Lecture Notes in Artificial Intelligence 4911

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Luc De Raedt Paolo Frasconi
Kristian Kersting Stephen Muggleton (Eds.)

Probabilistic Inductive Logic Programming

Theory and Applications

Springer

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Luc De Raedt
Katholieke Universiteit Leuven
Department of Computer Science, Belgium
E-mail: Luc.DeRaedt@cs.kuleuven.be

Paolo Frasconi
Università degli Studi di Firenze
Machine Learning and Neural Networks Group,
Dipartimento di Sistemi e Informatica, Italy
E-mail: p-f@dsi.unifi.it

Kristian Kersting
Massachusetts Institute of Technology, CSAIL
E-mail: kersting@csail.mit.edu

Stephen Muggleton
Imperial College London, Department of Computing
E-mail: shm@doc.ic.ac.uk

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.2.3, I.2.6, I.2, D.1.6, F.4.1, J.3

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-78651-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-78651-1 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12239573 06/3180 5 4 3 2 1 0

Preface

One of the key open questions within artificial intelligence is how to combine probability and logic with learning. This question is getting an increased attention in several disciplines such as knowledge representation, reasoning about uncertainty, data mining, and machine learning simulateously, resulting in the newly emerging subfield known as statistical relational learning and probabilistic inductive logic programming. A major driving force is the explosive growth in the amount of heterogeneous data that is being collected in the business and scientific world. Example domains include bioinformatics, chemoinformatics, transportation systems, communication networks, social network analysis, link analysis, robotics, among others. The structures encountered can be as simple as sequences and trees (such as those arising in protein secondary structure prediction and natural language parsing) or as complex as citation graphs, the World Wide Web, and relational databases.

This book provides an introduction to this field with an emphasis on those methods based on logic programming principles. The book is also the main result of the successful European IST FET project no. FP6-508861 on Application of Probabilistic Inductive Logic Programming (APRIL II, 2004-2007). This project was coordinated by the Albert Ludwigs University of Freiburg (Germany, Luc De Raedt) and the partners were Imperial College London (UK, Stephen Muggleton and Michael Sternberg), the Helsinki Institute of Information Technology (Finland, Heikki Mannila), the Università degli Studi di Florence (Italy, Paolo Frasconi), and the Institut National de Recherche en Informatique et Automatique Rocquencourt (France, Francois Fages). It was concerned with theory, implementations and applications of probabilistic inductive logic programming. This structure is also reflected in the book.

The book starts with an introductory chapter to "Probabilistic Inductive Logic Programming" by De Raedt and Kersting. In a second part, it provides a detailed overview of the most important probabilistic logic learning formalisms and systems. We are very pleased and proud that the scientists behind the key probabilistic inductive logic programming systems (also those developed outside the APRIL project) have kindly contributed a chapter providing an overview of their contributions. This includes: relational sequence learning techniques (Kersting et al.), using kernels with logical representations (Frasconi and Passerini), Markov Logic (Domingos et al.), the PRISM system (Sato and Kameya), CLP($\mathcal{BN}$) (Santos Costa et al.), Bayesian Logic Programs (Kersting and De Raedt), and the Independent Choice Logic (Poole). The third part then provides a detailed account of some show-case applications of probabilistic inductive logic programming, more specifically: in protein fold discovery (Chen et al.), haplotyping (Landwehr and Mielikäinen) and systems biology (Fages and Soliman). The final part touches upon some theoretical investigations and

includes chapters on behavioral comparison of probabilistic logic programming representations (Muggleton and Chen) and a model-theoretic expressivity analysis (Jaeger).

The editors would like to thank the EU (Future and Emerging Technology branch of the FP6 IST programme) for supporting the April II project as well as the partners in the consortium and all contributors to this book. We hope that you will enjoy reading this book as much as we enjoyed the process of producing it.

December 2007

Luc De Raedt
Paolo Frasconi
Kristian Kersting
Stephen H. Muggleton

Table of Contents

Introduction

Formalisms and Systems

Applications

Theory