

Contents

Part I Introduction

1	Inductive Databases and Constraint-based Data Mining: Introduction and Overview	3
	Sašo Džeroski	
1.1	Inductive Databases	3
1.2	Constraint-based Data Mining	7
1.3	Types of Constraints	9
1.4	Functions Used in Constraints	12
1.5	KDD Scenarios	14
1.6	A Brief Review of Literature Resources	15
1.7	The IQ (Inductive Queries for Mining Patterns and Models) Project	17
1.8	What's in this Book	22
2	Representing Entities in the OntoDM Data Mining Ontology	27
	Panče Panov, Larisa N. Soldatova, and Sašo Džeroski	
2.1	Introduction	27
2.2	Design Principles for the OntoDM ontology	29
2.3	OntoDM Structure and Implementation	33
2.4	Identification of Data Mining Entities	38
2.5	Representing Data Mining Entities in OntoDM	46
2.6	Related Work	52
2.7	Conclusion	54
3	A Practical Comparative Study Of Data Mining Query Languages . .	59
	Hendrik Blockeel, Toon Calders, Éliisa Fromont, Bart Goethals, Adriana Prado, and Céline Robardet	
3.1	Introduction	60
3.2	Data Mining Tasks	61
3.3	Comparison of Data Mining Query Languages	62
3.4	Summary of the Results	74
3.5	Conclusions	76

4	A Theory of Inductive Query Answering	79
	Luc De Raedt, Manfred Jaeger, Sau Dan Lee, and Heikki Mannila	
4.1	Introduction	80
4.2	Boolean Inductive Queries	81
4.3	Generalized Version Spaces	88
4.4	Query Decomposition	90
4.5	Normal Forms	98
4.6	Conclusions	100
Part II Constraint-based Mining: Selected Techniques		
5	Generalizing Itemset Mining in a Constraint Programming Setting .	107
	Jérémy Besson, Jean-François Boulicaut, Tias Guns, and Siegfried Nijssen	
5.1	Introduction	107
5.2	General Concepts	109
5.3	Specialized Approaches	111
5.4	A Generalized Algorithm	114
5.5	A Dedicated Solver	116
5.6	Using Constraint Programming Systems	120
5.7	Conclusions	124
6	From Local Patterns to Classification Models	127
	Björn Bringmann, Siegfried Nijssen, and Albrecht Zimmermann	
6.1	Introduction	127
6.2	Preliminaries	131
6.3	Correlated Patterns	132
6.4	Finding Pattern Sets	137
6.5	Direct Predictions from Patterns	142
6.6	Integrated Pattern Mining	146
6.7	Conclusions	152
7	Constrained Predictive Clustering	155
	Jan Struyf and Sašo Džeroski	
7.1	Introduction	155
7.2	Predictive Clustering Trees	156
7.3	Constrained Predictive Clustering Trees and Constraint Types	161
7.4	A Search Space of (Predictive) Clustering Trees	165
7.5	Algorithms for Enforcing Constraints	167
7.6	Conclusion	173
8	Finding Segmentations of Sequences	177
	Ella Bingham	
8.1	Introduction	177
8.2	Efficient Algorithms for Segmentation	182
8.3	Dimensionality Reduction	183

8.4	Recurrent Models	185
8.5	Unimodal Segmentation	188
8.6	Rearranging the Input Data Points	189
8.7	Aggregate Segmentation	190
8.8	Evaluating the Quality of a Segmentation: Randomization	191
8.9	Model Selection by BIC and Cross-validation	193
8.10	Bursty Sequences	193
8.11	Conclusion	194
9	Mining Constrained Cross-Graph Cliques in Dynamic Networks . . .	199
	Loïc Cerf, Bao Tran Nhan Nguyen, and Jean-François Boulicaut	
9.1	Introduction	199
9.2	Problem Setting	201
9.3	DATA-PEELER	205
9.4	Extracting δ -Contiguous Closed 3-Sets	208
9.5	Constraining the Enumeration to Extract 3-Cliques	212
9.6	Experimental Results	217
9.7	Related Work	224
9.8	Conclusion	226
10	Probabilistic Inductive Querying Using ProbLog	229
	Luc De Raedt, Angelika Kimmig, Bernd Gutmann, Kristian Kersting, Vítor Santos Costa, and Hannu Toivonen	
10.1	Introduction	229
10.2	ProbLog: Probabilistic Prolog	233
10.3	Probabilistic Inference	234
10.4	Implementation	238
10.5	Probabilistic Explanation Based Learning	243
10.6	Local Pattern Mining	245
10.7	Theory Compression	249
10.8	Parameter Estimation	252
10.9	Application	255
10.10	Related Work in Statistical Relational Learning	258
10.11	Conclusions	259

Part III Inductive Databases: Integration Approaches

11	Inductive Querying with Virtual Mining Views	265
	Hendrik Blockeel, Toon Calders, Éliisa Fromont, Bart Goethals, Adriana Prado, and Céline Robardet	
11.1	Introduction	266
11.2	The Mining Views Framework	267
11.3	An Illustrative Scenario	277
11.4	Conclusions and Future Work	285

12	SINDBAD and SiQL: Overview, Applications and Future Developments	289
	Jörg Wicker, Lothar Richter, and Stefan Kramer	
12.1	Introduction	289
12.2	SiQL	291
12.3	Example Applications	296
12.4	A Web Service Interface for SINDBAD	303
12.5	Future Developments	305
12.6	Conclusion	307
13	Patterns on Queries	311
	Arno Siebes and Diyah Puspitaningrum	
13.1	Introduction	311
13.2	Preliminaries	313
13.3	Frequent Item Set Mining	319
13.4	Transforming KRIMP	323
13.5	Comparing the two Approaches	331
13.6	Conclusions and Prospects for Further Research	333
14	Experiment Databases	335
	Joaquin Vanschoren and Hendrik Blockeel	
14.1	Introduction	336
14.2	Motivation	337
14.3	Related Work	341
14.4	A Pilot Experiment Database	343
14.5	Learning from the Past	350
14.6	Conclusions	358
Part IV Applications		
15	Predicting Gene Function using Predictive Clustering Trees	365
	Celine Vens, Leander Schietgat, Jan Struyf, Hendrik Blockeel, Dragi Kocев, and Sašo Džeroski	
15.1	Introduction	366
15.2	Related Work	367
15.3	Predictive Clustering Tree Approaches for HMC	369
15.4	Evaluation Measure	374
15.5	Datasets	375
15.6	Comparison of Clus-HMC/SC/HSC	378
15.7	Comparison of (Ensembles of) CLUS-HMC to State-of-the-art Methods	380
15.8	Conclusions	384

16 Analyzing Gene Expression Data with Predictive Clustering Trees . . . 389
 Ivica Slavkov and Sašo Džeroski

- 16.1 Introduction 389
- 16.2 Datasets 391
- 16.3 Predicting Multiple Clinical Parameters 392
- 16.4 Evaluating Gene Importance with Ensembles of PCTs 394
- 16.5 Constrained Clustering of Gene Expression Data 397
- 16.6 Clustering gene expression time series data 400
- 16.7 Conclusions 403

17 Using a Solver Over the String Pattern Domain to Analyze Gene Promoter Sequences 407
 Christophe Rigotti, Ieva Mitašiūnaitė, Jérémy Besson, Laurène Meyniel, Jean-François Boulicaut, and Olivier Gandrillon

- 17.1 Introduction 407
- 17.2 A Promoter Sequence Analysis Scenario 409
- 17.3 The *Marguerite* Solver 412
- 17.4 Tuning the Extraction Parameters 413
- 17.5 An Objective Interestingness Measure 415
- 17.6 Execution of the Scenario 418
- 17.7 Conclusion 422

18 Inductive Queries for a Drug Designing Robot Scientist 425
 Ross D. King, Amanda Schierz, Amanda Clare, Jem Rowland, Andrew Sparkes, Siegfried Nijssen, and Jan Ramon

- 18.1 Introduction 425
- 18.2 The Robot Scientist Eve 427
- 18.3 Representations of Molecular Data 430
- 18.4 Selecting Compounds for a Drug Screening Library 444
- 18.5 Active learning 446
- 18.6 Conclusions 448
- Appendix 452

Author index 455



<http://www.springer.com/978-1-4419-7737-3>

Inductive Databases and Constraint-Based Data Mining

Dzeroski, S.; Goethals, B.; Panov, P. (Eds.)

2010, XVII, 456 p., Hardcover

ISBN: 978-1-4419-7737-3