

THE APPLICATION OF GENETIC PROGRAMMING FOR FEATURE CONSTRUCTION IN CLASSIFICATION

A thesis

submitted to the School of Computing Sciences

at the University of East Anglia

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

By

Mohammed Ahmed Yahya Muharram

June 2005



Abstract

This Thesis addresses the task of feature construction for classification. The quality of the data is one of the most important factors influencing the performance of any classification algorithm. The attributes defining the feature space of a given data set can often be inadequate, making it difficult to discover interesting knowledge. However, even when the original attributes are individually inadequate, it is often possible to combine such attributes in order to construct new ones with greater predictive power.

The goal of this Thesis is to restructure the feature space in order to improve the performance of decision tree classification techniques on complex, real world data. The proposed framework involves the use of genetic programming to evolve (construct) new attributes, which are non-linear combinations of the original attributes. This approach incorporates a number of decision tree splitting mechanisms in the fitness measures of the genetic program.

The empirical results obtained are encouraging and show that classification techniques can definitely benefit from the inclusion of an evolved attribute in terms of the accuracy and model size (for decision tree classifiers). When compared to existing approaches, the use of a decision tree splitting criteria as the fitness of the genetic program prove to be competitive and robust in terms predictive accuracy. Additionally, some of the evolved attributes manage to uncover physical properties in the data.

UEA Print Thesis

The full text of this thesis is not currently available
in the repository

Please contact

lle.helpdesk@uea.ac.uk

to request access to the print copy