

# Bayesian Networks

## A Practical Guide to Applications

**Edited by**

**Dr Olivier Pourret**

*Electricité de France, France*

**Mr. Patrick Naim**

*ELSEWARE, France*

**Dr Bruce Marcot**

*USDA Forest Service, USA*



John Wiley & Sons, Ltd



# Bayesian Networks

## **STATISTICS IN PRACTICE**

*Series Advisory Editors*

**Marian Scott**

University of Glasgow, UK

**Stephen Senn**

University of Glasgow

*Founding Editor*

**Vic Barnett**

Nottingham Trent University, UK

*Statistics in Practice* is an important international series of texts which provide detailed coverage of statistical concepts, methods and worked case studies in specific fields of investigation and study.

With sound motivation and many worked practical examples, the books show in down-to-earth terms how to select and use an appropriate range of statistical techniques in a particular practical field within each title's special topic area.

The books provide statistical support for professionals and research workers across a range of employment fields and research environments. Subject areas covered include medicine and pharmaceuticals; industry, finance and commerce; public services; the earth and environmental sciences, and so on.

The books also provide support to students studying statistical courses applied to the above areas. The demand for graduates to be equipped for the work environment has led to such courses becoming increasingly prevalent at universities and colleges.

It is our aim to present judiciously chosen and well-written workbooks to meet everyday practical needs. Feedback of views from readers will be most valuable to monitor the success of this aim.

A complete list of titles in this series appears at the end of the volume.

# Bayesian Networks

## A Practical Guide to Applications

**Edited by**

**Dr Olivier Pourret**

*Electricité de France, France*

**Mr. Patrick Naim**

*ELSEWARE, France*

**Dr Bruce Marcot**

*USDA Forest Service, USA*



John Wiley & Sons, Ltd

Copyright © 2008

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester,  
West Sussex PO19 8SQ, England

Telephone (+44) 1243 779777

Email (for orders and customer service enquiries): [cs-books@wiley.co.uk](mailto:cs-books@wiley.co.uk)

Visit our Home Page on [www.wileyurope.com](http://www.wileyurope.com) or [www.wiley.com](http://www.wiley.com)

All Rights Reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except under the terms of the Copyright, Designs and Patents Act 1988 or under the terms of a licence issued by the Copyright Licensing Agency Ltd, 90 Tottenham Court Road, London W1T 4LP, UK, without the permission in writing of the Publisher. Requests to the Publisher should be addressed to the Permissions Department, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, or emailed to [permreq@wiley.co.uk](mailto:permreq@wiley.co.uk), or faxed to (+44) 1243 770620.

This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the Publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

### ***Other Wiley Editorial Offices***

John Wiley & Sons Inc., 111 River Street, Hoboken, NJ 07030, USA

Jossey-Bass, 989 Market Street, San Francisco, CA 94103-1741, USA

Wiley-VCH Verlag GmbH, Boschstr. 12, D-69469 Weinheim, Germany

John Wiley & Sons Australia Ltd, 42 McDougall Street, Milton, Queensland 4064, Australia

John Wiley & Sons (Asia) Pte Ltd, 2 Clementi Loop #02-01, Jin Xing Distripark, Singapore 129809

John Wiley & Sons Canada Ltd, 6045 Freemont Blvd, Mississauga, Ontario, L5R 4J3, Canada

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

### ***Library of Congress Cataloging-in-Publication Data***

Pourret, Olivier.

Bayesian networks : a practical guide to applications / Olivier Pourret and Patrick Naim.

p. cm.

Includes bibliographical references and index.

ISBN 978-0-470-06030-8 (cloth)

1. Bayesian statistical decision theory. 2. Mathematical models. I. Naïm. Patrick. II. Title. QA279.5.P68 2008

519.5'42 – dc22

2007045556

### ***British Library Cataloguing in Publication Data***

A catalogue record for this book is available from the British Library

ISBN: 978-0-470-06030-8

Typeset in 10/12pt Times by Laserwords Private Limited, Chennai, India

Printed and bound in Great Britain by TJ International, Padstow, Cornwall

This book is printed on acid-free paper responsibly manufactured from sustainable forestry in which at least two trees are planted for each one used for paper production.

# Contents

<b>Foreword</b>	<b>ix</b>
<b>Preface</b>	<b>xi</b>
<b>1 Introduction to Bayesian networks</b>	<b>1</b>
1.1 Models . . . . .	1
1.2 Probabilistic vs. deterministic models . . . . .	5
1.3 Unconditional and conditional independence . . . . .	9
1.4 Bayesian networks . . . . .	11
<b>2 Medical diagnosis</b>	<b>15</b>
2.1 Bayesian networks in medicine . . . . .	15
2.2 Context and history . . . . .	17
2.3 Model construction . . . . .	19
2.4 Inference . . . . .	26
2.5 Model validation . . . . .	28
2.6 Model use . . . . .	30
2.7 Comparison to other approaches . . . . .	31
2.8 Conclusions and perspectives . . . . .	32
<b>3 Clinical decision support</b>	<b>33</b>
3.1 Introduction . . . . .	33
3.2 Models and methodology . . . . .	34
3.3 The Busselton network . . . . .	35
3.4 The PROCAM network . . . . .	40
3.5 The PROCAM Busselton network . . . . .	44
3.6 Evaluation . . . . .	46
3.7 The clinical support tool: TakeHeartII . . . . .	47
3.8 Conclusion . . . . .	51
<b>4 Complex genetic models</b>	<b>53</b>
4.1 Introduction . . . . .	53
4.2 Historical perspectives . . . . .	54
4.3 Complex traits . . . . .	56
4.4 Bayesian networks to dissect complex traits . . . . .	59

4.5	Applications . . . . .	64
4.6	Future challenges . . . . .	71
<b>5</b>	<b>Crime risk factors analysis</b>	<b>73</b>
5.1	Introduction . . . . .	73
5.2	Analysis of the factors affecting crime risk . . . . .	74
5.3	Expert probabilities elicitation . . . . .	75
5.4	Data preprocessing . . . . .	76
5.5	A Bayesian network model . . . . .	78
5.6	Results . . . . .	80
5.7	Accuracy assessment . . . . .	83
5.8	Conclusions . . . . .	84
<b>6</b>	<b>Spatial dynamics in France</b>	<b>87</b>
6.1	Introduction . . . . .	87
6.2	An indicator-based analysis . . . . .	89
6.3	The Bayesian network model . . . . .	97
6.4	Conclusions . . . . .	109
<b>7</b>	<b>Inference problems in forensic science</b>	<b>113</b>
7.1	Introduction . . . . .	113
7.2	Building Bayesian networks for inference . . . . .	116
7.3	Applications of Bayesian networks in forensic science . . . . .	120
7.4	Conclusions . . . . .	126
<b>8</b>	<b>Conservation of marbled murrelets in British Columbia</b>	<b>127</b>
8.1	Context/history . . . . .	127
8.2	Model construction . . . . .	129
8.3	Model calibration, validation and use . . . . .	136
8.4	Conclusions/perspectives . . . . .	147
<b>9</b>	<b>Classifiers for modeling of mineral potential</b>	<b>149</b>
9.1	Mineral potential mapping . . . . .	149
9.2	Classifiers for mineral potential mapping . . . . .	151
9.3	Bayesian network mapping of base metal deposit . . . . .	157
9.4	Discussion . . . . .	166
9.5	Conclusions . . . . .	171
<b>10</b>	<b>Student modeling</b>	<b>173</b>
10.1	Introduction . . . . .	173
10.2	Probabilistic relational models . . . . .	175
10.3	Probabilistic relational student model . . . . .	176
10.4	Case study . . . . .	180
10.5	Experimental evaluation . . . . .	182
10.6	Conclusions and future directions . . . . .	185



<b>11 Sensor validation</b>	<b>187</b>
11.1 Introduction . . . . .	187
11.2 The problem of sensor validation . . . . .	188
11.3 Sensor validation algorithm . . . . .	191
11.4 Gas turbines . . . . .	197
11.5 Models learned and experimentation . . . . .	198
11.6 Discussion and conclusion . . . . .	202
<b>12 An information retrieval system</b>	<b>203</b>
12.1 Introduction . . . . .	203
12.2 Overview . . . . .	205
12.3 Bayesian networks and information retrieval . . . . .	206
12.4 Theoretical foundations . . . . .	207
12.5 Building the information retrieval system . . . . .	215
12.6 Conclusion . . . . .	223
<b>13 Reliability analysis of systems</b>	<b>225</b>
13.1 Introduction . . . . .	225
13.2 Dynamic fault trees . . . . .	227
13.3 Dynamic Bayesian networks . . . . .	228
13.4 A case study: The Hypothetical Sprinkler System . . . . .	230
13.5 Conclusions . . . . .	237
<b>14 Terrorism risk management</b>	<b>239</b>
14.1 Introduction . . . . .	240
14.2 The Risk Influence Network . . . . .	250
14.3 Software implementation . . . . .	254
14.4 Site Profiler deployment . . . . .	259
14.5 Conclusion . . . . .	261
<b>15 Credit-rating of companies</b>	<b>263</b>
15.1 Introduction . . . . .	263
15.2 Naive Bayesian classifiers . . . . .	264
15.3 Example of actual credit-ratings systems . . . . .	264
15.4 Credit-rating data of Japanese companies . . . . .	266
15.5 Numerical experiments . . . . .	267
15.6 Performance comparison of classifiers . . . . .	273
15.7 Conclusion . . . . .	276
<b>16 Classification of Chilean wines</b>	<b>279</b>
16.1 Introduction . . . . .	279
16.2 Experimental setup . . . . .	281
16.3 Feature extraction methods . . . . .	285
16.4 Classification results . . . . .	288
16.5 Conclusions . . . . .	298

<b>17 Pavement and bridge management</b>	<b>301</b>
17.1 Introduction . . . . .	301
17.2 Pavement management decisions . . . . .	302
17.3 Bridge management . . . . .	307
17.4 Bridge approach embankment – case study . . . . .	308
17.5 Conclusion . . . . .	312
<b>18 Complex industrial process operation</b>	<b>313</b>
18.1 Introduction . . . . .	313
18.2 A methodology for Root Cause Analysis . . . . .	314
18.3 Pulp and paper application . . . . .	321
18.4 The ABB Industrial IT platform . . . . .	325
18.5 Conclusion . . . . .	326
<b>19 Probability of default for large corporates</b>	<b>329</b>
19.1 Introduction . . . . .	329
19.2 Model construction . . . . .	332
19.3 BayesCredit . . . . .	335
19.4 Model benchmarking . . . . .	341
19.5 Benefits from technology and software . . . . .	342
19.6 Conclusion . . . . .	343
<b>20 Risk management in robotics</b>	<b>345</b>
20.1 Introduction . . . . .	345
20.2 DeepC . . . . .	346
20.3 The ADVOCATE II architecture . . . . .	352
20.4 Model development . . . . .	354
20.5 Model usage and examples . . . . .	360
20.6 Benefits from using probabilistic graphical models . . . . .	361
20.7 Conclusion . . . . .	362
<b>21 Enhancing Human Cognition</b>	<b>365</b>
21.1 Introduction . . . . .	365
21.2 Human foreknowledge in everyday settings . . . . .	366
21.3 Machine foreknowledge . . . . .	369
21.4 Current application and future research needs . . . . .	373
21.5 Conclusion . . . . .	375
<b>22 Conclusion</b>	<b>377</b>
22.1 An artificial intelligence perspective . . . . .	377
22.2 A rational approach of knowledge . . . . .	379
22.3 Future challenges . . . . .	384
<b>Bibliography</b>	<b>385</b>
<b>Index</b>	<b>427</b>

# Foreword

When we, in the late 1980s, worked in a European ESPRIT project on what later became the MUNIN network, expert systems and neural networks were the predominant new artificial intelligence techniques. However, we felt that the most important ingredient of medical diagnosis, *causality with inherent uncertainty*, could not be captured by these techniques.

Rather than trying to model the experts we felt that we should go back to the ancient scientific tradition of modeling the domain, and the new challenge was to incorporate causal uncertainty. We called our models *causal probabilistic networks* (CPNs). They are now called Bayesian networks. The task, we thought, is quite simple: determine a CPN through dialogues with the experts. The rest is just mathematics and computer power.

We were wrong in two ways. It is not ‘just’ mathematics and computer power. But even worse, to determine a CPN through dialogues with the experts is much more intriguing than we anticipated. Over the two decades since the revival of Bayesian networks, several books have addressed the first problem. Although the need is widely recognized, no book has so far focused on the second problem.

This book meets the demand for an aid in developing Bayesian network models in practice. The authors have done a great job in collecting a large sample of Bayesian network applications from a wide range of domains.

Each chapter tells a story about a particular application. However, they do more than that. By studying the various chapters, the reader can learn very much about how to collaborate with domain experts and how to combine domain knowledge with learning from databases. Furthermore, the reader will be presented to a long list of advantages, problems and shortcomings of Bayesian network modeling and inference.

The sample also reflects the two sides of Bayesian network. On the one hand, a Bayesian network is a causal probabilistic network. On the other hand, a Bayesian network is a way of decomposing a large joint probability distribution. In some of the applications, causality is an important part of the model construction, and in other applications, causality is not an issue.

I hope that this book will be studied by everyone who is about to model a domain containing causality with inherent uncertainty: this book will teach him/her if and how to use Bayesian networks.

Finn V. Jensen  
Aalborg University



# Preface

The spectacular improvements of the technologies to produce, transmit, store and retrieve information are leading to a paradox: in many circumstances, making the best use of the available information is much more difficult today than a few decades ago. Information is certainly abundant and easily accessible, but at the same time (and to a large extent, consequently) often inconsistent, contradictory, and of uncertain traceability and reliability. The process of interpreting information remains an essential one, because uninterpreted information is nothing else than noise, but becomes more and more delicate. To mention only one domain covered in this book, striking examples of this phenomenon are the famous criminal cases which remain unsolved, despite the accumulation over years of evidences, proofs and expert opinions.

Given this challenge of optimally using information, it is not surprising that a gain of interest for statistical approaches has appeared in many fields in recent years: the purpose of statistics is precisely to convert information into a usable form.

Bayesian networks, named after the works of Thomas Bayes (ca. 1702–1761) on the theory of probability, have emerged as the result of mathematical research carried out in the 1980s, notably by Judea Pearl at UCLA, and from that time on, have proved successful in a large variety of applications.

This book is intended for users, and also *potential* users of Bayesian networks: engineers, analysts, researchers, computer scientists, students and users of other modeling or statistical techniques. It has been written with a dual purpose in mind:

- highlight the versatility and modeling power of Bayesian networks, and also discuss their limitations and failures, in order to help potential users to assess the adequacy of Bayesian networks to their needs;
- provide practical guidance on constructing and using of Bayesian networks.

We felt that these goals would be better achieved by presenting real-world applications, i.e., models actually in use or that have been at least calibrated, tested, validated, and possibly updated from real-world data – rather than demonstration models, prototypes, or hypothetical models. Anyone who has constructed and used models to solve practical problems has learned that the process is never as straightforward as in textbook cases, due to some ever-present difficulties: inability of the model to capture some features of the problem, missing input data, untractability

(model size/computing time), and non-validable results. Our aim in the book is, also, to identify and document the techniques invented by practitioners to overcome or reduce the impact of these difficulties.

Besides a brief theoretical introduction to Bayesian networks, based on some basic, easily reproducible, examples (Chapter 1), the substance of this book is 20 application chapters (Chapters 2–21), written by invited authors.

In selecting the applications, we strove to achieve the following objectives:

1. cover the major types of applications of Bayesian networks: diagnosis, explanation, forecasting, classification, data mining, sensor validation, and risk analysis;
2. cover as many domains of applications as possible: industry (energy, defense, robotics), computing, natural and social sciences (medicine, biology, ecology, geology, geography), services (banking, business, law), while ensuring that each application is accessible and attractive for nonspecialists of the domain;
3. invite ‘famous names’ of the field of Bayesian networks, but also authors who are primarily known as experts of their field, rather than as Bayesian networks practitioners; find a balance between corporate and academic applications;
4. describe the main features of the most common Bayesian network software packages.

Chapter 22 is a synthesis of the application chapters, highlighting the most promising fields and types of applications, suggesting ways that useful lessons or applications in one field might be used in another field, and analysing, in the perspective of artificial intelligence, where the field of Bayesian networks as a whole is heading.

A companion website for this book can be found at: [www.wiley.com/go/pourret](http://www.wiley.com/go/pourret)

# Contributors

OLUFIKAYO ADERINLEWO Department of Civil and Environmental Engineering  
University of Delaware, Newark DE 19716, USA

NII O. ATTOH-OKINE Department of Civil and Environmental Engineering Univer-  
sity of Delaware, Newark DE 19716, USA

PHILIPPE BAUMARD Professor, University Paul Cézanne, France, IMPGT 21, rue  
Gaston Saporta, 13100 Aix-en-Provence, France

NICOLÁS H. BELTRÁN Electrical Engineering Department, University of Chile Av.  
Tupper 2007, Casilla 412-3, Santiago, Chile

ALEX BIEDERMANN The University of Lausanne – Faculté de Droit et des Sciences  
Criminelles – École des Sciences Criminelles, Institut de Police Scientifique, 1015  
Lausanne-Dorigny, Switzerland

ANDREA BOBBIO Dipartimento di Informatica, Università del Piemonte Orientale,  
Via Bellini 25 g, 15100 Alessandria, Italy

ROONGRASAMEE BOONDAO Faculty of Management Science, Ubon Rajathanee Uni-  
versity Warinchumrab, Ubon Ratchathani 34190, Thailand

LUIS M. DE CAMPOS Departamento de Ciencias de la Computación e Inteligencia  
Artificial, E.T.S.I. Informática y de Telecomunicaciones, Universidad de Granada,  
18071, Granada, Spain

E.J.M. CARRANZA International Institute for Geo-information Science and Earth  
Observation (ITC), Enschede, The Netherlands

DANIELE CODETTA-RAITERI Dipartimento di Informatica, Università del Piemonte  
Orientale, Via Bellini 25 g, 15100 Alessandria, Italy

ERIK DAHLQUIST Department of Public Technology, Mälardalen University, S-721  
78 Västerås, Sweden

DAVID C. DANIELS 8260 Greensboro Drive, Suite 200, McLean, VA 22102, USA

MANUEL A. DUARTE-MERMOUD Electrical Engineering Department, University of  
Chile Av. Tupper 2007, Casilla 412-3, Santiago, Chile

ESBEN EJSING Nykredit, Kalvebod Brygge 1-3, 1780 København V, Denmark