion. In the field of data mining, two rule evaluation criteria (i.e.,

confidence and support) have been often used [30], [31]. We have already shown the fuzzy version of the confidence criterion in (8). In the same manner, the support of the fuzzy rule "$\mathbf{A}_q \Rightarrow$ Class $h$" is calculated as follows [17]:

$$s(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{m}. \quad (12)$$

In our computational experiments, we use the following four rule evaluation criteria to extract a prespecified number of short fuzzy rules for each class from numerical data:

**Support with the minimum confidence level:** Each rule is evaluated based on its support when its confidence is larger than or equal to the prespecified minimum confidence level. This criterion never extracts unqualified rules whose confidence is smaller than the minimum confidence level. Various values of the minimum confidence level (e.g., 0.1, 0.2, ..., 0.9) are examined in computational experiments.

**Confidence with the minimum support level:** Each rule is evaluated based on its confidence when its support is larger than or equal to the prespecified minimum support level. This criterion never extracts unqualified rules whose support is smaller than the minimum support level. Various values of the minimum support level (e.g., 0.01, 0.02, ..., 0.09) are examined in computational experiments.

**Product of confidence and support:** Each rule is evaluated based on the product of its confidence and support.

**Difference in support:** Each rule is evaluated based on the difference between its support and the total support of the other rules with the same antecedent part and different consequent classes. More specifically, the rule $R_q$ with the antecedent fuzzy vector $\mathbf{A}_q$ and the consequent class $C_q$ is evaluated as

$$f(R_q) = s(\mathbf{A}_q \Rightarrow \text{Class } C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^{M} s(\mathbf{A}_q \Rightarrow \text{Class } h). \quad (13)$$

This criterion can be viewed as a simplified version of a rule evaluation criterion used in an iterative fuzzy genetics-based machine learning algorithm called SLAVE [32].

We generate a prespecified number of fuzzy rules with the largest values of each criterion in a greedy manner for each class. As we have already mentioned, only short fuzzy rules of length $L_{\max}$ or less are examined in heuristic rule extraction in order to find interpretable fuzzy rules.

## IV. Computational Experiment using Heuristic Fuzzy Rule Extraction

In this section, we perform computational experiments using heuristic fuzzy rule extraction based on rule evaluation criteria. Extracted fuzzy rules are used as candidate rules in multiobjective genetic fuzzy rule selection in the next section.

### A. Data Sets

We use six data sets in Table I: Wisconsin breast cancer (Breast W), diabetes (Diabetes), glass identification (Glass), Cleveland heart disease (Heart C), sonar (Sonar), and wine recognition (Wine) data sets. These six data sets are available from the UC Irvine machine learning repository. Data sets with missing values are marked by "*" in the third column of Table I. Since we do not use incomplete patterns with missing values, the number of patterns in the third column does not include those patterns with missing values. All attribute values are normalized into real numbers in the unit interval [0, 1]. For comparison, we show in the last two columns of Table I the reported results in Elomaa & Rousu [33] where six variants of the C4.5 algorithm were examined. The generalization ability of each variant was evaluated by ten independent runs (with different data partitions) of the whole ten-fold cross-validation (10CV) procedure (i.e., $10\times10$CV) in [33]. We show in the last two columns of Table I the best and worst error rates on test patterns among the six variants reported in [33] for each data set.

In our computational experiments, we also use the 10CV procedure. As in [33], 10CV is iterated ten times (i.e., $10\times10$CV). In each of 100 runs in $10\times10$CV, error rates are calculated on training patterns (90% of the given patterns) as well as test patterns (10% of the given patterns).

TABLE I
Data Sets Used in Our Computational Experiments

| Data set | Attributes | Patterns | Classes | C4.5 in [33] | |
|---|---|---|---|---|---|
| | | | | Best | Worst |
| Breast W | 9 | 683* | 2 | 5.1 | 6.0 |
| Diabetes | 8 | 768** | 2 | 25.0 | 27.2 |
| Glass | 9 | 214 | 6 | 27.3 | 32.2 |
| Heart C | 13 | 297* | 5 | 46.3 | 47.9 |
| Sonar | 60 | 208 | 2 | 24.6 | 35.8 |
| Wine | 13 | 178 | 3 | 5.6 | 8.8 |

* Incomplete patterns with missing values are not included.

** Some suspicious patterns with attribute value "0" are included.

### B. Performance on Training Patterns

In heuristic fuzzy rule extraction, various specifications are used as the number of extracted fuzzy rules in order to examine the relation between the accuracy and complexity of fuzzy rule-based systems. The number of extracted fuzzy rules for each class is specified as 1, 2, 3, 4, 5, 10, 20, 30, 40, 50, and 100. The four rule evaluation criteria in Section 3 are used in heuristic rule extraction. When multiple fuzzy rules have the same value of a rule evaluation criterion, those rules are randomly ordered (i.e., random tie break).

The maximum rule length $L_{\max}$ is specified as $L_{\max} = 2$ for the sonar data with 60 attributes and $L_{\max} = 3$ for the other data sets. That is, fuzzy rules of length 2 or less are examined for the sonar data while those of length 3 or less are examined for the other data sets. We use such a different specification because only the sonar data set involves a large number of attributes (i.e., it has a huge number of possible combinations of antecedent fuzzy sets).

Among the four heuristic rule evaluation criteria, good results are obtained from the support with the minimum confidence level, the product of confidence and support, and

the difference in support. Experimental results using these criteria are summarized in Tables II-VII where the average error rate on training patterns is calculated over $10\times10$CV. Five values of the minimum confidence level are examined in these tables. The best result (i.e., the lowest error rate) in each row is highlighted by underlined boldface.

From Tables II-VII, we can see that the increase in the number of extracted fuzzy rules does not always lead to the decrease in the error rates (e.g., see the last column of Table II). This observation suggests that the classification accuracy of heuristically extracted fuzzy rules can be improved by rule selection. We can also see that the choice of an appropriate rule evaluation criterion is problem-dependent.

TABLE II
ERROR RATES ON TRAINING PATTERNS (BREAST W)

| Number of rules | Support with the minimum confidence 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Product | Diff. |
|---|---|---|---|---|---|---|---|
| 1 | 9.47 | 9.47 | 8.99 | 7.24 | 9.33 | 5.89 | **5.81** |
| 2 | 9.08 | 9.08 | 7.53 | 7.77 | 8.24 | 6.41 | **6.32** |
| 3 | 7.60 | 7.60 | **5.29** | 6.69 | 6.45 | 6.39 | 6.47 |
| 4 | 5.30 | 5.30 | 5.20 | 6.48 | 6.60 | **5.18** | 5.64 |
| 5 | 5.27 | 5.27 | 5.43 | 6.46 | 6.93 | 4.81 | **4.72** |
| 10 | 6.37 | 5.78 | 5.38 | 5.20 | 6.08 | 4.29 | **4.20** |
| 20 | 5.12 | 4.84 | 4.29 | 4.10 | 4.45 | **3.61** | **3.61** |
| 30 | 4.26 | 4.12 | 4.23 | 4.35 | 4.44 | 3.78 | **3.75** |
| 40 | 4.46 | 4.44 | 4.43 | 4.39 | 4.41 | 3.82 | **3.76** |
| 50 | 4.42 | 4.34 | 4.29 | 4.18 | 4.37 | 3.83 | **3.73** |
| 100 | 3.94 | 3.94 | 3.94 | 3.94 | 4.26 | 3.81 | **3.70** |

TABLE III
ERROR RATES ON TRAINING PATTERNS (DIABETES)

| Number of rules | Support with the minimum confidence 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Product | Diff. |
|---|---|---|---|---|---|---|---|
| 1 | 34.43 | 33.53 | **26.39** | 26.73 | 46.44 | 34.27 | **26.39** |
| 2 | 34.96 | 33.02 | 26.48 | **26.44** | 41.65 | 34.13 | **26.44** |
| 3 | 34.96 | 31.28 | 28.57 | **26.42** | 38.21 | 34.59 | 27.45 |
| 4 | 34.90 | 31.55 | 29.04 | **26.34** | 35.63 | 33.57 | 28.02 |
| 5 | 34.90 | 31.48 | 29.11 | **26.11** | 34.07 | 32.44 | 28.79 |
| 10 | 34.90 | 30.40 | 29.13 | **26.15** | 31.14 | 30.17 | 29.72 |
| 20 | 34.90 | 30.23 | 29.78 | **26.18** | 28.33 | 30.63 | 30.22 |
| 30 | 34.66 | 30.34 | 30.13 | **26.41** | 26.96 | 30.47 | 30.34 |
| 40 | 33.06 | 30.29 | 30.31 | 26.59 | **25.79** | 30.47 | 30.42 |
| 50 | 31.74 | 30.29 | 30.37 | 26.72 | **24.61** | 30.61 | 30.51 |
| 100 | 31.14 | 30.75 | 30.32 | 26.86 | **22.77** | 30.78 | 30.65 |

TABLE IV
ERROR RATES ON TRAINING PATTERNS (GLASS)

| Number of rules | Support with the minimum confidence 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Product | Diff. |
|---|---|---|---|---|---|---|---|
| 1 | 47.49 | 43.51 | **38.53** | 63.65 | 76.77 | 45.30 | 39.81 |
| 2 | 44.06 | 42.76 | **37.58** | 61.63 | 76.43 | 44.67 | 39.87 |
| 3 | 42.54 | 42.15 | **37.08** | 59.78 | 76.02 | 44.52 | 39.70 |
| 4 | 42.67 | 41.92 | **36.76** | 58.46 | 75.78 | 44.22 | 39.68 |
| 5 | 42.92 | 41.53 | **36.55** | 57.63 | 75.69 | 43.75 | 39.29 |
| 10 | 40.97 | 39.84 | **36.18** | 55.14 | 74.40 | 40.08 | 38.64 |
| 20 | 39.70 | 38.09 | **34.76** | 52.64 | 72.54 | 38.37 | 38.25 |
| 30 | 38.41 | 37.41 | **33.97** | 51.96 | 71.07 | 37.69 | 37.78 |
| 40 | 38.00 | 36.68 | **33.41** | 51.42 | 70.18 | 37.06 | 37.18 |
| 50 | 37.91 | 35.76 | **33.07** | 50.68 | 68.84 | 36.67 | 36.43 |
| 100 | 35.92 | 33.16 | **32.79** | 49.39 | 66.96 | 35.16 | 34.72 |

TABLE V
ERROR RATES ON TRAINING PATTERNS (HEART C)

| Number of rules | Support with the minimum confidence 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Product | Diff. |
|---|---|---|---|---|---|---|---|
| 1 | 44.46 | **41.98** | 49.18 | 51.51 | 57.60 | 49.37 | 53.27 |
| 2 | 44.32 | **42.91** | 47.22 | 50.13 | 57.61 | 48.68 | 53.09 |
| 3 | 44.79 | **43.79** | 44.79 | 50.12 | 57.57 | 48.17 | 53.01 |
| 4 | 44.88 | **43.81** | 44.58 | 50.06 | 57.44 | 46.99 | 52.40 |
| 5 | 44.54 | 42.93 | **42.43** | 45.52 | 56.57 | 45.37 | 50.47 |
| 10 | 44.78 | 42.94 | **41.14** | 43.91 | 55.83 | 41.58 | 47.39 |
| 20 | 44.41 | 42.49 | **39.72** | 42.35 | 54.45 | 40.72 | 43.51 |
| 30 | 44.09 | 42.00 | **39.01** | 41.07 | 53.29 | 40.25 | 41.48 |
| 40 | 43.84 | 41.52 | **38.59** | 40.40 | 51.11 | 39.91 | 40.02 |
| 50 | 43.61 | 40.99 | **38.00** | 39.61 | 49.54 | 39.68 | 39.15 |
| 100 | 42.75 | 39.45 | 36.91 | 36.97 | 45.61 | 38.41 | **36.76** |

TABLE VI
ERROR RATES ON TRAINING PATTERNS (SONAR)

| Number of rules | Support with the minimum confidence 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Product | Diff. |
|---|---|---|---|---|---|---|---|
| 1 | 47.12 | **24.86** | 25.65 | 40.34 | 43.90 | 46.97 | 25.20 |
| 2 | 46.69 | 25.13 | **24.42** | 35.00 | 36.75 | 45.20 | 24.74 |
| 3 | 46.21 | 25.13 | **23.86** | 32.56 | 30.72 | 43.86 | 24.13 |
| 4 | 46.31 | 24.80 | **23.47** | 30.31 | 28.67 | 42.93 | 23.74 |
| 5 | 46.14 | 24.71 | **23.28** | 28.78 | 27.58 | 42.57 | 23.46 |
| 10 | 43.81 | 23.85 | **22.44** | 24.96 | 25.34 | 40.80 | 22.79 |
| 20 | 42.84 | 23.53 | **21.12** | 22.51 | 22.60 | 42.01 | 22.21 |
| 30 | 45.08 | 23.80 | **21.06** | 22.23 | 21.34 | 41.18 | 21.91 |
| 40 | 45.71 | 23.69 | 21.28 | 22.16 | **19.70** | 39.70 | 21.60 |
| 50 | 45.41 | 23.55 | 21.31 | 22.41 | **17.88** | 37.20 | 21.35 |
| 100 | 45.92 | 23.77 | 21.64 | 22.55 | **12.71** | 28.65 | 20.84 |

TABLE VII
ERROR RATES ON TRAINING PATTERNS (WINE)

| Number of rules | Support with the minimum confidence 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Product | Diff. |
|---|---|---|---|---|---|---|---|
| 1 | 32.46 | 14.36 | 11.14 | **9.60** | 9.81 | 10.03 | 10.15 |
| 2 | 25.05 | 12.36 | 9.26 | **5.65** | 7.32 | 7.32 | 6.80 |
| 3 | 14.03 | 13.24 | 8.10 | **5.44** | 6.01 | 6.21 | 6.77 |
| 4 | 13.03 | 12.62 | 7.10 | 5.53 | **5.39** | 5.44 | 5.65 |
| 5 | 13.42 | 11.76 | 5.96 | 5.56 | **5.32** | 5.54 | 5.38 |
| 10 | 12.86 | 7.38 | 5.49 | 5.17 | **3.53** | 5.82 | 4.90 |
| 20 | 9.47 | 5.07 | 5.11 | 4.91 | **3.45** | 4.90 | 4.11 |
| 30 | 6.25 | 5.14 | 5.11 | 4.64 | **3.56** | 4.44 | 3.86 |
| 40 | 5.03 | 4.96 | 5.01 | 4.31 | **3.46** | 4.05 | 3.63 |
| 50 | 5.03 | 4.86 | 4.80 | 4.07 | **3.28** | 3.85 | 3.55 |
| 100 | 4.43 | 4.69 | 4.09 | 3.34 | **3.09** | 3.46 | 3.22 |

### C. *Performance on Test Patterns*

Experimental results on test patterns are summarized in Tables VIII-XI (Due to the page limitation, we only show experimental results for the first four data sets). During ten iterations of the whole 10CV procedure, average error rates in Tables VIII-XI are calculated on test patterns while those in Tables II-VII in the previous subsection are calculated on training patterns. We have almost the same observations from Tables VIII-XI for test patterns as Tables II-VII for training patterns. That is, the choice of an appropriate rule evaluation criterion is problem-dependent. The increase in the number of extracted fuzzy rules does not always increase their classification accuracy.

TABLE VIII
ERROR RATES ON TEST PATTERNS (BREAST W)

| Number of rules | Support with the minimum confidence 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Product | Diff. |
|---|---|---|---|---|---|---|---|
| 1 | 9.60 | 9.60 | 9.07 | 7.35 | 9.35 | 6.59 | **6.49** |
| 2 | 9.16 | 9.16 | 8.30 | 8.13 | 8.70 | 7.01 | **6.90** |
| 3 | 8.34 | 8.34 | **5.62** | 7.09 | 6.65 | 7.07 | 7.06 |
| 4 | 5.71 | 5.71 | **5.46** | 6.87 | 6.68 | 5.95 | 6.38 |
| 5 | 5.54 | 5.54 | 5.46 | 6.62 | 7.06 | 5.42 | **5.27** |
| 10 | 6.50 | 6.02 | 5.58 | 5.33 | 6.31 | **4.53** | 4.54 |
| 20 | 5.41 | 5.26 | 4.77 | 4.42 | 4.72 | **3.76** | 3.81 |
| 30 | 4.45 | 4.28 | 4.47 | 4.55 | 4.64 | 3.97 | **3.95** |
| 40 | 4.69 | 4.61 | 4.58 | 4.53 | 4.58 | **4.03** | 4.04 |
| 50 | 4.51 | 4.45 | 4.44 | 4.31 | 4.48 | 4.14 | **4.06** |
| 100 | 4.00 | 4.00 | 3.98 | **3.97** | 4.45 | 4.17 | 4.04 |

TABLE IX
ERROR RATES ON TEST PATTERNS (DIABETES)

| Number of rules | Support with the minimum confidence 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Product | Diff. |
|---|---|---|---|---|---|---|---|
| 1 | 34.37 | 33.18 | **26.57** | 26.92 | 46.97 | 34.30 | **26.57** |
| 2 | 34.97 | 32.33 | 26.67 | **26.53** | 42.28 | 34.21 | 26.59 |
| 3 | 34.95 | 30.63 | 29.00 | **26.63** | 39.02 | 34.64 | 28.27 |
| 4 | 34.90 | 31.03 | 29.30 | **26.55** | 36.45 | 34.24 | 28.79 |
| 5 | 34.90 | 30.93 | 29.37 | **26.18** | 34.75 | 33.06 | 29.29 |
| 10 | 34.90 | 30.41 | 29.63 | **26.57** | 31.87 | 30.42 | 30.09 |
| 20 | 34.90 | 30.34 | 30.07 | **26.80** | 29.17 | 30.76 | 30.39 |
| 30 | 34.78 | 30.54 | 30.32 | **27.15** | 27.87 | 30.73 | 30.56 |
| 40 | 33.44 | 30.48 | 30.54 | 27.10 | **27.07** | 30.64 | 30.59 |
| 50 | 32.19 | 30.45 | 30.63 | 27.37 | **26.17** | 30.75 | 30.67 |
| 100 | 31.43 | 30.86 | 30.60 | 27.53 | **24.52** | 30.84 | 30.84 |

TABLE X
ERROR RATES ON TEST PATTERNS (GLASS)

| Number of rules | Support with the minimum confidence 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Product | Diff. |
|---|---|---|---|---|---|---|---|
| 1 | 52.24 | 49.57 | **42.59** | 71.58 | 81.75 | 49.05 | 46.86 |
| 2 | 49.14 | 49.10 | **41.94** | 70.79 | 81.61 | 48.16 | 47.24 |
| 3 | 47.08 | 48.76 | **41.28** | 69.05 | 81.19 | 47.98 | 47.34 |
| 4 | 47.40 | 48.76 | **41.09** | 68.07 | 81.05 | 48.26 | 47.63 |
| 5 | 47.32 | 48.11 | **41.42** | 67.37 | 80.86 | 48.21 | 47.29 |
| 10 | 46.19 | 47.13 | **42.46** | 65.48 | 80.01 | 47.34 | 47.72 |
| 20 | 46.44 | 47.27 | **41.90** | 63.14 | 78.44 | 47.09 | 47.38 |
| 30 | 46.53 | 47.04 | **41.38** | 62.30 | 76.48 | 47.71 | 46.82 |
| 40 | 46.80 | 46.37 | **40.86** | 61.59 | 75.88 | 47.29 | 47.06 |
| 50 | 47.04 | 46.14 | **41.39** | 60.44 | 74.42 | 47.15 | 46.18 |
| 100 | 46.03 | 45.55 | **41.21** | 58.87 | 72.79 | 45.76 | 44.31 |

TABLE XI
ERROR RATES ON TEST PATTERNS (HEART C)

| Number of rules | Support with the minimum confidence 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | Product | Diff. |
|---|---|---|---|---|---|---|---|
| 1 | **46.81** | 47.56 | 56.61 | 58.66 | 63.23 | 57.05 | 61.20 |
| 2 | **46.24** | 47.02 | 53.95 | 56.90 | 63.23 | 56.41 | 60.99 |
| 3 | **46.24** | 46.34 | 49.79 | 56.90 | 63.23 | 55.54 | 60.96 |
| 4 | **46.21** | 46.31 | 49.28 | 56.73 | 63.09 | 53.92 | 59.78 |
| 5 | **46.14** | 46.34 | 47.62 | 51.12 | 62.65 | 52.77 | 57.79 |
| 10 | 46.11 | **45.94** | 47.29 | 49.24 | 61.78 | 49.74 | 56.08 |
| 20 | 45.91 | **45.87** | 46.55 | 48.74 | 60.73 | 48.38 | 52.70 |
| 30 | 45.87 | **45.74** | 46.41 | 48.16 | 61.11 | 46.77 | 50.95 |
| 40 | **45.77** | 45.88 | 46.34 | 48.10 | 58.92 | 46.64 | 49.04 |
| 50 | **45.77** | 45.94 | 46.17 | 48.06 | 57.81 | 46.07 | 48.46 |
| 100 | **45.74** | 46.00 | 46.00 | 47.52 | 56.67 | 46.04 | 47.58 |

## V. COMPUTATIONAL EXPERIMENT USING MULTIOBJECTIVE GENETIC FUZZY RULE SELECTION

### A. Settings of Computational Experiments

For each data set, we choose a heuristic rule evaluation criterion from which the lowest error rate on test patterns was obtained for the case of 100 fuzzy rules for each class in the previous subsection. For example, the support criterion with the minimum confidence level 0.8 is chosen for the Wisconsin breast cancer data set (see Table VIII).

As in the previous section, we iterate the whole ten-fold cross-validation procedure ten times ($10\times10$CV). In each of 100 runs in $10\times10$CV, we generate 300 fuzzy rules for each class (i.e., $300M$ rules in total for an $M$-class classification problem) from training patterns using the chosen heuristic rule evaluation criterion for each data set. Multiobjective genetic rule selection based on the NSGA-II algorithm is applied to the extracted $300M$ fuzzy rules for each data set using the following parameter specifications:

Population size: 200 strings,
Crossover probability: 0.8 (uniform crossover),
Biased mutation probabilities:

$$p_{\mathrm{m}}(0\to1)=1/300M \quad \text{and} \quad p_{\mathrm{m}}(1\to0)=0.1,$$

Stopping condition: 5000 generations.

Multiple non-dominated rule sets are obtained by the NSGA-II algorithm in each of 100 runs in $10\times10$CV. We calculate the error rates of each rule set on training patterns and test patterns. Then the average error rates on training patterns and test patterns are calculated over obtained rule sets with the same number of fuzzy rules among 100 runs. Only when rule sets with the same number of fuzzy rules are found in all the 100 runs, we report the average error rates for that number of fuzzy rules.

### B. Performance on Training Patterns

Experimental results on training patterns are shown in Fig. 2 where open circles denote the performance of non-dominated rule sets obtained by multiobjective genetic rule selection. For comparison, we show in Fig. 2 experimental results of heuristic rule extraction using the same rule evaluation criterion as in the candidate rule generation in multiobjective genetic rule selection. For example, closed circles in Fig. 2 (a) correspond to the error rates in Table II by the support criterion with the minimum confidence level 0.8. This criterion is used in the candidate rule generation as explained in the previous subsection.

We can see from Fig. 2 that the accuracy of heuristically extracted fuzzy rules on training patterns is improved by multiobjective genetic rule selection for all the six data sets.

### C. Performance on Test Patterns

Experimental results on test patterns by multiobjective genetic rule selection are shown in Fig. 3 together with those by heuristic rule extraction. The reported results by the C4.5 algorithm in [33] are also shown in Fig. 3 for comparison (see the last two columns of Table I).

From Fig. 3, we can see that the generalization ability of

heuristically extracted fuzzy rules is improved for the five data sets (except for Heart C in Fig. 3 (d)) by multiobjective genetic rule selection. In Fig. 3 (d), we observe the increase in the error rates of obtained non-dominated fuzzy rule sets due to the increase in the number of fuzzy rules. We can also see that the generalization ability of obtained non-dominated fuzzy rule sets is comparable to the reported results by the C4.5 algorithm in many cases (except for Fig. 3 (c)).

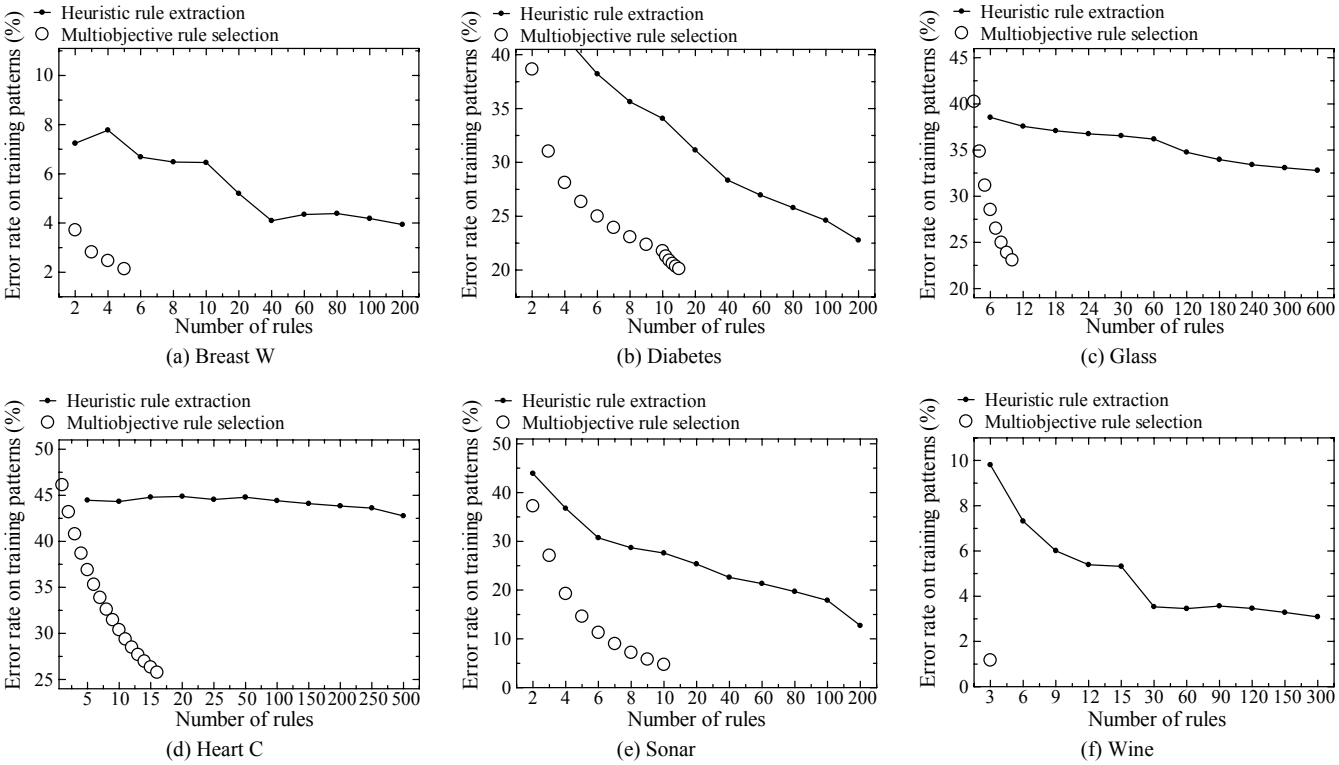


Fig. 2. Error rates on training patterns.

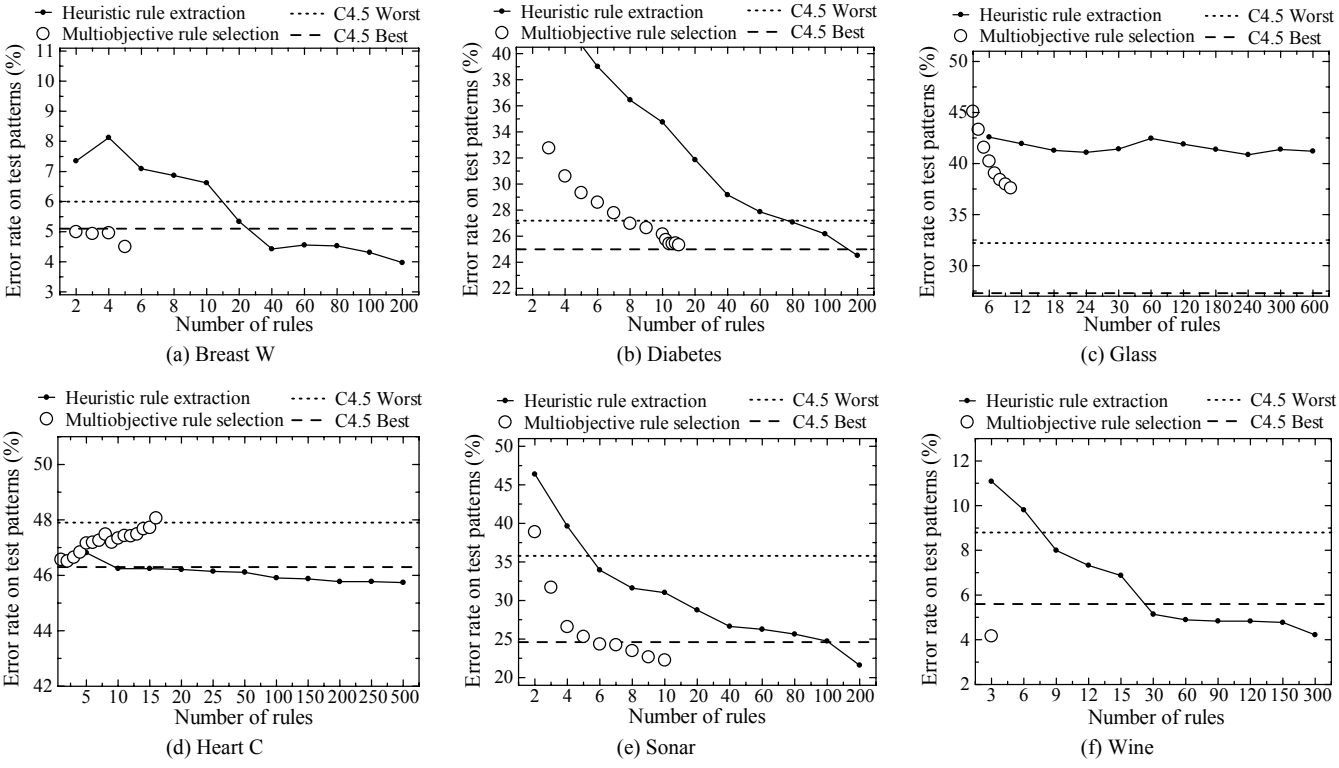


Fig. 3. Error rates on test patterns.

## VI. CONCLUSIONS

We showed that multiobjective genetic rule selection can decrease the number of heuristically extracted fuzzy rules while improving their classification accuracy on training patterns. Their generalization ability for test patterns was also improved by multiobjective genetic rule selection in many cases. Since a large number of fuzzy rules are usually extracted in a heuristic manner, our experimental results suggest the usefulness of multiobjective genetic fuzzy rule selection as a postprocessing procedure in fuzzy data mining with respect to the understandability of extracted knowledge.

## REFERENCES


[1] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*, John Wiley & Sons, Chichester, 2001.

[2] C. A. Coello Coello, D. A. van Veldhuizen, and G. B. Lamont, *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer Academic Publishers, Boston, 2002.

[3] C. A. Coello Coello and G. B. Lamont, *Applications of Multi-Objective Evolutionary Algorithms*, World Scientific, Singapore, 2004.

[4] M. A. Kupinski and M. A. Anastasio, "Multiobjective genetic optimization of diagnostic classifiers with implications for generating receiver operating characteristic curve," *IEEE Trans. on Medical Imaging*, vol. 18, no. 8, pp. 675-685, August 1999.

[5] J. Gonzalez, I. Rojas, J. Ortega, H. Pomares, F. J. Fernandez, and A. F. Diaz, "Multiobjective evolutionary optimization of the size, shape, and position parameters of radial basis function networks for function approximation," *IEEE Trans. on Neural Networks*, vol. 14, no. 6, pp. 1478-1495, November 2003.

[6] X. Llora and D. E. Goldberg, "Bounding the effect of noise in multiobjective learning classifier systems," *Evolutionary Computation*, vol. 11, no, 3, pp. 278-297, September 2003.

[7] H. A. Abbass, "Speeding up back-propagation using multiobjective evolutionary algorithms," *Neural Computation*, vol. 15, no. 11, pp. 2705-2726, November 2003.

[8] H. A. Abbass, "Pareto neuro-evolution: Constructing ensemble of neural networks using multi-objective optimization," *Proc. of Congress on Evolutionary Computation*, pp. 2074-2080, Canberra, Australia, December 8-12, 2003.

[9] A. Chandra and X. Yao, "DIVACE: Diverse and accurate ensemble learning algorithm," *Lecture Notes in Computer Science 3177: Intelligent Data Engineering and Automated Learning - IDEAL 2004*, Springer, Berlin, pp. 619-625, August 2004.

[10] A. Chandra and X. Yao, "Evolutionary framework for the construction of diverse hybrid ensemble," *Proc. of the 13th European Symposium on Artificial Neural Networks - ESANN 2005*, pp. 253-258, Brugge, Belgium, April 27-29, 2005.

[11] H. Ishibuchi and T. Yamamoto, "Evolutionary multiobjective optimization for generating an ensemble of fuzzy rule-based classifiers," *Lecture Notes in Computer Science*, vol. 2723, *Genetic and Evolutionary Computation - GECCO 2003*, pp. 1077-1088, Springer, Berlin, July 2003.

[12] 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, July 1997.

[13] 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, August 2001.

[14] O. Cordon, M. J. del Jesus, F. Herrera, L. Magdalena, and P. Villar, "A multiobjective genetic learning process for joint feature selection and granularity and contexts learning in fuzzy rule-based classification systems," in J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Interpretability Issues in Fuzzy Modeling*, pp. 79-99, Springer, Berlin, 2003.

[15] F. Jimenez, A. F. Gomez-Skarmeta, G. Sanchez, H. Roubos, and R. Babuska, "Accurate, transparent and compact fuzzy models by multi-objective evolutionary algorithms," in J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Interpretability Issues in Fuzzy Modeling*, pp. 431-451, Springer, Berlin, 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, January 2004.

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

[18] H. Wang, S. Kwong, Y. Jin, W. Wei, and K. F. Man, "Agent-based evolutionary approach for interpretable rule-based knowledge extraction," *IEEE Trans. on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, vol. 35, no. 2, pp. 143-155, May 2005.

[19] H. Wang, S. Kwong, Y. Jin, W. Wei, and K. F. Man, "Multi-objective hierarchical genetic algorithm for interpretable fuzzy rule-based knowledge extraction," *Fuzzy Sets and Systems*, vol. 149, no. 1, pp. 149-186, January 2005.

[20] 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* (in press).

[21] Y. Jin (ed.), *Multi-Objective Machine Learning*, Springer, Berlin, 2006.

[22] 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, August 1994.

[23] 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, August 1995.

[24] 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, April 2002.

[25] 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*, vol. 20, no. 1, pp. 21-45, January 1999.

[26] H. Ishibuchi, T. Nakashima, and T. Morisawa, "Voting in fuzzy rule-based systems for pattern classification problems," *Fuzzy Sets and Systems*, vol. 103, no. 2, pp. 223-238, April 1999.

[27] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets and Systems*, vol. 52, no. 1, pp. 21-32, November 1992.

[28] 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, August 2001.

[29] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. on Fuzzy Systems*, vol. 13, no. 4, pp. 428-435, August 2005.

[30] 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.

[31] F. Coenen, P. Leng, and L. Zhang, "Threshold tuning for improved classification association rule mining," *Lecture Notes in Computer Science 3518: Advances in Knowledge Discovery And Data Mining - PAKDD 2005*, pp. 216-225, Springer, Berlin, May 2005.

[32] 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, April 1999.

[33] T. Elomaa and J. Rousu, "General and efficient multisplitting of numerical attributes," *Machine Learning*, vol. 36, no. 3, pp. 201-244, September 1999.