Advances in Computer Vision and Pattern Recognition

More information about this series at http://www.springer.com/series/4205

Luis Enrique Sucar

Probabilistic Graphical Models

Principles and Applications

Springer

Luis Enrique Sucar
Instituto Nacional de Astrofísica,
Óptica y Electrónica (INAOE)
Santa María Tonantzintla
Puebla
Mexico

ISSN 2191-6586 ISSN 2191-6594 (electronic)
Advances in Computer Vision and Pattern Recognition
ISBN 978-1-4471-6698-6 ISBN 978-1-4471-6699-3 (eBook)
DOI 10.1007/978-1-4471-6699-3

Library of Congress Control Number: 2015939664

Springer London Heidelberg New York Dordrecht

Printed on acid-free paper

Springer-Verlag London Ltd. is part of Springer Science+Business Media (www.springer.com)

To my family, Doris, Edgar and Diana,
for their unconditional love and support

Foreword

Probabilistic graphical models (PGMs), and their use for reasoning intelligently under uncertainty, emerged in the 1980s within the statistical and artificial intelligence reasoning communities. The Uncertainty in Artificial Intelligence (UAI) conference became the premier forum for this blossoming research field. It was at UAI-92 in San Jose that I first met Enrique Sucar—both of us graduate students—where he presented his work on relational and temporal models for high-level vision reasoning. Enrique's impressive research contributions to our field over the past 25 years have ranged from foundational work on objective probabilities, to developing advanced forms of PGMS such as temporal and event Bayesian networks, to the learning of PGMs, for example his more recent work on Bayesian chain classifiers for multidimensional classification.

Probabilistic graphical models are now widely accepted as a powerful and mature technology for reasoning under uncertainty. Unlike some of the ad hoc approaches taken in early experts systems, PGMs are based on the strong mathematical foundations of graph and probability theory. They can be used for a wide range of reasoning tasks including prediction, monitoring, diagnosis, risk assessment and decision making. There are many efficient algorithms for both inference and learning available in open-source and commercial software. Moreover, their power and efficacy has been proven through their successful application to an enormous range of realworld problem domains. Enrique Sucar has been a leading contributor in this establishment of PGMs as practical and useful technology, with his work across a wide range of application areas. These include medicine, rehabilitation and care, robotics and vision, education, reliability analysis and industrial applications ranging from oil production to power plants.

The first authors to drawn upon the early research on Bayesian networks and craft it into compelling narratives in the book form were Judea Pearl in *Probabilistic Reasoning in Intelligent Systems* and Rich Neapolitan in *Probabilistic Reasoning in Expert Systems*. This monograph from Enrique Sucar is a timely addition to the body of literature following Pearl and Neapolitan, with up-to-date coverage of a broader range of PGMs than other recent texts in this area: various classifiers, hidden Markov models, Markov random fields, Bayesian networks and its dynamic, temporal and causal variants, relational PGMs, decision graphs and Markov decision process. It presents these PGMs, and the associated methods for reasoning (or inference) and learning, in a clear and accessible manner, making it suitable for advanced students as well as researchers or practitioners from other disciplines interested in using probabilistic models. The text is greatly enriched by the way Enrique has drawn upon his extensive practical experience in modelling with PGMs, illustrating their use across a diverse range of real-world applications from bioinformatics to air pollution to object recognition. I heartily congratulate Enrique on this book and commend it to potential readers.

Melbourne, Australia
May 2015

Ann E. Nicholson

Preface

Overview

Probabilistic graphical models have become a powerful set of techniques used in several domains. This book provides a general introduction to probabilistic graphical models (PGMs) from an engineering perspective. It covers the fundamentals of the main classes of PGMs: Bayesian classifiers, hidden Markov models, Bayesian networks, dynamic and temporal Bayesian networks, Markov random fields, influence diagrams, and Markov decision processes; including representation, inference, and learning principles for all the techniques. Realistic applications for each type of model are also covered in the book.

Some key features are:

- The main classes of PGMs are presented in a single monograph under a unified framework.
- The book covers the fundamental aspects: representation, inference, and learning for all the techniques.
- It illustrates the application of the different techniques in practical problems, an important feature for students and practitioners.
- It includes some of the latest developments in the field, such as multidimensional Bayesian classifiers, relational graphical models, and causal models.
- Each chapter has a set of exercises, including suggestions for research and programming projects.

Motivating the application of probabilistic graphical models to real-world problems is one of the goals of this book. This requires not only knowledge of the different models and techniques, but also some practical experience and domain knowledge. To help the professionals in different fields gain some insight into the use of PGMs for solving practical problems, the book includes many examples of the application of the different types of models in a wide range of domains, including:

- Computer vision.
- Biomedical applications.

- Industrial applications.
- Information retrieval.
- Intelligent tutoring systems.
- Bioinformatics.
- Environmental applications.
- Robotics.
- Human–computer interaction.
- Information validation.
- Caregiving.

Audience

This book can be used as a text book for an advanced undergraduate or a graduate course in probabilistic graphical models for students of computer science, engineering, physics, etc. It could also serve as a reference book for professionals that want to apply probabilistic graphical models in different areas, or anyone who is interested in knowing the basis of these techniques.

It does not have specific prerequisites, although some background in probability and statistics is recommended. It is assumed that the reader has a basic knowledge of mathematics at the high school level, as well as a certain background in computing and programming. The programming exercises require some knowledge and experience with any programming language, such as C, C++, JAVA, Matlab, etc.

Exercises

Each chapter (except the introduction) includes a set of exercises. Some of these exercises are questions and problems designed to reinforce the understanding of the concepts and techniques presented in the chapter. There are also a few suggestions for research or programming projects (marked with "***") in each chapter, which could be used as projects for a course.

Organization

The book is divided into four parts. The first part provides a general introduction and motivation for PGMs, and reviews the required background in probability and graph theory. The second part describes the models which do not consider decisions or utilities: Bayesian classifiers, hidden Markov models, Markov random fields, Bayesian networks, and dynamic and temporal Bayesian networks. The third part starts with a brief introduction to decision theory, and then describes the models which support decision making, including decision trees, influence diagrams, and Markov decision processes. Finally, the fourth part presents two extensions to the *standard* PGMs, one is relational probabilistic graphical models and the other causal models.

The *dependency relations* between the chapters are shown in Fig. 1. An arc from chapter X to chapter "Y", $X \rightarrow Y$, indicates that chapter X is required (or at least recommended) for understanding chapter Y. This graphical representation of the book gives a lot of information, in an analogous way to the graphical models that we will cover later.

From Fig. 1, we can deduce different ways of reading this book. First it is recommended that you read the introduction and the fundamental Chaps. 2 and 3. Then you can study relatively independently the different models in Part II: classification (Chap. 4), hidden Markov models (Chap. 5), Markov random fields (Chap. 6), and Bayesian networks (Chaps. 7–9). Before reading about learning Bayesian networks (Chap. 8), it is necessary to read Chap. 7—representation and inference; and both chapters are required before going into dynamic and temporal Bayesian networks.

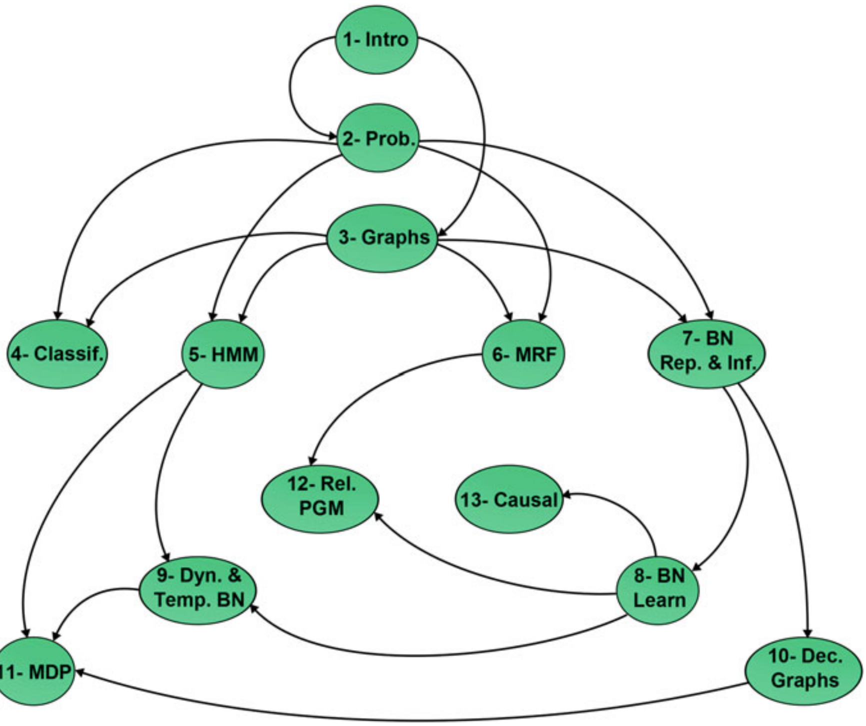

Fig. 1 This figure represents the structure of the book as a directed acyclic graph, showing which chapters are prerequisites for other chapters

The topics in Part III and IV require some of the chapters in Part II. For Chap. 10, which covers decision trees and influence diagrams, you should at least read the first chapter on Bayesian networks. For Chap. 11, which covers sequential decision making, it is recommended that you have covered hidden Markov models and dynamic and temporal Bayesian networks. Relational PGMs (Chap. 12) are based on Markov random fields and Bayesian networks; so Chaps. 6 and 8 are required. Finally, the causal models included in Chap. 13 are based on Bayesian networks including the learning techniques.

If there is not enough time in a course to cover all the book, there are several alternatives. One is to focus on probabilistic models without considering decisions or the more advanced extensions, covering Parts I and II. Another alternative is to focus on decision models, including Part I, the necessary prerequisites from Part II, and Part III. Or you can design your course a la carte, only respecting the dependencies in the graph. However, if you have the time and desire, I suggest you read all the book in order. Enjoy!

Puebla, Mexico Luis Enrique Sucar
February 2015

Acknowledgments

This book grew out of a course that I have been teaching for several years to graduate students. It initiated as a course in Uncertain Reasoning at Tec de Monterrey in Cuernavaca, and became a course on Probabilistic Graphical Models when I moved to INAOE, Puebla in 2006. During these years, my students have been the main motivation and the source of inspiration for writing this book. I will like to thank them all for their interest, questions, and frequent corrections to my notes. This book is dedicated to all my students, past, present, and future.

I will like to acknowledge those students with whom I have collaborated, usually in a small part, for their bachelor, master, or Ph.D. thesis. Some of the novel aspects of this book and most of the application examples originated form their work. I thank them all, and will mention just a few whose work was more influential for this manuscript: Gustavo Arroyo, Héctor Hugo Avilés, Leonardo Chang, Ricardo Omar Chávez, Francisco Elizalde, Hugo Jair Escalante, Lindsey Fiedler, Giovani Gómez, Carlos Hernández, Pablo Hernández, Yasmín Hernández, Pablo Ibargüengoytia, Roger Luis-Velásquez, Miriam Martínez, José Antonio Montero, Julieta Noguez, Annette Morales, Joaquín Pérez-Brito, Miguel Palacios, Mallinali Ramírez, Alberto Reyes, Andrés Rodríguez, Elías Ruiz, Gerardo Torres-Toledano, and Julio Zaragoza. Special thanks to Lindsey Fiedler who has helped me with all the figures in the book and has revised the English.

I will also like to thank my collaborators, the common research projects and technical discussions have enriched my knowledge of many topics, and have helped me write this manuscript. In particular, I will like to mention my colleagues and friends: Juan Manuel Ahuactzin, Olivier Aycard, Concha Bielza, Roberto Ley Borrás, Cristina Conati, Javier Díez, Hugo Jair Escalante, Duncan Gillies, Jesús González, Edel García, Jesse Hoey, Pablo Ibargüengoytia, Pedro Larrañaga, Ron Leder, Jim Little, José Luis Marroquín, Oscar Mayora, Manuel Montes, Eduardo Morales, Enrique Muñoz de Cote, Julieta Noguez, Felipe Orihuela, Luis Pineda, David Poole, Alberto Reyes, Carlos Ruiz, Sunil Vadera, and Luis Villaseñor. I appreciate the comments that Edel, Felipe, and Pablo have made of some earlier versions of this book. Thanks to Ann Nicholson for providing an excellent Foreword.

I want to acknowledge the support from my institution, Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE), which provides an excellent environment for doing research and teaching, and has given me all the facilities to dedicate part of my time to write this book.

Last, but not least, I will like that thank my family. My parents, Fuhed† and Aida, who promoted the desire to learn and to work hard, and supported my studies. In particular, my father, who wrote several wonderful books, and gave me the inspiration (and probably the genes) for writing. My brother Ricardo, and my sisters, Shafía and Beatriz, who have always supported and motivated my dreams. And especially to my wife Doris and my children, Edgar and Diana, who have suffered the long hours I have dedicated to the book instead of them, and whose love and support is what keeps me going.

Contents

Acronyms

ADD	Algebraic Decision Diagram
AI	Artificial Intelligence
BAN	Bayesian network Augmented Naive Bayes classifier
BCC	Bayesian Chain Classifier
BCCD	Bayesian Constraint-based Causal Discovery
BN	Bayesian Network
CBN	Causal Bayesian Network
CMI	Conditional Mutual Information
CPT	Conditional Probability Table
CRF	Conditional Random Field
DAG	Directed Acyclic Graph
DBN	Dynamic Bayesian Network
DBNC	Dynamic Bayesian Network Classifier
DD	Decision Diagram
DDN	Dynamic Decision Network
DT	Decision Tree
EC	Expected Cost
EM	Expectation–Maximization
FN	False Negative
FP	False Positive
GRM	Gibbs Random Field
HMM	Hidden Markov Model
ICM	Iterative Conditional Modes
ID	Influence Diagram
ILP	Inductive Logic Programming
KB	Knowledge Base
LIMID	Limited Memory Influence Diagram
MAG	Maximal Ancestral Graph
MAP	Maximum a Posteriori
MB	Markov Blanket
MBC	Multidimensional Bayesian network Classifier

MC	Markov Chain
MDL	Minimum Description Length
MDP	Markov Decision Process
MLN	Markov Logic Network
MN	Markov Network
MPE	Most Probable Explanation
MRF	Markov Random Field
NBC	Naïve Bayes Classifier
PAG	Parental Ancestral Graph
PGM	Probabilistic Graphical Model
PL	Pseudolikelihood
POMDP	Partially Observable Markov Decision Process
PRM	Probabilistic Relational Model
RPGM	Relational Probabilistic Graphical Model
SNBC	Semi-Naïve Bayesian Classifier
TAN	Tree Augmented Naive Bayes classifier
TEN	Temporal Event Network
TN	Temporal Node
TNBN	Temporal Nodes Bayesian Network

Notations

T	True
F	False
A, B, C, …	Propositions (binary variables)
$\neg A$	Not A (negation)
$A \wedge B$	A and B (conjunction)
$A \vee B$	A or B (disjunction)
$A \rightarrow B$	B if A (implication)
$A \leftrightarrow B$	A if B and B if A (double implication)
$X \in A$	X is an element of A
$\forall(X)$	Universal quantifier: for all X
$\exists(X)$	Existential quantifier: exists an X
$C \cup D$	Union of two sets
$C \cap D$	Intersection of two sets
Ω	Sample space
X	A random variable
x	A particular value of a random variable, $X = x$
$\mathbf{X}$	A vector of random variables, $\mathbf{X} = \mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_\mathrm{N}$
$\mathbf{x}$	A particular realization of vector $\mathbf{X}, \mathbf{x} = \mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_N$
$X_{1:T}$	Vector of variable X from $t = 1$ to $t = T$, $X_{1:T} = X_1, X_2, \ldots, X_T$
$P(X = x)$	Probability of X being in state x; for short $P(x)$
$P(\mathbf{X} = \mathbf{x})$	Probability of $\mathbf{X}$ being in state $\mathbf{x}$; for short $P(\mathbf{x})$
$P(x, y)$	Probability of x and y
$P(x \vee y)$	Probability of x or y
$P(x \mid y)$	Conditional probability of x given y
$P(x) \sim y$	The probability of x is *proportional* to y, that is $P(x) = k \times y$
$\mathbf{P}(\mathbf{X})$	Cumulative distribution function of a discrete variable X
$P(X)$	Probability function of a discrete variable X
$F(X)$	Cumulative distribution function of a continuous variable X
$f(X)$	Probability density function of a continuous variable X
$I(X, Y, Z)$	X independent of Z given Y
$G(V, E)$	Graph G with set of vertices V and set of edges E

$Pa(X)$	Parents of node X in a directed graph
$Nei(X)$	Neighbors of node X in a graph
$n!$	Factorial of $n, n! = n \times (n-1) \times (n-2) \times \ldots 1$
$\binom{n}{r}$	Combinations of r from n, $\binom{n}{r} = \frac{n!}{r!(n-r)!}$
$exp(x)$	Exponential of $x, exp(x) = e^x$
$\lvert X \rvert$	The dimension or number of states of a discrete variable X
μ	Mean
σ^2	Variance
σ	Standard deviation
$N(\mu, \sigma^2)$	Normal distribution with mean μ and standard deviation σ
$I(m)$	Information
$H(M)$	Entropy
$E(X)$	Expected value of a random variable X
$ArgMax_x F(X)$	The value of X for which the function F is maximum