

Contents

1	Introduction	1
1.1	The Problem and Its Various Formulations	1
1.1.1	Mathematical Versus Computer Science Perspectives	1
1.1.2	Different Kinds of Output	2
1.1.3	Representing Languages	3
1.1.4	Complexity Issues	5
1.1.5	Summary	6
1.2	Assessing Algorithms' Performance	7
1.2.1	Measuring Classifier Performance	7
1.2.2	McNemar's Test	8
1.2.3	5×2 Cross-Validated Paired t Test	9
1.3	Exemplary Applications	9
1.3.1	Peg Solitaire	10
1.3.2	Classification of Proteins	12
1.4	Bibliographical Background	15
	References	16
2	State Merging Algorithms	19
2.1	Preliminaries	19
2.2	Evidence Driven State Merging	21
2.3	Gold's Idea	23
2.4	Grammatical Inference with MDL Principle	27
2.4.1	The Motivation and Appropriate Measures	28
2.4.2	The Proposed Algorithm	28
2.5	Bibliographical Background	30
	References	31
3	Partition-Based Algorithms	33
3.1	Preliminaries	33
3.2	The k -tails Method	36
3.3	Grammatical Inference by Genetic Search	37
3.3.1	What Are Genetic Algorithms?	37

3.3.2	Basic Notions of the Genetic Algorithm for GI	37
3.3.3	Our Implementation	39
3.4	CFG Inference Using Tabular Representations	40
3.4.1	Basic Definitions	41
3.4.2	The Algorithm	41
3.4.3	Our Implementation	43
3.5	Bibliographical Background	45
	References	45
4	Substring-Based Algorithms	47
4.1	Error-Correcting Grammatical Inference	47
4.1.1	The GI Algorithm	47
4.1.2	Our Implementation	49
4.2	Alignment-Based Learning	50
4.2.1	Alignment Learning	51
4.2.2	Selection Learning	54
4.2.3	Our Implementation	54
4.3	Bibliographical Background	55
	References	56
5	Identification Using Mathematical Modeling	57
5.1	From DFA Identification to Graph Coloring	57
5.1.1	Encoding	57
5.1.2	Our Implementation	58
5.2	From NFA Identification to a Satisfiability Problem	61
5.2.1	Encoding	61
5.2.2	Our Implementation	62
5.3	From CFG Identification to a CSP	64
5.3.1	Encoding	64
5.3.2	Our Implementation	65
5.4	Bibliographical Background	67
	References	67
6	A Decomposition-Based Algorithm	69
6.1	Prime and Decomposable Languages	69
6.1.1	Preliminaries	69
6.1.2	Cliques and Decompositions	70
6.2	CFG Inference	71
6.2.1	The GI Algorithm	71
6.2.2	Our Implementation	72
6.3	Bibliographical Background	75
	References	75
7	An Algorithm Based on a Directed Acyclic Word Graph	77
7.1	Definitions	77
7.2	Constructing a DAWG From a Sample	78

7.3	Our Implementation	79
7.4	Bibliographical Background	81
	References.	81
8	Applications of GI Methods in Selected Fields	83
8.1	Discovery of Generating Functions.	83
8.1.1	Generating Functions	83
8.1.2	The Schützenberger Methodology.	84
8.1.3	Applications	85
8.2	Minimizing Boolean Functions.	93
8.2.1	Background and Terminology.	94
8.2.2	The Algorithm	96
8.2.3	Our Implementation	97
8.2.4	Examples	98
8.3	Use of Induced Star-Free Regular Expressions.	100
8.3.1	Definitions and an Algorithm	100
8.3.2	An Application in Classification of Amyloidogenic Hexapeptides.	103
8.3.3	An Application in the Construction of Opening Books. . . .	105
8.4	Bibliographical Background	108
	References.	109
	Appendix A: A Quick Introduction to Python	111
	Appendix B: Python’s Tools for Automata, Networks, Genetic Algorithms, and SAT Solving.	129
	Appendix C: OML and its Usage in IronPython	139

Grammatical Inference
Algorithms, Routines and Applications
Wieczorek, W.
2017, XI, 145 p. 18 illus., Hardcover
ISBN: 978-3-319-46800-6