Enhancing the Performance of GP Using an Ancestry-Based Mate Selection Scheme

Rodney Fry and Andy Tyrrell

Department of Electronics
University of York, UK
{rjlf101,amt}@ohm.york.ac.uk

Abstract. The performance of genetic programming relies mostly on population-contained variation. If the population diversity is low then there will be a greater chance of the algorithm being unable to find the global optimum. We present a new method of approximating the genetic similarity between two individuals using ancestry information. We define a new diversity-preserving selection scheme, based on standard tournament selection, which encourages genetically dissimilar individuals to undergo genetic operation. The new method is illustrated by assessing its performance in a well-known problem domain: algebraic symbolic regression.

1   Ancestry-Based Tournament Selection

Genetic programming [1] in its most conventional form, suffers from a well-known problem: the search becomes localised around a point in the search space which is not the global optimum. This may be a result of not taking steps to preserve the genetic diversity of the population.

A selection scheme in GP may be made to promote diversity if one could select individuals not only based on their fitness, but also based on how dissimilar the individuals' genotypes are. Previous work in measuring how genotypically similar individuals are has concentrated their phenotypic relationship [2] and structural difference [3]. A good measure of how similar two individuals are can be attained by looking at the individuals' ancestries. For example, siblings are far more likely to have similar genotypes than two randomly chosen individuals. If each individual in a GP population had the capability of knowing who its parents and grandparents were, and chose a mate based partly on dissimilarity between their respective ancestries, then search of deceptive local optima would be reduced. Previous work in the area includes the incest prevention scheme [4] and crossover prohibition based on ancestry [5].

In our scheme, a binary tree of ancestry is associated with each individual. The information stored relates to which previous individuals were used in the creation of the current solution and the proportion of each parent individual that was donated to create the current solution: the *Parental Contribution Fraction* or PCF. The first of the two individuals that will take part in the crossover

operation is selected using standard tournament selection. A second parent individual must then be selected based on how genetically dissimilar it is when compared to the first. This can be achieved by modifying the fitness values of the candidate individuals to reflect their *apparent* fitness from the first parent's point of view, before they enter the tournament.

2   Results

The modified tournament selection mechanism was tested and results were compared with the standard tournament selection scheme and the ancestry-based crossover prohibition scheme. Tests were carried out on the symbolic regression problem outlined in Section 10.2 of Koza [1]. As Fig. 1 shows, our 'PCF Tree' implementation outperforms the standard tournament selection scheme and the crossover prohibition scheme.
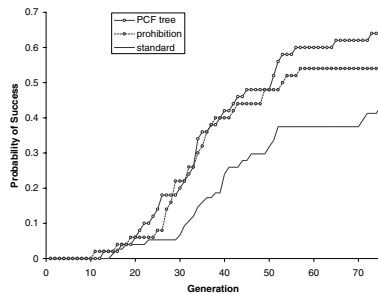

Fig. 1. Comparison of the performance of the PCF ancestry tree, the crossover prohibition scheme and standard tournament selection

References

1. John R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection.* MIT Press, 1992.
2. S.W. Mahfoud. Crowding and preselection revisited. In Reinhard Männer and Bernard Manderick, editors, *Parallel problem solving from nature 2*, pages 27–36, Amsterdam, 1992. North-Holland.
3. R. Keller and W. Banzhaf. Explicit maintenance of genetic diversity on genospaces, 1994.
4. L.J. Eshelman and J.D. Schaffer. Preventing premature convergence in genetic algorithms by preventing incest. In *Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 115–122, San Diego, CA, 1991. Morgan Kaufmann.
5. Rob Craighurst and Worthy Martin. Enhancing ga performance through crossover prohibitions based on ancestry. In Larry J. Eshelman, editor, *Proceedings of the Sixth International Conference on Genetic Algorithms*, pages 130–135. Morgan Kaufmann, 1995.