Fitness Clouds and Problem Hardness in Genetic Programming

Leonardo Vanneschi[2], Manuel Clergue[1], Philippe Collard[1], Marco Tomassini[2], and Sébastien Vérel[1]

[1] I3S Laboratory, University of Nice, Sophia Antipolis, France
[2] Information Systems Department, University of Lausanne, Lausanne, Switzerland

Abstract. This paper presents an investigation of genetic programming fitness landscapes. We propose a new indicator of problem hardness for tree-based genetic programming, called *negative slope coefficient*, based on the concept of fitness cloud. The negative slope coefficient is a predictive measure, i.e. it can be calculated without prior knowledge of the global optima. The fitness cloud is generated via a sampling of individuals obtained with the Metropolis-Hastings method. The reliability of the negative slope coefficient is tested on a set of well known and representative genetic programming benchmarks, comprising the binomial-3 problem, the even parity problem and the artificial ant on the Santa Fe trail.

1 Introduction

Genetic Programming (GP) has had an indeniable practical success in its fifteen years of existence [13,14]. However, it is still difficult to understand why some problems are easily solved by GP, while others resist solution or require massive amounts of computational effort. It would thus be of interest if we were able to somehow classify problems according to some measure of difficulty. To start with, it might be useful to take a look at what has been done in the older field of *genetic algorithms* (GAs). Difficulty studies in GAs have been pioneered by Goldberg and coworkers [4,6,8]. Their approach is focused on the construction of functions that should *a priori* be easy or hard for GAs to solve. These ideas have been followed by many others, for instance [18,5] and have been at least partly successful in the sense that they have been the source of many ideas as to what makes a problem easy or difficult for GAs. One concept that underlies many of these approaches is based on the notion of *fitness landscape*. The metaphor of a fitness landscape [22], although not without faults, has been a fruitful one in several fields. In particular, a statistic called *fitness distance correlation* (FDC) [10] has been studied in detail in the past in the context of GAs. Its suitability for GP has been investigated in [2,24,23]. As far as GP is concerned, but also in the GA field, the general conclusion of these studies was that the FDC can be considered as a rather reliable indicator of problem hardness. However, it has some severe drawbacks: sometimes the measure does not give any indication, and problems can be constructed for which the FDC leads to contradictory conclusions. The first consideration is not really serious since it manifests itself rarely and other tools, such as the analysis of the fitness-distance scatterplot can be brought to bear in these cases. On the other hand, the existence of counterexamples casts a shadow on the

K. Deb et al. (Eds.): GECCO 2004, LNCS 3103, pp. 690–701, 2004.

usefulness of the FDC, although such cases are typically contrived ones and they do not seem to appear often among "natural" problems. But the really annoying fact about FDC, and its main weakness in our opinion, is that the optimal solution (or solutions) must be known beforehand, which is obviously unrealistic in applied search and optimization problems, and prevents us from applying FDC to many GP benchmarks and real-life applications. Thus, although the study of FDC is an useful first step, we present another approach based on quantities that can be measured without any explicit knowledge of the genotype of optimal solutions, such as different kinds of fitness distributions. A second consideration concerns the way in which space is sampled: uniform random sampling has the merit of being simple and algorithm-independent (only random search is implied), but it would be more useful to have a sample of the landscape as "seen" by a specific algorithm for it is the latter the one that is really relevant. In this way, more weight can be given to particular points in the space. Thus, sampling according to a given stationary probability distribution, like Markov chain Monte Carlo, appears to be more appropriate.

This paper is structured as follows: section 2 summarizes some techniques used by other researchers in the past few years, which inspired this work. Section 3 introduces the concept of fitness cloud. This concept is used in section 4 to define a new measure, called *negative slope coefficient* (NSC), that is proposed as an indicator of problem hardness. Section 5 briefly introduces the test problems used to verify the reliability of NSC. Section 6 shows experimental results. Finally, section 7 offers our conclusions and hints for future work.

2 Fitness-Fitness Correlation: Previous Work

In genetic algorithms, plotting fitness against some features is not a new idea. Manderick *et al*. [17] study the correlation coefficient of genetic operators: they compute the correlation between the fitnesses of a number of parents and the fitnesses of their offspring. Grefenstette [7] uses fitness distribution of genetic operators to predict GA behaviour. Rosé *et al.* [20] develop the *density of states* approach by plotting the number of genotypes with the same fitness value. Smith *et al.* [21] focus on notions of *evolvability* and *neutrality*; they plot the average fitness of offspring over fitness according to Hamming neighbouring. *Evolvability* refers to the efficiency of evolutionary search. It is defined by Altenberg as "the ability of an operator/representation scheme to produce offspring that are fitter than their parents" [1].

Fitness correlation measures are almost absent in GP but there are a few precursors. The first work we are aware of is the article of Kinnear [12] in which GP difficulty is analysed through the use of the fitness *autocorrelation function*, as first proposed by Weinberger [26]. While fitness autocorrelation analysis has been useful in the study of NK landscapes [26], Kinnear found his results inconclusive and difficult to interpret: essentially no simple relationship was found between correlation length values and GP hardness. The work of Nikolaev and Slavov [19] also makes use of the fitness autocorrelation function with the main purpose of determining which mutation operator, among a few that they propose, "sees" a smoother landscape on a particular problem. Their analysis however, does not lead to a general study of problem difficulty in GP, which is our aim here. More recently, a much more detailed study of fitness correlation in GP has