Size Fair Tree Genetic Programming Crossover

W.B.Langdon@cwi.nl

Centrum voor Wiskunde en Informatica, Kruislaan 413, Amsterdam

Abstract

Size fair crossover genetic operator for tree based genetic programming is described and tested. It produces considerably reduced increases in program size and no detrimental effect on GP performance. GP search spaces are partitioned by the ridge in the number of program v. their size and depth. A ramped uniform random initialisation is described which straddles the ridge. With subtree crossover trees increase about one level per generation leading to sub-quadratic bloat in length.

1 Introduction

It has been known for some time that programs within GP populations tend to rapidly increase in size as the population evolves [1]. We suggest stochastic search techniques tend to find the most common programs in the search space of the current best fitness. Since in general there are more of these which are long than there are which are short (but GP starts with the shorter ones) the population tends to be filled with longer and longer programs. Using this argument we devise an unbiased tree crossover operator which carefully controls variation in size and produces much less bloat (Section 2). Section 3 extends it to produce a homologous crossover operator.

We compare the evolution of tree size and depth for the new crossover operators with standard subtree crossover starting from three types of initial random populations: standard "ramped half-and-half", "ramped half-and-half" with bigger initial trees and ramped uniform random (Section 4). We compare both new operators with standard subtree crossover on two continuous domain problems (symbolic regression of the quintic and sextic polynomials) and two discrete problems (Boolean 6 multiplexor and 11 multiplexor, [3] runs these $\approx \times 32$ faster).

2 Size Fair Crossover

We select in the normal way two parents and the crossover point in the first parent. The difference is the choice of the second crossover point. The size of the subtree to be deleted from the first parent is used to guide the random choice of the second crossover point so that a subtree of a similar size will be inserted. On average, like subtree crossover, removed and added subtrees are the same size.

3 Homologous Crossover

Homologous crossover is identical to size fair except instead of randomly choosing a subtree of the desired size it deterministically chooses the one syntactically closest to the subtree in the first parent.

4 Ramped Uniform Initialisation

The standard techniques for producing initial populations have biases towards full trees. Ramped uniform initialisation uniformly samples from a range of sizes.

5 Linear Increase in Depth

In standard crossover runs programs' depths increase rapidly at about one level per generation (this is about 5 times faster than size fair or homologous runs).

Most programs are neither full nor sparse but of intermediate shapes. The depths of most binary trees lie near $2\sqrt{\pi(\text{internal nodes})}$. Linear increase in depth together with remaining close to this ridge in the number of programs versus their shape leads to a prediction of sub-quadratic growth in program length [1].

6 Conclusions

We present and demonstrate on four bench mark problems two new bloat reduced crossover operators and a new mechanism for creating random populations for tree based genetic programming [2]. The results in terms of reduction in growth of both mean and maximum program and solution sizes are impressive and are achieved without reduction in search efficiency.

Growth in program sizes was found not to depend overly on the initial population however it does have a dominant role in the evolution of program shapes. Populations tend to evolve towards the ridge in the distribution of programs for each size and shape. They don't cross it. I.e. it divides the search space. "Ramped half-and-half" does not search a large part of the search space corresponding to long thin trees.

Average growth in program depth when using standard subtree crossover is near linear in these problems. When combined with the parabolic distribution of programs, yields a prediction of sub-quadratic growth in program size. GP populations using standard crossover (and no parsimony techniques) quickly reach bounds on size or depth commonly used.

References

[1] W. B. Langdon, T. Soule, R. Poli, and J. A. Foster. The evolution of size and shape. In L. Spector, W. B. Langdon, U.-M. O'Reilly, and P. J. Angeline, editors, *Advances in Genetic Programming 3*, chapter 8, MIT Press, 1999.

[2] W. B. Langdon. Size fair and homologous tree genetic programming crossovers. In W. Banzhaf *et. al.* editors, *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*, Orlando, 1999. Morgan Kaufmann.

[3] R. Poli and W. B. Langdon. Sub-machine-code genetic programming. In L. Spector *et. al.* editors, *Advances in Genetic Programming 3*, chapter 13.