Multiobjective Association Rule Mining

Hisao Ishibuchi, Isao Kuwajima, and Yusuke Nojima

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, kuwajima@ci.cs.osakafu-u.ac.jp, nojima@cs.osakafu-u.ac.jp
http://www.ie.osakafu-u.ac.jp/~hisaoi/ci_lab_e

Abstract. This paper discusses the application of evolutionary multiobjective optimization (EMO) to association rule mining. We focus our attention especially on classification rule mining where the consequent part of each rule is a class label. First we briefly explain evolutionary multiobjective classification rule mining techniques. Those techniques are roughly categorized into two approaches. In one approach, each classification rule is handled as an individual. An EMO algorithm is used to search for Pareto-optimal rules with respect to some rule evaluation criteria such as support and confidence. In the other approach, each rule set is handled as an individual. An EMO algorithm is used to search for Pareto-optimal rule sets with respect to some rule set evaluation criteria such as accuracy and complexity. Next we explain evolutionary multiobjective rule selection as a post-processing procedure in classification rule mining. Pareto-optimal rule sets are found from a large number of candidate classification rules, which are extracted from a database using an association rule mining technique. Finally we report experimental results where the effect of evolutionary multiobjective rule selection is examined. We also examine the relation between Pareto-optimal rules and Pareto-optimal rule sets.

1 Introduction

Data mining is a very active and rapidly growing research area in the field of computer science. The task of data mining is to extract useful knowledge for human users from a database. Whereas the application of evolutionary computation to data mining is not always easy due to its heavy computation load especially in the case of a large database [3], [4], [9], many evolutionary approaches have been proposed in the literature [6], [16], [31], [33], [36]. Evolutionary multiobjective optimization (EMO) has also been applied to data mining in some studies [10]-[12], [17], [21], [22], [35]. In the field of fuzzy logic, multiobjective formulations have frequently been used for knowledge extraction [5], [18]-[20], [24], [26], [38], [39]. This is because the interpretability-accuracy tradeoff analysis is a very important research issue in the design of fuzzy rule-based systems [5]. Multiobjective formulations have also been used in non-fuzzy genetics-based machine learning [25], [27], [30].

Association rule mining [1] is one of the most well-known data mining techniques. In its basic form [1], all association rules satisfying the minimum support and confidence are efficiently extracted from a database. The application of association rule mining to classification problems is often referred to as classification rule mining or associative classification [28], [29], [32], [37]. Classification rule mining usually consists of two phases: rule discovery and rule selection. In the rule discovery phase, a large number of classification rules are extracted from a database using an association rule mining technique. All classification rules satisfying the minimum support and confidence are usually extracted from a database. A part of extracted classification rules are selected to design a classifier in the rule selection phase using a heuristic rule sorting criterion. The accuracy of the designed classifier usually depends on the specification of the minimum support and confidence. Their tuning was discussed for classification data mining in [7], [8].

Whereas the basic form of association rule mining is to extract all association rules that satisfy the minimum support and confidence [1], other rule evaluation measures have been proposed to qualify the *interestingness* or *goodness* of an association rule. Among them are gain, variance, chi-squared value, entropy gain, gini, laplace, lift, and conviction [2]. It is shown in [2] that the best rule according to any of the abovementioned measures is a Pareto-optimal rule with respect to support and confidence. Motivated by this study, the use of an EMO algorithm was proposed to search for Pareto-optimal classification rules with respect to support and confidence for partial classification [10]-[12], [35]. Similar formulations were used to search for Pareto-optimal association rules [17] and Pareto-optimal fuzzy association rules [26]. EMO algorithms were also used to search for Pareto-optimal rule sets in classification rule mining [21], [22] where the accuracy of rule sets was maximized and their complexity was minimized. The same idea was also used in the multiobjective design of fuzzy rule-based classifiers [18], [19], [24].

In this paper, we empirically examine the effect of evolutionary multiobjective rule selection through computational experiments for some well-known benchmark data sets from the UCI machine learning repository. We also examine the relation between Pareto-optimal rules and Pareto-optimal rule sets in the classifier design. This examination is performed by depicting selected rules in Pareto-optimal rule sets together with candidate classification rules in the support-confidence plain. Our interest is to check whether selected rules in Pareto-optimal rule sets are close to the Pareto front with respect to support and confidence.

This paper is organized as follows. First we briefly explain some basic concepts in classification rule mining in Section 2. Next we briefly explain two approaches in evolutionary multiobjective classification rule mining in Section 3. One approach handles each classification rule as an individual to search for Pareto-optimal rules. In the other approach, each rule set is handled as an individual. An EMO algorithm is used to search for Pareto-optimal rule sets. Then we explain evolutionary multiobjective rule selection as a post-processing procedure in the rule selection phase of classification rule mining in Section 4. Pareto-optimal rule sets are found from a large number of candidate classification rules, which are extracted from a database using an association rule mining technique in the rule discovery phase. Finally we report experimental results on some well-known benchmark data sets. Experimental results demonstrate the effect of evolutionary multiobjective rule selection. The relation between Pareto-optimal rules and Pareto-optimal rule sets is also demonstrated.

2 Classification Rule Mining

Let us assume that we have m training patterns $\mathbf{x}_p = (x_{p1}, x_{p2}, ..., x_{pn})$, $p = 1, 2, ..., m$ from M classes in the n-dimensional continuous pattern space where x_{pi} is the attribute value of the p-th training pattern for the i-th attribute. We denote the set of these m training patterns by D. For our pattern classification problem, we use classification rules of the following type:

$$\text{Rule } R_q : \text{If } x_1 \text{ is } A_{q1} \text{ and ... and } x_n \text{ is } A_{qn} \text{ then Class } C_q \text{ with } CF_q , \tag{1}$$

where R_q is the label of the q-th rule, $\mathbf{x} = (x_1, x_2, ..., x_n)$ is an n-dimensional pattern vector, A_{qi} is an antecedent interval for the i-th attribute, C_q is a class label, and CF_q is a rule weight (i.e., certainty grade). We denote the classification rule R_q in (1) as "$\mathbf{A}_q \Rightarrow C_q$" where $\mathbf{A}_q = (A_{q1}, A_{q2}, ..., A_{qn})$. Each antecedent condition "x_i is A_{qi}" in (1) means the inclusion relation "$x_i \in A_{qi}$". It should be noted that classification rules of the form in (1) do not always have n antecedent conditions. Some rules may have only a few conditions while others may have many conditions.

In the field of association rule mining, two rule evaluation measures called *support* and *confidence* have often been used [1], [2]. Let us denote the support count of the classification rule $\mathbf{A}_q \Rightarrow C_q$ by $SUP(\mathbf{A}_q \Rightarrow C_q)$, which is the number of patterns compatible with both the antecedent part $\mathbf{A}_q$ and the consequent class C_q. $SUP(\mathbf{A}_q)$ and $SUP(C_q)$ are also defined in the same manner, which are the number of patterns compatible with $\mathbf{A}_q$ and C_q, respectively. The support of the classification rule $\mathbf{A}_q \Rightarrow C_q$ is defined as

$$Support(\mathbf{A}_q \Rightarrow C_q) = \frac{SUP(\mathbf{A}_q \Rightarrow C_q)}{|D|}, \tag{2}$$

where $|D|$ is the cardinality of the data set D (i.e., $|D| = m$). On the other hand, the confidence of $\mathbf{A}_q \Rightarrow C_q$ is defined as

$$Confidence(\mathbf{A}_q \Rightarrow C_q) = \frac{SUP(\mathbf{A}_q \Rightarrow C_q)}{SUP(\mathbf{A}_q)}. \tag{3}$$

In partial classification [10]-[12], [35], the coverage is often used instead of the support:

$$Coverage(\mathbf{A}_q \Rightarrow C_q) = \frac{SUP(\mathbf{A}_q \Rightarrow C_q)}{SUP(C_q)}. \tag{4}$$

Since the consequent class is fixed in partial classification (i.e., since the denominator of (4) is constant), the maximization of the coverage is the same as that of the support.

In classification rule mining [28], [29], [32], [37], an association rule mining technique such as Apriori [1] is used in the rule discovery phase to efficiently extract all classification rules that satisfy the minimum support and confidence. These two parameters are prespecified by users. Then a part of extracted classification rules are selected to design a classifier in the rule selection phase.

Let S be a set of selected classification rules. That is, S is a classifier. When a new pattern $\mathbf{x}_p$ is to be classified by S, we choose a single winner rule with the maximum rule weight among compatible rules with $\mathbf{x}_p$ in S. The consequent class of the winner rule is assigned to $\mathbf{x}_p$. When multiple compatible rules with different consequent classes have the same maximum rule weight, the classification of $\mathbf{x}_p$ is rejected in evolutionary multiobjective rule selection in this paper. Only when the accuracy of the finally obtained rule set is to be evaluated, we use a random tiebreak among those classes with the same maximum rule weight in computational experiments.

3 Evolutionary Multiobjective Classification Rule Mining

Evolutionary multiobjective techniques in classification rule mining can be roughly categorized into two approaches. In one approach, each rule is evaluated according to multiple rule evaluation criteria such as support and confidence. An EMO algorithm is used to search for Pareto-optimal classification rules. In the other approach, each rule set is evaluated according to multiple rule set evaluation criteria such as accuracy and complexity. An EMO algorithm is used to search for Pareto-optimal rule sets. In this section, we briefly explain these two approaches.

3.1 Techniques to Search for Pareto-Optimal Classification Rules

It is shown in [2] that the set of Pareto-optimal rules with respect to support and confidence includes the best rule according to any of the following rule evaluation criteria: gain, variance, chi-squared value, entropy gain, gini, laplace, lift, and conviction. Thus it is an important research issue to search for Pareto-optimal rules with respect to support and confidence in association rule mining. The use of NSGA-II [13], [14] for this task was proposed by de la Iglesia et al. [10], [12] where they applied NSGA-II to the following two-objective optimization problem for partial classification.

$$\text{Maximize } \{ \mathit{Coverage}(R) \,,\; \mathit{Confidence}(R) \}, \tag{5}$$

where R denotes a classification rule. It should be noted that the maximization of the coverage means that of the support since the consequent class is fixed in partial classification. The use of a dissimilarity measure between classification rules instead of the crowding distance in NSGA-II was examined in [11] in order to search for a set of Pareto-optimal classification rules with a large diversity. The Pareto-dominance relation in NSGA-II was modified in [35] in order to search for not only Pareto-optimal classification rules but also dominated (but interesting) classification rules.

Ghosh and Nath [17] used an EMO algorithm to search for Pareto-optimal association rules with respect to *confidence*, *comprehensibility* and *interestingness*. That is, association rule mining was formulated as a three-objective optimization problem in [17]. A similar three-objective optimization problem was formulated in Kaya [26] where an EMO algorithm was used to search for Pareto-optimal fuzzy association rules with respect to *support*, *confidence* and *comprehensibility*.

3.2 Techniques to Search for Pareto-Optimal Rule Sets

In classification rule mining [28], [29], [32], [37], first an association mining technique such as Apriori [1] is used in the rule discovery phase to efficiently extract all classification rules that satisfy the minimum support and confidence. Then a part of extracted classification rules are selected using a heuristic rule sorting criterion in the rule selection phase to design a classifier. Evolutionary multiobjective rule selection was proposed in [21], [22] to search for Pareto-optimal rule sets with respect to accuracy and complexity in the rule selection phase of classification rule mining.

Genetic algorithm-based rule selection was first proposed for the design of accurate and comprehensible fuzzy rule-based classifiers in [23] where a weighted sum fitness function was used to maximize the classification accuracy and minimize the number of fuzzy rules. An EMO algorithm was used to search for Pareto-optimal fuzzy rule-based classifiers with respect to these two objectives in [18]. The total number of antecedent conditions was introduced as the third objective in [19] to minimize not only the number of fuzzy rules but also their length while maximizing the classification accuracy of fuzzy rule-based classifies. The use of a memetic EMO algorithm was examined to search for Pareto-optimal fuzzy rule-based classifiers with respect to these three objectives in [24]. Fuzzy rule selection techniques in these studies were used for non-fuzzy classification rule mining in [21], [22].

4 Evolutionary Multiobjective Rule Selection

Let us assume that we have already extracted *N* classification rules in the rule discovery phase of classification rule mining. These *N* classification rules are used as candidate rules in rule selection. Let *S* be a subset of the *N* candidate rules (i.e., *S* is a classifier). We use a binary string of length *N* to represent *S* where "1" and "0" mean the inclusion in *S* and the exclusion from *S* of the corresponding candidate rule.

As in our former studies [21], [22], we use the following three objectives:

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

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

$f_3(S)$: The total number of antecedent conditions over selected rules in *S*.

The first objective is maximized while the second and third objectives are minimized. The third objective can be viewed as the minimization of the total rule length since the number of antecedent conditions of each rule is often referred to as the rule length.

We use NSGA-II [13], [14] to search for Pareto-optimal rule sets (i.e., Pareto-optimal subsets of the *N* candidate rules) with respect to these three objectives. We also use a single-objective genetic algorithm (SOGA) with the $(\mu+\lambda)$-ES generation update mechanism to optimize the weighted sum fitness function of the three objectives for comparison in our computational experiments.

5 Computational Experiments

In this section, we demonstrate how SOGA and NSGA-II can decrease the number of extracted rules and their rule length without severely degrading their classification accuracy through computational experiments on some well-known benchmark data sets with continuous attributes in the UCI machine learning repository. We also examine the relation between Pareto-optimal rules and Pareto-optimal rule sets by depicting selected rules in the support-confidence plain.

5.1 Conditions of Computational Experiments

We used seven data sets in Table 1 (whereas we do not report experimental results on all of these data sets in this paper due to the page limitation). We did not use incomplete patterns with missing values. All attribute values were handled as real numbers. The domain of each attribute was divided into multiple intervals using an optimal splitting method [15] based on the class entropy measure [34]. Since the choice of an appropriate number of intervals is not easy, we simultaneously used four different partitions with two, three, four, and five intervals (i.e., 14 antecedent intervals in total for each attribute). As a result, various candidate classification rules were examined in the rule discovery phase using overlapping antecedent intervals of various widths for each attribute.

Table 1. Data sets used in computational experiments.

Data set	Attributes	Patterns	Classes
Breast W	9	683*	2
Glass	9	214	6
Heart C	13	297*	5
Iris	4	150	3
Pendig	16	10992	10
Shuttle	9	58000	7
Wine	13	178	3

* Incomplete patterns with missing values are not included.

We extracted candidate classification rules with three or less antecedent conditions using prespecified values of the minimum support and confidence. This restriction on the number of antecedent conditions is to find rule sets with high understandability (i.e., because it is very difficult for human users to intuitively understand long classification rules with many antecedent conditions). We examined 4×4 combinations of the following four specifications of each threshold for the seven data sets in Table 1:

Minimum support: 1%, 2%, 5%, 10%,
Minimum confidence: 60%, 70%, 80%, 90%.

All the extracted classification rules for each combination of the two threshold values were used in evolutionary rule selection as candidate rules. NSGA-II was executed with the following parameter values:

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

We also used SOGA to maximize the following weighted sum fitness function:

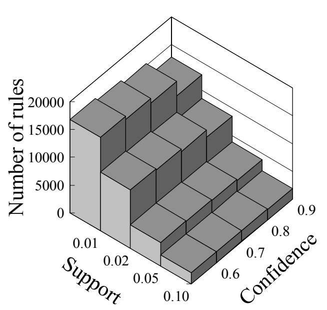
$$\text{Maximize } f(S) = w_1 \cdot f_1(S) - w_2 \cdot f_2(S) - w_3 \cdot f_3(S), \tag{6}$$

where $\mathbf{w} = (w_1, w_2, w_3)$ is a non-negative weight vector, which was specified as $\mathbf{w} = (2, 1, 1)$ in our computational experiments.

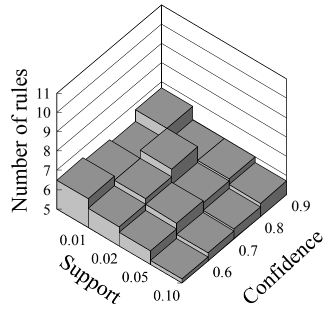
The classification accuracy on test patterns of candidate rules and selected rules was examined by iterating the two-fold cross-validation procedure with 50% training patterns and 50% test patterns five times for each data set. We report average results over its five iterations in the next subsection. In some computational experiments, we show experimental results of only a single run of NSGA-II. In other computational experiments, all the given patterns in each data set were used as training patterns for examining the relation between Pareto-optimal rules and Pareto-optimal rule sets.

5.2 Experimental Results

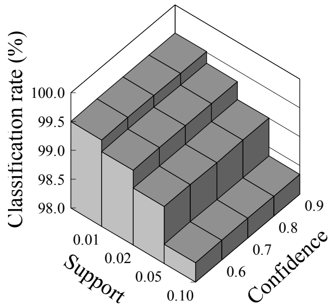
First we show some experimental results by SOGA to clearly demonstrate the effect of genetic rule selection. Experimental results on the Wisconsin breast cancer data set were summarized in Fig. 1. Each plot in the right-hand side was obtained by applying SOGA to candidate classification rules in the corresponding plot in the left-hand side. For example, about six rules were selected by SOGA in Fig. 1 (b) from thousands of candidate rules in Fig. 1 (a). The deterioration in the classification rates on training patterns by genetic rule selection from Fig. 1 (c) to Fig. 1 (d) was less than 1%. When the minimum support was 0.10 (i.e., the right-most row), the classification rates on training patterns were increased by genetic rule selection from Fig. 1 (c) to Fig. 1 (d). The deterioration in the classification rates on test patterns by genetic rule selection from Fig. 1 (e) to Fig. 1 (f) was about 1% - 2%. The average rule length was decreased by genetic rule selection from about 3 in Fig. 1 (g) to less than 2 in Fig. 1 (h). These observations show that only a small number of simple classification rules were selected by SOGA without severely deteriorating the classification accuracy.



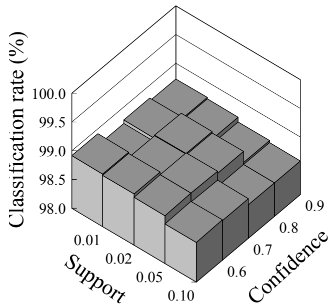
(a) Number of rules before rule selection.



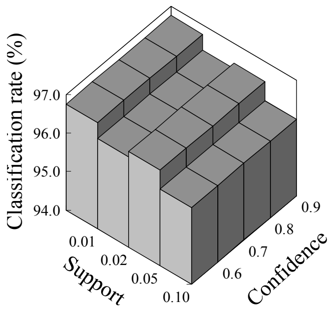
(b) Number of rules after rule selection.



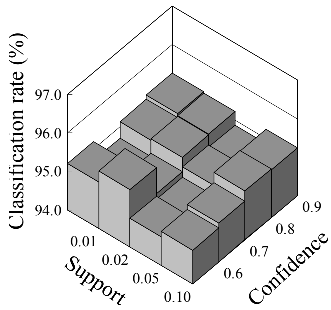
(c) Training data accuracy before rule selection.



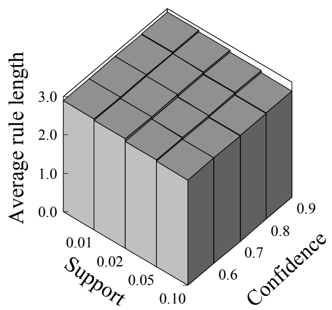
(d) Training data accuracy after rule selection.



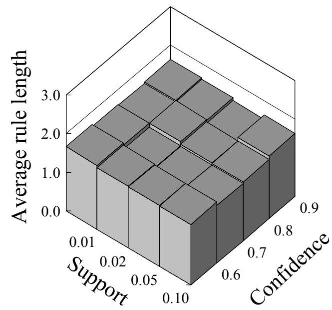
(e) Test data accuracy before rule selection.



(f) Test data accuracy after rule selection.



(g) Average rule length before rule selection.

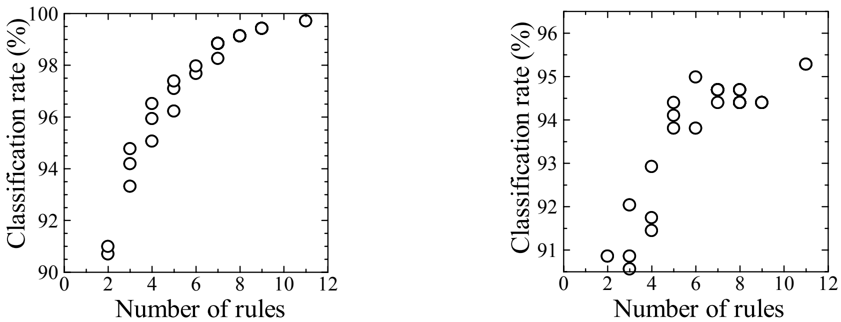


(h) Average rule length after rule selection.

Fig. 1. Experimental results by SOGA on the Wisconsin breast cancer data set.

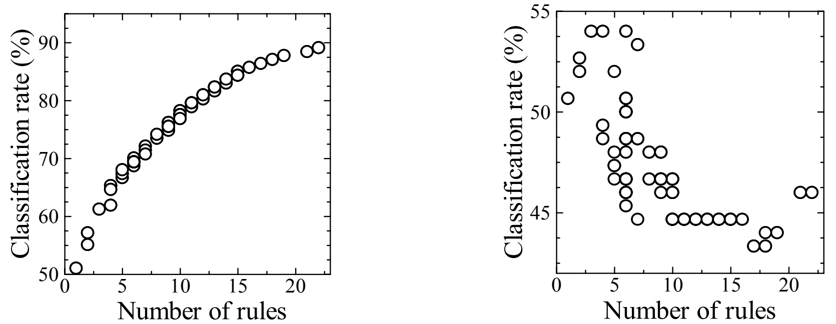
We also applied evolutionary multiobjective rule selection to the Wisconsin breast cancer data set in the following manner. First the given 683 patterns were randomly divided into 342 training patterns and 341 test patterns. Next candidate rules were extracted from the 342 training patterns using the minimum support 0.01 and the minimum confidence 0.6. As a result, 17070 classification rules were extracted. Then NSGA-II was applied to the extracted classification rules. From its single run, 15 non-dominated rule sets were obtained. Finally each of the obtained rule sets was evaluated for the training and test patterns. The classification rates of obtained rule sets are shown in Fig. 2 (a) for the training patterns and Fig. 2 (b) for the test patterns. Some of the obtained rule sets (i.e., a rule set with only a single rule) are not shown because their classification rates are out of the range of the vertical axis of each plot in Fig. 2. We can observe a clear tradeoff relation between the number of selected rules and the classification rates on the training patterns in Fig. 2 (a). A similar tradeoff relation is also observed for the test patterns in Fig. 2 (b).

While we observed very similar tradeoff relations between the accuracy on training patterns and the number of selected rules for all the seven data sets, we obtained totally different results on test patterns. For example, experimental results on the Cleveland heart disease data set are shown in Fig. 3 where we observe a clear degrade in the accuracy on test patterns due to the increase in the number of selected rules.



(a) Classification rates on training patterns. (b) Classification rates on test patterns.

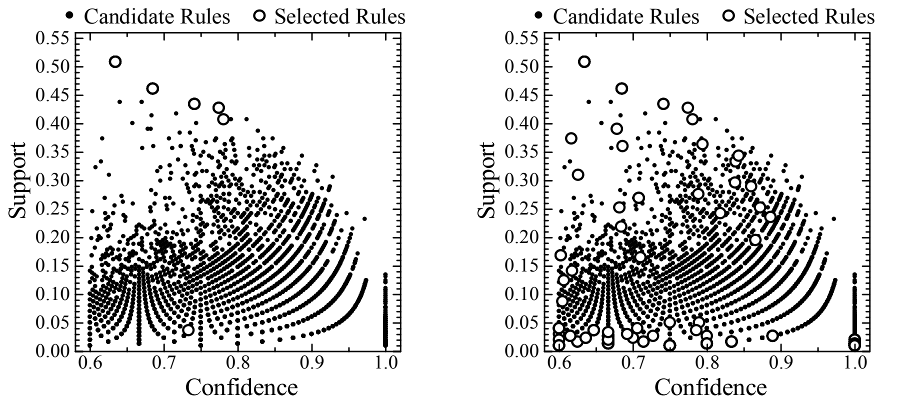
Fig. 2. Experimental results by NSGA-II on the Wisconsin breast cancer data set.



(a) Classification rates on training patterns. (b) Classification rates on test patterns.

Fig. 3. Experimental results by NSGA-II on the Cleveland heart disease data set.

Finally we examined the relation between Pareto-optimal rules and Pareto-optimal rule sets for the Cleveland heart disease data set. First we extracted 12206 classification rules from all the 297 patterns using the minimum support 0.01 and the minimum confidence 0.6. Next we applied NSGA-II to the extracted classification rules. Then we chose two rule sets from the obtained non-dominated rule sets. One is the most complicated rule set with the highest accuracy on the training patterns. The other is the simplest rule set among those rule sets with only two rules. This computational experiment was iterated ten times. Candidate classification rules and selected rules in each rule set are shown in Fig. 4. We can see that Pareto-optimal rule sets do not necessarily consist of Pareto-optimal rules with respect to support and confidence.



(a) Rules in the simplest rule sets. (b) Rules in the most complicated rule sets.

Fig. 4. Locations of selected rules in the support-confidence plain (results of ten runs).

6 Conclusions

In this paper, first we briefly explained two approaches in evolutionary multiobjective classification rule mining. One is to search for Pareto-optimal rules and the other is to search for Pareto-optimal rule sets. Next we demonstrated the effect of GA-based rule selection as a post-processing procedure in the second phase of classification rule mining. Then we showed the accuracy-complexity tradeoff of non-dominated rule sets obtained by evolutionary multiobjective rule selection. Finally we examined the relation between Pareto-optimal rules and Pareto-optimal rule sets.

This work was partially supported by Grant-in-Aid for Scientific Research on Priority Areas: KAKENHI (18049065).

References

1. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., Verkamo, A. I.: Fast Discovery of Association Rules. In Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R.

(eds.) Advances in Knowledge Discovery and Data Mining. AAAI Press, Menlo Park (1996) 307-328
2. Bayardo Jr., R. J., Agrawal, R.: Mining the Most Interesting Rules. Proc. of 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (1999) 145-153
3. Cano, J. R., Herrera, F., Lozano, M.: Stratification for Scaling up Evolutionary Prototype Selection. Pattern Recognition Letters 26 (2005) 953-963
4. Cano, J. R., Herrera, F., Lozano, M.: On the Combination of Evolutionary Algorithms and Stratified Strategies for Training Set Selection in Data Mining. Applied Soft Computing 6 (2006) 323-332
5. Casillas, J., Cordon, O., Herrera, F., Magdalena, L. (eds.): Interpretability Issues in Fuzzy Modeling, Springer, Berlin (2003)
6. Chiu, C. C., Hsu, P. L.: A Constraint-Based Genetic Algorithm Approach for Mining Classification Rules. IEEE Trans. on Systems, Man, and Cybernetics: Part C - Applications and Reviews 35 (2005) 205-220
7. Coenen F., Leng, P.: Obtaining Best Parameter Values for Accurate Classification. Proc. of 5th IEEE International Conference on Data Mining (2005) 549-552
8. Coenen, F., Leng, P., Zhang, L.: Threshold Tuning for Improved Classification Association Rule Mining. Lecture Notes in Artificial Intelligence, Vol. 3518: Advances in Knowledge Discovery and Data Mining - PAKDD 2005. Springer, Berlin (2005) 216-225
9. Curry, R., Heywood, M. I.: Towards Efficient Training on Large Datasets for Genetic Programming. Lecture Notes in Artificial Intelligence, Vol. 3060: Advances in Artificial Intelligence - Canadian AI 2004. Springer, Berlin (2004) 161-174
10. de la Iglesia, B., Philpott, M. S., Bagnall, A. J., Rayward-Smith, V. J.: Data Mining Rules using Multi-Objective Evolutionary Algorithms. Proc. of 2003 Congress on Evolutionary Computation (2003) 1552-1559
11. de la Iglesia, B., Reynolds, A., Rayward-Smith, V. J.: Developments on a Multi-Objective Metaheuristic (MOMH) Algorithm for Finding Interesting Sets of Classification Rules. Lecture Notes in Computer Science, Vol. 3410: Evolutionary Multi-Criterion Optimization - EMO 2005. Springer, Berlin (2005) 826-840
12. de la Iglesia, B., Richards, G., Philpott, M. S., Rayward-Smith, V. J.: The Application and Effectiveness of a Multi-Objective Metaheuristic Algorithm for Partial Classification. European Journal of Operational Research 169 (2006) 898-917
13. Deb, K.: Multi-Objective Optimization Using Evolutionary Algorithms. John Wiley & Sons, Chichester (2001)
14. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II. IEEE Trans. on Evolutionary Computation 6 (2002) 182-197
15. Elomaa, T., Rousu, J.: General and Efficient Multisplitting of Numerical Attributes. Machine Learning 36 (1999) 201-244
16. Freitas, A. A.: Data Mining and Knowledge Discovery with Evolutionary Algorithms. Springer, Berlin (2002)
17. Ghosh, A., Nath, B. T.: Multi-Objective Rule Mining using Genetic Algorithms, Information Sciences 163 (2004) 123-133
18. Ishibuchi, H., Murata, T., Turksen, I. B.: Single-Objective and Two-Objective Genetic Algorithms for Selecting Linguistic Rules for Pattern Classification Problems. Fuzzy Sets and Systems 89 (1997) 135-150
19. Ishibuchi, H., Nakashima, T., Murata, T.: Three-Objective Genetics-Based Machine Learning for Linguistic Rule Extraction. Information Sciences 136 (2001) 109-133
20. Ishibuchi, H., Nakashima, T., Nii, M.: Classification and Modeling with Linguistic Information Granules: Advanced Approaches to Linguistic Data Mining. Springer, Berlin (2004)

21. Ishibuchi, H., Namba, S.: Evolutionary Multiobjective Knowledge Extraction for High-Dimensional Pattern Classification Problems. Lecture Notes in Computer Science, Vol. 3242: Parallel Problem Solving from Nature - PPSN VIII. Springer, Berlin (2004) 1123-1132
22. Ishibuchi, H., Nojima, Y.: Accuracy-Complexity Tradeoff Analysis by Multiobjective Rule Selection. Proc. of ICDM 2005 Workshop on Computational Intelligence in Data Mining (2005) 39-48
23. Ishibuchi, H., Nozaki, K., Yamamoto, N., Tanaka, H.: Selecting Fuzzy If-Then Rules for Classification Problems using Genetic Algorithms. IEEE Trans. on Fuzzy Systems 3 (1995) 260-270
24. Ishibuchi, H., Yamamoto, T.: Fuzzy Rule Selection by Multi-Objective Genetic Local Search Algorithms and Rule Evaluation Measures in Data Mining. Fuzzy Sets and Systems 141 (2004) 59-88
25. Jin, Y. (ed.) Multi-Objective Machine Learning, Springer, Berlin (2006)
26. Kaya, M.: Multi-Objective Genetic Algorithm based Approaches for Mining Optimized Fuzzy Association Rules, Soft Computing 10 (2006) 578-586
27. Kupinski, M. A., Anastasio, M. A.: Multiobjective Genetic Optimization of Diagnostic Classifiers with Implications for Generating Receiver Operating Characteristic Curve. IEEE Trans. on Medical Imaging 18 (1999) 675-685
28. Li, W., Han, J., Pei, J.: CMAR: Accurate and Efficient Classification based on Multiple Class-association Rules. Proc. of 1st IEEE International Conference on Data Mining (2001) 369-376
29. Liu, B., Hsu, W., Ma, Y.: Integrating Classification and Association Rule Mining. Proc. of 4th International Conference on Knowledge Discovery and Data Mining (1998) 80-86
30. Llora, X., Goldberg, D. E.: Bounding the Effect of Noise in Multiobjective Learning Classifier Systems. Evolutionary Computation 11 (2003) 278-297
31. Mitra, S., Pal, S. K., Mitra, P.: Data Mining in Soft Computing Framework: A Survey. IEEE Trans. on Neural Networks 13 (2002) 3-14
32. Mutter, S., Hall, M., Frank, E.: Using Classification to Evaluate the Output of Confidence-based Association Rule Mining. Lecture Notes in Artificial Intelligence, Vol. 3339: Advances in Artificial Intelligence - AI 2004. Springer, Berlin (2004) 538-549
33. Pal, S. K., Talwar, V., Mitra, P.: Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions. IEEE Trans. on Neural Networks 13 (2002) 1163-1177
34. Quinlan, J. R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo (1993)
35. Reynolds, A., de la Iglesia, B.: Rule Induction using Multi-Objective Metaheuristics: Encouraging Rule Diversity. Proc. of 2006 International Joint Conference on Neural Networks (2006) 6375-6382
36. Tan, K. C., Yu, Q., Lee, T. H.: A Distributed Evolutionary Classifier for Knowledge Discovery in Data Mining. IEEE Trans. on Systems, Man, and Cybernetics: Part C - Applications and Reviews 35 (2005) 131-142
37. Thabtah, F., Cowling, P., Hammoud, S.: Improving Rule Sorting, Predictive Accuracy and Training Time in Associative Classification. Expert Systems with Applications 31 (2006) 414-426
38. Wang, H., Kwong, S., Jin, Y., Wei, W., Man, K. F.: Multi-Objective Hierarchical Genetic Algorithm for Interpretable Fuzzy Rule-Based Knowledge Extraction. Fuzzy Sets and Systems 149 (2005) 149-186
39. Wang, H., Kwong, S., Jin, Y., Wei, W., Man, K. F.: Agent-Based Evolutionary Approach for Interpretable Rule-Based Knowledge Extraction. IEEE Trans. on Systems, Man, and Cybernetics: Part C - Applications and Reviews 35 (2005) 143-155