

Table of Contents

Part I. Introduction

1. Data Mining in a Nutshell

Sašo Džeroski	3
1.1 Introduction	3
1.2 Data mining tasks	4
1.3 Patterns	8
1.4 Basic algorithms	16
1.5 Relational data mining	22
1.6 Data mining literature and Internet resources	25
1.7 Summary	26

2. Knowledge Discovery in Databases: An Overview

Usama Fayyad	28
2.1 Introduction	28
2.2 From transactions to warehouses to KDD	29
2.3 Why data mining?	31
2.4 KDD and data mining	32
2.5 Data mining methods: An overview	34
2.6 Applications in science data analysis	37
2.7 Research challenges for KDD	41
2.8 ILP and KDD: Prospects and challenges	43
2.9 Concluding remarks	45

3. An Introduction to Inductive Logic Programming

Sašo Džeroski and Nada Lavrač	48
3.1 Introduction	48
3.2 Logic programming and databases	49
3.3 Logic programming in a nutshell	51
3.4 The basic ILP task: Relational rule induction	55
3.5 Structuring the space of clauses	58
3.6 Searching the space of clauses	60
3.7 Bounding the search for clauses	62

3.8	Transforming ILP problems to propositional form	67
3.9	Relational data mining tasks addressed by ILP	69
3.10	ILP literature	70
3.11	Summary	71
4.	Inductive Logic Programming for Knowledge Discovery in Databases	
	Stefan Wrobel	74
4.1	Introduction	74
4.2	ILP: Relational analysis technology	81
4.3	ILP subgroup discovery: MIDOS	84
4.4	Using MIDOS and other ILP methods in KEPLER	93
4.5	Conclusion	99
<hr/>		
Part II. Techniques		
<hr/>		
5.	Three Companions for Data Mining in First Order Logic	
	Luc De Raedt, Hendrik Blockeel, Luc Dehaspe, and Wim Van Laer	105
5.1	Introduction	105
5.2	Representation	106
5.3	ICL: Inductive classification logic	114
5.4	TILDE: Top-down induction of logical decision trees	116
5.5	CLAUDIEN: Clausal discovery	117
5.6	Practical use: Getting started	119
5.7	Sample application: Mutagenesis	133
5.8	An exercise	136
5.9	Conclusions and practical info	137
6.	Inducing Classification and Regression Trees in First Order Logic	
	Stefan Kramer and Gerhard Widmer	140
6.1	Introduction	140
6.2	Tree induction in logic	142
6.3	Structural classification and regression trees (S-CART): The top level algorithm	145
6.4	Growing a tree in first-order logic	145
6.5	Model selection by error/cost complexity pruning	150
6.6	First-order model trees	152
6.7	Applications	153
6.8	Related work	154
6.9	Conclusion	156

7. Relational Rule Induction with CPROGOL4.4:**A Tutorial Introduction**

Stephen Muggleton and John Firth 160

7.1 Introduction 160

7.2 How to obtain CPROGOL4.4 161

7.3 Developing an input file for CPROGOL4.4 162

7.4 The theory 167

7.5 Estimating accuracy and significance 179

7.6 Declarative bias 182

7.7 Setting resource bounds 185

7.8 Debugging PROLOG input files 186

7.9 Summary 187

8. Discovery of Relational Association Rules

Luc Dehaspe and Hannu Toivonen 189

8.1 Introduction 189

8.2 From association rules to query extensions 190

8.3 Evaluation measures 194

8.4 Declarative language bias 198

8.5 Query (extension) discovery with WARMR 201

8.6 A sample run 206

8.7 Discussion 208

9. Distance Based Approaches**to Relational Learning and Clustering**

Mathias Kirsten, Stefan Wrobel, and Tamás Horváth 213

9.1 Introduction 213

9.2 A first-order distance measure 215

9.3 Instance-based learning with RIBL2 220

9.4 Hierarchical agglomerative clustering with RDBC 221

9.5 FORC - k -means for multirelational data 223

9.6 A case study in mRNA signal structures 225

9.7 Conclusion 230

Part III. From Propositional to Relational Data Mining

10. How to Upgrade Propositional Learners to First Order**Logic: A Case Study**

Wim Van Laer and Luc De Raedt 235

10.1 Introduction 235

10.2 Knowledge representation 236

10.3 The propositional learner CN2 240

10.4	Upgrading CN2	241
10.5	Some experimental results with ICL	252
10.6	Related work and conclusions	256
11.	Propositionalization Approaches to Relational Data Mining	
	Stefan Kramer, Nada Lavrač, and Peter Flach	262
11.1	Introduction	262
11.2	Background and definition of terms	265
11.3	An example illustrating a simple propositionalization	267
11.4	Feature construction for general-purpose propositionalization	271
11.5	Special-purpose feature construction	274
11.6	Related transformation approaches	277
11.7	A sample propositionalization method: Extending LINUS to handle non-determinate literals	279
11.8	Concluding remarks	286
12.	Relational Learning and Boosting	
	Ross Quinlan	292
12.1	Introduction	292
12.2	Boosting	293
12.3	FOIL	294
12.4	Overview of FFOIL	297
12.5	Boosting FFOIL	299
12.6	Experiments	300
12.7	Summary	304
13.	Learning Probabilistic Relational Models	
	Lise Getoor, Nir Friedman, Daphne Koller, and Avi Pfeffer	307
13.1	Introduction	307
13.2	Probabilistic models	309
13.3	Relational models	312
13.4	Probabilistic relational models	315
13.5	Learning PRMs	321
13.6	Experimental results	325
13.7	Discussion and related work	329
13.8	Extensions	331
13.9	Conclusions	333

Part IV. Applications and Web Resources

14. Relational Data Mining Applications:**An Overview**

Sašo Džeroski 339

14.1 Introduction 339

14.2 Drug design 340

14.3 Predicting mutagenicity and carcinogenicity 343

14.4 Predicting protein structure and function 346

14.5 Medical applications 349

14.6 Environmental applications 350

14.7 Mechanical engineering applications 353

14.8 Traffic engineering applications 356

14.9 Text mining, Web mining, and natural language processing . 357

14.10 Business data analysis 358

14.11 Miscellaneous applications 359

14.12 Summary and discussion 360

15. Four Suggestions and a Rule**Concerning the Application of ILP**

Ashwin Srinivasan 365

15.1 Introduction 365

15.2 Background 366

15.3 When and why ILP? 367

15.4 Encoding background knowledge 368

15.5 Utility mismatch 371

15.6 Comprehensibility 371

15.7 From nursery slopes to Darwin's rule 372

16. Internet Resources on ILP for KDDLjupčo Todorovski, Irene Weber, Nada Lavrač, Olga Stěpánková,
Sašo Džeroski, Dimitar Kazakov, Darko Zupanič, and Peter Flach 375

16.1 Introduction 375

16.2 Brief history on ILP Internet resources 376

16.3 ILPnet2 Internet resources 377

16.4 Other ILP-related Internet resources 383

16.5 KDD related Internet resources 385

16.6 Conclusion 385

Author Index 389**Subject Index** 391

Relational Data Mining

Dzeroski, S.; Lavrac, N. (Eds.)

2001, XIX, 398 p., Hardcover

ISBN: 978-3-540-42289-1