

Luc De Raedt

Logical and Relational Learning

 Springer

Cognitive Technologies

H. Prendinger, M. Ishizuka (Eds.)

Life-Like Characters

Tools, Affective Functions, and Applications

IX, 477 pages. 2004

H. Helbig

Knowledge Representation and the Semantics of Natural Language

XVIII, 646 pages. 2006

P.M. Nugues

An Introduction to Language Processing with Perl and Prolog

An Outline of Theories, Implementation, and Application with Special Consideration of English, French, and German

XX, 513 pages. 2006

W. Wahlster (Ed.)

SmartKom: Foundations of Multimodal Dialogue Systems

XVIII, 644 pages. 2006

B. Goertzel, C. Pennachin (Eds.)

Artificial General Intelligence

XVI, 509 pages. 2007

O. Stock, M. Zancanaro (Eds.)

PEACH — Intelligent Interfaces for Museum Visits

XVIII, 316 pages. 2007

V. Torra, Y. Narukawa

Modeling Decisions: Information Fusion and Aggregation Operators

XIV, 284 pages. 2007

P. Manoonpong

Neural Preprocessing and Control of Reactive Walking Machines

Towards Versatile Artificial Perception–Action Systems

XVI, 185 pages. 2007

S. Patnaik

Robot Cognition and Navigation

An Experiment with Mobile Robots

XVI, 290 pages. 2007

M. Cord, P. Cunningham (Eds.)

Machine Learning Techniques for Multimedia

Case Studies on Organization and Retrieval

XVI, 290 pages. 2008

L. De Raedt

Logical and Relational Learning

XVI, 388 pages. 2008

Cognitive Technologies

Managing Editors: D. M. Gabbay J. Siekmann

Editorial Board: A. Bundy J. G. Carbonell
M. Pinkal H. Uszkoreit M. Veloso W. Wahlster
M. J. Wooldridge

Advisory Board:

Luigia Carlucci Aiello
Franz Baader
Wolfgang Bibel
Leonard Bolc
Craig Boutilier
Ron Brachman
Bruce G. Buchanan
Anthony Cohn
Artur d'Avila Garcez
Luis Fariñas del Cerro
Koichi Furukawa
Georg Gottlob
Patrick J. Hayes
James A. Hendler
Anthony Jameson
Nick Jennings
Aravind K. Joshi
Hans Kamp
Martin Kay
Hiroaki Kitano
Robert Kowalski
Sarit Kraus
Maurizio Lenzerini
Hector Levesque
John Lloyd

Alan Mackworth
Mark Maybury
Tom Mitchell
Johanna D. Moore
Stephen H. Muggleton
Bernhard Nebel
Sharon Oviatt
Luis Pereira
Lu Ruqian
Stuart Russell
Erik Sandewall
Luc Steels
Oliviero Stock
Peter Stone
Gerhard Strube
Katia Sycara
Milind Tambe
Hidehiko Tanaka
Sebastian Thrun
Junichi Tsujii
Kurt VanLehn
Andrei Voronkov
Toby Walsh
Bonnie Webber

Luc De Raedt

Logical and Relational Learning

With 77 Figures and 10 Tables

 Springer

Author:

Prof. Dr. Luc De Raedt
Dept. of Computer Science
Katholieke Universiteit Leuven
Celestijnenlaan 200A
3001 Heverlee
Belgium
luc.deraedt@cs.kuleuven.be

Managing Editors:

Prof. Dov M. Gabbay
Augustus De Morgan Professor of Logic
Department of Computer Science
King's College London
Strand, London WC2R 2LS, UK

Prof. Dr. Jörg Siekmann
Forschungsbereich Deduktions- und
Multiagentensysteme, DFKI
Stuhlsatzenweg 3, Geb. 43
66123 Saarbrücken, Germany

ISBN: 978-3-540-20040-6

e-ISBN: 978-3-540-68856-3

Cognitive Technologies ISSN: 1611-2482

Library of Congress Control Number: 2008933706

ACM Computing Classification (1998): I.2.6, I.2.4, H.2.8, I.5.1

© 2008 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Cover design: KuenkelLopka GmbH

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

To Lieve, Maarten and Soetkin

Preface

I use the term *logical and relational learning* to refer to the subfield of artificial intelligence, machine learning and data mining that is concerned with learning in expressive logical or relational representations. It is the union of inductive logic programming, (statistical) relational learning and multi-relational data mining, which all have contributed techniques for learning from data in relational form. Even though some early contributions to logical and relational learning are about forty years old now, it was only with the advent of inductive logic programming in the early 1990s that the field became popular. Whereas initial work was often concerned with logical (or logic programming) issues, the focus has rapidly changed to the discovery of new and interpretable knowledge from structured data, often in the form of rules, and soon important successes in applications in domains such as bio- and chemo-informatics and computational linguistics were realized. Today, the challenges and opportunities of dealing with structured data and knowledge have been taken up by the artificial intelligence community at large and form the motivation for a lot of ongoing research. Indeed, graph, network and multi-relational data mining are now popular themes in data mining, and statistical relational learning is receiving a lot of attention in the machine learning and uncertainty in artificial intelligence communities. In addition, the range of tasks for which logical and relational techniques have been developed now covers almost all machine learning and data mining tasks. On the one hand these developments have resulted in a new role and novel views on logical and relational learning, but on the other hand have also made it increasingly difficult to obtain an overview of the field as a whole.

This book wants to address these needs by providing a new synthesis of logical and relational learning. It constitutes an attempt to summarize some of the key results about logical and relational learning, it covers a wide range of topics and techniques, and it describes them in a uniform and accessible manner. While the author has tried to select a representative set of topics and techniques from the field of logical and relational learning, he also realizes that he is probably biased by his own research interests and views on the

field. Furthermore, rather than providing detailed accounts of the many specific systems and techniques, the book focuses on the underlying principles, which should enable the reader to easily get access to and understand the relevant literature on logical and relational learning. Actually, at the end of each chapter, suggestions for further reading are provided.

The book is intended for graduate students and researchers in artificial intelligence and computer science, especially those in machine learning, data mining, uncertainty in artificial intelligence, and computational logic, with an interest in learning from relational data. The book is the first textbook on logical and relational learning and is suitable for use in graduate courses, though it can also be used for self-study and as a reference. It contains many different examples and exercises. Teaching material will become available from the author's website. The author would also appreciate receiving feedback, suggestions for improvement and needed corrections by email to luc.deraedt@cs.kuleuven.be.

The book starts with an introductory chapter clarifying the nature, motivations and history of logical and relational learning. Chapter 2 provides a gentle introduction to logic and logic programming, which will be used throughout the book as the representation language. Chapter 3 introduces the idea of learning as search and provides a detailed account of some fundamental machine learning algorithms that will play an important role in later chapters. In Chapter 4, a detailed study of a hierarchy of different representations that are used in machine learning and data mining is given, and two techniques (propositionalization and aggregation) for transforming expressive representations into simpler ones are introduced. Chapter 5 is concerned with the theoretical basis of the field. It studies the generality relation in logic, the relation between induction and deduction, and introduces the most important framework and operators for generality. In Chapter 6, a methodology for developing logical and relational learning systems is presented and illustrated using a number of well-known case studies that learn relational rules, decision trees and frequent queries. The methodology starts from existing learning approaches and upgrades them towards the use of rich representations. Whereas the first six chapters are concerned with the foundations of logical and relational learning, the chapters that follow introduce more advanced techniques. Chapter 7 focuses on learning the definition of multiple relations, that is, on learning theories. This chapter covers abductive reasoning, using integrity constraints, program synthesis, and the use of an oracle. Chapter 8 covers statistical relational learning, which combines probabilistic models with logical and relational learning. The chapter starts with a gentle introduction to graphical models before turning towards probabilistic logics. The use of kernels and distances for logical and relational learning is addressed in Chapter 9, and in Chapter 10 computational issues such as efficiency considerations and learnability results are discussed. Finally, Chapter 11 summarizes the most important lessons learned about logical and relational learning. The author suggests to read it early on, possibly even directly after Chapter 1.

An introductory course to logical and relational learning covers most of the materials in Chapters 1 to 4, Sects. 5.1 – 5.4, 5.9, and Chapters 6 and 11. The other chapters do not depend on one another, and, hence, further chapters can be selected according to the interests and the preferences of the reader. Given the interests in statistical relational learning, the author certainly recommends Chapter 8. Advanced sections and exercises are marked with a * or even with **. They are more challenging, but can be skipped without loss of continuity.

This book could not have been written without the help and encouragement of many persons. The author is indebted to a number of co-workers who contributed ideas, techniques, surveys and views that have found their way into this book, including: Maurice Bruynooghe for influencing the use of logic in this book and numerous suggestions for improvement; Hendrik Blockeel, Luc Dehaspe and Wim Van Laer for contributions to the upgrading methodology described in Chapter 6, Kristian Kersting for joint work on statistical relational learning presented in Chapter 8, and Jan Ramon for his work on distances in Chapter 9. This book has also taken inspiration from a number of joint overview papers and tutorials that the author delivered in collaboration with Hendrik Blockeel, Sašo Džeroski, Kristian Kersting, Nada Lavrač and Stephen Muggleton. The author would also like to thank the editor at Springer, Ronan Nugent, for his patience, help, and support during all phases of this book-writing project.

The author is grateful for the feedback and encouragement on the many earlier versions of this book he received. He would like to thank especially: the reading clubs at the University of Bristol (headed by Peter Flach, and involving Kerstin Eder, Robert Egginton, Steve Gregory, Susanne Hoche, Simon Price, Simon Rawles and Ksenia Shalnova) and at the Katholieke Universiteit Leuven (Hendrik Blockeel, Björn Bringmann, Maurice Bruynooghe, Fabrizio Costa, Tom Croonenborghs, Anton Dries, Kurt Driessens, Daan Fierens, Christophe Costa Florencio, Elisa Fromont, Robby Goetschalkx, Bernd Gutmann, Angelika Kimmig, Niels Landwehr, Wannes Meert, Siegfried Nijssen, Stefan Raeymaekers, Leander Schietgat, Jan Struyf, Ingo Thon, Anneleen van Assche, Joaquin Vanschoren, Celine Vens, and Albrecht Zimmermann), several colleagues who gave feedback or discussed ideas in this book (James Cussens, Paolo Frasconi, Thomas Gaertner, Tamas Horvath, Manfred Jaeger, Martijn van Otterlo), and the students in Freiburg who used several previous versions of this book.

Last but not least I would like to thank my wife, Lieve, and my children, Soetkin and Maarten, for their patience and love during the many years it took to write this book. I dedicate this book to them.

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