
Encyclopedia of Machine Learning and Data Mining

Claude Sammut • Geoffrey I. Webb
Editors

Encyclopedia of Machine Learning and Data Mining

Second Edition

With 263 Figures and 34 Tables

 Springer

Editors

Claude Sammut
The University of New South Wales
Sydney, NSW
Australia

Geoffrey I. Webb
Faculty of Information Technology
Monash University
Melbourne, VIC, Australia

ISBN 978-1-4899-7685-7 ISBN 978-1-4899-7687-1 (eBook)
ISBN 978-1-4899-7686-4 (print and electronic bundle)
DOI 10.1007/978-1-4899-7687-1

Library of Congress Control Number: 2016958560

© Springer Science+Business Media New York 2011, 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, 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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Science+Business Media LLC
The registered company address is: 233 Spring Street, New York, NY 10013, U.S.A.

Preface

Machine learning and data mining are rapidly developing fields. Following the success of the first edition of the *Encyclopedia of Machine Learning*, we are delighted to bring you this updated and expanded edition. We have expanded the scope, as reflected in the revised title *Encyclopedia of Machine Learning and Data Mining*, to encompass more of the broader activity that surrounds the machine learning process. This includes new articles in such diverse areas as anomaly detection, online controlled experiments, and record linkage as well as substantial expansion of existing entries such as data preparation. We have also included new entries on key recent developments in core machine learning, such as deep learning. A thorough review has also led to updating of much of the existing content.

This substantial tome is the product of an intense effort by many individuals. We thank the Editorial Board and the numerous contributors who have provided the content. We are grateful to the Springer team of Andrew Spencer, Michael Hermann, and Melissa Fearon who have shepherded us through the long process of bringing this second edition to print. We are also grateful to the production staff who have turned the content into its final form.

We are confident that this revised encyclopedia will consolidate the first edition's place as a key reference source for the machine learning and data mining communities.

Contributors

Pieter Abbeel EECS Department, UC Berkeley, Stanford, CA, USA

Zahraa S. Abdallah Faculty of Information Technology, Monash University, Clayton, VIC, Australia

Charu C. Aggarwal IBM T. J. Watson Research Center, Hawthorne, NY, USA

Biliana Alexandrova-Kabadjova Banco de México, Mexico City, Mexico

Periklis Andritsos Faculty of Information, University of Toronto, Toronto, ON, Canada

Peter Auer Department of Information Technology, University of Leoben, Leoben, Austria

J. Andrew Bagnell Carnegie Mellon University, Pittsburgh, PA, USA

Michael Bain University of New South Wales, Sydney, NSW, Australia

Arindam Banerjee University of Minnesota, Minneapolis, MN, USA

Andrew G. Barto University of Massachusetts, Amherst, MA, USA

Rohan A. Baxter Australian Taxation Office, Sydney, NSW, Australia

Bettina Berendt KU Leuven, Leuven, Belgium

Indrajit Bhattacharya IBM India Research Laboratory, New Delhi, India

Mustafa Bilgic University of Maryland, College Park, MD, USA

Mauro Birattari Université Libre de Bruxelles, Brussels, Belgium

Hendrik Blockeel Katholieke Universiteit Leuven, Heverlee, Leuven, Belgium

Leiden Institute of Advanced Computer Science, Heverlee, Belgium

Shawn Bohn Pacific Northwest National Laboratory, Richland, WA, USA

Antal van den Bosch Centre for Language Studies, Radboud University, Nijmegen, The Netherlands

Luka Bradesko Jožef Stefan Institute, Ljubljana, Slovenia

Janez Brank Jožef Stefan Institute, Ljubljana, Slovenia

-
- Jürgen Branke** University of Warwick, Coventry, UK
- Pavel Brazdil** LIAAD-INESC Tec/Faculdade de Economia, University of Porto, Porto, Portugal
- Gavin Brown** The University of Manchester, Manchester, UK
- Ivan Bruha** McMaster University, Hamilton, ON, Canada
- Dariusz Brzezinski** Institute of Computing Sciences, Poznan University of Technology, Poznan, Poland
- Martin D. Buhmann** Justus-Liebig University, Gießen, Germany
- Wray L. Buntine** Statistical Machine Learning Program, NICTA, Canberra, ACT, Australia
Faculty of Information Technology, Monash University, Clayton, VIC, Australia
- Tibério Caetano** Statistical Machine Learning Program, NICTA, Canberra, ACT, Australia
- Nicola Cancedda** Xerox Research Centre Europe, Meylan, France
- Gail A. Carpenter** Department of Mathematics & Center for Adaptive Systems, Boston University, Boston, MA, USA
- John Case** University of Delaware, Newark, DE, USA
- Tonatiuh Peña Centeno** German Center for Neurodegenerative Diseases, Banco de México, Mexico City, Mexico
- Deepayan Chakrabarti** Yahoo! Research, Sunnyvale, CA, USA
- Philip K. Chan** Florida Institute of Technology, Melbourne, FL, USA
- Varun Chandola** State University of New York at Buffalo, Buffalo, NY, USA
- Zhiyuan Chen** University of Illinois at Chicago, Chicago, IL, USA
- Peter Christen** Research School of Computer Science, The Australian National University, Canberra, ACT, Australia
- Massimiliano Ciaramita** Yahoo! Research Barcelona, Barcelona, Spain
- Adam Coates** Stanford University, Stanford, CA, USA
- David Cohn** Mountain View, CA, USA
Edinburgh, UK
- David Corne** Herriot-Watt University, Edinburgh, UK
- Susan Crow** Robert Gordon University, Aberdeen, UK
- James Cussens** University of York, Heslington, UK
- Artur Czumaj** University of Warwick, Coventry, UK

-
- Walter Daelemans** CLIPS University of Antwerp, Antwerpen, Belgium
- Sanjoy Dasgupta** University of California, San Diego, La Jolla, CA, USA
- Gerald DeJong** University of Illinois at Urbana, Urbana, IL, USA
- Marco Dorigo** Université Libre de Bruxelles, Brussels, Belgium
- Kurt Driessens** Maastricht University, Maastricht, The Netherlands
- Chris Drummond** National Research Council of Canada, Ottawa, ON, Canada
- Lan Du** Faculty of Information Technology, Monash University, Clayton, VIC, Australia
- Yaakov Engel** University of Alberta, Edmonton, AB, Canada
- Scott E. Fahlman** Carnegie Mellon University, Pittsburgh, PA, USA
- Alan Fern** Science, Oregon State University, Corvallis, OR, USA
- Peter A. Flach** Department of Computer Science, University of Bristol, Bristol, UK
- Pierre Flener** Department of Information Technology, Uppsala University, Uppsala, Sweden
- Blaž Fortuna** Jozef Stefan Institute, Ljubljana, Slovenia
- Johannes Fürnkranz** Knowledge Engineering Group, TU Darmstadt, Darmstadt, Deutschland
Department of Information Technology, University of Leoben, Leoben, Austria
- Thomas Gärtner** Fraunhofer IAIS, Schloss Birlinghoven, University of Bonn, Sankt Augustin, Germany
- João Gama** University of Porto, Porto, Portugal
- Alma Lilia García-Almanza** Directorate of Regulation and Supervision, Banco de México, Mexico City, Mexico
- Gemma C. Garriga** Université Pierre et Marie Curie, Paris, France
- Wulfram Gerstner** Brain Mind Institute, Lausanne EPFL, Lausanne, Switzerland
- Lise Getoor** University of Maryland, College Park, MD, USA
- Christophe Giraud-Carrier** Department of Computer Science, Brigham Young University, Provo, UT, USA
- Marko Grobelnik** Artificial Intelligence Laboratory, Jožef Stefan Institute, Ljubljana, Slovenia
- Stephen Grossberg** Center for Adaptive Systems, Graduate Program in Cognitive and Neural Systems, Department of Mathematics, Boston University, Boston, MA, USA

Eyke Hüllermeier Department of Computer Science, Paderborn University, Paderborn, Germany

Jiawei Han University of Illinois at Urbana-Champaign, Urbana, IL, USA

Julia Handl University of Manchester, Manchester, UK

Michael Harries Citrix Labs, Advanced Products Group, North Ryde, NSW, Australia

Jun He Aberystwyth University, Aberystwyth, UK

Bernhard Hengst University of New South Wales, Sydney, NSW, Australia

Tom Heskes Institute for Computing and Information Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands

Geoffrey Hinton University of Toronto, Toronto, ON, Canada

James Hodson AI for Good Foundation, New York, NY, USA

Lawrence Holder Washington State University, Pullman, WA, USA

Tamás Horváth Fraunhofer IAIS, Schloss Birlinghoven, University of Bonn, Sankt Augustin, Germany

Phil Husbands Department of Informatics, Centre for Computational Neuroscience and Robotics, University of Sussex, Brighton, UK

Marcus Hutter Research School of Computer Science, Australian National University, Canberra, ACT, Australia

Christian Igel Department of Computer Science, University of Copenhagen, Copenhagen, Denmark

Sanjay Jain School of Computing, National University of Singapore, Singapore, Singapore

Szymon Jaroszewicz Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland

Tommy R. Jensen Alpen-Adria-Universität Klagenfurt, Klagenfurt, Austria

Xin Jin PayPal Inc., San Jose, CA, USA

Antonis C. Kakas University of Cyprus, Nicosia, Cyprus

Subbarao Kambhampati Arizona State University, Tempe, AZ, USA

Anne Kao Boeing Phantom Works, Seattle, WA, USA

Samuel Kaski Helsinki University of Technology, Helsinki, Finland

Carlos Kavka University of Trieste, Trieste, Italy

James Kennedy U.S. Bureau of Labor Statistics, Washington, DC, USA

Eamonn Keogh University of California-Riverside, Riverside, CA, USA

- Kristian Kersting** Technische Universität Dortmund, Dortmund, Germany
Knowledge Discovery, Fraunhofer IAIS, Sankt Augustin, Germany
- Joshua Knowles** University of Manchester, Manchester, UK
- Aleksander Kołcz** Microsoft One Microsoft Way, Redmond, WA, USA
- Ron Kohavi** Application Services Group, Microsoft, Bellevue, WA, USA
- Kevin B. Korb** Clayton School of Information Technology, Monash University, Clayton, VIC, Australia
- Petra Kralj Novak** Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia
- Stefan Kramer** Technische Universität München, Garching b. München, Germany
- Krzysztof Krawiec** Poznan University of Technology, Poznan, Poland
- Vipin Kumar** University of Minnesota, Minneapolis, MN, USA
- Nicolas Lachiche** University of Strasbourg, Strasbourg, France
- Michail G. Lagoudakis** Technical University of Crete, Chania, Greece
- John Langford** Microsoft Research, New York, NY, USA
- Pier Luca Lanzi** Politecnico di Milano, Milano, Italy
- Nada Lavrač** Department of Knowledge Technologies, Jožef Stefan Institute, Ljubljana, Slovenia
University of Nova Gorica, Nova Gorica, Slovenia
- Gregor Leban** Jozef Stefan Institute, Ljubljana, Slovenia
- Christina Leslie** Memorial Sloan Kettering Cancer Research Center, New York, NY, USA
- Hang Li** Huawei Technologies, Hong Kong, China
- Shiau Hong Lim** University of Illinois, Champaign, IL, USA
- Charles X. Ling** The University of Western Ontario, London, ON, Canada
- Bin Liu** Monash University, Clayton, VIC, Australia
- Bing Liu** University of Illinois at Chicago, Chicago, IL, USA
- Huan Liu** Arizona State University, Tempe, AZ, USA
- John Lloyd** The Australian National University, Canberra, ACT, Australia
- Roger Longbotham** Data and Decision Sciences Group, Microsoft, Redmond, WA, USA
- Shie Mannor** Israel Institute of Technology, Haifa, Israel
- Serafín Martínez-Jaramillo** Directorate of Financial System Risk Analysis, Banco de México, Mexico City, Mexico

Eric Martin University of New South Wales, Sydney, NSW, Australia

Stan Matwin University of Ottawa, Ottawa, ON, Canada
Polish Academy of Sciences, Warsaw, Poland

Julian McAuley Computer Science Department, University of California,
San Diego, CA, USA
Statistical Machine Learning Program, NICTA, Canberra, ACT, Australia

Franziska Meier University of Southern California, Los Angeles, CA, USA

Prem Melville IBM T. J. Watson Research Center, Yorktown Heights, NY,
USA

Pietro Michelucci Strategic Analysis, Inc., Arlington, VA, USA

Rada Mihalcea University of North Texas, Denton, TX, USA

Risto Miikkulainen Department of Computer Science, The University of
Texas at Austin, Austin, TX, USA

Dunja Mladenić Artificial Intelligence Laboratory, Jožef Stefan Institute,
Ljubljana, Slovenia

Katharina Morik Technische Universität Dortmund, Dortmund, Germany

Jun Morimoto Advanced Telecommunication Research Institute Interna-
tional (ATR), Kyoto, Japan

Abdullah Mueen University California-Riverside, Riverside, CA, USA

Paul Munro University of Pittsburgh, Pittsburgh, PA, USA

Ion Muslea Language Weaver, Inc., Marina del Rey, CA, USA

Galileo Namata University of Maryland, College Park, MD, USA

Sriraam Natarajan Department of Computer Science, University of
Wisconsin Medical School, Madison, WI, USA

School of Informatics and Computing, Indiana University, Bloomington, IN,
USA

Andrew Y. Ng Computer Science Department, Stanford University,
Stanford, CA, USA

Stanford University, Stanford, CA, USA

Siegfried Nijssen Katholieke Universiteit Leuven, Leuven, Belgium

William Stafford Noble Department of Genome Science/Department of
Computer Science and Engineering, University of Washington, Seattle, WA,
USA

Eirini Ntoutsi Leibniz Universität Hannover, Hannover, Germany
Ludwig Maximilians Universität München, Munich, Germany

Daniel Oblinger DARPA/IPTO, Arlington, VA, USA

- Peter Orbanz** Cambridge University, Cambridge, UK
- Miles Osborne** University of Edinburgh, Edinburgh, UK
- Stefano Pacifico** Jožef Stefan Institute, Ljubljana, Slovenia
- C. David Page** Department of Biostatistics and Medical Informatics, University of Wisconsin Medical School, Madison, WI, USA
- Jonathan Patrick** University of Ottawa, Ottawa, ON, Canada
- Claudia Perlich** IBM T.J. Watson Research Center, Yorktown Heights, NY, USA
- Jan Peters** Department of Empirical Inference, Max-Planck Institute for Intelligent Systems, Tübingen, Germany
Intelligent Autonomous Systems, Computer Science Department, Technische Universität Darmstadt, Darmstadt, Hessen, Germany
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
- Francesco Petruccione** National Institute of Theoretical Physics (NITheP), KwaZulu-Natal, South Africa
- Bernhard Pfahringer** University of Waikato, Hamilton, New Zealand
- Steve Poteet** Boeing Phantom Works, Seattle, WA, USA
- Pascal Poupart** University of Waterloo, Waterloo, ON, Canada
- Rob Powers** Stanford University, Stanford, CA, USA
- Cecilia M. Procopiuc** AT&T Labs, NJ, USA
- Martin L. Puterman** University of British Columbia, Vancouver, BC, Canada
- Lesley Quach** Boeing Phantom Works, Seattle, WA, USA
- Novi Quadrianto** Department of Informatics, SMiLe CLiNiC, University of Sussex, Brighton, UK
- Luc De Raedt** Department of Computer Science, Katholieke Universiteit Leuven, Heverlee, Leuven, Belgium
- Dev Rajnarayan** NASA Ames Research Center, Moffett Field, CA, USA
- Adwait Ratnaparkhi** Yahoo!, Sunnyvale, CA, USA
- Soumya Ray** Case Western Reserve University, Cleveland, OH, USA
- Mark Reid** The Australian National University, Canberra, ACT, Australia
- Jean-Michel Renders** Xerox Research Centre Europe, Meylan, France
- John Risch** Pacific Northwest National Laboratory, Richland, WA, USA
- Teemu Roos** Department of Computer Science, Helsinki Institute for Information Technology, University of Helsinki, Helsinki, Finland

- Nick Roy** Massachusetts Institute of Technology, Cambridge, MA, USA
- Lorenza Saitta** Università del Piemonte Orientale, Alessandria, Italy
- Yasubumi Sakakibara** Keio University, Hiyoshi, Kohoku-ku, Japan
- Claude Sammut** The University of New South Wales, Sydney, NSW, Australia
- Joerg Sander** University of Alberta, Edmonton, AB, Canada
Statistical Machine Learning Group, NICTA, Canberra, ACT, Australia
- Scott Sanner** Statistical Machine Learning Group, NICTA, Canberra, ACT, Australia
- Stefan Schaal** Max Planck Institute for Intelligent Systems, Stuttgart, Germany
Computer Science, University of Southern California, Los Angeles, CA, USA
- Ute Schmid** Faculty of Information Systems and Applied Computer Science, University of Bamberg, Bamberg, Germany
- Jürgen Schmidhuber** The Swiss AI Lab, IDSIA, USI & SUPSI, Manno & Lugano, Switzerland
- Maria Schuld** Quantum Research Group, School of Chemistry & Physics, University of KwaZulu-Natal, Durban, South Africa
- Stephen Scott** University of Nebraska, Lincoln, NE, USA
- Michele Sebag** CNRS – INRIA – Université Paris-Sud, Orsay, France
- Prithviraj Sen** University of Maryland, College Park, MD, USA
- Hanhuai Shan** University of Minnesota, Minneapolis, MN, USA
- Hossam Sharara** University of Maryland, College Park, MD, USA
- Viktoriia Sharmanska** Department of Informatics, University of Sussex, SMiLe CLiNiC, Falmer, UK
- Victor S. Sheng** The University of Western Ontario, London, ON, Canada
- Jelber Sayyad Shirabad** University of Ottawa, Ottawa, ON, Canada
- Yoav Shoham** Stanford University, Stanford, CA, USA
- Thomas R. Shultz** McGill University, Montréal, QC, Canada
- Ricardo Silva** Centre for Computational Statistics and Machine Learning, University College London, London, UK
- Vikas Sindhwani** IBM T. J. Watson Research Center, Yorktown Heights, NY, USA
- Moshe Sipper** Ben-Gurion University, Beer-Sheva, Israel
- William D. Smart** Washington University in St. Louis, St. Louis, MO, USA

Carlos Soares LIAAD-INESC Porto L.A./Faculdade de Economia, University of Porto, Porto, Portugal

LIAAD-INESC Tec/Faculdade de Economia, University of Porto, Porto, Portugal

Christian Sohler University of Paderborn, Paderborn, Germany

Myra Spiliopoulou Otto-von-Guericke University-Magdeburg, Magdeburg, Germany

Thomas Stütze Université libre de Bruxelles (ULB), Brussels, Belgium

Janez Starc Jožef Stefan Institute, Ljubljana, Slovenia

Jerzy Stefanowski Institute of Computing Sciences, Poznan University of Technology, Poznan, Poland

Frank Stephan Department of Mathematics, National University of Singapore, Singapore, Singapore

Peter Stone Department of Computer Science, The University of Texas at Austin, Austin, TX, USA

Alexander L. Strehl Rutgers University, New Brunswick, NJ, USA

Jan Struyf Katholieke Universiteit Leuven, Heverlee, Leuven, Belgium

Prasad Tadepalli School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA

Jiliang Tang Michigan State University, East Lansing, MI, USA

Russ Tedrake Massachusetts Institute of Technology, Cambridge, MA, USA

Yee Whye Teh University College London, London, UK

Jon Timmis University of York, Heslington, North Yorkshire, UK

Jo-Anne Ting University of Edinburgh, Edinburgh, UK

Kai Ming Ting Federation University, Mount Helen, VIC, Australia

Ljupčo Todorovski University of Ljubljana, Ljubljana, Slovenia

Hannu Toivonen University of Helsinki, Helsinki, Finland

Luís Torgo University of Porto, Porto, Portugal

Panayiotis Tsaparas Department of Computer Science & Engineering, University of Ioannina, Ioannina, Greece

Paul E. Utgoff University of Massachusetts, Amherst, MA, USA

William Uther NICTA and The University of New South Wales, Sydney, NSW, Australia

Sethu Vijayakumar University of Edinburgh, Edinburgh, UK
University of Southern California, Los Angeles, CA, USA

Ricardo Vilalta Department of Computer Science, University of Houston,
Houston, TX, USA

Michail Vlachos IBM Research, Zurich, Switzerland

Kiri L. Wagstaff Pasadena, CA, USA

Suhang Wang Arizona State University, Tempe, AZ, USA

Geoffrey I. Webb Faculty of Information Technology, Monash University,
Victoria, Australia

R. Paul Wiegand University of Central Florida, Orlando, FL, USA

Eric Wiewiora University of California, Sydney, NSW, Australia

William E. Winkler US Census Bureau, Suitland, MD, USA

Anthony Wirth The University of Melbourne, Melbourne, VIC, Australia

Michael Witbrock Cypcorp Inc, Austin, TX, USA

David Wolpert NASA Ames Research Center, Moffett Field, CA, USA
Santa Fe Institute, Santa Fe, NM, USA

Stefan Wrobel Fraunhofer IAIS, Schloss Birlinghoven, University of Bonn,
Sankt Augustin, Germany

Jason Wu Boeing Phantom Works, Seattle, WA, USA

Zhao Xu Fraunhofer IAIS, Sankt Augustin, Germany

Ying Yang Australian Taxation Office, Box Hill, VIC, Australia

Sungwook Yoon MapR, San Jose, CA, USA

Thomas Zeugmann Hokkaido University, Sapporo, Japan

Xinhua Zhang NICTA, Australian National University, Canberra, ACT,
Australia

School of Computer Science, Australian National University, Canberra, ACT,
Australia

NICTA London Circuit, Canberra, ACT, Australia

Lei Zhang LinkedIn, San Francisco, CA, USA

Min-Ling Zhang School of Computer Science and Engineering, Southeast
University, Nanjing, China

Fei Zheng Monash University, Sydney, NSW, Australia

Monash University, Clayton, Melbourne, VIC, Australia

Zhi-Hua Zhou National Key Laboratory for Novel Software Technology,
Nanjing University, Nanjing, China

Xiaojin Zhu University of Wisconsin-Madison, Madison, WI, USA

Max Zimmermann Swedish Institute of Computer Science (SICS Swedish
ICT), Kista, Sweden