Chapter 7

WHAT MAKES A PROBLEM GP-HARD?
A Look at How Structure Affects Content

Jason M. Daida
The University of Michigan, Center for the Study of Complex Systems and Space Physics Research Laboratory, 2455 Hayward Avenue, Ann Arbor, Michigan USA 48109-2143, daida@umich.edu

Abstract: This chapter summarizes theoretical work at The University of Michigan concerning the question: "What makes a problem difficult for genetic programming to solve?" It specifically describes linkages between content, tree structures, and problem difficulty in genetic programming. It focuses on the significance of structure in influencing problem difficulty.

Key words: GP theory, tree structures, problem difficulty, GP-hard, test problems

1. Introduction

Most of what I have learned about doing large, significant problems with GP has come from working on real-world applications. In particular, if I wanted to do a significant (and presumably difficult) problem nowadays, I would look to Koza (Koza et al., 1999). A short list of what I get from that work would be the following:

- Use large populations
- Use structure-altering operations
- Specify GP solutions as meta-programs and choose functions and terminals accordingly.

My list might differ from anyone else's, if only because my take on Koza's work has been influenced by a structuralist view of complex systems. I define a structuralist as one who examines the relationship of parts to a thing. I would subsequently be interested in the structure and topology of interconnections between nodes in GP. With this view, my
group has focused on the theoretical aspects of why some GP problems are more difficult than others. Still, it is ironic that I would turn to applications for understanding on how to really solve problems.

While GP theory has come a long way in the past decade, there is still much that theory has not explained. Even simple questions have yet to be answered. For example, although we have hypotheses, we still do not know why some problems are more difficult than others for GP. We also do not have a bound for what can or cannot be solved with even the most basic of GP systems. We have no theory that would tell us what to do if GP were to get stuck on a problem.

On the other hand, there are many success stories, anecdotes, and ad hoc methods that have been proven to work—if only for a particular instance. Do X, because we have shown it to work for problem Y. For a while, this approach resulted in one paper per application; it was justifiable, too. Every new application (seemingly) resulted in a new operator or a major adjustment to GP. I should know... guilty as charged.

I would suggest that the situation between theory and application in GP is like navigating the eastern seaboard of the United States. In the late 1800s, travel there was treacherous and difficult. Good maps (theory) were absent. A person relied on the experience (application) of navigators just to remain alive. An observer wrote in 1888, “...so little was known of the dangers attending navigation along our extensive seaboard, that those who engaged in commercial ventures were constrained to rely upon local knowledge and the reports of the hardy navigators who might carry their venture to success (Ogden, 1888).” I would say that the success of GP lies now with the practitioners, the “hardy” that make a living with GP. The “mapmaker” theoreticians are playing catch-up...

The purpose of this chapter is to highlight what our theory says in relationship to the question “What makes a problem GP-difficult?” It is not meant to be a rigorous treatment of theory. Instead, it summarizes what my group has learned. Much of what we know supports what practitioners have implicitly known for years. This is not to say that there is no room for surprises. In this exercise of making implicit knowledge explicit, surprises can and still do happen. Counterintuitive findings can occur by logically extending explicit knowledge. (My college professors liked to point this out when I took physics as an undergraduate.) Explicit knowledge also defines what we do not know—but would be good to know to improve the practice of GP.

This chapter specifically discusses three claims. The first is “content drives difficulty” (Section 2). The second is “structure matters” (Section 3). The third is “structure determines limits” (Section 4). Each claim is followed by an overview, followed by a summary of what our group knows, followed