Solving Classification Problems through Automatic Programming

Project Details:

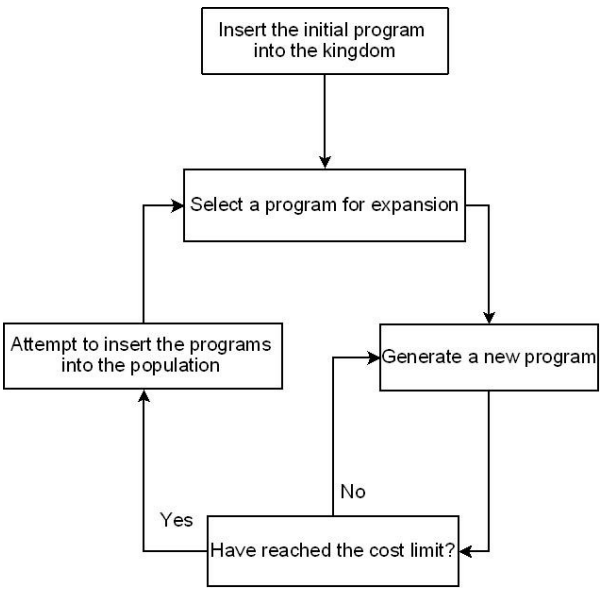
MSc Student: Stig-Erland Hansen
Supervisor: Associate Professor Roland Olsson

Introduction

The problem in classification is how to predict which class an instance belongs to based on its attributes, for example predicting whether a mushroom is poisonous by considering its color and height.

Another paradigm in machine learning is automatic programming where a system automatically creates programs to solve a specific problem.

This project presents two approaches to how automatic programming can be applied to classification through the use of the automatic programming system, Automatic Design of Algorithms Through Evolution (ADATE).



Graphical representation of the overall search algorithm in ADATE.

Improving Decision Tree Learning

ADATE was used to improve a specific classification algorithm called decision tree learning by implementing the algorithm in ADATE-ML and targeting different parts of it. This methodology is certainly not restricted to decision tree learning and can be applied to virtually any classification algorithm.

We targeted decision tree pruning, a specific part of decision tree learning, and showed on two occasions that ADATE was able to improve the pruning algorithm chosen as the start program with statistical significance for a fairly general domain.

```
fun estimateError( c, n ) =
let
    val v1 = tanh( tanh( tanh( (n - c) /  n ) ) )
    val v2 = sqrt( tanh( tanh( sqrt( n ) ) ) )
    val v3 = tanh( sqrt( sqrt( c ) ) )
in
    (v1 + v2) / v3
end
```

Excerpt from the pruning algorithm synthesized by ADATE

Synthesizing Classifiers

ADATE was applied directly to 10 classification problems in order to synthesize classifiers. The specifications were generated automatically and the traditional attribute-value encoding was used.

The execution of the 10 specifications resulted in accurate classifiers, and ADATE proved to be competitive with state-of-the-art classification algorithms. It was found, along with Logical Model Trees, to produce on average the most accurate classifiers, although it was only significantly better than 8 of the 32 classification algorithms executed.