Rule Induction for Classification Using Multi-objective Genetic Programming

A.P. Reynolds and B. de la Iglesia

School of Computing Sciences,
University of East Anglia,
Norwich

Abstract. Multi-objective metaheuristics have previously been applied to partial classification, where the objective is to produce simple, easy to understand rules that describe subsets of a class of interest. While this provides a useful aid in descriptive data mining, it is difficult to see how the rules produced can be combined usefully to make a predictive classifier. This paper describes how, by using a more complex representation of the rules, it is possible to produce effective classifiers for two class problems. Furthermore, through the use of multi-objective genetic programming, the user can be provided with a selection of classifiers providing different trade-offs between the misclassification costs and the overall model complexity.

1 Introduction

Earlier work by the authors [1,2,3,4] described the application of multi-objective metaheuristics to the problem of partial classification [5]. This problem is the search for simple rules, that represent 'strong' or 'interesting' descriptions of a specified class, or subsets of the specified class, even when that class has few representative cases in the data. These rules are of the form

- `if age` $\geq$ `28 and firstDegree = mathematics and attendance` $\geq$ `90% then result = distinction`

where the antecedent is a conjunction of simple attribute tests and the consequent, describing the class of interest, is the same for all rules generated.

Such simple rules may have high confidence, in that the rule produces few false positives. They may have high coverage, in that they describe a high proportion of the class of interest. Multi-objective metaheuristics can be used to produce different trade-offs between confidence and coverage. However, this simple rule representation is insufficiently descriptive to produce an individual rule with both high confidence and coverage.

In other work, Ghosh and Nath [6] used a multi-objective genetic algorithm for association rule mining, optimizing the accuracy, comprehensibility and interestingness of the rules produced. Association rules are similar to that shown above, but with tests usually limited to equalities and with an unconstrained consequent that may be any conjunction of such tests.

S. Obayashi et al. (Eds.): EMO 2007, LNCS 4403, pp. 516–530, 2007.

Both partial classification and association rule mining fall primarily into the category of *descriptive* data mining. A natural question is, can this work with simple descriptive rules be extended or modified to create *understandable* and highly *predictive* models that can classify previously unseen records? There are two approaches to take to this task: select a subset of the simple rules created to act as a classifier or increase the expressiveness of the rule representation.

Ishibuchi et al. [7,8] take the first of these approaches. In their work, a multi-objective algorithm is used to select a small subset of association rules produced by another algorithm, minimizing rule set complexity and error rate. However, this approach has a number of disadvantages:

- A good rule set may contain individuals that are far from the Pareto-front, according to the objectives of rule confidence and support [8]. Hence a very large set of simple rules must be created. For example, from a small training set of 342 records, 17070 classification rules were extracted, from which the multi-objective metaheuristic selected no more than 25 [8].
- Ideally, a record should assigned to a class if *any* of the rules in the rule set make this prediction. Then each rule provides a useful description of a subset of the data. However, in practice rules may make conflicting predictions. To handle conflicts, Ishibuchi et al. essentially create a decision list rather than a simple rule set [9], with rules lower on the list being used only when none of the higher rules apply. Converting such a list to a simple rule set reveals added complexity hidden in the decision list representation.
- A set of rules may not be the simplest way in which to represent a model of the data. This is illustrated by the example given in section 2.

In this paper we take the alternative approach of using a more expressive rule representation, specifically by using expression trees. While there is much literature on the use of genetic programming to optimize trees for the purposes of classification, this mostly concentrates on the optimization of decision trees, e.g. [10,11,12,13]. In particular, Mugambi and Hunter [14] apply multi-objective genetic programming to decision tree induction, optimizing both tree accuracy and tree simplicity. However, decision trees are different to the expression trees developed in this paper, with internal nodes that define partitions of the data and leaf nodes that indicate class membership. While rules may easily be extracted from such decision trees, we concentrate in this paper on the direct production and optimization of rules.

The work of Setzkorn and Paton [15,16] is perhaps more relevant to this paper, applying multi-objective genetic programming directly to fuzzy rule induction. However, internal nodes are restricted to two fuzzy forms of the boolean 'and' operation and the algorithm optimizes sets of these rules.

Section 2 provides details of the expression tree representation used. This representation is manipulated by a multi-objective metaheuristic to produce models with different trade-offs between model complexity and model accuracy (section 3). Section 4 describes the experiments performed and presents the results of using this approach. Finally, section 5 presents some conclusions and section 6 describes areas of further research.