Introduction to Statistical Relational Learning

Adaptive Computation and Machine Learning

Thomas Dietterich, Editor
Christopher M. Bishop, David Heckerman, Michael I. Jordan, and Michael Kearns, Associate Editors

Bioinformatics: The Machine Learning Approach
Pierre Baldi and Søren Brunak, 1998

Reinforcement Learning: An Introduction
Richard S. Sutton and Andrew G. Barto, 1998

Graphical Models for Machine Learning and Digital Communication
Brendan J. Frey, 1998

Learning in Graphical Models
Michael I. Jordan, ed., 1998

Causation, Prediction, and Search, 2nd Edition
Peter Spirtes, Clark Glymour, and Richard Scheines, 2001

Principles of Data Mining
David Hand, Heikki Mannila, and Padhraic Smyth, 2001

Bioinformatics: The Machine Learning Approach, 2nd Edition
Pierre Baldi and Søren Brunak, 2001

Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond
Bernhard Schölkopf and Alexander J. Smola, 2001

Learning Kernel Classifiers: Theory and Algorithms
Ralf Herbrich, 2001

Introduction to Machine Learning
Ethem Alpaydin, 2004

Gaussian Processes for Machine Learning
Carl Edward Rasmussen and Christopher K. I. Williams, 2005

Semi-Supervised Learning
Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien, eds. 2006

The Minimum Description Length Principle
Peter D. Grünwald, 2007

Introduction to Statistical Relational Learning
Lise Getoor and Ben Taskar, eds., 2007

Introduction to Statistical Relational Learning

edited by
Lise Getoor
Ben Taskar

The MIT Press
Cambridge, Massachusetts
London, England

Typeset by the authors using LaTeX 2ε
Printed and bound in the United States of America

Library of Congress Cataloging-in-Publication Data
Introduction to statistical relational learning / edited by Lise Getoor, Ben Taskar.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-07288-5 (hardcover : alk. paper)
1. Relational databases. 2. Machine learning–Statistical methods 3. Computer algorithms. I. Getoor, Lise. II. Taskar, Ben.
QA76.9.D3I68 2007 006.3'1–dc22 2007000951
10 9 8 7 6 5 4 3 2 1

Contents

Series Foreword **xi**

Preface **xiii**

1 Introduction **1**
Lise Getoor, Ben Taskar
1.1 Overview . . . 1
1.2 Brief History of Relational Learning . . . 2
1.3 Emerging Trends . . . 3
1.4 Statistical Relational Learning . . . 3
1.5 Chapter Map . . . 5
1.6 Outlook . . . 8

2 Graphical Models in a Nutshell **13**
Daphne Koller, Nir Friedman, Lise Getoor, Ben Taskar
2.1 Introduction . . . 13
2.2 Representation . . . 14
2.3 Inference . . . 22
2.4 Learning . . . 42
2.5 Conclusion . . . 54

3 Inductive Logic Programming in a Nutshell **57**
Sašo Džeroski
3.1 Introduction . . . 57
3.2 Logic Programming . . . 58
3.3 Inductive Logic Programming: Settings and Approaches . . . 64
3.4 Relational Classification Rules . . . 71
3.5 Relational Decision Trees . . . 75
3.6 Relational Association Rules . . . 80
3.7 Relational Distance-Based Methods . . . 84
3.8 Recent Trends in ILP and RDM . . . 89

4 An Introduction to Conditional Random Fields for Relational Learning **93**
Charles Sutton, Andrew McCallum
4.1 Introduction . . . 93
4.2 Graphical Models . . . 94
4.3 Linear-Chain Conditional Random Fields . . . 100
4.4 CRFs in General . . . 108
4.5 Skip-Chain CRFs . . . 116
4.6 Conclusion . . . 122

5 Probabilistic Relational Models **129**
Lise Getoor, Nir Friedman, Daphne Koller, Avi Pfeffer, Ben Taskar
5.1 Introduction . . . 129
5.2 PRM Representation . . . 130
5.3 The Difference between PRMs and Bayesian Networks . . . 140
5.4 PRMs with Structural Uncertainty . . . 141
5.5 Probabilistic Model of Link Structure . . . 141
5.6 PRMs with Class Hierarchies . . . 151
5.7 Inference in PRMs . . . 159
5.8 Learning . . . 161
5.9 Conclusion . . . 173

6 Relational Markov Networks **175**
Ben Taskar, Pieter Abbeel, Ming-Fai Wong, Daphne Koller
6.1 Introduction . . . 175
6.2 Relational Classification and Link Prediction . . . 177
6.3 Graph Structure and Subgraph Templates . . . 178
6.4 Undirected Models for Classification . . . 180
6.5 Learning the Models . . . 184
6.6 Experimental Results . . . 187
6.7 Discussion and Conclusions . . . 197

7 Probabilistic Entity-Relationship Models, PRMs, and Plate Models **201**
David Heckerman, Chris Meek, Daphne Koller
7.1 Introduction . . . 201
7.2 Background: Graphical Models . . . 202
7.3 The Basic Ideas . . . 204
7.4 Probabilistic Entity-Relationship Models . . . 210
7.5 Plate Models . . . 226
7.6 Probabilistic Relational Models . . . 228
7.7 Technical Details . . . 229
7.8 Extensions and Future Work . . . 233

8 Relational Dependency Networks **239**
Jennifer Neville, David Jensen
8.1 Introduction . . . 239
8.2 Dependency Networks . . . 242
8.3 Relational Dependency Networks . . . 243
8.4 Experiments . . . 252
8.5 Related Work . . . 262
8.6 Discussion and Future Work . . . 264

9 Logic-based Formalisms for Statistical Relational Learning **269**
James Cussens
9.1 Introduction . . . 269
9.2 Representation . . . 271
9.3 Inference . . . 278
9.4 Learning . . . 281
9.5 Conclusion . . . 287

10 Bayesian Logic Programming: Theory and Tool **291**
Kristian Kersting, Luc De Raedt

10.1 Introduction . . . 291
10.2 On Bayesian Networks and Logic Programs . . . 293
10.3 Bayesian Logic Programs . . . 296
10.4 Extensions of the Basic Framework . . . 304
10.5 Learning Bayesian Logic Programs . . . 311
10.6 BALIOS – The Engine for Basic Logic Programs . . . 315
10.7 Related Work . . . 315
10.8 Conclusions . . . 318

11 Stochastic Logic Programs: A Tutorial **323**
Stephen Muggleton, Niels Pahlavi
11.1 Introduction . . . 323
11.2 Mixing Deterministic and Probabilistic Choice . . . 324
11.3 Stochastic Grammars . . . 330
11.4 Stochastic Logic Programs . . . 333
11.5 Learning Techniques . . . 335
11.6 Conclusion . . . 337

12 Markov Logic: A Unifying Framework for Statistical Relational Learning **339**
Pedro Domingos, Matthew Richardson
12.1 The Need for a Unifying Framework . . . 339
12.2 Markov Networks . . . 341
12.3 First-Order Logic . . . 342
12.4 Markov Logic . . . 344
12.5 SRL Approaches . . . 350
12.6 SRL Tasks . . . 354
12.7 Inference . . . 356
12.8 Learning . . . 358
12.9 Experiments . . . 360
12.10 Conclusion . . . 367

13 BLOG: Probabilistic Models with Unknown Objects **373**
Brian Milch, Bhaskara Marthi, Stuart Russell, David Sontag, Daniel L. Ong, Andrey Kolobov
13.1 Introduction . . . 373
13.2 Examples . . . 375
13.3 Syntax and Semantics: Possible Worlds . . . 378
13.4 Syntax and Semantics: Probabilities . . . 383
13.5 Evidence and Queries . . . 388
13.6 Inference . . . 388
13.7 Related Work . . . 393
13.8 Conclusions and Future Work . . . 394

14 The Design and Implementation of IBAL: A General-Purpose Probabilistic Language **399**
Avi Pfeffer
14.1 Introduction . . . 399
14.2 The IBAL Language . . . 401
14.3 Examples . . . 407
14.4 Semantics . . . 411
14.5 Desiderata for Inference . . . 415
14.6 Related Approaches . . . 416
14.7 Inference . . . 419

14.8 Lessons Learned and Conclusion 429

15 Lifted First-Order Probabilistic Inference 433
Rodrigo de Salvo Braz, Eyal Amir, Dan Roth
15.1 Introduction 433
15.2 Language, Semantics and Inference problem 435
15.3 The First-Order Variable Elimination (FOVE) algorithm 437
15.4 An experiment 444
15.5 Auxiliary operations 446
15.6 Applicability of lifted inference 448
15.7 Future Directions 449
15.8 Conclusion 449

16 Feature Generation and Selection in Multi-Relational Statistical Learning 453
Alexandrin Popescul, Lyle H. Ungar
16.1 Introduction 453
16.2 Detailed Methodology 458
16.3 Experimental Evaluation 463
16.4 Related Work and Discussion 471
16.5 Conclusion 472

17 Learning a New View of a Database: With an Application in Mammography 477
Jesse Davis, Elizabeth Burnside, Inês Dutra, David Page, Raghu Ramakrishnan, Jude Shavlik, Vítor Santos Costa
17.1 Introduction 477
17.2 View Learning for Mammography 478
17.3 Naive View Learning Framework 482
17.4 Initial Experiments 483
17.5 Integrated View Learning Framework 490
17.6 Further Experiments and Results 491
17.7 Related Work 493
17.8 Conclusions and Future Work 494

18 Reinforcement Learning in Relational Domains: A Policy-Language Approach 499
Alan Fern, SungWook Yoon, Robert Givan
18.1 Introduction 499
18.2 Problem Setup 502
18.3 Approximate Policy Iteration with a Policy Language Bias 503
18.4 API for Relational Planning 507
18.5 Bootstrapping 516
18.6 Relational Planning Experiments 520
18.7 Related Work 527
18.8 Summary and Future Work 530

19 Statistical Relational Learning for Natural Language Information Extraction 535
Razvan C. Bunescu, Raymond J. Mooney
19.1 Introduction 535
19.2 Background on Natural Language Processing 536
19.3 Information Extraction 537
19.4 Collective Information Extraction with RMNs 538
19.5 Future Research on SRL for NLP 549
19.6 Conclusions 550

20 Global Inference for Entity and Relation Identification via a Linear Programming Formulation **553**
Dan Roth, Wen-tau Yih
20.1 Introduction . . . 553
20.2 The Relational Inference Problem . . . 556
20.3 Integer Linear Programming Inference . . . 560
20.4 Solving Integer Linear Programming . . . 562
20.5 Experiments . . . 563
20.6 Comparison with Other Inference Methods . . . 570
20.7 Conclusion . . . 576

Contributors **581**

Index **587**

Series Foreword

The goal of building systems that can adapt to their environments and learn from their experience has attracted researchers from many fields, including computer science, engineering, mathematics, physics, neuroscience, and cognitive science. Out of this research has come a wide variety of learning techniques that have the potential to transform many scientific and industrial fields. Recently, several research communities have converged on a common set of issues surrounding supervised, unsupervised, and reinforcement learning problems. The MIT Press series on Adaptive Computation and Machine Learning seeks to unify the many diverse strands of machine learning research and to foster high quality research and innovative applications.

Thomas Dietterich

Preface

The goal of this book is to bring together important research at the intersection of statistical, logical and relational learning. The material in the collection is aimed at graduate students and researchers in machine learning and artificial intelligence. While by no means exhaustive, the articles introduce a wide variety of recent approaches to combining expressive knowledge representation and statistical learning.

The idea for this book emerged from a series of successful workshops addressing these issues:

- *Learning Statistical Models from Relational Data (SRL2000)* at the National Conference on Artificial Intelligence, AAAI-2000, organized by Lise Getoor and David Jensen.
- *Learning Statistical Models from Relational Data (SRL2003)* at the International Joint Conference on Artificial Intelligence, (IJCAI-2003), organized by Lise Getoor and David Jensen.
- *Statistical Relational Learning and its Connections to Other Fields (SRL2004)* at the International Conference on Machine Learning, (ICML2004), organized by Tom Dietterich, Lise Getoor and Kevin Murphy.
- *Probabilistic, Logical and Relational Learning - Towards a Synthesis*, Dagstuhl Seminar 2005, organized by Luc De Raedt, Thomas Dietterich, Lise Getoor and Stephen Muggleton.
- *Open Problems in Statistical Relational Learning (SRL2006)* at the International Conference on Machine Learning, (ICML2006), organized by Alan Fern, Lise Getoor, and Brian Milch.

We would like to thank all of the participants at these workshops for their intellectual contributions and also for creating a warm and welcoming research community coming together from several distinct research areas.

In addition, there have been several other closely related workshops, including the series of workshops on Multi-Relational Data Mining held in conjunction with the Knowledge Discovery and Data Mining Conference beginning in 2002 organized by Sašo Džeroski, Luc De Raedt, Stefan Wrobel, and Hendrik Blockeel.

This volume contains invited contributions from leading researchers in this new research area. Each chapter has been reviewed by at least two anonymous reviewers. We are very grateful to all the authors for their high quality contributions and to all the reviewers for helping to clarify and improve this work.

In addition to thanking the workshop participants, book contributors and reviewers, we would like to thank our advisors: Daphne Koller, our PhD advisor; Stuart Russell, Lise Getoor's MS advisor; and Michael Jordan, Ben Taskar's Postdoctoral advisor. Lise Getoor would also like to thank David Jensen; besides being one of the people responsible for the name "Statistical Relational Learning," David has been a great mentor, workshop co-organizer and friend. We would also like to thank Tom Dietterich, Pedro Domingos, and David Heckerman, who have been very encouraging in developing this book. Luc De Raedt, Kristian Kersting, Stephen Muggleton, Sašo Džeroski and Hendrik Blockeel have been especially encouraging members from the inductive logic programming and relational learning community. Lise would also like to thank her inquisitive graduate students, members of the LINQs group at the University of Maryland, College Park, for their participation in this project. Finally, on a more personal note, Lise would like to thank Pete for his unwavering support and Ben would like to thank Anat for being his rock.