Genetic Rule Selection as a Postprocessing Procedure in Fuzzy Data Mining

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

***Abstract*— We examine the effect of genetic rule selection as a postprocessing procedure in fuzzy data mining. Usually a large number of fuzzy rules are extracted in a heuristic manner from numerical data using a rule evaluation criterion in fuzzy data mining. It is, however, very difficult for human users to understand thousands of fuzzy rules. Thus it is necessary to decrease the number of extracted fuzzy rules when our task is to present understandable knowledge to human users. In this paper, we use genetic rule selection to decrease the number of extracted fuzzy rules. Through computational experiments, we examine the effect of genetic rule selection. First we extract fuzzy rules that satisfy minimum support and confidence levels. Thousands of fuzzy rules are extracted from numerical data in a heuristic manner. Then we apply genetic rule selection to extracted fuzzy rules. Experimental results show that genetic rule selection significantly decreases the number of extracted fuzzy rules without degrading their classification accuracy.**

I. Introduction

Fuzzy rule-based systems have been successfully applied to various application areas such as control, modeling and classification. In the design of fuzzy rule-based systems, not only their accuracy but also their interpretability has been taken into account [1]-[7]. Multiobjective optimization has been used in some studies [8]-[15] to find a number of non-dominated fuzzy rule-based systems along the accuracy-interpretability tradeoff surface. In the field of data mining, emphasis has been usually placed on the interpretability of extracted rules rather than their accuracy. In this sense, fuzzy logic has a high potential ability to play an important role in data mining. Fuzzy rules have been used in some studies [16]-[20] on data mining under the name of fuzzy data mining or linguistic data mining.

Usually a large number of fuzzy rules are extracted in a heuristic manner using rule evaluation criteria (e.g., see [16]). It is, however, very difficult for human users to understand thousands of fuzzy rules. Thus the number of extracted fuzzy rules should be significantly decreased when our task is to present understandable knowledge to human users.

In this paper, we use genetic rule selection to decrease the number of extracted fuzzy rules as a postprocessing procedure in fuzzy data mining. Genetic rule selection was first formulated as an optimization problem for the design of fuzzy rule-based classification systems in Ishibuchi et al. [1], [2] where the following weighted sum fitness function was maximized:

$$f(S) = w_1 \cdot f_1(S) - w_2 \cdot f_2(S), \qquad (1)$$

In this formulation, S is a subset of candidate fuzzy rules, $f_1(S)$ is the number of correctly classified training patterns by S, $f_2(S)$ is the number of fuzzy rules in S, and w_1 and w_2 are positive weights. A large number of extracted fuzzy rules in fuzzy data mining are used as candidate fuzzy rules in genetic rule selection in this paper.

A single-objective genetic algorithm was used in [1], [2] to find the optimal rule set of the rule selection problem in (1). The single-objective formulation in (1) was extended to the case of two-objective rule selection [8] where a two-objective genetic algorithm was used to find a number of non-dominated rule sets with respect to $f_1(S)$ and $f_2(S)$. The two-objective formulation in [8] was further extended to three-objective rule selection [9] where the minimization of the total number of antecedent conditions (i.e., the total rule length) was introduced as the third objective function $f_3(S)$.

In this paper, we examine the effect of genetic rule selection as a postprocessing procedure in fuzzy data mining for pattern classification problems. It is shown that genetic rule selection significantly decreases the number of extracted fuzzy rules without severely degrading their classification accuracy. First we explain fuzzy rule-based classification and fuzzy rule extraction in Section II. Next we explain genetic rule selection in Section III. Then we demonstrate the effect of genetic rule selection through computational experiments on some benchmark data sets from the UC Irvine machine learning repository in Section IV. Finally Section V concludes this paper.

II. Fuzzy Rule-Based Classification

A. Pattern 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].

B. Fuzzy Rules for Pattern Classification Problems

For our n-dimensional pattern classification problem, we use fuzzy rules of the following form [21]:

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, Sakai, Osaka 599-8531, Japan (hisaoi@cs.osakafu-u.ac.jp, nojima@cs.osakafu-u.ac.jp, kuwajima@ci.cs.osakafu-u.ac.jp).

0-7803-9719-3/06/$20.00 ©2006 IEEE

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

where R_q is the label of the q-th fuzzy rule, $\mathbf{x} = (x_1, ..., x_n)$ is an n-dimensional pattern vector, A_{qi} is an antecedent fuzzy set, 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 (2) 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 classification systems as shown in [22], [23]. For other types of fuzzy rules for pattern classification problems, see [20], [24], [25].

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 as shown in Fig. 1. In addition to the 14 fuzzy sets in Fig. 1, 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 antecedent fuzzy sets for each attribute in our computational experiments.

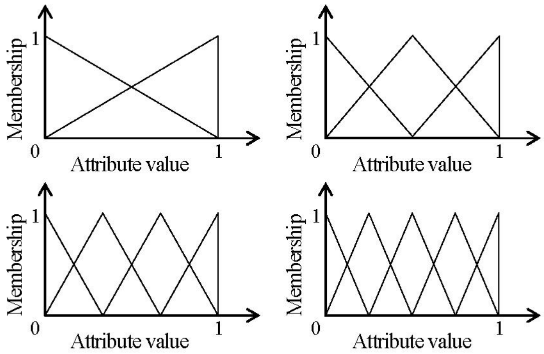


Fig. 1. Four fuzzy partitions used in our computational experiments.

C. Fuzzy Rule Generation

Since we 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 in the antecedent part of the fuzzy rule in (2). 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

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

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 [16], [20]:

$$c(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{\sum_{p=1}^{m} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}. \quad (4)$$

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

$$c(\mathbf{A}_q \Rightarrow \text{Class } C_q) = \max_{h=1,2,...,M} \{c(\mathbf{A}_q \Rightarrow \text{Class } h)\}. \quad (5)$$

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 [21].

Different specifications of the rule weight CF_q have been proposed and examined in the literature. We use the following specification because good results were reported by this specification in the literature [20], [23]:

$$CF_q = c(\mathbf{A}_q \Rightarrow \text{Class } C_q) - \sum_{\substack{h=1 \\ h \neq C_q}}^{M} c(\mathbf{A}_q \Rightarrow \text{Class } h). \quad (6)$$

D. Rule Extraction 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 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 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 extract fuzzy rules that satisfy both minimum confidence and support levels. In the field of data mining, these two rule extraction criteria have been often used [26]-[28]. In the same manner as the fuzzy version of confidence in (4), the support of the fuzzy rule $\mathbf{A}_q \Rightarrow \text{Class } h$ is calculated as follows [16], [20]:

$$s(\mathbf{A}_q \Rightarrow \text{Class } h) = \frac{\sum_{\mathbf{x}_p \in \text{Class } h} \mu_{\mathbf{A}_q}(\mathbf{x}_p)}{m}. \quad (7)$$

E. Fuzzy Rule-based Classification Systems

Let S be a set of fuzzy rules of the form in (2). That is, S is a fuzzy rule-based classification system. 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:

$$\mu_{\mathbf{A}_w}(\mathbf{x}_p) \cdot CF_w = \max\{\mu_{\mathbf{A}_q}(\mathbf{x}_p) \cdot CF_q \mid R_q \in S\}. \quad (8)$$

III. Genetic Rule Selection

Let us assume that N fuzzy rules have already been extracted from numerical data using the heuristic rule extraction procedure in Section II. We denote a subset of the extracted N fuzzy rules by a binary string of length N as

$$S = s_1 s_2 \cdots s_N, \quad (9)$$

where $s_j = 1$ and $s_j = 0$ mean the inclusion of the j-th rule in S and its exclusion from S, respectively. Such a binary string is handled as an individual in genetic rule selection.

The fitness value of S is calculated as follows:

$$f(S) = w_1 \cdot f_1(S) - w_2 \cdot f_2(S) - w_3 \cdot f_3(S), \qquad (10)$$

where $f_1(S)$ is the number of correctly classified training patterns by S, $f_2(S)$ is the number of fuzzy rules in S, $f_3(S)$ is the sum of the rule lengths of all fuzzy rules in S, and w_1, w_2 and w_3 are prespecified (i.e., user-definable) positive weights for $f_1(S)$, $f_2(S)$ and $f_3(S)$, respectively.

We use a single-objective genetic algorithm with the $(\mu + \lambda)$-ES generation update mechanism to find the optimal rule set of our rule selection problem in (10). First an initial population is randomly generated. Then a pair of parent strings are chosen from the current population by binary tournament selection. An offspring string is generated from the selected pair of parent strings by uniform crossover and bit-flip mutation. By iterating selection, crossover and mutation, we generate an offspring population. The next population is constructed by choosing the best strings from the current population and the offspring population.

As in our former studies [9], [12], we used two problem-specific heuristics tricks: biased mutation probabilities (i.e., 0.1 for the mutation from 1 to 0, and 0.001 for that from 0 to 1) and unnecessary rule removal (for details, see [9], [12]).

IV. Computational Experiments

A. Data Sets

We use seven data sets in Table I: Wisconsin breast cancer (Breast W), diabetes (Diabetes), glass identification (Glass), Cleveland heart disease (Heart C), iris (Iris), sonar (Sonar), and wine recognition (Wine). These data sets are available from the UC Irvine machine learning repository. Data sets with missing values are marked by "*" and "**" in the third column of Table I. 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 [29] where six variants of C4.5 [30] were examined. The generalization ability of each variant was evaluated by ten independent runs of the ten-fold cross-validation procedure (i.e., 10×10CV) in [29]. We show the best and worst classification rates on test patterns among the six variants in [29] for each data set in Table I.

TABLE I
Data Sets Used in Our Computational Experiments

Data set	Attributes	Patterns	Classes	C4.5 in [29]	
				Best	Worst
Breast W	9	683*	2	94.9	94.0
Diabetes	8	768**	2	75.0	72.8
Glass	9	214	6	72.7	67.8
Heart C	13	297*	5	53.7	52.1
Iris	4	150	3	94.3	92.5
Sonar	60	208	2	75.4	64.2
Wine	13	178	3	94.4	91.2

* Incomplete patterns with missing values are not included.
** Some suspicious patterns with attribute value "0" are included.

B. Conditions of Computational Experiments

In heuristic extraction of fuzzy rules (i.e., in the fuzzy data mining phase), the upper bound of the rule length is specified as $L_{\max} = 2$ for the sonar data set with 60 attributes and $L_{\max} = 3$ 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).

We examine the 16 combinations of the following four values of the minimum confidence and support levels:

Minimum confidence level: 0.6, 0.7, 0.8, 0.9,
Minimum support level: 0.01, 0.02, 0.05, 0.10.

Fuzzy rules are extracted using each of the 16 combinations of the minimum confidence and support levels.

A standard single-objective genetic algorithm is applied to extracted fuzzy rules as a postprocessing procedure. We use the following parameter values in our genetic algorithm:

Weight vector: $(w_1, w_2, w_3) = (10, 1, 1)$,
Population size: 200,
Stopping condition: 1000 generations,
Crossover probability: 0.9 (Uniform crossover),
Mutation probability: $1/N$,

where N is the number of extracted fuzzy rules.

Average classification rates on training and test patterns are calculated over five independent runs of the two-fold cross-validation procedure (i.e., 5×2CV) before and after genetic rule selection for each data set.

C. Experimental Results

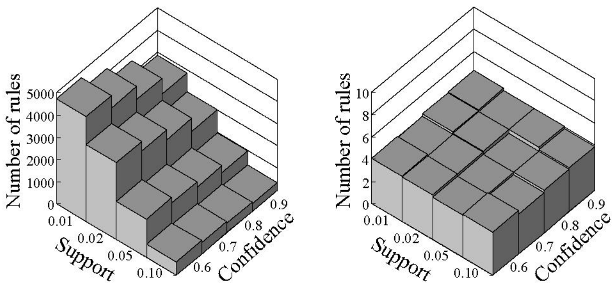
Due to the page limitation, we report only a part of experimental results. More details will be reported at the conference presentation.

In Figs. 2-5, we show average results on the iris data set. The iris data set is one of the most frequently-used data sets in the literature. Fig. 2 (a) shows the average number of extracted fuzzy rules for each combination of the minimum confidence and support levels. Thousands of fuzzy rules are extracted by the heuristic rule extraction procedure. We can see from Fig. 2 (a) that the number of extracted fuzzy rules strongly depends on the values of the minimum confidence and support levels. On the other hand, Fig. 2 (b) shows the average number of selected fuzzy rules after genetic rule selection. Four fuzzy rules are selected from thousands of extracted rules on average in all cases in Fig. 2 (b).

Fig. 3 shows the average classification rates on training patterns. The average classification rates on training patterns are improved by genetic rule selection in Fig. 3 whereas the number of fuzzy rules is significantly decreased in Fig. 2. On the other hand, Fig. 4 shows the average classification rates on test patterns. Almost the same generalization ability is obtained in Fig. 4 before and after genetic rule selection.

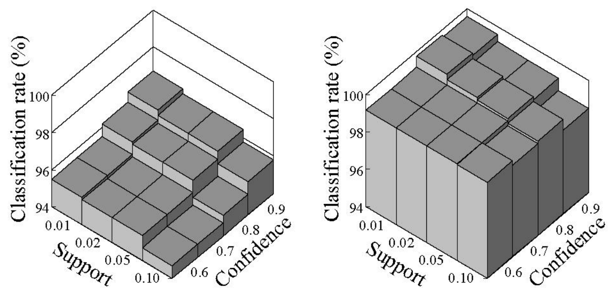
Fig. 5 shows the average rule length of fuzzy rules. From Fig. 5, we can see that short fuzzy rules are selected by genetic rule selection. This is because short fuzzy rules are compatible with many training patterns.

In Figs. 6-8, we show average results on the diabetes data set. The number of fuzzy rules is significantly decreased by genetic rule selection from Fig. 6 (a) to Fig. 6 (b). The average classification rates on training patterns are improved by genetic rule selection in Fig. 7. An interesting observation is that the average classification rates on test patterns are also improved by genetic rule selection in Fig. 8. Whereas we do not show experimental results, the average rule length is decreased by genetic rule selection for the diabetes data as in Fig. 5 for the iris data.



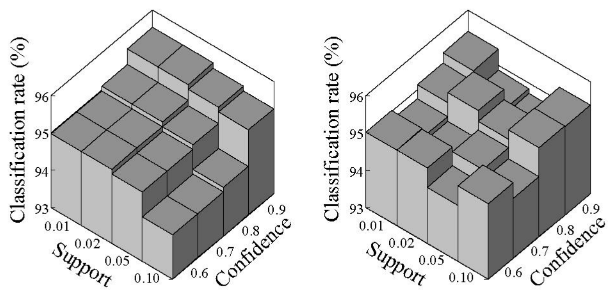
(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 2. The number of fuzzy rules (Iris).



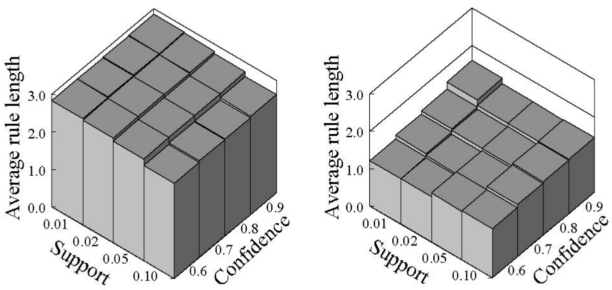
(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 3. Average classification rates on training patterns (Iris).



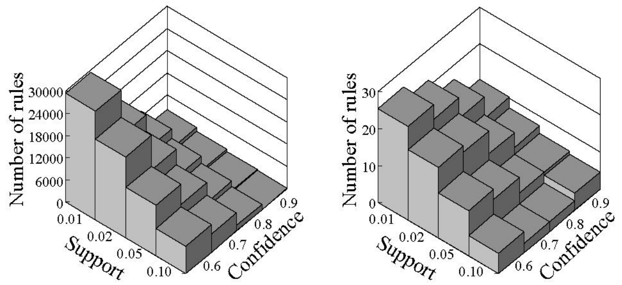
(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 4. Average classification rates on test patterns (Iris).



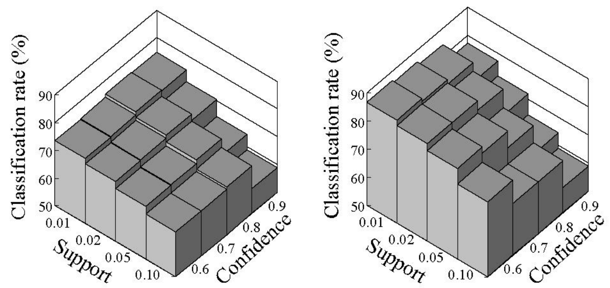
(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 5. Average rule length of fuzzy rules (Iris).



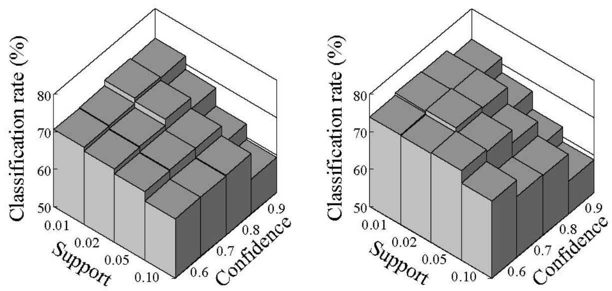
(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 6. The number of fuzzy rules (Diabetes).



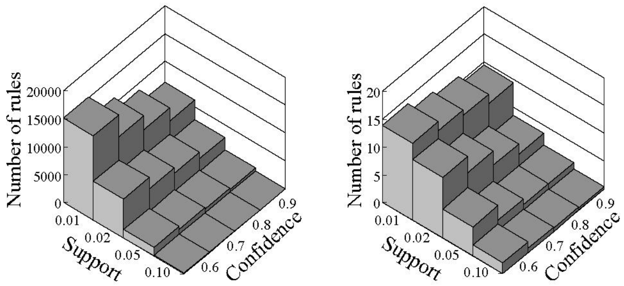
(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 7. Average classification rates on training patterns (Diabetes).



(a) Before genetic rule selection. (b) After genetic rule selection.

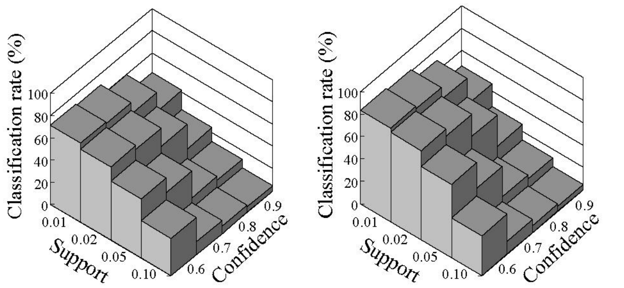
Fig. 8. Average classification rates on test patterns (Diabetes).



(a) Before genetic rule selection. (b) After genetic rule selection.

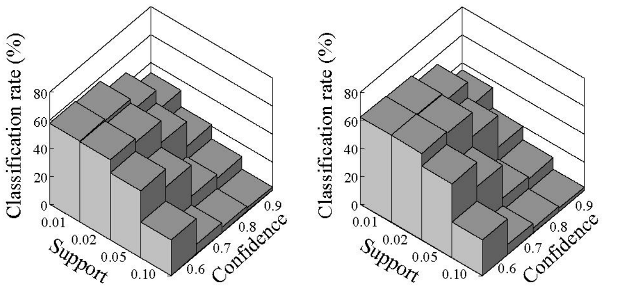
Fig. 9. The number of fuzzy rules (Glass).

In Figs. 9-11, we show average results on the glass data set. In Fig. 9 in the previous page, the number of fuzzy rules is significantly decreased by genetic rule selection. The average classification rates on training and test patterns are improved in many cases in Fig. 10 and Fig. 11.



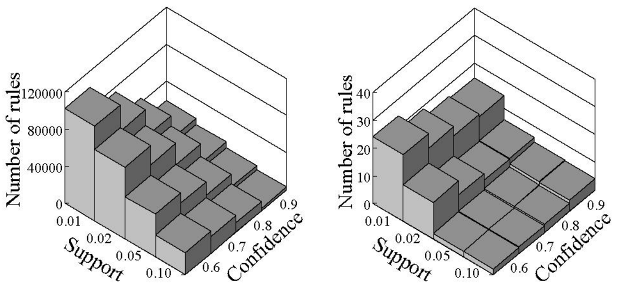
(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 10. Average classification rates on training patterns (Glass).



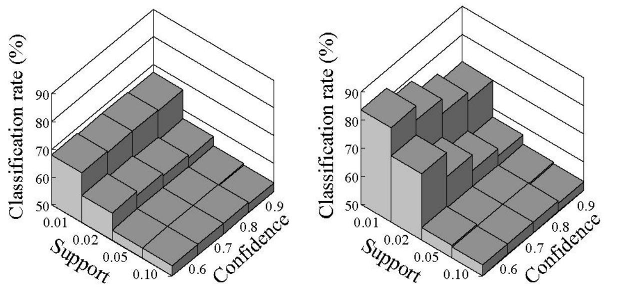
(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 11. Average classification rates on test patterns (Glass).



(a) Before genetic rule selection. (b) After genetic rule selection.

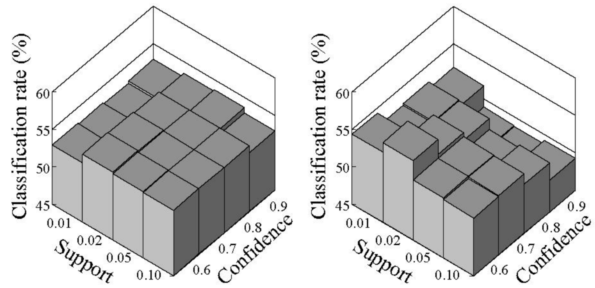
Fig. 12. The number of fuzzy rules (Heart C).



(a) Before genetic rule selection. (b) After genetic rule selection.

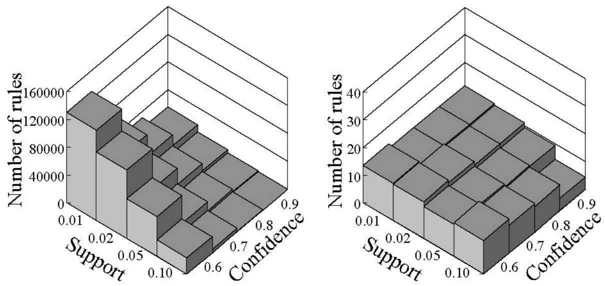
Fig. 13. Average classification rates on training patterns (Heart C).

Average results on the Cleveland heart disease and sonar data sets are shown in Figs. 12-17. We can observe almost the same effect of genetic rule selection in these figures as in the above-mentioned results in Figs. 2-11. Its effect on the generalization ability seems to be problem-dependent.



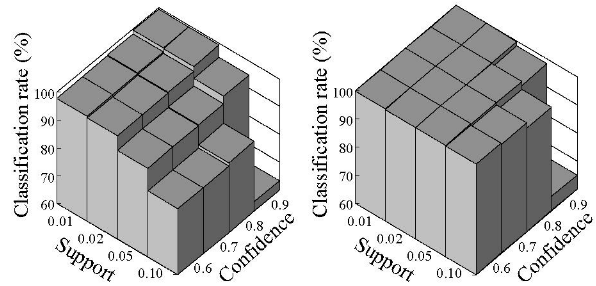
(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 14. Average classification rates on test patterns (Heart C).



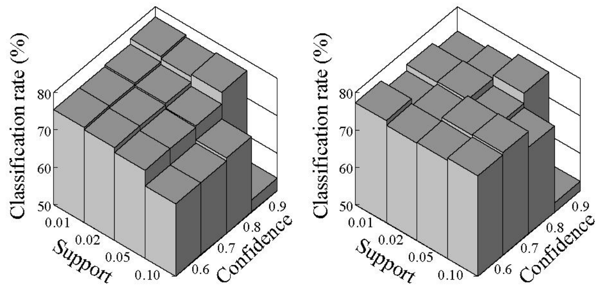
(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 15. The number of fuzzy rules (Sonar).



(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 16. Average classification rates on training patterns (Sonar).



(a) Before genetic rule selection. (b) After genetic rule selection.

Fig. 17. Average classification rates on test patterns (Sonar).

V. CONCLUSIONS

We showed that the number of heuristically extracted fuzzy rules in fuzzy data mining was significantly decreased by genetic rule selection without severely degrading their classification accuracy through computational experiments on some benchmark data sets. That is, the understandability of extracted knowledge was significantly improved. This observation clearly shows potential usefulness of genetic rule selection in fuzzy data mining as a postprocessing procedure. Genetic rule selections is a general scheme, which is applicable to both fuzzy and non-fuzzy rules generated in various manners (e.g., [16], [26]-[28], [31]).

Classification rates on training patterns were improved in almost all cases in our computational experiments by genetic rule selection whereas its effect on the generalization ability was problem-dependent. The main drawback of genetic rule selection is its large computation load especially when it is applied to large data sets. One promising trick is to divide a data set into multiple subsets of small size. A different subset is assigned to each individual for its fitness evaluation.

REFERENCES

[1] 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.

[2] 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.

[3] 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*, vol. 29, no. 6, pp. 829-845, December 1999.

[4] Y. Jin, "Fuzzy modeling of high-dimensional systems: Complexity reduction and interpretability improvement," *IEEE Trans. on Fuzzy Systems*, vol. 8, no. 2, pp. 212-221, April 2000.

[5] M. Setnes and H. Roubos, "GA-based modeling and classification: Complexity and performance," *IEEE Trans. on Fuzzy Systems*, vol. 8, no. 5, pp. 509-522, October 2000.

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

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

[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, July 1997.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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).

[16] 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*, vol. 9, no. 5, pp. 587-604, October 2001.

[17] H. Ishibuchi, T. Yamamoto, and T. Nakashima, "Fuzzy data mining: Effect of fuzzy discretization," *Proc. of 2001 IEEE International Conference on Data Mining*, pp. 241-248, San Jose, USA, November 29 - December 2, 2001.

[18] G. Q. Chen and Q. Wei, "Fuzzy association rules and the extended mining algorithms," *Information Sciences*, vol. 147, no. 1-4, pp. 201-228, November 2002.

[19] Y. C. Hu, R. S. Chen, and G. H. Tzeng, "Mining fuzzy association rules for classification problems," *Computers & Industrial Engineering*, vol. 43, no. 4, pp. 735-750, September 2002.

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

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

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

[27] 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.

[28] F. Coenen and P. Leng, "Obtaining best parameter values for accurate classification," *Proc. of the 5th IEEE International Conference on Data Mining*, pp. 549-552, Houston, Texas, November 27-30, 2005.

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

[30] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA, 1993.

[31] L. X. Wang and J. M. Mendel, "Generating fuzzy rules by learning from examples," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 22, no. 6, pp. 1414-1427, December 1992.