

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Developing and Evaluating
Incremental Evolution using
High Quality Performance Measures
for Genetic Programming

A thesis presented in partial fulfillment of the requirements for the
degree of Doctor of Philosophy in Computer Science
at Massey University, Albany, Auckland,
New Zealand

Matthew Garry William Walker
2007

To Mum and Dad

Abstract

This thesis is divided into two parts. The first part considers and develops some of the statistics used in genetic programming (GP) while the second uses those statistics to study and develop a form of incremental evolution and an early termination heuristic for GP.

The first part looks in detail at success proportion, Koza's minimum computational effort, and a measure we rename "success effort". We describe and develop methods to produce confidence intervals for these measures as well as confidence intervals for the difference and ratio of these measures.

The second part studies Jackson's fitness-based incremental evolution. If the number of fitness evaluations are considered (rather than the number of generations) then we find some potential benefit through reduction in the effort required to find a solution. We then automate the incremental evolution method and show a statistically significant improvement compared to GP with automatically defined functions (ADFs).

The success effort measure is shown to have the critical advantage over Koza's measure as it has the ability to include a decreasing cost of failure. We capitalise on this advantage by demonstrating an early termination heuristic that again offers a statistically significant advantage.

Preface

Acknowledgements

First, I'd like to thank Mum and Dad. Without their support and—equally importantly—encouragement, I wouldn't be here at the end of this experience. Thank you for your enthusiasm, for putting up with draft chapters being read to you, and for listening to ideas when mere mortals would have fallen asleep.

Chris Messom, my supervisor, has been great. My PhD-studying friends have had some horror stories to tell regarding their supervisors, but Chris was fantastic. He organised funding (thank you to the Government for their “Top Doctoral” programme), was ridiculously patient, always supportive and thorough and motivational when things weren't working. Another of Chris's wonders is his ability to respond. The university recommends forming an agreement with supervisors to return work within a specific time. They recommended a turn-around time of two weeks. I think Chris might have averaged two hours. Thanks Chris.

Howard Edwards has been a huge resource for the statistics in this thesis. He provided excellent assistance and discussion. He also provided great motivation as it was fantastic to work with someone who was excited by the ideas. Although all the work presented in this thesis is my own, it is Howard and Chris and I who are the “we” in this thesis.

Hopefully we can convince Bruce Mills to return from the United Arab Emirates. Bruce provided the most entertaining conversations (that could easily eliminate an afternoon) introducing me to evolutionary ideas that I still think about. But possibly most importantly to me was that it was his suggestion that gave me a concrete starting point. The ideas transitioned unrecognisably but finally they formed this thesis. His influence was an excellent catalyst just when I needed it.

Christian Gagné produced Open BEAGLE, GP software that I have con-

tributed to and have used almost exclusively throughout the thesis. He provided excellent practical discussions surrounding GP and was very helpful when I requested large datasets of GP experiments.

Steffen Christensen was also of considerable assistance when he provided his absolutely enormous artificial ant datasets.

I'd also like to say a very big thank you to my little sister and my friends. You provided excellent conversations and emails, or lovely distracting phone calls and good vibes, or excellent dinners, or just a break away from this work. You also applied what was highly necessary pressure to get this work done. Thanks.

Finally, I'd like to thank the anonymous reviewers of papers I submitted, especially the reviewer who voted as best paper the work that turned into chapter 3.

Publications

Parts of this thesis have already been published. The initial work surrounding the Wilson-Dependent method in chapter 3 was published at *EuroGP 2007* [114]. The discussion of its reliability (section 3.2) was published at *GECCO* a few months later [115], as was a summary of the confidence interval method for the success effort statistic (chapter 4) [117]. The more detailed analysis of success effort that appears in chapter 4 along with parts of the literature review in chapter 2 and the comparisons that appear in chapter 5 were later published at *CEC* [116].

Table of Contents

Abstract	v
Preface	vii
1 Introduction	1
1.1 Genetic Programming	1
1.2 Performance Measurements	2
1.3 Performance Comparisons	3
1.4 Incremental Evolution and Cost of Failure	3
1.5 Simple Problems	4
1.6 Objectives and Motivation	5
1.7 How to Read this Thesis	6
I High Quality Performance Measures	7
2 Review: Performance Comparison Measures	9
2.1 The Importance of Statistics	9
2.1.1 Comparison of Two Results	10
2.2 Success Proportion	11
2.2.1 Dependence	11
2.2.2 Confidence Intervals	12
2.2.3 Confidence Intervals for Differences	14
2.2.4 Confidence Intervals for Ratios	14
2.2.5 Effective Success Probability	14
2.3 Minimum Computational Effort	17
2.3.1 Koza's Original Definition	17
2.3.2 Dropping the Ceiling Operator	18
2.3.3 Underestimating the True Computational Effort	18

2.3.4	Influence of Probability of Success	19
2.3.5	Number of Runs	20
2.3.6	Confidence Intervals	20
2.4	Mean Fitness	20
2.4.1	Confidence Intervals	21
2.4.2	Variations	22
2.5	Mean Generation	22
2.5.1	Confidence Intervals	22
2.6	Average Evaluations to Success	23
2.7	The Y-Test	23
2.8	Success Effort	24
2.9	Summary and Further Reading	25
3	Minimum Computational Effort	27
3.1	Defining Confidence Intervals	27
3.1.1	Normal Approximation Method	27
3.1.2	Wilson-Dependent Method	28
3.1.3	Resampling Statistics Method	29
3.1.4	When Minimum Generation is Known	29
3.1.5	When Minimum Generation is Unknown	32
3.1.6	Further Analysis of the Wilson-Dependent Method	35
3.1.7	Summary	36
3.2	Reliability of Confidence Intervals	37
3.2.1	Dependence Issues	37
3.2.2	Varying Alpha Values	39
3.2.3	Large Minimum Generations	43
3.2.4	More Large Datasets	50
3.2.5	Summary	52
3.3	Computational Effort near 100% Success	52
3.3.1	Coverage with True Success Rate Near 100%	53
3.3.2	Coverage with Observed Success Greater than z	58
3.4	Comparing Two Minimum Computational Efforts	61
3.4.1	Simulating Minimum Computational Effort	61
3.4.2	Minimum Computational Effort Differences	62
3.4.3	Minimum Computational Effort Ratios	62
3.5	Computational Effort versus Success Proportion	65
3.5.1	Non-Crossing Success Proportions	66

3.5.2	Crossing Success Proportions	68
3.5.3	Summary	70
4	Success Effort	73
4.1	Confidence Intervals	73
4.1.1	Coverage	76
4.1.2	Conclusions	78
4.2	Comparing Two Success Efforts	78
4.3	Success Effort versus Success Proportion	79
4.3.1	Non-Crossing Success Proportions	81
4.3.2	Crossing Success Proportions	82
4.3.3	Summary	84
5	Comparison of the Statistics	87
5.1	Coverage	87
5.2	Power	91
5.3	A Better Statistic?	93
5.4	Summary	94
II	Developing Incremental Evolution	97
6	Review: Incremental Evolution	99
6.1	Introduction	99
6.2	Detail of Incremental Evolution	100
6.2.1	Terminology	100
6.2.2	Functional versus Environmental	101
6.2.3	Transferring the Population Between Stages	101
6.2.4	Hierarchies, Directions and Concurrency	101
6.2.5	Diversity and Overtraining	102
6.2.6	Manual Incremental Evolution	103
6.2.7	Automatic Incremental Evolution	104
6.2.8	Success Rates	104
6.2.9	Naysayers	104
6.3	Related Techniques	105
6.3.1	Layered Learning	105
6.3.2	Shaping	107
6.3.3	Seeding	107

6.3.4	Co-evolution	108
6.4	Incremental versus Direct Evolution	108
6.4.1	Success Proportion	109
6.4.2	Quality of a Solution	109
6.4.3	Cost of Finding a Solution	113
6.5	Summary	114
7	Direct Evolution	115
7.1	Even- n -Parity	115
7.2	Method	116
7.3	Results without ADFs	117
7.4	Results with ADFs	118
7.5	Efficiency Ratios	121
7.6	Results to 50 Generations	126
8	Manual Fitness-Based Incremental Evolution	129
8.1	Related Work	129
8.1.1	Saturation	130
8.1.2	Parameterless Functions	133
8.1.3	Incremental Evolution with Simplified Problems	135
8.1.4	Summary	136
8.2	Method	136
8.3	Results: Unit-Cost Generations	139
8.3.1	Addition of ADFs	139
8.3.2	Allocation of Generations	142
8.3.3	Fitness Cases in Stage One	147
8.3.4	Problem Difficulty	150
8.3.5	Comparison to Direct Evolution	153
8.3.6	Summary	153
8.4	Defending Adjusted Generations	154
8.4.1	Related Work	155
8.4.2	Calculating Adjusted Generations	156
8.5	Results: Adjusted Generations	157
8.5.1	Addition of ADFs	157
8.5.2	Allocation of Generations	157
8.5.3	Fitness Cases in Stage One	161
8.5.4	Problem Difficulty	162
8.5.5	Comparison to Direct Evolution	164

8.5.6	Summary	164
8.6	Future Research	165
9	Automatic Fitness-Based Incremental Evolution	167
9.1	Motivation	167
9.2	Related Work	168
9.3	The Algorithms	170
9.4	Method	175
9.5	Results: Unit-Cost Generations	176
9.5.1	Aggression	176
9.5.2	Mutation	178
9.5.3	Generations Before Step	178
9.5.4	Problem Difficulty	178
9.5.5	Run Length	179
9.5.6	Summary	179
9.6	Results: Adjusted Generations	179
9.6.1	Aggression	180
9.6.2	Mutation	180
9.6.3	Generations Before Step	181
9.6.4	Problem Difficulty	182
9.6.5	Run Length	182
9.6.6	Summary	183
9.7	Analysis of the Three Measures	183
9.8	Summary	183
9.9	Future Research	184
10	Early Termination	187
10.1	Motivation	187
10.2	Defending the Early Termination Heuristic	188
10.2.1	Four Datasets	188
10.2.2	Koza's Books	190
10.3	The Heuristic	191
10.3.1	Related Work	191
10.4	Direct Evolution Experiments	193
10.5	Automatic Incremental Evolution Experiments	194
10.5.1	Results: Unit-Cost Generations with Early Termination . .	195
10.5.2	Results: Adjusted Generations with Early Termination . .	197
10.6	Summary	199

10.7 Future Research	199
III Finale	201
11 Conclusions, Contributions, Limitations, Future Research	203
11.1 Conclusions	203
11.2 Contributions	204
11.3 Limitations	204
11.4 Future Research	205
IV Appendices	207
A Success Proportion versus Fitness	209
A.1 Statistical Concerns	209
A.1.1 A Point Statistic	209
A.1.2 Statistical Independence	210
A.1.3 Koza's Minimum Computational Effort	211
A.2 Empirical Evidence	211
A.2.1 Re-Using the Incremental Evolution Experiments	211
A.2.2 Re-Analysing their Results	215
A.2.3 Re-Considering Symbolic Regression	215
A.2.4 Rationalising Negative Correlations	220
A.3 Philosophical Motivation	220
A.3.1 Pro Fitness	221
A.3.2 Anti Fitness	221
A.3.3 Future Research: A Compromise	222
A.4 Summary	223
B Proof of Equation 4.2	225
C Proof of Equation 4.4	227
D Electronic Appendix	231

List of Tables

2.1	Success proportion effort confidence intervals: simulation method	13
2.2	Success prop. confidence intervals: difference of two	15
2.3	Success prop. confidence intervals: ratio of two	16
3.1	Comp. effort confidence intervals: resampling method	29
3.2	Comp. effort coverage: by runs, known min. gen.	31
3.3	Comp. effort coverage: by problem, known min. gen.	31
3.4	Comp. effort confidence intervals: Wilson-Dependent method . . .	33
3.5	Comp. effort coverage: by runs, estimated min. gen.	34
3.6	Comp. effort coverage: by problem, estimated min. gen.	34
3.7	Comp. effort confidence intervals: Resampling-Independent method	38
3.8	Comp. effort coverage: by target coverage	41
3.9	Comp. effort width: by target coverage	42
3.10	Comp. effort coverage: log-normal model	46
3.11	Comp. effort coverage: normal model	47
3.12	Comp. effort coverage: rectangle model	49
3.13	Comp. effort coverage: right-triangle model	49
3.14	Comp. effort coverage: left-triangle model	50
3.15	Comp. effort coverage: semi-ellipse model	50
3.16	Comp. effort coverage: extra ant datasets	51
3.17	Comp. effort: random number, likelihood distribution	62
3.18	Comp. effort confidence intervals: difference of two	63
3.19	Comp. effort confidence intervals: ratio of two	64
3.20	Comp. effort, success prop. comparison: specifications	68
3.21	Comp. effort, success prop. comparison: performance	70
4.1	Success effort confidence intervals: resampling method	74
4.2	Success effort confidence intervals: simulation method	75
4.3	Success effort coverage: by problem	77

4.4	Success effort coverage: by runs	77
4.5	Success effort: random numbers, likelihood distribution	79
4.6	Success effort confidence intervals: difference of two	80
4.7	Success effort confidence intervals: ratio of two	80
4.8	Success effort, success prop. comparison: specifications	82
4.9	Success effort, success prop. comparison: performance	84
5.1	Statistics' best estimate of true value: by problem	88
5.2	Statistics' coverage: by problem	89
5.3	Statistics' coverage: by runs	89
5.4	Statistics' widths: by problem	92
5.5	Statistics' widths: by runs	92
7.1	ADF specification	116
7.2	Direct evolution statistics: by problem size, without ADFs	119
7.3	Direct evolution statistics: by problem size, with ADFs	122
7.4	Direct evolution stat's, 50 gen's: by problem size, with ADFs	126
7.5	Direct evolution stat's, 50 gen's: by problem size, without ADFs	127
8.1	Confidence intervals: efficiency ratio of two efficiency ratios	141
8.2	Efficiency ratios: even-5, 8 versus 16 fitness cases	149
8.3	Efficiency ratios: even-6, by fitness cases	149
8.4	Efficiency ratios: even-7, by fitness cases	150
9.1	Automatic inc. evol.: aggressive strategy	171
9.2	Automatic inc. evol: less-aggressive strategy	172
9.3	Efficiency ratios: aggression compared to direct evolution	177
9.4	Efficiency ratios: aggression	177
10.1	Successful learning curves compared to a straight line	190
10.2	Statistics comparing early termination to standard	194
A.1	Success prop. versus fitness: correlation, manual inc. evol.	212
A.2	Success prop. versus fitness: correlation, auto inc. evol.	213
A.3	Success prop. versus fitness: correlation, Luke and Panait	215

List of Figures

3.1	Comp. effort confidence intervals: widths	35
3.2	Comp. effort coverage: three methods	40
3.3	Comp. effort coverage: target coverage	42
3.4	Log-normal model of Ant dataset	44
3.5	The four arbitrarily-selected distributions	48
3.6	Success factor for normal and log-normal experiments	55
3.7	Coverage by success factor for normal and log-normal	56
3.8	Coverage by success factor and run size for normal and log-normal	57
3.9	Coverage and likelihood by success factor for normal	59
3.10	Coverage and likelihood by success factor for log-normal	60
3.11	Comp. effort: overlapping confidence intervals	65
3.12	Comp. effort given success prop. domination	66
3.13	Comp. effort given crossing success proportions	69
4.1	Success effort given crossing success proportions	83
5.1	Mean best fitness zero-width confidence intervals	90
7.1	Direct evolution performance curves: even-4 without ADFs	118
7.2	Direct evolution performance curves: even-5 without ADFs	119
7.3	Direct evolution performance curves: even-6 without ADFs	120
7.4	Direct evolution performance curves: even-4 with ADFs	121
7.5	Direct evolution performance curves: even-5 with ADFs	122
7.6	Direct evolution performance curves: even-6 with ADFs	123
7.7	Direct evolution performance curves: even-7 with ADFs	124
7.8	Efficiency ratios of ADFs	125
8.1	Jackson's results with confidence intervals	131
8.2	Comp. effort, success prop. curves for even-4, 8 fitness cases	133
8.3	Comp. effort, success prop. curves for even-4, 4 fitness cases	138

8.4	Typical performance under unit-cost generations	143
8.5	Potential problematic convergence during stage one	145
8.6	Genetic convergence unlikely to close route “through failure” . . .	148
8.7	Efficiency ratios: by problem difficulty	152
8.8	Performance: comp. effort versus success effort	158
8.9	Reduced cost of failure using adjusted generations	159
8.10	Efficiency ratios: by problem difficulty	163
9.1	Example run: aggressive	173
9.2	Example run: less-aggressive	174
9.3	Efficiency ratios: by gen’s before automatic step	181
9.4	Efficiency ratios: by problem difficulty	182
10.1	Straight lines used to analyse performance of successful runs . . .	189
10.2	Fitness cases correct: auto inc. evol. versus standard	196
A.1	Success prop. vs. fitness: correlation, manual inc. evol.	214
A.2	Success prop. vs fitness: correlation, Luke & Panit’s symb. reg. . .	216
A.3	Success prop. vs. fitness: correlation, Luke & Panait’s multiplexor	217
A.4	Success prop. vs. fitness: correlation, Luke & Panait’s ant	218
A.5	Success prop. vs. fitness: correlation, symbolic regression	219