
Contents

1	Introduction	1
1.1	What Is Logical and Relational Learning?	1
1.2	Why Is Logical and Relational Learning Important?	2
1.2.1	Structure Activity Relationship Prediction	3
1.2.2	A Web Mining Example	5
1.2.3	A Language Learning Example	7
1.3	How Does Relational and Logical Learning Work?	8
1.4	A Brief History	11
2	An Introduction to Logic	17
2.1	A Relational Database Example	17
2.2	The Syntax of Clausal Logic	20
2.3	The Semantics of Clausal Logic — Model Theory	22
2.4	Inference with Clausal Logic — Proof Theory	28
2.5	Prolog and SLD-resolution	35
2.6	Historical and Bibliographic Remarks	39
3	An Introduction to Learning and Search	41
3.1	Representing Hypotheses and Instances	41
3.2	Boolean Data	43
3.3	Machine Learning	44
3.4	Data Mining	45
3.5	A Generate-and-Test Algorithm	47
3.6	Structuring the Search Space	48
3.7	Monotonicity	50
3.8	Borders	53
3.9	Refinement Operators	56
3.10	A Generic Algorithm for Mining and Learning	58
3.11	A Complete General-to-Specific Algorithm	59
3.12	A Heuristic General-to-Specific Algorithm	60
3.13	A Branch-and-Bound Algorithm	62

3.14	A Specific-to-General Algorithm	63
3.15	Working with Borders*	64
3.15.1	Computing a Single Border	64
3.15.2	Computing Two Borders	65
3.15.3	Computing Two Borders Incrementally	66
3.15.4	Operations on Borders	68
3.16	Conclusions	69
3.17	Bibliographical Notes	69
4	Representations for Mining and Learning	71
4.1	Representing Data and Hypotheses	71
4.2	Attribute-Value Learning	73
4.3	Multiple-Instance Learning: Dealing With Sets	76
4.4	Relational Learning	79
4.5	Logic Programs	84
4.6	Sequences, Lists, and Grammars	85
4.7	Trees and Terms	87
4.8	Graphs	89
4.9	Background Knowledge	91
4.10	Designing It Yourself	95
4.11	A Hierarchy of Representations*	97
4.11.1	From <i>AV</i> to <i>BL</i>	99
4.11.2	From <i>MI</i> to <i>AV</i>	100
4.11.3	From <i>RL</i> to <i>MI</i>	102
4.11.4	From <i>LP</i> to <i>RL</i>	103
4.12	Propositionalization	106
4.12.1	A Table-Based Approach	106
4.12.2	A Query-Based Approach	108
4.13	Aggregation	109
4.14	Conclusions	112
4.15	Historical and Bibliographical Remarks	113
5	Generality and Logical Entailment	115
5.1	Generality and Logical Entailment Coincide	115
5.2	Propositional Subsumption	118
5.3	Subsumption in Logical Atoms	119
5.3.1	Specialization Operators	121
5.3.2	Generalization Operators*	123
5.3.3	Computing the lgg and the glb	125
5.4	Θ -Subsumption	127
5.4.1	Soundness and Completeness	128
5.4.2	Deciding Θ -Subsumption	128
5.4.3	Equivalence Classes	131
5.5	Variants of Θ -Subsumption*	135
5.5.1	Object Identity*	135

5.5.2	Inverse Implication*	137
5.6	Using Background Knowledge	138
5.6.1	Saturation and Bottom Clauses	139
5.6.2	Relative Least General Generalization*	141
5.6.3	Semantic Refinement*	143
5.7	Aggregation*	145
5.8	Inverse Resolution	147
5.9	A Note on Graphs, Trees, and Sequences	152
5.10	Conclusions	154
5.11	Bibliographic Notes	154
6	The Upgrading Story	157
6.1	Motivation for a Methodology	157
6.2	Methodological Issues	159
6.2.1	Representing the Examples	159
6.2.2	Representing the Hypotheses	160
6.2.3	Adapting the Algorithm	161
6.2.4	Adding Features	161
6.3	Case Study 1: Rule Learning and FOIL	161
6.3.1	FOIL's Problem Setting	162
6.3.2	FOIL's Algorithm	164
6.4	Case Study 2: Decision Tree Learning and TILDE	168
6.4.1	The Problem Setting	168
6.4.2	Inducing Logical Decision Trees	172
6.5	Case Study 3: Frequent Item-Set Mining and WARMR	174
6.5.1	Relational Association Rules and Local Patterns	174
6.5.2	Computing Frequent Queries	177
6.6	Language Bias	179
6.6.1	Syntactic Bias	180
6.6.2	Semantic Bias	183
6.7	Conclusions	184
6.8	Bibliographic Notes	184
7	Inducing Theories	187
7.1	Introduction to Theory Revision	188
7.1.1	Theories and Model Inference	188
7.1.2	Theory Revision	190
7.1.3	Overview of the Rest of This Chapter	192
7.2	Towards Abductive Logic Programming	193
7.2.1	Abduction	193
7.2.2	Integrity Constraints	194
7.2.3	Abductive Logic Programming	196
7.3	Shapiro's Theory Revision System	199
7.3.1	Interaction	199
7.3.2	The Model Inference System	203

7.4	Two Propositional Theory Revision Systems*	208
7.4.1	Learning a Propositional Horn Theory Efficiently	208
7.4.2	Heuristic Search in Theory Revision	212
7.5	Inducing Constraints	213
7.5.1	Problem Specification	214
7.5.2	An Algorithm for Inducing Integrity Constraints	215
7.6	Conclusions	220
7.7	Bibliographic Notes	220
8	Probabilistic Logic Learning	223
8.1	Probability Theory Review	224
8.2	Probabilistic Logics	225
8.2.1	Probabilities on Interpretations	226
8.2.2	Probabilities on Proofs	232
8.3	Probabilistic Learning	238
8.3.1	Parameter Estimation	238
8.3.2	Structure Learning	246
8.4	First-Order Probabilistic Logics	247
8.4.1	Probabilistic Interpretations	248
8.4.2	Probabilistic Proofs	255
8.5	Probabilistic Logic Learning	267
8.5.1	Learning from Interpretations	267
8.5.2	Learning from Entailment	270
8.5.3	Learning from Proof Trees and Traces	271
8.6	Relational Reinforcement Learning*	274
8.6.1	Markov Decision Processes	274
8.6.2	Solving Markov Decision Processes	277
8.6.3	Relational Markov Decision Processes	280
8.6.4	Solving Relational Markov Decision Processes	282
8.7	Conclusions	287
8.8	Bibliographic Notes	287
9	Kernels and Distances for Structured Data	289
9.1	A Simple Kernel and Distance	289
9.2	Kernel Methods	291
9.2.1	The Max Margin Approach	291
9.2.2	Support Vector Machines	292
9.2.3	The Kernel Trick	294
9.3	Distance-Based Learning	296
9.3.1	Distance Functions	296
9.3.2	The k -Nearest Neighbor Algorithm	297
9.3.3	The k -Means Algorithm	297
9.4	Kernels for Structured Data	298
9.4.1	Convolution and Decomposition	299
9.4.2	Vectors and Tuples	299

9.4.3	Sets and Multi-sets	300
9.4.4	Strings	301
9.4.5	Trees and Atoms	302
9.4.6	Graph Kernels*	303
9.5	Distances and Metrics	307
9.5.1	Generalization and Metrics	308
9.5.2	Vectors and Tuples	309
9.5.3	Sets	310
9.5.4	Strings	315
9.5.5	Atoms and Trees	318
9.5.6	Graphs	319
9.6	Relational Kernels and Distances	321
9.7	Conclusions	323
9.8	Bibliographical and Historical Notes	323
10	Computational Aspects of Logical and Relational Learning	325
10.1	Efficiency of Relational Learning	325
10.1.1	Coverage as θ -Subsumption	326
10.1.2	θ -Subsumption Empirically	327
10.1.3	Optimizing the Learner for θ -subsumption	328
10.2	Computational Learning Theory*	333
10.2.1	Notions of Learnability	334
10.2.2	Positive Results	336
10.2.3	Negative Results	338
10.3	Conclusions	342
10.4	Historical and Bibliographic Notes	342
11	Lessons Learned	345
11.1	A Hierarchy of Representations	345
11.2	From Upgrading to Downgrading	346
11.3	Propositionalization and Aggregation	346
11.4	Learning Tasks	347
11.5	Operators and Generality	347
11.6	Unification and Variables	348
11.7	Three Learning Settings	349
11.8	Knowledge and Background Knowledge	350
11.9	Applications	350
	References	351
	Author Index	375
	Index	381