Chapter 2

A POPULATION BASED STUDY OF EVOLUTIONARY DYNAMICS IN GENETIC PROGRAMMING

A.A. Almal¹, C.D. MacLean¹ and W.P. Worzel¹

¹Genetics Squared Inc.

Abstract Understanding the cause-and-effect relationship of the control parameters to the end result can significantly improve the efficiency of any machine learning algorithm. Genetic Programming (GP), even in its simplest form, exhibits complex evolutionary dynamics. This makes the task of determining the relationship of control parameter to the evolved population in a genetic programming system decidedly non-trivial. This chapter describes an investigation into the dynamics of GP. We present several different observations made on the relationship of parameters to evolution in a GP system with the help of a method for visualizing a GP population’s fitness, structure and content simultaneously.

Keywords: evolutionary dynamics, visualization, structure, fitness

1. Introduction

The evolutionary dynamics of genetic programming are surprising both in their complexity and in their similarity to natural evolution despite the utter simplicity of the evolutionary mechanisms used compared to those present in natural evolution. This paper draws from work done by the authors in (Almal et al., 2007) regarding a method for visualizing the structure and fitness of all of the individuals in a population. This is done by assigning a point on a circle for each terminal and then, starting from the center of the circle, moving $1/(n + 1)$ distance towards the point of the next terminal on the circle. The endpoints arrived at in this way are plotted and the fitness of the individual associated with the endpoint is colored in heat map form where the “heat” of
Figure 2-1. Example heat map for a population.

the point plotted is associated with the fitness of the individual. Figure 2-1\(^1\) shows a black and white plot with the boxes highlighting the most fit areas and demonstrating the presence of two distinctly different areas of roughly co-equal fitness. For further details on this technique, the reader is referred to (Almal et al., 2007).

In (Almal et al., 2007) the authors discussed the relationship between genotype and phenotype in GP. We concluded that because the genetic operators of crossover and mutation were genotypic operations, while selection was a phenotypic operation, there was a disconnect between movement through “fitness space” and “structure-content space” that interacted in ways that suggest some areas of structure space to be easier to reach than others, thereby influencing the search path of GP. Moreover, the behaviors observed using this method of visualizing a GP population suggested that GP population behavior was more similar to natural evolution than might be expected.

In this paper, we have visualized the evolution of a number of populations during evolution when the GP control parameters are varied and the GP system is applied to the same problem. Differences between the evolutionary process and outcome are examined and the effect of changing the parameters is discussed.

2. Methods

We used a classification problem from molecular biology as the basis for this study. The input data are normalized gene expression values generated from

\(^1\)Complete sets of color plots are available at http://www.cscs.umich.edu/gptp-workshops/gptp2008 or at http://genetics2.com/Partners\%20&\%20Collaborators/partners.html