Bio-Inspired Computation: Coursework 3  
Experimenting with Genetic Programming

Genetic programming (GP) can be applied to a wide variety of problems, and often finds good solutions in comparison with other computational methods. However, in practice good performance requires a certain amount of parameter tuning to be done. A typical GP system has many parameters, and many values that can be chosen for these. These choices affect how quickly and, in the case of multiple runs, how often a good solution will be found.

The aim of this coursework is for you to gain an understanding of how parameter choices affect the performance of GP in practice. You will be using ECJ to do this. ECJ provides a mature implementation of GP. Using parameter files, it allows you to experiment with a range of different parameter settings. ECJ also comes with a selection of different benchmarking problems, which you can use to measure the effect of different parameter settings.

What you are asked to do:

1. Familiarise yourself with ECJ and its GP facilities.
2. Investigate how different parameter choices affect the performance of GP upon a selection of benchmark problems.
3. Write a report, of up to 5 pages, describing what you have learnt.

Again, in more detail

1. ECJ

ECJ can be downloaded from: 
http://cs.gmu.edu/~eclab/projects/ecj/

Usage instructions, including a manual and tutorials, can be found on the ECJ website: 
http://cs.gmu.edu/~eclab/projects/ecj/docs/

If you have not already done so, it is recommended that you work through the tutorials, since these contain information about using ECJ and its parameter files. An overview of parameters and parameter files can also be found at: 
http://cs.gmu.edu/~eclab/projects/ecj/docs/parameters.html

2. Experimentation

The aim of this coursework is for you to gain an understanding of how parameter choices affect the performance of GP in practice. To begin with, you should familiarise yourself with ECJ, and identify parameters which you think will have a significant impact upon the ability of GP to solve problems. ECJ allows you to investigate many parameters, but it will not be feasible for you to experiment with all of these. You should therefore choose parameters that you think will have a significant impact upon the performance of GP, based upon knowledge you have gained whilst studying on this course, or knowledge you have picked up whilst doing wider reading.

You should investigate at least three significant parameters. These should be chosen from at least three of the following categories:
1. Solution representation, e.g. choice of function sets, use of ADFs and types
2. Initialisation, e.g. population size and constitution
3. Variation, e.g. mutation and crossover operators
4. Selection mechanisms

You should also choose at least three problems on which to evaluate the effects of your parameter choices. ECJ provides a number of benchmark problems which you can use. These are located in the ec/app directory within the ECJ distribution. You have been introduced to some of these in the lectures. However, you are also encouraged to investigate the other problems, many of which are standard benchmarks for GP systems. Note: Not all the problems in this directory are suitable for use with GP—read the comments in the java files, or the readme files, as appropriate.

Having identified the parameters and problems that you wish to investigate, you should then design suitable experiments and collect appropriate results.

Bear in mind that GP is a non-deterministic algorithm. This means that it will generate different results in different runs, depending on the initial random seed. Hence, you should use different random seeds in each run, and consider performance measures that summarise behaviour across a number of runs.

Also bear in mind that there are various measures of GP performance, for instance the correctness of a solution, how often a solution is found, how interpretable a solution is, and the amount of computation that needs to be done to find a solution.

3. Report

Your report (up to 5 pages in length) should:
  • Describe the parameters you chose to investigate. Explain why you think these are important, and describe the different choices that are available for these parameters.
  • Briefly describe the problems you chose to use, and explain why you think these are useful problems with which to evaluate GP.
  • Summarise the experiments you carried out and present results using appropriate tables and graphs.
  • Discuss the findings of your experiments, highlighting any insights that you have made.
  • Include references to any relevant material.

Whilst you are not expected to include whole parameter files with your report, you should include enough information so that it is obvious which ECJ parameters you experimented with and the values you used. Be concise wherever possible. You are not expected to write your own code. However, if you choose to do so, attach the code to the report (it will not be included within the 5 page limit).

Hand-in: Email your report to me at M.Lones@hw.ac.uk with subject line “BIC GP” by 23:59pm Wednesday 3rd December 2014.

Marking scheme
This is worth 20% of the module. Marking will take into account the quality and quantity of work done. You are encouraged to be adventurous, and extra marks will be awarded for going beyond the material covered in the lectures.