

Multiobjective Genetic Rule Selection as a Data Mining Postprocessing Procedure

Hisao Ishibuchi
Graduate School of Engineering
Osaka Prefecture University
Sakai, Osaka 599-8531, Japan
+81-72-254-9350
hisaoi@cs.osakafu-u.ac.jp

Yusuke Nojima
Graduate School of Engineering
Osaka Prefecture University
Sakai, Osaka 599-8531, Japan
+81-72-254-9351
nojima@cs.osakafu-u.ac.jp

Isao Kuwajima
Graduate School of Engineering
Osaka Prefecture University
Sakai, Osaka 599-8531, Japan
+81-72-254-9351
kuwajima@ci.cs.osakafu-u.ac.jp

ABSTRACT

In this paper, we show the usefulness of multiobjective genetic rule selection as a postprocessing procedure in data mining for pattern classification problems. First we extract a prespecified number of rules using a data mining technique. Then we apply multiobjective genetic rule selection to the extracted rules. Experimental results show that multiobjective genetic rule selection significantly decreases the number of extracted rules while improving their classification accuracy.

Categories and Subject Descriptors

I.2.6 [Artificial Intelligence]: Learning – Knowledge Acquisition.

General Terms

Algorithms.

Keywords

Data mining, knowledge extraction, rule selection, multiobjective rule selection, evolutionary multiobjective optimization.

1. INTRODUCTION

Evolutionary multiobjective optimization (EMO) is a very active research area in the field of evolutionary computation [2]. The main advantage of EMO algorithms is that multiple tradeoff solutions are simultaneously obtained by their single run. Recently EMO algorithms have been employed in some studies on modeling and classification. For example, an EMO algorithm was used to generate multiple neural networks on a receiver operating characteristic curve [8]. Gonzalez et al. [5] generated multiple radial basis function networks of different sizes. In [9], an EMO algorithm was used in Pittsburgh-style learning classifier systems. In some studies on fuzzy systems [1], [6], [7], EMO algorithms were used to analyze the tradeoff structure between the accuracy and the interpretability of fuzzy rule-based systems. In this paper, we examine the usefulness of multiobjective genetic rule selection as a postprocessing procedure in data mining.

2. MULTIOBJECTIVE RULE SELECTION

Let us assume that we have N classification rules extracted from numerical data by a data mining procedure. Multiobjective rule selection is to find Pareto-optimal rule sets from the given N rules

with respect to the two main goals of rule extraction: accuracy maximization and complexity minimization.

Let S be a subset of the given N rules. The accuracy of the rule set S is measured by the error rate when all the training patterns are classified by S . We use a single winner rule-based method to classify each training pattern by S . That is, each pattern is classified by the single winner rule that has the maximum rule weight among compatible rules with that pattern. On the other hand, we measure the complexity of S by the number of rules in S . Thus our multiobjective rule selection problem is formulated as

Minimize the error rate by S and the number of rules in S . (1)

Any subset S of the N rules can be represented by a binary string of length N as $S = s_1s_2 \cdots s_N$, where $s_j = 1$ and $s_j = 0$ mean that the j th rule is included in S and excluded from S , respectively. Such a binary string is handled as an individual in our approach.

Since feasible solutions are represented by binary strings, we can directly apply almost all EMO algorithms to our multiobjective rule selection problem in (1) using standard genetic operations (e.g., uniform crossover and bit-flip mutation). In this paper, we use the NSGA-II algorithm [3]. Two domain-specific tricks are used as in our former study on fuzzy rule selection [7].

3. HEURISTIC RULE EXTRACTION

For n -dimensional pattern classification problems with continuous attributes, we use classification rules of the following type:

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

where R_q is the label of the q th rule, $\mathbf{x} = (x_1, \dots, x_n)$ is an n -dimensional pattern vector, A_{qi} is an antecedent interval, C_q is a class label, and CF_q is a rule weight (i.e., certainty grade). Each antecedent condition “ x_i is A_{qi} ” means the inclusion $x_i \in A_{qi}$.

The first step of heuristic rule extraction is the discretization of each continuous attribute. Since we usually have no *a priori* information about an appropriate granularity of the discretization for each attribute, we simultaneously use multiple partitions with different granularities (i.e., from coarse partitions into a few intervals to fine partitions into many intervals). This is one characteristic feature of our approach to knowledge extraction. Since we simultaneously use multiple partitions with different granularities, we do not have to determine the number of intervals for each attribute. In computational experiments, we use five partitions into K intervals where $K = 1, 2, 3, 4, 5$. It should be

Copyright is held by the author/owner(s).

GECCO '06, July 8–12, 2006, Seattle, Washington, USA.

ACM 1-59593-186-4/06/0007.

noted that $K = 1$ corresponds to the whole domain interval. To find $(K-1)$ cutting points for the discretization of each attribute into K intervals, we use an optimal splitting method [4] based on the class entropy measure [10].

The next step is to determine the consequent class and the rule weight for each combination of the antecedent intervals. The consequent class is specified as the class with the maximum confidence for each antecedent combination among all possible classes. The confidence of the rule is used as its rule weight.

In this manner, we can generate a large number of classification rules. It is, however, very difficult for human users to handle such a large number of rules. It is also difficult to understand long rules with many antecedent conditions. Thus we generate only short rules with a small number of antecedent conditions.

We further decrease the number of rules by choosing only good rules with respect to a heuristic rule evaluation criterion. That is, we choose a prespecified number of short rules for each class using a heuristic rule evaluation criterion. In our computational experiments, we examine the performance of the following four heuristic rule evaluation criteria:

- (1) The support criterion with the minimum confidence level.
- (2) The confidence criterion with the minimum support level.
- (3) The product criterion of confidence and support.
- (4) The difference criterion between the support and the sum of the supports over the other classes.

4. COMPUTATIONAL EXPERIMENTS

We use six data sets from the UCI repository: Wisconsin breast cancer, diabetes, glass, Cleveland heart disease, sonar, and wine. First we extract 100 rules for each class using each rule evaluation criterion. Then we apply multiobjective genetic rule selection to the extracted rules. We use the ten-fold cross-validation (10-CV) procedure to evaluate the average error rates on test patterns. An example of experimental results is shown in Figure 1 where 200 rules (100 rules for each class) are extracted by the support criterion with the minimum confidence level 0.6. The right-most closed circle shows the performance of those 200 rules. The results of C4.5 reported in [4] are also shown in Figure 1. From Figure 1, we can see that multiobjective genetic rule selection significantly decreases the number of rules while improving their classification accuracy.

5. CONCLUSIONS

In this paper, we demonstrated that multiobjective genetic rule selection improved the accuracy-complexity tradeoff curve of heuristically extracted classification rules. This improvement was observed in our computational experiments on all the six data sets for training patterns and the five data sets for test patterns except for the Cleveland heart disease data. Since a large number of rules are usually extracted in data mining, multiobjective genetic rule selection seems to play a significant role as a postprocessing procedure to decrease the complexity of extracted rules.

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

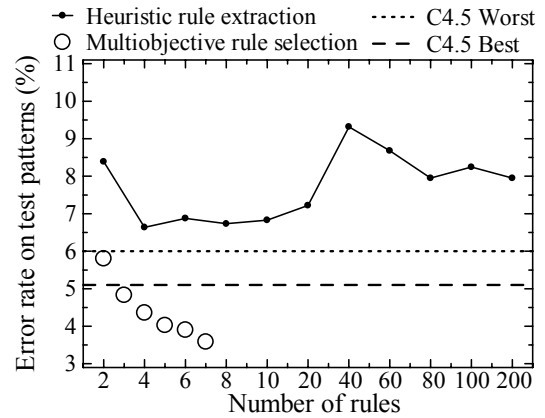


Figure 1. Results on the Wisconsin breast cancer data set.

6. REFERENCES

- [1] Casillas, J., Cordon, O., Herrera, F., and Magdalena, L. (eds.) *Interpretability Issues in Fuzzy Modeling*. Springer, Berlin, 2003.
- [2] Deb, K. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley & Sons, Chichester, 2001.
- [3] Deb, K., Pratap, A., Agrawal, S., and Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. on Evolutionary Computation* 6, 2 (2002) 182-197.
- [4] Elomaa, T., and Rousu, J. General and efficient multisplitting of numerical attributes. *Machine Learning* 36, 3 (1999) 201-244.
- [5] Gonzalez, J., Rojas, I., Ortega, J., Pomares, H., Fernandez, F. J., and Diaz, A. F. Multiobjective evolutionary optimization of the size, shape, and position parameters of radial basis function networks for function approximation. *IEEE Trans. on Neural Networks* 14, 6 (2003) 1478-1495.
- [6] Ishibuchi, H., Nakashima, T., and Nii, M. *Classification and Modeling with Linguistic Information Granules*. Springer, Berlin, 2004.
- [7] Ishibuchi, H., and Yamamoto, T. Fuzzy rule selection by multi-objective genetic local search algorithms and rule evaluation measures in data mining. *Fuzzy Sets and Systems* 141, 1 (2004) 59-88.
- [8] Kupinski, M. A., and Anastasio, M. A. Multiobjective genetic optimization of diagnostic classifiers with implications for generating receiver operating characteristic curve. *IEEE Trans. on Medical Imaging* 18, 8 (1999) 675-685.
- [9] Llorca, X., and Goldberg, D. E. Bounding the effect of noise in multiobjective learning classifier systems. *Evolutionary Computation* 11, 3 (2003) 278-297.
- [10] Quinlan, J. R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Mateo, CA, 1993.