Intelligent Systems Reference Library, Volume 15

Editors-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

Prof. Lakhmi C. Jain
University of South Australia
Adelaide
Mawson Lakes Campus
South Australia 5095
Australia
E-mail: Lakhmi.jain@unisa.edu.au

Further volumes of this series can be found on our homepage: springer.com

Vol. 1. Christine L. Mumford and Lakhmi C. Jain (Eds.)
Computational Intelligence: Collaboration, Fusion and Emergence, 2009
ISBN 978-3-642-01798-8

Vol. 2. Yuehui Chen and Ajith Abraham
Tree-Structure Based Hybrid Computational Intelligence, 2009
ISBN 978-3-642-04738-1

Vol. 3. Anthony Finn and Steve Scheding
Developments and Challenges for Autonomous Unmanned Vehicles, 2010
ISBN 978-3-642-10703-0

Vol. 4. Lakhmi C. Jain and Chee Peng Lim (Eds.)
Handbook on Decision Making: Techniques and Applications, 2010
ISBN 978-3-642-13638-2

Vol. 5. George A. Anastassiou
Intelligent Mathematics: Computational Analysis, 2010
ISBN 978-3-642-17097-3

Vol. 6. Ludmila Dymowa
Soft Computing in Economics and Finance, 2011
ISBN 978-3-642-17718-7

Vol. 7. Gerasimos G. Rigatos
Modelling and Control for Intelligent Industrial Systems, 2011
ISBN 978-3-642-17874-0

Vol. 8. Edward H.Y. Lim, James N.K. Liu, and Raymond S.T. Lee
Knowledge Seeker – Ontology Modelling for Information Search and Management, 2011
ISBN 978-3-642-17915-0

Vol. 9. Menahem Friedman and Abraham Kandel
Calculus Light, 2011
ISBN 978-3-642-17847-4

Vol. 10. Andreas Tolk and Lakhmi C. Jain
Intelligence-Based Systems Engineering, 2011
ISBN 978-3-642-17930-3

Vol. 11. Samuli Niiranen and Andre Ribeiro (Eds.)
Information Processing and Biological Systems, 2011
ISBN 978-3-642-19620-1

Vol. 12. Florin Gorunescu
Data Mining, 2011
ISBN 978-3-642-19720-8

Vol. 13. Witold Pedrycz and Shyi-Ming Chen (Eds.)
Granular Computing and Intelligent Systems, 2011
ISBN 978-3-642-19819-9

Vol. 14. George A. Anastassiou and Oktay Duman
Towards Intelligent Modeling: Statistical Approximation Theory, 2011
ISBN 978-3-642-19825-0

Vol. 15. Antonino Freno and Edmondo Trentin
Hybrid Random Fields, 2011
ISBN 978-3-642-20307-7

Antonino Freno and Edmondo Trentin

Hybrid Random Fields

A Scalable Approach to Structure and Parameter Learning in Probabilistic Graphical Models

Springer

Dr. Antonino Freno
Università degli Studi di Siena
Dipartimento di Ingegneria dell'Informazione
Via Roma 56
53100 Siena (SI)
Italy
E-mail: freno@dii.unisi.it
http://www.dii.unisi.it/∼freno/

Dr. Edmondo Trentin
Università degli Studi di Siena
Dipartimento di Ingegneria dell'Informazione
Via Roma 56
53100 Siena (SI)
Italy
E-mail: trentin@dii.unisi.it
http://www.dii.unisi.it/∼trentin/HomePage.html

ISBN 978-3-642-20307-7 e-ISBN 978-3-642-20308-4

DOI 10.1007/978-3-642-20308-4

Intelligent Systems Reference Library ISSN 1868-4394

Library of Congress Control Number: 2011925689

© 2011 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.

Typeset & *Cover Design:* Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

A Mario Freno, per avermi insegnato la differenza tra essere onesti ed essere fessi; e a Maria Scaramozzino, per avermi insegnato ad apprezzarla.

–nino

Alla mia famiglia

Edmondo

Foreword

The field of graphical models has been growing significantly since the pioneering studies on probabilistic reasoning in intelligent systems. Beginning from Bayesian networks and Markov random fields, this book offers a new perspective deriving from their nice integration, which results in the new framework of hybrid random fields. While reading the book, one early realizes that there is a unifying approach, that seems to be motivated by the question on how existing types of probabilistic graphical models can be properly combined in such a way to obtain model classes that are rich enough to express a wide variety of conditional independence structures. The authors provide evidence that this combination exhibits scalable behavior in parameter and structure learning. This is also supported by massive experimental results to support the claims on the performance of hybrid random fields in different topics, ranging from bioinformatics to information retrieval. The clean integration idea behind hybrid random fields gives rise to the distinguishing feature of the model, namely the dramatic reduction of complexity that opens the doors to large scale real-world problems. The book not only marks an effective direction of investigation with significant experimental advances, but it is also—and perhaps primarily—a guide for the reader through an original trip in the space of probabilistic modeling.

Interestingly, even though the main subject of investigation is quite specific, while digesting the book, one is enriched with a very open view of the field, with full of stimulating connections. The reader finds a vivid presentation, well rooted in the literature, with inspiring historical references. I very much like the philosophical framework to embed scientific issues given at the end of the book. The authors clear the ground for a view of AI as a science that investigates cognitive technologies. While their investigation does not necessarily provide evidence on natural cognition, they offer indeed a number of intriguing insights. Statistical learning methods are labeled as cognitive technologies, rather than cognitive models. Machines equipped with these methods are viewed as tools for extending human cognition to novel domains, thus offering a nice perspective on the somehow ill-posed question on whether or not machines are intelligent.

I was partially involved in some of the advances on hybrid random fields with the authors, but I frankly find that the book goes well beyond the systematic treatment of the subject. Everyone specifically interested in Bayesian networks and Markov random fields should not miss it.

Siena, January 2011 Marco Gori

Acknowledgments

We are profoundly grateful to Manfred Jaeger, whose continuous advice and in-depth reviews of the evolving drafts of the manuscript turned up invaluable to us, and essential to the pursuance of the ultimate fit of this book.

We feel beholden to our friend and advisor Marco Gori, whose far-reaching and keen support to our work is gratefully acknowledged.

We wish to express our gratitude to Lakhmi Jain, Editor of the present Series, who enthusiastically reacted to the proposal of this editorial project, rendering it possible.

Friends and colleagues who reviewed portions of the manuscript according to their respective competences are thankfully credited, namely (in lexicographic order) Giuseppe Bevilacqua, Valerio Biancalana, John Bickle, Stefania Biscetti, Ilaria "Hillary Castles" Castelli, Michael Glodek, Lorenzo Menconi, Sandro Nannini, Duccio Papini, Claudio Saccà, and Giuseppe Varnier.

Siena,
January 2011

Antonino Freno
Edmondo Trentin

Contents

1 Introduction 1
1.1 Manifesto 1
1.2 Statistics, Graphs, and Beyond 3
1.3 Probabilistic Graphical Models 6
1.4 A Piece of History 9
1.5 Overview of the Book 10

2 Bayesian Networks 15
2.1 Introduction 15
2.2 Representation of Probabilities 21
2.2.1 d-Separation 22
2.2.2 Markov Blankets 24
2.3 Parameter Learning 26
2.4 Structure Learning 28
2.4.1 Evaluation Function 28
2.4.2 Search Strategy 31
2.5 The Naive Bayes Classifier 33
2.6 Final Remarks 35
2.6.1 Extensions of Bayesian Networks 36
2.6.2 Applications of Bayesian Networks 40
2.6.3 Strengths and Weaknesses of Bayesian Networks 40

3 Markov Random Fields 43
3.1 Introduction 43
3.2 Representation of Probabilities 48
3.2.1 The Hammersley-Clifford Theorem 49
3.2.2 A Pseudo-Likelihood Measure 51
3.2.3 Markov Blankets 52

3.3 Parameter Learning 53
3.3.1 Optimizing the Weights 54
3.3.2 Finding the Maximal Cliques 58
3.4 Structure Learning 59
3.5 Final Remarks 60
3.5.1 Generalizations and Variations on the Theme 62
3.5.2 Applications of Markov Random Fields 66
3.5.3 Points of Strength and Limitations of Markov Random Fields 67

4 Introducing Hybrid Random Fields: Discrete-Valued Variables 69
4.1 Introduction 69
4.2 Representation of Probabilities 71
4.3 Formal Properties 74
4.4 Inference 77
4.5 Parameter Learning 79
4.6 Structure Learning 79
4.6.1 Model Initialization 80
4.6.2 Search Operator 80
4.6.3 Evaluation Function 81
4.6.4 Discussion 82
4.7 Related Work 84
4.8 Final Remarks 85

5 Extending Hybrid Random Fields: Continuous-Valued Variables 87
5.1 Introduction 87
5.1.1 An Abiding Debate: Discrete vs Continuous Variables 88
5.1.2 Optimal Convergence to the Real Probability Density Function: The Parzen Window Estimator 90
5.1.3 How the Parzen Window Throws Light on the Debate 91
5.1.4 Chapter Outline 92
5.2 Conditional Density Estimation 93
5.3 Parametric Hybrid Random Fields 94
5.3.1 Normal Distributions 94
5.3.2 Gaussian Mixture Models 100
5.4 Semiparametric Hybrid Random Fields 104
5.4.1 Change of Variables 104
5.4.2 The Nonparanormal 104
5.5 Nonparametric Hybrid Random Fields 108
5.5.1 Kernel-Based Conditional Density Estimation 109
5.5.2 Bandwidth Selection 109
5.5.3 Dual-Tree Recursion 110

5.6 Structure Learning . . . 114
5.6.1 Model Initialization . . . 115
5.6.2 Learning the Local Structures . . . 116
5.6.3 Learning the Global Structure . . . 117
5.7 Final Remarks . . . 117

6 Applications . . . 121
6.1 Introduction . . . 121
6.2 Selecting Features by Learning Markov Blankets . . . 123
6.2.1 A Feature Selection Technique . . . 123
6.2.2 Related Work . . . 124
6.2.3 Results . . . 125
6.3 Application to Discrete Domains . . . 127
6.3.1 Setting Up the Markov Random Field and Dependency Network Models . . . 127
6.3.2 Computational Burden of Structure Learning . . . 128
6.3.3 Pattern Classification . . . 130
6.3.4 Link Prediction . . . 135
6.4 Pattern Classification in Continuous Domains . . . 143
6.4.1 Random Data Generation . . . 143
6.4.2 Results . . . 146
6.5 Final Remarks . . . 150

7 Probabilistic Graphical Models: Cognitive Science or Cognitive Technology? . . . 151
7.1 Introduction . . . 151
7.2 A Philosophical View of Artificial Intelligence . . . 152
7.2.1 The Argument from Authority . . . 152
7.2.2 The Argument from Scientific Practice . . . 156
7.3 From Cognitive Science to Cognitive Technology . . . 156
7.4 Statistical Machine Learning and the Philosophy of Science . . . 158
7.4.1 Machine Learning as a Logic of Scientific Discovery . . . 159
7.4.2 Simplicity Reconsidered . . . 160
7.4.3 Scalability as an Epistemic Virtue . . . 161
7.5 Final Remarks . . . 162

8 Conclusions . . . 163
8.1 Hybrid Random Fields: Where Are We Now? . . . 163
8.2 Future Research: Where Do We Go from Here? . . . 164
8.2.1 Statistical Relational Learning: Some Open Questions . . . 164
8.2.2 Nonparametric Density Estimation: Beyond Kernel Machines . . . 167

A Probability Theory 169
A.1 Random Variables 169
A.2 Discrete Probability Distributions 170
A.3 Continuous Probability Distributions 173

B Graph Theory 177
B.1 Definitions 177
B.2 Algorithms 179

Afterword 183

References 185

Index 201

Acronyms

ANN	Artificial neural network
BN	Bayesian network
BIC	Bayesian information criterion
cdf	Cumulative distribution function
CPT	Conditional probability table
CRF	Conditional random field
CVLL	Cross-validated log-likelihood
DAG	Directed acyclic graph
DBN	Dynamic Bayesian network
DN	Dependency network
DOA	Degree of agreement
edf	Empirical (cumulative) distribution function
FOL	First-order logic
GMM	Gaussian mixture model
GMRF	Gaussian Markov random field
HMM	Hidden Markov model
HMRF	Hidden Markov random field
HRF	Hybrid random field
iid	Independent and identically distributed
ILP	Inductive logic programming
IR	ItemRank
KBN	Kernel-based Bayesian networks
KHRF	Kernel-based hybrid random field
KMRF	Kernel-based Markov random field
MAP	Maximum *a posteriori*
MB	Markov blanket
MBF	Markov blanket filter
MBM	Markov Blanket Merging
MCMC	Markov chain Monte Carlo
MDL	Minimum description length
ML	Maximum likelihood

MLN	Markov logic network
MMMB	Max-Min Markov Blanket
MRF	Markov random field
MRR	Mean reciprocal rank
NB	Naive Bayes
NPMRF	Nonparanormal Markov random field
NW	Nadaraya-Watson
OOBN	Object oriented Bayesian network
pdf	Probability density function
PW	Parzen window
SR	Success rate
SRL	Statistical relational learning

List of Algorithms

2.1 `HCLearnBN` : Hill-climbing structure learning in Bayesian networks 32
2.2 `K2LearnBN`: K2 structure learning in Bayesian networks 33
3.1 `findMaximalCliques`: Finding all maximal cliques in undirected graphs 58
3.2 `HCLearnMRF`: Hill-climbing structure learning in Markov random fields 60
4.1 `GSInference`: Gibbs sampling-based inference for probabilistic graphical models 78
4.2 `MBMLearnHRF`: MBM structure learning in hybrid random fields . . . 82
5.1 `DDSSelectBandwidths`: Bandwidth selection by double dichotomic search 111
5.2 `buildKDTree`: *k*d-tree construction for dual-tree recursion 113
5.3 `MCDualTreeCVLL`: Monte Carlo dual-tree recursion for CVLL estimation 115
6.1 `generateRandomDAG`: Random DAG generation 144
6.2 `PDBNGenerateRandomDistribution`: Generating a random distribution with DAG-shaped polynomial dependencies and beta-distributed noise 145
6.3 `PDBNGenerateRandomData`: Generating random data from a DAG-shaped distribution 146
B.1 `Kahn`: Topological sorting of nodes in directed acyclic graphs 180
B.2 `BronKerbosch`: Finding all maximal cliques in undirected graphs 181