Proceedings of the 2007 IEEE Symposium on
Foundations of Computational Intelligence (FOCI 2007)

Evolutionary Multiobjective Design of Fuzzy Rule-Based Systems

Hisao Ishibuchi

Department of Computer Science and Intelligent Systems, Osaka Prefecture University

hisaoi@cs.osakafu-u.ac.jp

Abstract – The main advantage of fuzzy rule-based systems over other non-linear models such as neural networks is their high interpretability. Fuzzy rules can be usually interpreted in a linguistic manner because they are described by linguistic values such as *small* and *large*. Fuzzy rule-based systems have high accuracy as well as high interpretability. A large number of tuning methods have been proposed to improve their accuracy. Most of those tuning methods are based on learning algorithms of neural networks and/or evolutionary optimization techniques. Accuracy improvement of fuzzy rule-based systems, however, is usually achieved at the cost of interpretability. This is because the accuracy improvement often increases the complexity of fuzzy rule-based systems. Thus one important issue in the design of fuzzy rule-based systems is to find a good tradeoff between the accuracy and the complexity. The importance of finding a good accuracy-complexity tradeoff has been pointed out in some studies in the late 1990s. Recently evolutionary multiobjective optimization algorithms were used to search for various fuzzy rule-based systems with different accuracy-complexity tradeoffs. Users are supposed to choose a final model based on their preference from the obtained fuzzy rule-based systems. Some users may prefer a simple one with high interpretability. Other users may prefer a complicated one with high accuracy. In this paper, we explain evolutionary multiobjective approaches to the design of accurate and interpretable fuzzy rule-based systems. We also suggest some future research directions related to the evolutionary multiobjective design of fuzzy rule-based systems.

I. INTRODUCTION

A large number of tuning methods of fuzzy rule-based systems have been proposed to improve their accuracy in the literature [1]-[7]. Most of those tuning methods are based on evolutionary algorithms [8], [9] and/or learning schemes of neural networks such as the back-propagation algorithm [10]. Since fuzzy rule-based systems are universal approximators of nonlinear systems [11]-[13] as neural networks [14]-[16], we can improve the accuracy of fuzzy rule-based systems on training data by increasing their complexity. Complicated fuzzy rule-based systems with high accuracy on training data, however, do not necessarily have high generalization ability for unseen test data. In Fig. 1, we show a typical tradeoff relation between accuracy and complexity. As shown by the dotted curve in Fig. 1, training data accuracy of fuzzy rule-based systems is monotonically improved by increasing their complexity. On the other hand, test data accuracy is first improved to its maximum point then degraded by the increase in their complexity as shown by the continuous curve in Fig. 1. Such an undesirable deterioration in test data accuracy is known as the overfitting to training data [17]. Finding the optimal complexity with the maximum accuracy on test data (i.e., finding the optimal complexity S^* in Fig. 1) is one of the main research issues in machine learning especially in the field of statistical learning theory [18].

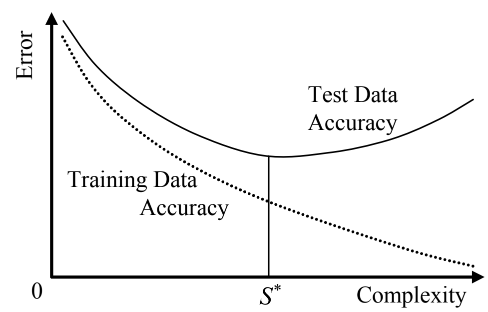


Fig. 1. A typical tradeoff relation between accuracy and complexity.

The main advantage of fuzzy rule-based systems over other non-linear models such as neural networks is their high interpretability. From the point of view of interpretability, the optimal complexity with the maximum accuracy on test data (i.e., S^* in Fig. 1) is not necessarily preferable. Some human users may prefer simpler fuzzy rule-based systems with higher interpretability than S^* in Fig. 1 whereas others may prefer the optimal complexity S^* with the maximum test data accuracy.

From the above discussions, we can see that the design of fuzzy rule-based systems can be written as the following two-objective optimization problem:

$$\text{Maximize } Accuracy(S) \text{ and minimize } Complexity(S), \quad (1)$$

where S, $Accuracy(S)$, and $Complexity(S)$ are a fuzzy rule-based system, an accuracy measure, and a complexity measure, respectively. It should be noted that $Accuracy(S)$ is usually calculated using training data.

Many tuning methods of fuzzy rule-based systems were proposed to maximize the accuracy on training data in the 1990s [1]-[7]. Those tuning methods can be viewed as optimizing the following single-objective problem:

$$\text{Maximize } \mathit{Accuracy}(S). \tag{2}$$

Since the late 1990s, the importance of interpretability in the design of fuzzy rule-based systems has been pointed out in a number of studies [19]-[31]. Whereas both accuracy and interpretability were considered, the finally obtained solution was a single fuzzy rule-based system in [19]-[31]. Those studies can be viewed as optimizing the following single-objective problem:

$$\text{Optimize } f(S) = f(\mathit{Accuracy}(S), \mathit{Complexity}(S)), \tag{3}$$

where $f(\cdot)$ is a scalarizing function, which combines an accuracy measure *Accuracy*(S) and a complexity measure *Complexity*(S) into a single scalar objective function. In some of those studies, the objective function in (3) can be more appropriately written as

$$\text{Optimize } f(S) = f(\mathit{Accuracy}(S), \mathit{Interpretability}(S)), \tag{4}$$

where *Interpretability*(S) is an interpretability measure.

An example of the scalarizing objective function $f(\cdot)$ in (3) is the following weighted sum:

$$\text{Maximize } f(S) = w_1 \cdot NCP(S) - w_2 \cdot Card(S), \tag{5}$$

where $\mathbf{w} = (w_1, \mathrm{w}_2)$, $NCP(S)$ and $Card(S)$ are a weight vector, the number of correctly classified training patterns by S, and the cardinality of S (i.e., the number of fuzzy rules in S), respectively. The weighted sum objective function in (5) was used in [19] to search for an accurate and interpretable fuzzy rule-based classifier.

Single-objective search using the weighted sum objective function is illustrated in Fig. 2. As shown in Fig. 2, a single fuzzy rule-based system is obtained as a final solution by maximizing the weighted sum objective function. The search direction represented by the arrow in Fig. 2 is specified by the weight vector $\mathbf{w} = (w_1, \mathrm{w}_2)$. The main difficulty of this approach is that the specification of the weight vector is not easy and problem-dependent whereas the finally obtained fuzzy rule-based system strongly depends on it.

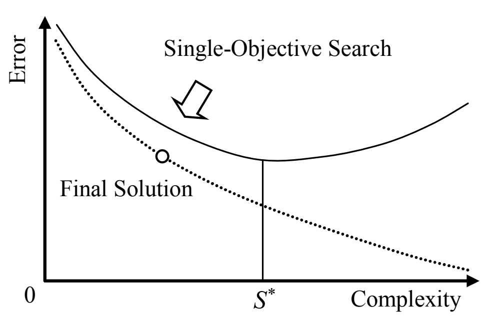


Fig. 2. Single-objective search based on a weighted sum objective function.

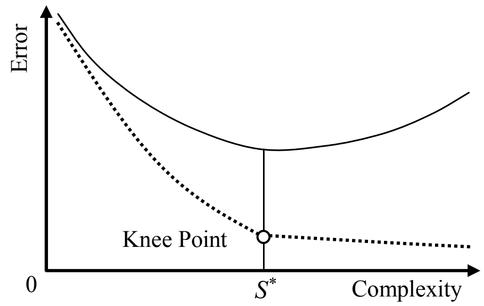


Fig. 3. An accuracy-complexity tradeoff relation with a knee solution.

If there exists a clear knee solution as shown in Fig. 3 (see Deb [32] for the concept of knee solutions), the finally obtained fuzzy rule-based system is the knee solution for a wide range of weight vectors. Unfortunately, the existence of a knee solution is problem-dependent. Some problems (e.g., with artificially given noise [33]) have clear knee solutions as in Fig. 3, other problems have no knee solutions as in Fig. 2.

Whereas the finally obtained solution is a single fuzzy rule-based system in single-objective approaches based on scalarizing functions, evolutionary multiobjective approaches [34]-[40] search for a number of fuzzy rule-based systems with different accuracy-complexity tradeoffs as shown in Fig. 4. More specifically, multiobjective approaches search for non-dominated fuzzy rule-based systems by solving the two-objective optimization problem in (1). In some studies (e.g., [35], [37], [40]), two complexity measures were used. In those studies, non-dominated fuzzy rule-based systems were found by solving the following three-objective problem:

$$\begin{aligned}&\text{Maximize } \mathit{Accuracy}(S) \text{ and}\\&\text{minimize } \mathit{Complexity}_1(S) \text{ and } \mathit{Complexity}_2(S),\end{aligned} \tag{6}$$

where $\mathit{Complexity}_1(S)$ and $\mathit{Complexity}_2(S)$ are complexity measures.

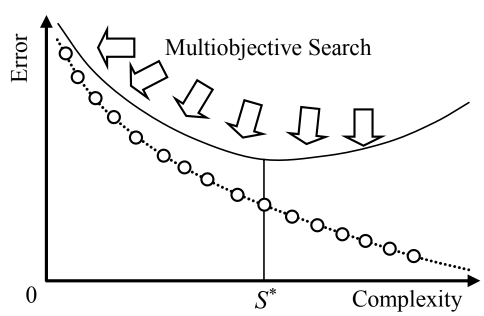


Fig. 4. Evolutionary multiobjective approaches that search for a number of fuzzy rule-based systems with different accuracy-complexity tradeoffs.

The following two-objective problem was used in [34] to search for a number of non-dominated fuzzy rule-based classifiers with respect to accuracy and complexity:

$$\text{Maximize } NCP(S) \text{ and minimize } Card(S). \tag{7}$$

This is a two-objective version of the weighted sum objective function in (5).

It should be noted that a large number of non-dominated fuzzy rule-based systems are obtained from a single run of evolutionary multiobjective approaches. The obtained fuzzy rule-based systems help human users to understand the accuracy-complexity tradeoff relation (see [41], [42] for further discussions on the accuracy-complexity tradeoff).

II. EVOLUTIONARY MULTIOBJECTIVE OPTIMIZATION

Since Schaffer's pioneering study [43], a large number of evolutionary multiobjective optimization (EMO) algorithms have been proposed and applied to various application tasks [32], [44], [45]. Currently EMO is one of the most active research areas in the field of evolutionary computation.

A k-objective maximization problem is written as follows:

$$\text{Maximize } \mathbf{f}(\mathbf{y}) = (f_1(\mathbf{y}),\ f_2(\mathbf{y}),\ ...,\ f_k(\mathbf{y}))\,, \tag{8}$$

$$\text{subject to } \mathbf{y} \in \mathbf{Y}\,, \tag{9}$$

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

Let $\mathbf{y}$ and $\mathbf{z}$ be two feasible solutions of the k-objective maximization problem in (8)-(9). If the following conditions hold, $\mathbf{z}$ can be viewed as being better than $\mathbf{y}$:

$$\forall\, i\,,\ f_i(\mathbf{y}) \le f_i(\mathbf{z}) \quad \text{and} \quad \exists\, j\,,\ f_j(\mathbf{y}) < f_j(\mathbf{z})\,. \tag{10}$$

In this case, we say that $\mathbf{z}$ dominates $\mathbf{y}$ (equivalently $\mathbf{y}$ is dominated by $\mathbf{z}$: $\mathbf{z}$ is better than $\mathbf{y}$).

When $\mathbf{y}$ is not dominated by any other feasible solutions (i.e., when there exists no feasible solution $\mathbf{z}$ that dominates $\mathbf{y}$), the solution $\mathbf{y}$ is referred to as a Pareto-optimal solution of the k-objective maximization problem in (8)-(9). 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.

One of the most well-known and frequently-used EMO algorithms is NSGA-II of Deb et al. [46]. In this paper, we use NSGA-II for multiobjective genetic fuzzy rule selection.

Let P and N_{pop} be the current population in NSGA-II and the population size, respectively (i.e., $N_{\text{pop}} = |P|$). Then 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 N_{pop} solutions are randomly generated to form an initial population P in Step 1 in the same manner as in standard single-objective genetic algorithms (SOGAs). Next N_{pop} pairs of parent solutions are selected from the current population P to form a parent population P' in Step 3. Then an offspring population P'' is constructed in Step 4 by generating a single offspring solution from each pair of parent solutions in P' by crossover and mutation. Genetic operations in Step 4 are the same as those in SOGAs. The next population is constructed in Step 5 by choosing the best N_{pop} solutions from the $2 \cdot N_{\text{pop}}$ solutions in the current population P and the offspring population P''. The parent selection in Step 3 and the generation update in Step 5 of NSGA-II are different from those in SOGAs. Pareto ranking and a crowding measure are used to evaluate each solution in each step. Binary tournament selection is used in Step 3 to choose parent solutions. For details of NSGA-II, see Deb [32] and Deb et al. [46].

III. MULTIOBJECTIVE GENETIC FUZZY RULE SELECTION

As an example of an evolutionary multiobjective approach to the design of fuzzy rule-based systems, we explain a basic form of multiobjective genetic fuzzy rule selection [35], [37]. A number of non-dominated fuzzy rule-based classifiers are found from a large number of candidate fuzzy rules. That is, multiobjective genetic fuzzy rule selection is performed as the following two-phase method:

Phase 1: Heuristic candidate rule extraction (data mining)
Phase 2: Rule selection (multiobjective optimization)

A. 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. 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 our classification problem is an M-class problem with m training patterns in the n-dimensional unit hypercube $[0, 1]^n$.

B. Fuzzy rules

For our n-dimensional pattern classification problem, we use fuzzy rules of the following form [47]:

$$\text{Rule } R_q : \text{If } x_1 \text{ is } A_{q1} \text{ and ... and } x_n \text{ is } A_{qn} \text{ then Class } C_q \text{ with } CF_q\,, \tag{11}$$

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, C_q is a class label, and CF_q is a rule weight (i.e., certainty grade). We also denote the fuzzy rule R_q in (11) as $\mathbf{A}_q \Rightarrow \text{Class } C_q$. The rule weight CF_q has a large effect on the accuracy of fuzzy rule-based classifiers as shown in [48], [49]. For other types of fuzzy rules for pattern classification problems, see [50]-[53].

Since we usually have no *a priori* information about an appropriate granularity of discretization (i.e., the number of antecedent fuzzy sets) for each attribute, we simultaneously use multiple fuzzy partitions with different granularities as shown in Fig. 5. In addition to the 14 fuzzy sets in Fig. 5, we also use the domain interval [0, 1] itself as an antecedent fuzzy set in order to represent a *don't care* condition. Thus we have the 15 possible antecedent fuzzy sets as A_{qi} for each attribute.

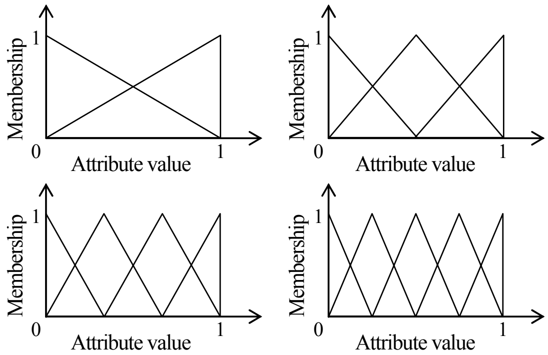


Fig. 5. Four fuzzy partitions used in our computational experiments.

C. Fuzzy rule generation

Since we have 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 as the antecedent part $\mathbf{A}_q$ of the fuzzy rule R_q in (11). Its consequent class C_q and rule weight CF_q are specified from compatible training patterns with $\mathbf{A}_q$ in the following heuristic manner.

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

$$\mu_{\mathbf{A}_q}(\mathbf{x}_p) = \mu_{A_{q1}}(x_{p1}) \cdot \ldots \cdot \mu_{A_{qn}}(x_{pn}), \tag{12}$$

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

Next the confidence of the fuzzy rule $\mathbf{A}_q \Rightarrow \text{Class } h$ is calculated for each class h as follows [53]-[55]:

$$c(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{\sum_{p=1}^{m} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}. \tag{13}$$

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

$$c(\mathbf{A}_q \Rightarrow \text{Class } C_q) = \max_{h=1,2,\ldots,M} \{c(\mathbf{A}_q \Rightarrow \text{Class } h)\}. \tag{14}$$

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 the early 1990s [47].

It should be noted that the same consequent class as in (13)-(14) is obtained when we use the support of the fuzzy rule $\mathbf{A}_q \Rightarrow \text{Class } h$ instead of the confidence. The support is calculated as follows [53]-[55]:

$$s(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{m}. \tag{15}$$

Different specifications of the rule weight CF_q have been proposed and examined. We use the following specification because good results were reported in the literature [49]:

$$CF_q = c(\mathbf{A}_q \Rightarrow \text{Class } C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^{M} c(\mathbf{A}_q \Rightarrow \text{Class } h). \tag{16}$$

D. Rule discovery criteria for candidate rule extraction

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 a few antecedent conditions. It should be noted that *don't care* conditions can be omitted from fuzzy rules. So 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 compact set of fuzzy rules with high interpretability.

Among short fuzzy rules, we only use fuzzy rules that satisfy both the minimum confidence and support as candidate rules in multiobjective genetic fuzzy rule selection. In the field of data mining (especially association rule mining [56] and classification rule mining [57]), these two rule evaluation criteria have been frequently used to generate meaningful association and classification rules.

E. Multiobjective genetic fuzzy rule selection

Let S be a subset of candidate fuzzy rules of the form in (11). That is, S is a fuzzy rule-based classifier. Each pattern $\mathbf{x}_p$ is classified by a single winner rule R_w, which is chosen from the rule set S as follows:

$$\mu_{\mathbf{A}_w}(\mathbf{x}_p) \cdot CF_w = \max\{\mu_{\mathbf{A}_q}(\mathbf{x}_p) \cdot CF_q \mid R_q \in S\}. \tag{17}$$

As in our former studies [35], [37], we use the following three objectives in multiobjective genetic fuzzy rule selection:

$f_1(S)$: The number of correctly classified training patterns by S,

$f_2(S)$: The number of selected fuzzy rules in S,

$f_3(S)$: The total number of antecedent conditions in S (i.e., the total rule length in S).

The first objective is maximized while the second and third ones are minimized. That is, our three-objective fuzzy rule selection problem is written as follows:

$$\text{Maximize } f_1(S), \text{ and minimize } f_2(S) \text{ and } f_3(S). \tag{18}$$

We apply NSGA-II [32], [46] to the three-objective fuzzy rule selection problem in (18). For the implementation of three-objective genetic fuzzy rule selection, see [35], [37].

F. Computational experiments

We applied multiobjective genetic fuzzy rule selection to some problems in the UCI machine learning repository. Incomplete patterns with missing values were not used.

We divided each data set into two subsets of the same size: training data and test data. Using training data, first we extracted fuzzy rules satisfying the minimum confidence 0.6 and the minimum support 0.01. The maximum rule length was specified as three in the rule extraction phase. All the extracted fuzzy rules were used as candidate rules.

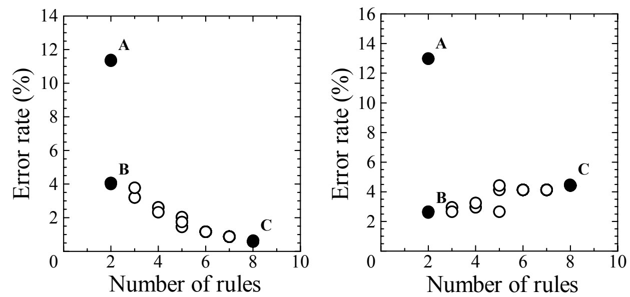
Then we applied NSGA-II to the extracted candidate rules to search for Pareto-optimal rule sets (i.e., Pareto-optimal subsets of the candidate rules) with respect to the three objectives using the following parameter specifications:

Population size: 200 strings,
Crossover probability: 0.9 (uniform crossover),
Mutation probability: 0.05 ($1 \rightarrow 0$),
$1/N$ ($0 \rightarrow 1$, N: string length),
Termination condition: 1000 generations.

In our multiobjective genetic fuzzy rule selection, string length N is the same as the number of the candidate rules because their subsets are represented by binary strings of length N. We use biased mutation where changes from 1 to 0 and from 0 to 1 have different mutation probabilities. We also use a hill-climbing procedure to remove unnecessary rules from each string (for details, see [35], [37]).

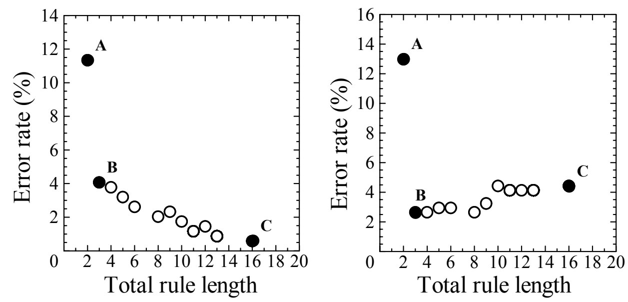
Experimental results by a single run on the Wisconsin breast cancer data set (Breast W) are summarized in Fig. 6 where error rates on training data and test data are depicted in each plot. In Fig. 6 (a), we can observe a clear accuracy-complexity tradeoff relation on training data. On the other hand, the overfitting to test data is clearly observed in Fig. 6 (b). The same rule sets are also shown in Fig. 7 using the total rule length as the horizontal axis. Two fuzzy rules in a rule set marked by A in Fig. 6 and Fig. 7 are shown in Fig. 8 where real numbers in parentheses in the last column denote rule weights. Fig. 8 is an example of a very simple fuzzy rule-based classifier with high interpretability. Another rule set with two fuzzy rules marked by B is shown in Fig. 9. On the other hand, a rule set with eight rules marked by C in Fig. 6 is shown in Fig. 10. The main advantage of multiobjective approaches over single-objective ones is that a number of non-dominated fuzzy rule-based systems are obtained by their single run as shown in Fig. 6. Some are very simple and interpretable (Fig. 8 and Fig. 9), and others are complicated and accurate on training data (Fig. 10).

On the other hand, experimental results by a single run on the Cleveland heart disease data set (Heart C) are summarized in Fig. 11 in the same manner as Fig. 6. As in Fig. 6 (a), we can observe a clear accuracy-complexity tradeoff relation for training data in Fig. 11 (a). The overfitting to training data in Fig. 11 (b), however, seems to be less severe than Fig. 6 (b).



(a) Training data accuracy. (b) Test data accuracy.

Fig. 6. Obtained non-dominated fuzzy rule sets (Breast W).



(a) Training data accuracy. (b) Test data accuracy.

Fig. 7. The same data sets as Fig. 6 with a different horizontal axis.

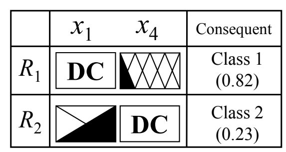


Fig. 8. Rule set A in Fig. 6 and Fig. 7.

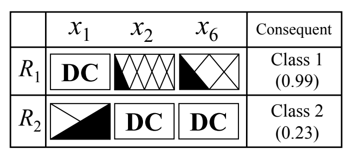


Fig. 9. Rule set B in Fig. 6 and Fig. 7.

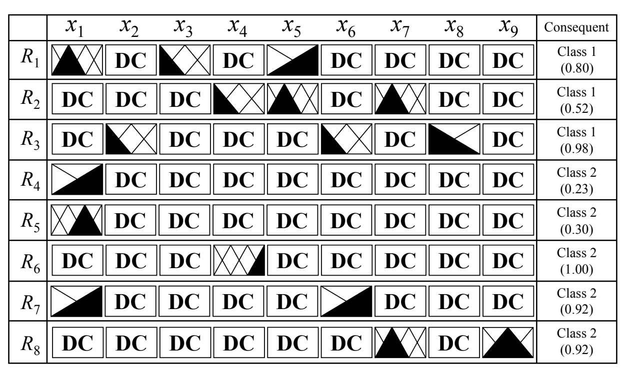
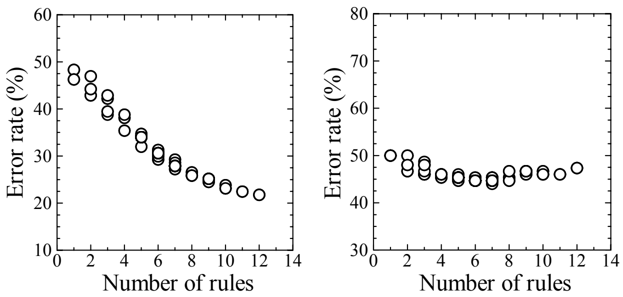


Fig. 10. Rule set C in Fig. 6 and Fig. 7.



(a) Training data accuracy. (b) Test data accuracy.

Fig. 11. Obtained non-dominated fuzzy rule sets (Heart C).

G. Some extensions

The choice of candidate rules has a dominant effect on the performance of multiobjective genetic fuzzy rule selection. For example, good rule sets with high accuracy are not likely to be found when good fuzzy rules are not used as candidate rules. The computation load for multiobjective genetic fuzzy rule selection directly depends on the number of candidate rules as well as the size of data sets. Thus the choice of a tractable number of promising candidate rules is an important issue for the success of rule selection.

For low-dimensional classification problems with four or less attributes, we can use all combinations of antecedent fuzzy sets to extract candidate rules. On the other hand, the prescreening of only promising candidate rules is necessary for high-dimensional problems. One idea is to choose a pre-specified number of the best candidate rules for each class with respect a pre-specified rule evaluation criterion. We used this idea in our former studies [35], [37]. It is, however, not easy to choose an appropriate rule evaluation criterion. This is because the choice of an appropriate rule evaluation criterion is problem-dependent [58].

In this paper, we use the minimum support and confidence for candidate rule prescreening as in data mining [56], [57]. The choice of appropriate threshold values of the two rule evaluation criteria is not easy. The number of extracted rules and their accuracy strongly depend on these two parameter values [58]-[60]. An idea of extracting only Pareto-optimal rules with respect to the support and the confidence [61]-[63] can be used in multiobjective genetic fuzzy rule selection as a prescreening procedure [64].

For accuracy improvement, asymmetric antecedent fuzzy sets in inhomogeneous fuzzy partitions [65] can be used for candidate rule extraction. Learning algorithms of antecedent fuzzy sets and/or rule weights [4]-[6], [53], [66] can be also incorporated in the candidate rule extraction phase as well as the multiobjective genetic rule selection phase.

IV. Future Research Directions

The following seem to be interesting future research directions related to evolutionary multiobjective design of fuzzy rule-based systems:

(1) Interpretability measures: It is an interesting and challenging issue to mathematically formulate various aspects of the interpretability of fuzzy rule-based systems (for further discussions on interpretability, see [41], [42], [67], [68]).

(2) Accuracy improvement: Various tuning methods can be incorporated into evolutionary multiobjective design methods. Evolutionary multiobjective clustering [69] can be also used to generate candidate rules or initial fuzzy rules.

(3) Theoretical tradeoff analysis: Theoretical tradeoff analysis between complexity and accuracy on test data seems to be needed to find a fuzzy rule-based system with high generalization ability (for such a theoretical tradeoff analysis for non-fuzzy systems, see [18]).

(4) Handling of large data sets: Data mining is a very active research area. Only a few approaches, however, have been proposed to evolutionary multiobjective fuzzy data mining [70], [71]. Those approaches were proposed to search for Pareto-optimal fuzzy rules. Whereas a large number of evolutionary approaches have already been proposed for data mining [72], it is not easy to apply evolutionary algorithms to large data sets due to their large computation cost. Some tricks for decreasing computational costs such as data set subdivisions [73] seem to be needed in the handling of large data sets.

(5) Applications to non-fuzzy machine learning and data mining: Recently the concept of multiobjective optimization has been used to develop multiobjective machine learning and data mining methods [74], [75]. Evolutionary techniques developed for the multiobjective design of fuzzy rule-based systems can be utilized as multiobjective machine learning and data mining techniques.

(6) Incorporation of users' preference: It is not easy for evolutionary multiobjective optimization algorithms to find a good non-dominated solution set that approximates the entire Pareto front of a large-scale multiobjective combinatorial optimization problem [76], [77]. It is a good idea to focus the multiobjective search on a particular area of the Pareto front using users' preference [78], [79].

V. Concluding Remarks

We first explained the basic idea of evolutionary multi-objective design of fuzzy rule-based systems. Then we demonstrated that a number of non-dominated fuzzy rule-based classifiers with different accuracy-complexity tradeoffs were obtained by a single run of multiobjective genetic fuzzy rule selection. Finally we pointed out some future research directions related to multiobjective fuzzy system design.

Acknowledgement

This work was partially supported by Grant-in-Aid for Scientific Research on Priority Areas (18049065) and for Scientific Research (B) (17300075). The author would like to thank Mr. Isao Kuwajima for his help with computational experiments in this paper.

REFERENCES

[1] S. Horikawa, T. Furuhashi, and Y. Uchikawa, "On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm," *IEEE Trans. on Neural Networks* 3 (5) 801-806, 1992.

[2] J. -S. R. Jang, "ANFIS: Adaptive-network-based fuzzy inference system," *IEEE Trans. on Systems, Man, and Cybernetics* 23 (3) 665-685, 1993.

[3] C. L. Karr and E. J. Gentry, "Fuzzy control of pH using genetic algorithms," *IEEE Trans. on Fuzzy Systems* 1 (1) 46-53, 1993.

[4] D. Nauck, F. Klawonn, and R. Kruse, *Foundations of Neuro-Fuzzy Systems*, John Wiley & Sons, 1997.

[5] S. Abe, *Pattern Classification: Neuro-fuzzy Methods and Their Comparison*, Springer, 2001.

[6] O. Cordon, F. Herrera, F. Hoffman, and L. Magdalena, *Genetic Fuzzy Systems*, World Scientific, 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) 5-31, 2004.

[8] J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, 1975.

[9] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, MA, 1989.

[10] D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing*, MIT Press, 1986.

[11] B. Kosko, "Fuzzy systems as universal approximators," *Proc. of FUZZ-IEEE 1992*, 1153-1162.

[12] L. X. Wang and J. M. Mendel, "Fuzzy basis functions, universal approximation, and orthogonal least-squares learning," *IEEE Trans. on Neural Networks* 3 (5) 807-814, 1992.

[13] J. M. Mendel, "Fuzzy-logic systems for engineering - A tutorial," *Proceedings of IEEE* 83 (3) 345-377, 1995.

[14] K. Funahashi, "On the approximate realization of continuous-mappings by neural networks," *Neural Networks* 2 (3) 183-192, 1989.

[15] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks* 2 (5) 359-366, 1989.

[16] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks* 4 (2) 251-257, 1991.

[17] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2001.

[18] V. Cherkassky and F. Mulier, *Learning from Data: Concepts, Theory, and Methods*, John Wiley & Sons, New York, 1998.

[19] 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) 260-270, 1995.

[20] M. Setnes, R. Babuska, and B. Verbruggen, "Rule-based modeling: Precision and transparency," *IEEE Trans. on Systems, Man, and Cybernetics - Part C* 28 (1) 165-169, 1998.

[21] J. Yen, L. Wang, and G. W. Gillespie, "Improving the interpretability of TSK fuzzy models by combining global learning and local learning," *IEEE Trans. on Fuzzy Systems* 6 (4) 530-537, 1998.

[22] Y. Jin, W. von Seelen, and B. Sendhoff, "On generating FC^3 fuzzy rule systems from data using evolution strategies," *IEEE Trans. on Systems, Man, and Cybernetics - Part B* 29 (6) 829-845, 1999.

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

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

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

[26] L. Castillo, A. Gonzalez, and R. Perez, "Including a simplicity criterion in the selection of the best rule in a genetic fuzzy learning algorithm," *Fuzzy Sets and Systems* 120 (2) 309-321, 2001.

[27] H. Roubos and M. Setnes, "Compact and transparent fuzzy models and classifiers through iterative complexity reduction," *IEEE Trans. on Fuzzy Systems* 9 (4) 516-524, 2001.

[28] J. Abonyi, J. A. Roubos, and F. Szeifert, "Data-driven generation of compact, accurate, and linguistically sound fuzzy classifiers based on a decision-tree initialization," *International Journal of Approximate Reasoning* 32 (1) 1-21, 2003.

[29] R. Alcala, J. R. Cano, O. Cordon, F. Herrera, P. Villar, and I. Zwir, "Linguistic modeling with hierarchical systems of weighted linguistic rules," *International Journal of Approximate Reasoning* 32 (2-3) 187-215, 2003.

[30] Y. C. Hu, R. S. Chen, and G. H. Tzeng, "Finding fuzzy classification rules using data mining techniques," *Pattern Recognition Letters* 24 (1-3) 509-519, 2003.

[31] J. Casillas, O. Cordon, M. J. del Jesus, and F. Herrera, "Genetic tuning of fuzzy rule deep structures preserving interpretability and its interaction with fuzzy rule set reduction," *IEEE Trans. on Fuzzy Systems* 13 (1) 13-29, 2005.

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

[33] X. Llora and D. E. Goldberg, "Bounding the effect of noise in multiobjective learning classifier systems," *Evolutionary Computation* 11 (3) 278-297, 2003.

[34] 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) 135-150, 1997.

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

[36] F. Jimenez, A. F. Gomez-Skarmeta, G. Sanchez, H. Roubos, and R. Babuska, "Accurate, transparent and compact fuzzy models for function approximation and dynamic modeling through multi-objective evolutionary optimization," *Proc. of EMO 2001*, 653-667.

[37] 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) 59-88, 2004.

[38] 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* 35 (2) 143-155, 2005.

[39] 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* 149 (1) 149-186, 2005.

[40] 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) 4-31, 2007.
[41] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Interpretability Issues in Fuzzy Modeling*, Springer, 2003.
[42] J. Casillas, O. Cordon, F. Herrera, and L. Magdalena (eds.), *Accuracy Improvements in Linguistic Fuzzy Modeling*, Springer, 2003.
[43] J. D. Schaffer, "Multiple objective optimization with vector evaluated genetic algorithms," *Proc. of ICGA 1985*, 93-100.
[44] C. A. Coello Coello, D. A. van Veldhuizen, and G. B. Lamont, *Evolutionary Algorithms for Solving Multi-Objective Problems*, Kluwer, 2002.
[45] C. A. Coello Coello and G. B. Lamont, *Applications of Multi-Objective Evolutionary Algorithms*, World Scientific, 2004.
[46] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. on Evolutionary Computation* 6 (2) 182-197, 2002.
[47] H. Ishibuchi, K. Nozaki, and H. Tanaka, "Distributed representation of fuzzy rules and its application to pattern classification," *Fuzzy Sets and Systems* 52 (1) 21-32, 1992.
[48] H. Ishibuchi and T. Nakashima, "Effect of rule weights in fuzzy rule-based classification systems," *IEEE Trans. on Fuzzy Systems* 9 (4) 506-515, 2001.
[49] H. Ishibuchi and T. Yamamoto, "Rule weight specification in fuzzy rule-based classification systems," *IEEE Trans. on Fuzzy Systems* 13 (4) 428-435, 2005.
[50] 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) 21-45, 1999.
[51] H. Ishibuchi, T. Nakashima, and T. Morisawa, "Voting in fuzzy rule-based systems for pattern classification problems," *Fuzzy Sets and Systems* 103 (2) 223-238, 1999.
[52] L. I. Kuncheva, *Fuzzy Classifier Design*, Physica-Verlag, 2000.
[53] H. Ishibuchi, T. Nakashima, and M. Nii, *Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining*, Springer, 2004.
[54] T. P. Hong, C. S. Kuo, and S. C. Chi, "Trade-off between computation time and number of rules for fuzzy mining from quantitative data," *International Journal of Uncertainty Fuzziness and Knowledge-based Systems* 9 (5) 587-604, 2001.
[55] H. Ishibuchi, T. Yamamoto, and T. Nakashima, "Fuzzy data mining: Effect of fuzzy discretization," *Proc. of ICDM 2001*, 241-248.
[56] 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, 307-328, 1996.
[57] B. Liu, W. Hsu, and Y. Ma, "Integrating classification and association rule mining," *Proc. of KDD 1998*, 80-86.
[58] H. Ishibuchi and T. Yamamoto, "Comparison of heuristic criteria for fuzzy rule selection in classification problems," *Fuzzy Optimization and Decision Making* 3 (2) 119-139, 2004.
[59] F. Coenen, P. Leng, and L. Zhang, "Threshold tuning for improved classification association rule mining," *Proc. of PAKDD 2005*, 216-225.
[60] F. Coenen and P. Leng, "Obtaining best parameter values for accurate classification," *Proc. of ICDM 2005*, 549-552.
[61] B. de la Iglesia, M. S. Philpott, A. J. Bagnall, and V. J. Rayward-Smith, "Data mining rules using multi-objective evolutionary algorithms," *Proc. of CEC 2003*, 1552-1559.
[62] B. de la Iglesia, A. Reynolds, and V. J. Rayward-Smith, "Developments on a multi-objective metaheuristic (MOMH) algorithm for finding interesting sets of classification rules," *Proc. of EMO 2005*, 826-840.
[63] A. Reynolds and B. de la Iglesia, "Rule induction using multi-objective metaheuristics: Encouraging rule diversity," *Proc. of IJCNN 2006*, 6375-6382.
[64] H. Ishibuchi, I. Kuwajima, and Y. Nojima, "Relation between Pareto-optimal fuzzy rules and Pareto-optimal fuzzy rule set," *Proc. of IEEE MCDM 2007* (in press).
[65] H. Ishibuchi and T. Yamamoto, "Effect of fuzzy discretization in fuzzy rule-based systems for classification problems with continuous attributes," *Archives of Control Sciences* 12 (4) 351-378, 2002.
[66] K. Nozaki, H. Ishibuchi, and H. Tanaka, "Adaptive fuzzy rule-based classification systems," *IEEE Trans. on Fuzzy Systems* 4 (3) 238-250, 1996.
[67] S. Guillaume, "Designing fuzzy inference systems from data: An interpretability-oriented review," *IEEE Trans. on Fuzzy Systems* 9 (3) 426-443, 2001.
[68] R. Mikut, J. Jakel, and L. Groll, "Interpretability issues in data-based learning of fuzzy systems," *Fuzzy Sets and Systems* 150 (2) 179-197, 2005.
[69] J. Handl and J. Knowles, "Evolutionary multiobjective clustering," *Proc. of PPSN 2004*, 1081-1091.
[70] A. Ghosh and B. T. Nath, "Multi-objective rule mining using genetic algorithms," *Information Sciences* 163 (1-3) 123-133, 2004.
[71] M. Kaya, "Multi-objective genetic algorithm based approaches for mining optimized fuzzy association rules," *Soft Computing* 10 (7) 578-586, 2006.
[72] A. A. Freitas, *Data Mining and Knowledge Discovery with Evolutionary Algorithms*, Springer, 2002.
[73] J. R. Cano, F. Herrera, and M. Lozano, "Stratification for scaling up evolutionary prototype selection," *Pattern Recognition Letters* 26 (7) 953-963, 2005.
[74] Y. Jin (ed.), *Multi-Objective Machine Learning*, Springer, Berlin, 2006.
[75] A. Ghosh, S. Dehuri and S. Ghosh (Eds.), *Multi-Objective Evolutionary Algorithms for Knowledge Discovery from Data Bases*, Springer, Berlin (to appear).
[76] A. Jaszkiewicz, "On the performance of multiple-objective genetic local search on the 0/1 knapsack problem - A comparative experiment," *IEEE Trans. on Evolutionary Computation* 6 (4) 402-412, 2002.
[77] A. Jaszkiewicz, "On the computational efficiency of multiple objective metaheuristics: The knapsack problem case study," *European Journal of Operational Research* 158 (2) 418-433, 2004.
[78] K. Deb and J. Sundar, "Reference point based multi-objective optimization using evolutionary algorithms," *Proc. of GECCO 2006*, 635-642.
[79] H. Ishibuchi and Y. Nojima, "Optimization of scalarizing functions through evolutionary multiobjective optimization," *Proc. of EMO 2007* (in press).