

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	Book Outline	4
	References.	5
<b>2</b>	<b>Decision-Tree Induction</b>	7
2.1	Origins	7
2.2	Basic Concepts.	8
2.3	Top-Down Induction.	9
2.3.1	Selecting Splits.	11
2.3.2	Stopping Criteria	29
2.3.3	Pruning	30
2.3.4	Missing Values.	36
2.4	Other Induction Strategies	37
2.5	Chapter Remarks	40
	References.	40
<b>3</b>	<b>Evolutionary Algorithms and Hyper-Heuristics</b>	47
3.1	Evolutionary Algorithms	47
3.1.1	Individual Representation and Population Initialization.	49
3.1.2	Fitness Function	51
3.1.3	Selection Methods and Genetic Operators	52
3.2	Hyper-Heuristics.	54
3.3	Chapter Remarks	56
	References.	56
<b>4</b>	<b>HEAD-DT: Automatic Design of Decision-Tree Algorithms.</b>	59
4.1	Introduction	60
4.2	Individual Representation.	61
4.2.1	Split Genes	61
4.2.2	Stopping Criteria Genes.	63

4.2.3	Missing Values Genes . . . . .	63
4.2.4	Pruning Genes . . . . .	64
4.2.5	Example of Algorithm Evolved by HEAD-DT . . . . .	66
4.3	Evolution . . . . .	67
4.4	Fitness Evaluation . . . . .	69
4.5	Search Space . . . . .	72
4.6	Related Work . . . . .	73
4.7	Chapter Remarks . . . . .	74
	References . . . . .	75
<b>5</b>	<b>HEAD-DT: Experimental Analysis . . . . .</b>	<b>77</b>
5.1	Evolving Algorithms Tailored to One Specific Data Set . . . . .	78
5.2	Evolving Algorithms from Multiple Data Sets . . . . .	83
5.2.1	The Homogeneous Approach . . . . .	84
5.2.2	The Heterogeneous Approach . . . . .	99
5.2.3	The Case of Meta-Overfitting . . . . .	121
5.3	HEAD-DT's Time Complexity . . . . .	123
5.4	Cost-Effectiveness of Automated Versus Manual Algorithm Design . . . . .	123
5.5	Examples of Automatically-Designed Algorithms . . . . .	125
5.6	Is the Genetic Search Worthwhile? . . . . .	126
5.7	Chapter Remarks . . . . .	127
	References . . . . .	139
<b>6</b>	<b>HEAD-DT: Fitness Function Analysis . . . . .</b>	<b>141</b>
6.1	Performance Measures . . . . .	141
6.1.1	Accuracy . . . . .	142
6.1.2	F-Measure . . . . .	142
6.1.3	Area Under the ROC Curve . . . . .	143
6.1.4	Relative Accuracy Improvement . . . . .	143
6.1.5	Recall . . . . .	144
6.2	Aggregation Schemes . . . . .	144
6.3	Experimental Evaluation . . . . .	145
6.3.1	Results for the Balanced Meta-Training Set . . . . .	146
6.3.2	Results for the Imbalanced Meta-Training Set . . . . .	156
6.3.3	Experiments with the Best-Performing Strategy . . . . .	164
6.4	Chapter Remarks . . . . .	169
	References . . . . .	170
<b>7</b>	<b>Conclusions . . . . .</b>	<b>171</b>
7.1	Limitations . . . . .	172
7.2	Opportunities for Future Work . . . . .	173
7.2.1	Extending HEAD-DT's Genome: New Induction Strategies, Oblique Splits, Regression Problems . . . . .	173

7.2.2	Multi-objective Fitness Function . . . . .	173
7.2.3	Automatic Selection of the Meta-Training Set . . . . .	174
7.2.4	Parameter-Free Evolutionary Search . . . . .	174
7.2.5	Solving the Meta-Overfitting Problem . . . . .	175
7.2.6	Ensemble of Automatically-Designed Algorithms . . . . .	175
7.2.7	Grammar-Based Genetic Programming . . . . .	176
References. . . . .		176

Automatic Design of Decision-Tree Induction Algorithms

Barros, R.C.; de Carvalho, A.C.P.L.F.; Freitas, A.A.

2015, XII, 176 p. 18 illus., Softcover

ISBN: 978-3-319-14230-2