An Analysis of Genetic Programming

by

Una-May O'Reilly

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Ottawa-Carleton Institute for Computer Science

School of Computer Science

Carleton University, Ottawa, Ontario

Sept 22 , 1995

The undersigned hereby recommend
to the Faculty of Graduate Studies and Research
acceptance of the thesis

An Analysis of Genetic Programming

submitted by Una-May O'Reilly, B.Sc. M.C.S.
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Director, School of Computer Science

Thesis Supervisor

External Examiner

Carleton University

September 22, 1995

An Analysis of Genetic Programming

This thesis analyzes Koza's Genetic Programming (GP) paradigm, a genetic algorithm for program discovery. In order to improve upon our understanding of GP and to improve GP, it provides a systematic analysis of GP that is based upon experimentation and theory.

We assess the role of designer expertise in successfully using GP. Our experiments show that its performance is influenced by propitious designer choices of the test suite and primitive set. We also appraise whether GP proceeds in a hierarchical manner. In experiments with the canonical earliest version of GP, GP did not appear to exploit a hierarchical process.

The theoretical analysis develops a schema-based framework for describing GP search behaviour. We formally develop a Schema Theorem for GP, define building blocks and state a GP Building Block Hypothesis. We proceed to methodically question the plausibility that GP exploits a building block process while searching.

We conduct further experimental analysis by comparing GP to alternative algorithms. A mutation-based operator, HVL-Mutate, that generates a syntactically valid and possibly structurally different program from another is introduced. Two adaptive search algorithms, Stochastic Iterated Hill Climbing and Simulated Annealing, which use either HVL-Mutate or GP crossover are implemented to solve exactly the same class of program discovery problems as GP. The resulting algorithms are comparable to GP and sometimes even outperform it on a small suite of these problems.

Because these algorithms are relatively successful at solving the same problems GP solves, we conjecture that synthesizing a localized search strategy into GP will complement its global, population-based search and improve it. Our experiments with our problem suite confirm this insight. When we hybridize GP by adding a hill climbing component, various versions of the hybrid algorithm achieve higher likelihood of success and process less candidate programs than GP.

To my sister Glenn Crowder
for showing me courage and strength

Acknowledgments

I wish to thank Ken De Jong, Jean-Pierre Corriveau, John Goldak and Stan Matwin for agreeing to serve on my thesis committee.

I would like to thank Franz Oppacher for acting as my supervisor. Franz is an intellectual magnet. I will always be in awe of his creativity, analysis, and comprehension. For seven years he has been a great source of ideas, knowledge and feedback for me. I like to brain-storm with him, argue with him, subject my writing to his critical eye, and simply hang around with him. If I had spent more time with him on this work, it would only have improved.

I would like to thank my family for loving me through all the ups and downs of these five years. My parents, Bob and Glenda invited me into their home again and gave me tremendously valuable support while I restructured my life in 1992-93. When my courage and resolve returned, they confidently waved me off to Santa Fe. They have always encouraged me to engage in activities that test me.

Unfortunately, working at SFI has meant more time away from my brother and sister and their families. I miss them all and thank them for being supportive and patient. My sister Glenn deserves special mention. She is presently engaged in the toughest test of all - fighting for her life, and I am incredibly proud of the tenacity and courage she has shown. Her illness forced me to place my life in perspective at a time when I could have comfortably slipped into the narrow, intensive, isolating endeavors of academics. She has been an inspiration to me and I would trade my Ph.D. for a cure for her cancer.

I started my Ph.D. in Ottawa at Carleton University's School of Computer Science. Among the graduate students, I would like to especially thank Dwight Deugo and Andrew Rau-Chaplin. I don't know if I simply tried to keep up with Dwight and Andrew or whether each of us was equally influential in pushing the others. Whatever, Andrew and Dwight have been true friends. I also especially thank Mark Wineberg and wish him the best of luck with finishing his dissertation.

The School of Computer Science is staffed by excellent people to whom I am indebted for efficiently managing teaching assignments, technical reports, computer administration, etc. as well as patiently replacing my lost keys and forwarding my mail! I especially thank Marlene Wilson, Barbara Coleman, and Rosemary Carter.

Other influential Ottawa people were Nicola Santoro, Daryl Graf, Laura and Larry House, and Tony White. Thank you.

I thank Bell Northern Research for two years of financial assistance.

I spent the greater part of two years conducting my thesis research at the Santa Fe Institute in Santa Fe, New Mexico. It is an unsurmountable task to state how influential SFI has been. I would like to thank Melanie Mitchell and Stephanie Forrest for considering my first (and second) request to visit. I would like to thank Mike Simmons, Vice President of Research at SFI, for agreeing that I should stay on as a graduate fellow once I got there.

SFI has an awesome visitor's program. Crucial interactions with Peter Stadler and Richard Palmer came this way. I profited tremendously from many debates with Bill Macready. How ironic that I had to move to New Mexico to work with someone from Kanata.

Terry Jones first greeted me standing atop a concrete slab, simultaneously juggling and teasing passers by. His unicycle was parked around the corner. He has always offered his help, good cheer and attention. I value it highly. He was a crucial link to a group of Stephanie Forrest's students. I would like to thank them, in particular Ron Hightower and Derek Smith, for allowing me to occasionally join their meetings.

I felt at home at SFI almost from the moment I arrived. I would like to thank Ginger Richardson, Andi Sutherland and Deborah Smith for their warm and friendly welcome. I would also like to thank Andi Sutherland for something extra. Marita Prandoni's cheerful daily greeting was wonderful.

Among the many helpful people at SFI I especially thank Walter Fontana, Raja Das, Nick Vriend, David Wolpert, and Emily Dickinson. I owe a great deal to dear friends Mihaela Oprea (my first in Santa Fe!), Martijn Huynen, Melanie Mitchell, Ann Bell and Kai Nagel.

Last of all, I would like to thank Blake LeBaron. Blake doggedly attacked the technical issues that were obstacles to me working in Madison. He taught me that love doesn't have to always be romantically expressed. Instead, it can be communicated by technical support, objectivity, patience, and the relinquishment of a CPU!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

The following abbreviations are used in this dissertation:

§x(y) Section x on page y.

AI Artificial Intelligence.

ADF Automatically Defined Functions. §2.5(68)

AR-GP Adaptive Representation Genetic Programming. §3.2(110)

GA Genetic Algorithm. §1.2.1(8)

GLiB Genetic Library Builder §2.5(66)

GP Genetic Programming. §2.2(40)

HVL-Mutate Hierarchical Variable Length Mutate §5.3(145)

SIHC Stochastic Iterated Hill Climbing. §5.1(138)

SA Simulated Annealing. §5.2(141)

XO-SA Crossover Simulated Annealing. §6.2.2(167)

XO-SIHC Crossover Stochastic Iterated Hill Climbing. §6.2.1(167)