

University of Groningen

Genetic algorithms in data analysis

Lankhorst, Marc Martijn

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

1996

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Lankhorst, M. M. (1996). Genetic algorithms in data analysis Groningen: s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Genetic Algorithms in Data Analysis

Marc M. Lankhorst

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Lankhorst, Marc M.

Genetic algorithms in data analysis /

Marc M. Lankhorst. – [S.l. : s.n.]. – Ill.

Thesis Rijksuniversiteit Groningen. – With index, ref.

With summary in Dutch.

ISBN 90-367-0559-2

Subject headings: genetic algorithms, data analysis, optimization.

Printed by: Universiteitsdrukkerij Groningen.

RIJKSUNIVERSITEIT GRONINGEN

Genetic Algorithms in Data Analysis

Proefschrift

ter verkrijging van het doctoraat in de
Wiskunde en Natuurwetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus Dr. F. van der Woude
in het openbaar te verdedigen op
vrijdag 12 januari 1996
des namiddags te 4.00 uur

door

Marc Martijn Lankhorst

geboren op 19 december 1968
te Naaldwijk

Promotor: prof. dr. N. Petkov

Acknowledgments

I would like to express my thanks to all people who have contributed, directly or indirectly, to the realization of this thesis.

First of all, I want to thank my promotor, Nikolay Petkov, for letting me ‘do my own thing’ and supporting me although I did not exactly choose the direction he had in mind for me. Furthermore, I thank the members of the supervisory committee, W.A. Halang, J.N. Kok, and A. Nijholt, for carefully reading the manuscript and improving it with their comments.

Many of the computations in the course of this research were carried out on the Connection Machine CM-5 of the University of Groningen, the investments in which were partly supported by the Netherlands Computer Science Research Foundation (SION) and the Netherlands Organization for Scientific Research (NWO). NWO grants also facilitated a number of conference visits.

I would like to thank my colleagues at the Department for the pleasant working environment. Special thanks go to my roommates Ronald Tol (“Hmmm!”) and Marten van der Laan (“Een nulmodemkabel is vies!”) for putting up with me and making coffee. I am also very grateful to Marten for his contribution to the work described in Chapter 6, and for his lessons in sailing. Furthermore, I want to thank Rix Groenboom (“Schoenen van de baas!”) and Rudi van Drunen (“We have the technology!”) for their company on various trips abroad and for their friendship in general.

Finally, I am greatly indebted to my parents for their continuing support and encouragement even if they had no idea what I was doing.

Contents

Acknowledgments	v
1 Introduction	1
1.1 Data and Structure	1
1.2 Evolutionary Algorithms	2
1.3 Objectives	2
1.4 Overview	3
2 Genetic Algorithms	5
2.1 Traditional Search and Optimization Methods	6
2.1.1 Problem Solving = Optimization	6
2.1.2 Different Types of Optimization Methods	6
2.1.3 Differences with Genetic Algorithms	8
2.2 The Standard Genetic Algorithm	9
2.2.1 Selection	9
2.2.2 Crossover	11
2.2.3 Mutation	11
2.2.4 Convergence	12
2.2.5 Termination	13
2.2.6 An Example	13
2.3 Mathematical Foundations	13
2.3.1 Schemata	14
2.3.2 The Schema Theorem	15
2.3.3 Counting Schemata	18
2.3.4 Deception	20
2.4 Improving the Standard Genetic Algorithm	21
2.4.1 Other Selection Algorithms	21
2.4.2 Other Binary Crossover Operators	23
2.4.3 Non-Binary Encodings and Operators	24
2.5 Parallel Genetic Algorithms	26
2.5.1 Parallel Genetic Algorithm Models	26
2.5.2 A Centralized, Message-Passing Genetic Algorithm	28
2.5.3 Data-Parallel Genetic Algorithms	29
2.5.4 A Distributed, Message-Passing Genetic Algorithm	29
2.6 Other Evolutionary Algorithms	31
2.6.1 Evolution Strategies	31
2.6.2 Evolutionary Programming	32
2.6.3 Genetic Programming	32
3 Word Categorization	35
3.1 Related Work	35

3.2	Information Theory	37
3.3	A Genetic Algorithm for Word Categorization	38
3.3.1	Representation	38
3.3.2	Genetic Operators	39
3.3.3	Fitness Evaluation	40
3.4	Experiments	40
3.4.1	Test Data	40
3.4.2	Binary versus Integer Representation	40
3.4.3	Two-Point versus Uniform Crossover	41
3.4.4	Results	42
3.5	Discussion	43
4	Grammatical Inference I: Pushdown Automata	45
4.1	Formal Language Theory	45
4.1.1	Basic Definitions	46
4.1.2	Finite Automata	46
4.1.3	Pushdown Automata	47
4.1.4	Grammars	49
4.1.5	The Chomsky Hierarchy of Languages	50
4.2	Grammatical Inference	51
4.3	Related Work	52
4.4	A Genetic Algorithm for the Induction of Automata	54
4.4.1	Type of Pushdown Automaton	54
4.4.2	Representation	54
4.4.3	Genetic Operators	55
4.4.4	Fitness Evaluation	55
4.5	Experiments	57
4.5.1	Test Data	57
4.5.2	Results	59
4.6	Discussion	61
5	Grammatical Inference II: Context-Free Grammars	63
5.1	A Genetic Algorithm for the Induction of Grammars	63
5.1.1	Representation	63
5.1.2	Genetic Operators	64
5.1.3	Fitness Evaluation	65
5.2	Experiments	66
5.2.1	Test Data	66
5.2.2	Results	66
5.2.3	Comparison with Results on Pushdown Automata	68
5.3	Discussion	70
6	Signal Decomposition	73
6.1	Time-Frequency Decompositions	73
6.1.1	Short-Time Fourier Transform	74
6.1.2	Wavelet Transform	75

6.1.3	Wave Packets	76
6.1.4	Matching Pursuit Algorithm	76
6.2	Related Work	77
6.3	A Genetic Approach to Signal Decomposition	78
6.3.1	Choice of the Elementary Function	78
6.3.2	Representation	80
6.3.3	Genetic Operators	81
6.3.4	Fitness Evaluation	82
6.4	Experiments	82
6.4.1	Motivation	82
6.4.2	Test Data	83
6.4.3	Results Regarding Representation	84
6.4.4	Approximation Results	88
6.4.5	Results of the Matching Pursuit Algorithm	91
6.5	Discussion	91
7	Iterated Function Systems Optimization	93
7.1	Iterated Function Systems	93
7.1.1	IFS Approximation of 1D Signals	94
7.1.2	IFS Approximation of 2D Images	95
7.2	Related Work	96
7.3	A Genetic Algorithm for Fractal Approximation	96
7.3.1	Representation	96
7.3.2	Genetic Operators	98
7.3.3	Fitness Evaluation	99
7.4	Experiments	99
7.4.1	Test Data	99
7.4.2	Experiments Regarding Parameters	102
7.4.3	Results on 1D Signals	105
7.4.4	Results on 2D Images	107
7.5	Discussion	109
8	Fuzzy Systems Inference	111
8.1	Fuzzy Logic	111
8.1.1	Fuzzy Subsets	111
8.1.2	Logic Operations	112
8.1.3	Fuzzy Rules	113
8.1.4	Constructing a Fuzzy System	114
8.2	Related Work	114
8.3	A Genetic Algorithm for Fuzzy Systems Discovery	115
8.3.1	Representation	115
8.3.2	Genetic Operators	118
8.3.3	Fitness Evaluation	119
8.4	Experiments	119
8.4.1	Test Data	119

8.4.2	Results	120
8.5	Discussion	123
9	Concluding Remarks	125
9.1	Summary and Conclusions	125
9.2	Research Contributions of this Thesis	127
9.3	Further Research	128
	Bibliography	129
	Index	137
	Samenvatting	141