
GENETIC PROGRAMMING THEORY AND PRACTICE

GENETIC PROGRAMMING SERIES

Series Editor

John Koza
Stanford University

Also in the series:

GENETIC PROGRAMMING AND DATA STRUCTURES: Genetic Programming + Data Structures = Automatic Programming! *William B. Langdon*; ISBN: 0-7923-8135-1

AUTOMATIC RE-ENGINEERING OF SOFTWARE USING GENETIC PROGRAMMING, *Conor Ryan*; ISBN: 0-7923-8653-1

DATA MINING USING GRAMMAR BASED GENETIC PROGRAMMING AND APPLICATIONS, *Man Leung Wong and Kwong Sak Leung*; ISBN: 0-7923-7746-X

GRAMMATICAL EVOLUTION: *Evolutionary Automatic Programming in an Arbitrary Language*, *Michael O'Neill and Conor Ryan*; ISBN: 1-4020-7444-1

GENETIC PROGRAMMING IV: *Routine Human-Computer Machine Intelligence*, *John R. Koza, Martin A. Keane, Matthew J. Streeter, William Mydlowec, Jessen Yu, Guido Lanza*; ISBN 1-4020-7446-8

The cover art was created by Leslie Sobel in Photoshop from an original photomicrograph of plant cells and genetic programming code. More of Sobel's artwork can be seen at www.lesliesobel.com

GENETIC PROGRAMMING THEORY AND PRACTICE

edited by

Rick Riolo

*Center for the Study of Complex Systems
University of Michigan*

Bill Worzel

Genetics Squared



SPRINGER SCIENCE+BUSINESS MEDIA, LLC

Library of Congress Cataloging-in-Publication Data

GENETIC PROGRAMMING THEORY AND PRACTICE

Edited by Rick Riolo and Bill Worzel

ISBN 978-1-4613-4747-7 ISBN 978-1-4419-8983-3 (eBook)

DOI 10.1007/978-1-4419-8983-3

Copyright © 2003 by Springer Science+Business Media New York

Originally published by Kluwer Academic Publishers in 2003

Softcover reprint of the hardcover 1st edition 2003

All rights reserved. No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording, or otherwise, without the written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Contents

Contributing Authors	ix
Preface	xiii
Foreword	xv
1	
Genetic Programming: Theory and Practice <i>Bill Worzel, and Rick Riolo</i>	1
2	
An Essay Concerning Human Understanding of Genetic Programming <i>Lee Spector</i>	11
3	
Classification of Gene Expression Data with Genetic Programming <i>Joseph A. Driscoll, Bill Worzel, and Duncan MacLean</i>	25
4	
Artificial Regulatory Networks and Genetic Programming <i>W. Banzhaf</i>	43
5	
Using Software Engineering Knowledge to Drive Genetic Program Design Using Cultural Algorithms <i>David A. Ostrowski, and Robert G. Reynolds</i>	63
6	
Continuous Hierarchical Fair Competition Model for Sustainable Innovation in Genetic Programming <i>Jianjun Hu, Erik D. Goodman, and Kisung Seo</i>	81
7	
What Makes a Problem GP-Hard? <i>Jason M. Daida</i>	99

8

A Probabilistic Model of Size Drift	119
<i>Justinian Rosca</i>	

9

Building-Block Supply in Genetic Programming	137
<i>Kumara Sastry, Una-May O'Reilly, David E. Goldberg, David Hill</i>	

10

Modularization by Multi-Run Frequency Driven Subtree Encapsulation	155
<i>Daniel Howard</i>	

11

The Distribution of Reversible Functions is Normal	173
<i>W. B. Langdon</i>	

12

Doing Genetic Algorithms the Genetic Programming Way	189
<i>Conor Ryan and Miguel Nicolau</i>	

13

Probabilistic Model Building and Competent Genetic Programming	205
<i>Kumara Sastry, David E. Goldberg</i>	

14

Automated Synthesis by Means of Genetic Programming of Complex Structures Incorporating Reuse, Parameterized Reuse, Hierarchies, and Development	221
<i>John R. Koza, Matthew J. Streeter, and Martin A. Keane</i>	

15

Industrial Strength Genetic Programming	239
<i>Mark Kotanchek, Guido Smits, and Arthur Kordon</i>	

16

Operator Choice and the Evolution of Robust Solutions	257
<i>Terence Soule</i>	

17

A Hybrid GP-Fuzzy Approach for Reservoir Characterization	271
<i>Tina Yu, Dave Wilkinson, and Deyi Xie</i>	

18

Enhanced Emerging Market Stock Selection	291
<i>Anjun Zhou</i>	

<i>Contents</i>	vii
19	
Three Fundamentals of the Biological Genetic Algorithm	303
<i>Stephen Freeland</i>	
Index	313

Contributing Authors

Wolfgang Banzhaf is Associate Professor of Computer Science at the University of Dortmund, Germany.

Jason M. Daida is an Associate Research Scientist in the Space Physics Research Laboratory at The University of Michigan, Ann Arbor.

Joseph A. Driscoll is an Assistant Professor in the Department of Computer Science at Middle Tennessee State University.

Stephen Freeland is an Assistant Professor of Bioinformatics within the Biology Department of UMBC (University of Maryland, Baltimore County).

David E. Goldberg is the Jerry S. Dobrovolsky Distinguished Professor in Entrepreneurial Engineering and Director of the Illinois Genetic Algorithms Laboratory at the University of Illinois at Urbana-Champaign.

Erik D. Goodman is Professor of Electrical and Computer Engineering and of Mechanical Engineering at Michigan State University.

David Hill is a graduate student in the Department of Civil and Environmental Engineering at the University of Illinois at Urbana-Champaign.

Daniel Howard is a QinetiQ Fellow and Head of the Software Evolution Centre at QinetiQ in Malvern, UK.

Jianjun Hu is a Ph.D student in Computer Science at Michigan State University.

Martin A. Keane is a consultant to the gaming industry and works with Genetic Programming Inc.

Mark Kotanchek is the group leader for Applied Math and Physics within Physical Sciences Research and Development of the Dow Chemical Company, Midland, MI, USA.

Arthur K. Kordon is a Research and Development Leader in the Applied Math and Physics Group within the Physical Sciences Research and Development of the Dow Chemical Company.

John Koza is Consulting Professor at Stanford University in the Biomedical Informatics Program in the Department of Medicine and in the Department of Electrical Engineering.

Duncan MacLean is a founder of and Chief Engineer for Genetics Squared, Inc, a computational discovery company working in the pharmaceutical industry.

Miguel Nicolau is a researcher and part-time PhD student at the University of Limerick, Ireland.

David A. Ostrowski works in the Business Systems Research group at the Ford Research Laboratory. Dr. Ostrowski is also an adjunct faculty member in the Department of Computer Science at the University of Michigan-Dearborn.

Una-May O'Reilly is a research member of the Living Machines and Humanoid Robotics group in the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology.

Robert G. Reynolds is a professor in the Department of Computer Science at Wayne State University and the director of the Artificial Intelligence Laboratory. He is also an Associate Adjunct Research Scientist with the Museum of Anthropology and a faculty member of the Center for the Study of Complex Systems at the University of Michigan-Ann Arbor.

Rick Riolo is Director of the Computer Lab and Associate Research Scientist in the Center for the Study of Complex Systems at the University of Michigan.

Conor Ryan is a senior lecturer and University Fellow at the University of Limerick, Ireland.

Kumara Sastry is a member of Illinois Genetic Algorithms Laboratory and a graduate student of Material Science and Engineering at the University of Illinois at Urbana-Champaign.

Kisung Seo is Visiting Assistant Professor of ECE at Michigan State University and also Assistant Professor of EE at Seokyoung University in Seoul.

Guido Smits is a Research and Development Leader in the Applied Math and Physics group within Physical Sciences Research and Development of Dow Benelux, Terneuzen, NV.

Terence Soule is a Professor at the University of Idaho, where is a member of the Computer Science Department and of the Bioinformatics and Computational Biology Program.

Lee Spector is Dean of the School of Cognitive Science and Associate Professor of Computer Science at Hampshire College in Amherst, Massachusetts.

Matthew Streeter is a Systems and Research Programmer at Genetic Programming, Inc. in Mountain View, California.

Dave Wilkinson is a geophysicist in the Seismic Analysis and Reservoir Property Estimation Team at ChevronTexaco Exploration and Production Technology Company.

Bill Worzel is the President and co-founder of Genetics Squared, Inc, a computational discovery company working in the pharmaceutical industry.

Deyi Xie is a geologist in the Seismic Analysis and Reservoir Property Estimation Team at ChevronTexaco Exploration and Production Technology Company.

Tina Yu is a computer scientist in the Mathematical Modeling Team at Chevron-Texaco Information Technology Company.

Anjun Zhou is a principal of the Advanced Research Center at State Street Global Advisors.

Preface

The work described in the chapters in this book was first presented at a workshop titled “Genetic Program, Theory and Practice,” organized by the Center for the Study of Complex Systems at the University of Michigan, Ann Arbor, 15-17 May 2003. The goal of this workshop was to start an exchange of research results and ideas between those who focus on Genetic Programming (GP) theory and those who focus on the application of GP to various real-world problems. In order to facilitate these interactions, the number of talks and participants was small, and the time for discussion was large. Further, participants were asked to review each other chapters *before* the workshop, with each chapter reviewed by both a “theory” and an “applications” researcher. Those reviewer comments, as well as discussion at the workshop, are reflected in the chapters presented in this book. Additional information about the Workshop, addendum’s to chapters, and a site for continuing discussions by participants and by others can be found at <http://cscs.umich.edu:8000/GPTP-2003/>.

We thank all the Workshop participants for making the Workshop an exciting and productive three days. In particular we thank all the authors, without whose hard work and creative talents Workshops and books like this one would not be possible. We also thank the keynote speakers: Dr. John Holland, Dr. Stephen Freeland, Ms. Lynne Ellyn and Mr. Tom Chase who delivered three thought provoking speeches that inspired a great deal of discussion among the participants.

The Workshop received support from several sources, including:

- The Center for the Study of Complex Systems (CSCS);
- State Street Global Advisors, Boston, MA;
- Christopher T. May, RedQueen Capital Management; and
- DTE Energy Foundation, Michigan.

We thank all of the contributors for their kind and generous support for the Workshop and for GP research in general.

A number of people made key contributions to running the Workshop and assisting the attendees while they were in Ann Arbor. Foremost among them was

Howard Oishi, assisted by Jesse Buck and Mike Charter. We thank them all for running yet another very successful workshop. We also thank those who helped with reading and copy-editing chapters, including Jim Odell, Bill Rand, Bill Tozier, and Eric Wollesen. Jesse Buck also put in many hard hours working on assembling this book. Melissa Fearon's editorial efforts were invaluable, from the initial plans for the book through its final publication. Thanks also to Deborah Doherty of Kluwer for helping with various technical publishing issues. Finally, we thank Carl Simon, Director of CSCS, for his support for this endeavor from its very inception, and to James Odell for serving on the Organizing Committee and helping out in various ways throughout.

RICK RIOLO AND BILL WORZEL

Foreword

I was very happy and excited when I was invited to participate to the workshop on Genetic Programming (GP) Theory and Practice organised by the Center for the Study of Complex Systems (CSCS) held at the University of Michigan, Ann Arbor, in May 2003. The focus of this workshop was explicitly on how theory and practice should interact and what they can teach each other: an important and timely topic, particularly in the light of the many, but still very recent and not widely understood, developments seen in GP theory the last few years.

The participants' list was impressive, including many of the most influential and respected researchers in genetic programming and evolutionary computation. I was really looking forward to attend this workshop and learn and interact with them all, when "disaster" struck: I discovered that our new baby would be born only days before the workshop, which made it impossible for me to leave my family and attend the event. To punish me for my desertion the organisers of the event, Rick Riolo and Bill Worzel, very kindly asked me whether I would write a foreword for this volume gathering the proceedings of the workshop — an invitation I was very glad to accept.

I am very impressed by the quality of the papers presented at the workshop and now printed in this volume, both in terms of the novelty and importance of theoretical results and the variety and credibility of real-world applications of genetic programming. In terms of both contributions to science and inspiration for future practical applications this workshop appears to have been a great success. Knowing the depth and breadth of many of the participants I have no doubts that the discussion of all contributions has been lively and enormously productive.

Many events take place regularly in evolutionary computation. However, very few are specifically devoted to genetic programming, and definitely none focuses on the "theory meets practice" idea behind this workshop, which is an urgent topic, not just for genetic programming, but for the whole field of genetic and evolutionary computation. I hope this event will be the first of a series, where theoreticians and practitioners will be able to exchange and generate new ideas, look back critically at what has been achieved and develop

a vision of the things to come, thereby guiding and inspiring future research in genetic programming.

I am sure I will learn a lot from studying this volume in detail, and I am sure many others, practitioners and theoreticians alike, can gain much from doing the same — Thank you Rick and Bill for having put together such a high profile and inspiring event.

Professor Ricardo Poli
Department of Computer Science
University of Essex, May 1999