

# Studies in Big Data

Volume 32

## Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland  
e-mail: [kacprzyk@ibspan.waw.pl](mailto:kacprzyk@ibspan.waw.pl)

### *About this Series*

The series “Studies in Big Data” (SBD) publishes new developments and advances in the various areas of Big Data-quickly and with a high quality. The intent is to cover the theory, research, development, and applications of Big Data, as embedded in the fields of engineering, computer science, physics, economics and life sciences. The books of the series refer to the analysis and understanding of large, complex, and/or distributed data sets generated from recent digital sources coming from sensors or other physical instruments as well as simulations, crowd sourcing, social networks or other internet transactions, such as emails or video click streams and other. The series contains monographs, lecture notes and edited volumes in Big Data spanning the areas of computational intelligence incl. neural networks, evolutionary computation, soft computing, fuzzy systems, as well as artificial intelligence, data mining, modern statistics and Operations research, as well as self-organizing systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

More information about this series at <http://www.springer.com/series/11970>

Tania Cerquitelli • Daniele Quercia • Frank Pasquale  
Editors

# Transparent Data Mining for Big and Small Data

 Springer

*Editors*

Tania Cerquitelli  
Department of Control  
and Computer Engineering  
Politecnico di Torino  
Torino, Italy

Daniele Quercia  
Bell Laboratories  
Cambridge, UK

Frank Pasquale  
Carey School of Law  
University of Maryland  
Baltimore, MD, USA

ISSN 2197-6503

Studies in Big Data

ISBN 978-3-319-54023-8

DOI 10.1007/978-3-319-54024-5

ISSN 2197-6511 (electronic)

ISBN 978-3-319-54024-5 (eBook)

Library of Congress Control Number: 2017936756

© Springer International Publishing AG 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. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

Algorithms are increasingly impacting our lives. They promote healthy habits by recommending activities that minimize risks, facilitate financial transactions by estimating credit scores from multiple sources, and recommend what to buy by profiling purchasing patterns. They do all that based on data that is not only directly disclosed by people but also inferred from patterns of behavior and social networks.

Algorithms affect us, yet the processes behind them are hidden. They often work as black boxes. With little transparency, wrongdoing is possible. Algorithms could recommend activities that minimize health risks only for a subset of the population because of biased training data. They could perpetuate racial discrimination by refusing mortgages based on factors imperfectly tied to race. They could promote unfair price discrimination by offering higher online shopping prices to those who are able to pay them. Shrouded in secrecy and complexity, algorithmic decisions might well perpetuate bias and prejudice.

This book offers design principles for better algorithms. To ease readability, the book is divided into three parts, which are tailored to readers of different backgrounds. To ensure transparent mining, solutions should first and foremost increase transparency (Part I), plus they should not only be algorithmic (Part II) but also regulatory (Part III).

To begin with Part I, algorithms are increasingly used to make better decisions about public goods (e.g., health, safety, finance, employment), and requirements such as transparency and accountability are badly needed. In Chapter “The Tyranny of Data? The Bright and Dark Sides of Data-Driven Decision-Making for Social Good”, Lepri et al. present some key ideas on how algorithms could meet those requirements without compromising predictive power. In times of “post-truth” politics—the political use of assertions that “feel true” but have no factual basis—also news media might benefit from transparency. Nowadays, algorithms are used to produce, distribute, and filter news articles. In Chapter “Enabling Accountability of Algorithmic Media: Transparency as a Constructive and Critical Lens”, Diakopoulos introduces a model that enumerates different types of information

that might be disclosed about such algorithms. In so doing, the model enables transparency and media accountability. More generally, to support transparency on the entire Web, the Princeton Web Transparency and Accountability Project (Chapter “The Princeton Web Transparency and Accountability Project”) has continuously monitored thousands of web sites to uncover how user data is collected and used, potentially reducing information asymmetry.

Design principles for better algorithms are also of algorithmic nature, and that is why Part II focuses on algorithmic solutions. Datta et al. introduce a family of measures that quantify the degree of influence exerted by different input data on the output (Chapter “Algorithmic Transparency via Quantitative Input Influence”). These measures are called quantitative input influence (QII) measures and help identify discrimination and biases built in a variety of algorithms, including black-boxes ones (only full control of the input and full observability of the output are needed). But not all algorithms are black boxes. Rule-based classifiers could be easily interpreted by humans, yet they have been proven to be less accurate than state-of-the-art algorithms. That is also because of ineffective traditional training methods. To partly fix that, in Chapter “Learning Interpretable Classification Rules with Boolean Compressed Sensing”, Malioutov et al. propose new approaches for training Boolean rule-based classifiers. These approaches not only are well-grounded in theory but also have been shown to be accurate in practice. Still, the accuracy achieved by deep neural networks has been so far unbeaten. Huge amounts of training data are fed into an input layer of neurons, information is processed into a few (middle) hidden layers, and results come out of an output layer. To shed light on those hidden layers, visualization approaches of the inner functioning of neural networks have been recently proposed. Seifert et al. provide a comprehensive overview of these approaches, and they do so in the context of computer vision (Chapter “Visualizations of Deep Neural Networks in Computer Vision: A Survey”).

Finally, Part III dwells on regulatory solutions that concern data release and processing—upon private data, models are created, and those models, in turn, produce algorithmic decisions. Here there are three steps. The first concerns data release. Current privacy regulations (including the “end-user license agreement”) do not provide sufficient protection to individuals. Hutton and Henderson introduce new approaches for obtaining sustained and meaningful consent (Chapter “Beyond the EULA: Improving Consent for Data Mining”). The second step concerns data models. Despite being generated from private data, algorithm-generated models are not personal data in the strict meaning of law. To extend privacy protections to those emerging models, Giovanni Comandè proposes a new regulatory approach (Chapter “Regulating Algorithms’ Regulation? First Ethico-Legal Principles, Problems, and Opportunities of Algorithms”). Finally, the third step concerns algorithmic decisions. In Chapter “What Role Can a Watchdog Organization Play in Ensuring Algorithmic Accountability?”, AlgorithmWatch is presented. This is a watchdog and advocacy initiative that analyzes the effects of algorithmic decisions on human behavior and makes them more transparent and understandable.

There is huge potential for data mining in our society, but more transparency and accountability are needed. This book has introduced only a few of the encouraging initiatives that are beginning to emerge.

Torino, Italy,  
Cambridge, UK  
Baltimore, MD, USA  
January 2017

Tania Cerquitelli  
Daniele Quercia  
Frank Pasquale

# Contents

## Part I Transparent Mining

<b>The Tyranny of Data? The Bright and Dark Sides of Data-Driven Decision-Making for Social Good</b> .....	3
Bruno Lepri, Jacopo Staiano, David Sangokoya, Emmanuel Letouzé, and Nuria Oliver	
<b>Enabling Accountability of Algorithmic Media: Transparency as a Constructive and Critical Lens</b> .....	25
Nicholas Diakopoulos	
<b>The Princeton Web Transparency and Accountability Project</b> .....	45
Arvind Narayanan and Dillon Reisman	

## Part II Algorithmic Solutions

<b>Algorithmic Transparency via Quantitative Input Influence</b> .....	71
Anupam Datta, Shayak Sen, and Yair Zick	
<b>Learning Interpretable Classification Rules with Boolean Compressed Sensing</b> .....	95
Dmitry M. Malioutov, Kush R. Varshney, Amin Emad, and Sanjeeb Dash	
<b>Visualizations of Deep Neural Networks in Computer Vision: A Survey</b> ..	123
Christin Seifert, Aisha Aamir, Aparna Balagopalan, Dhruv Jain, Abhinav Sharma, Sebastian Grottel, and Stefan Gumhold	

## Part III Regulatory Solutions

<b>Beyond the EULA: Improving Consent for Data Mining</b> .....	147
Luke Hutton and Tristan Henderson	



**Regulating Algorithms' Regulation? First Ethico-Legal Principles, Problems, and Opportunities of Algorithms** ..... 169  
Giovanni Comandè

**AlgorithmWatch: What Role Can a Watchdog Organization Play in Ensuring Algorithmic Accountability?** ..... 207  
Matthias Spielkamp

**Erratum** ..... E1

# List of Contributors

**Aisha Aamir** Technische Universität Dresden, Dresden, Germany

**Aparna Balagopalan** Technische Universität Dresden, Dresden, Germany

**Giovanni Comandé** Scuola Superiore Sant'Anna Pisa, Pisa, Italy

**Sanjeeb Dash** IBM T. J. Watson Research Center, Yorktown Heights, NY, USA

**Anupam Datta** Carnegie Mellon University, Pittsburgh, PA, USA

**Nicholas Diakopoulos** Philip Merrill College of Journalism, University of Maryland, College Park, MD, USA

**Amin Emad** Institute for Genomic Biology, University of Illinois, Urbana Champaign, Urbana, IL, USA

1218 Thomas M. Siebel Center for Computer Science, University of Illinois, Urbana, IL, USA

**Sebastian Grottel** Technische Universität Dresden, Dresden, Germany

**Stefan Gumhold** Technische Universität Dresden, Dresden, Germany

**Tristan Henderson** School of Computer Science, University of St Andrews, St Andrews, UK

**Luke Hutton** Centre for Research in Computing, The Open University, Milton Keynes, UK

**Dhruv Jain** Technische Universität Dresden, Dresden, Germany

**Bruno Lepri** Fondazione Bruno Kessler, Trento Italy

**Emmanuel Letouzé** Data-Pop Alliance, New York, NY, USA

MIT Media Lab, Cambridge, MA, USA

**Dmitry M. Malioutov** IBM T. J. Watson Research Center, Yorktown Heights, NY, USA

- Arvind Narayanan** Princeton University, Princeton, NJ, USA
- Nuria Oliver** Data-Pop Alliance, New York, NY, USA
- Dillon Reisman** Princeton University, Princeton, NJ, USA
- David Sangokoya** Data-Pop Alliance, New York, NY, USA
- Christin Seifert** Technische Universität Dresden, Dresden, Germany
- Shayak Sen** Carnegie Mellon University, Pittsburgh, PA, USA
- Abhinav Sharma** Technische Universität Dresden, Dresden, Germany
- Matthias Spielkamp** AlgorithmWatch, Berlin, Germany
- Jacopo Staiano** Fortia Financial Solutions, Paris, France
- Kush R. Varshney** IBM T. J. Watson Research Center, Yorktown Heights, NY, USA
- Yair Zick** School of Computing, National University of Singapore, Singapore, Singapore

# Acronyms

<i>A</i>	Algorithm
<i>Q<sub>A</sub></i>	Quantity of interest
<i>l</i>	Influence
1Rule	Boolean compressed sensing-based single rule learner
ADM	Automated decision making
AKI	Acute kidney injury
AI	Artificial intelligence
API	Application Program Interface
C5.0	C5.0 Release 2.06 algorithm with rule set option in SPSS
CAL. BUS & PROF. CODE	California Business and Professions Code
CAL. CIV. CODE	California Civil Code
CASIA-HWDB	Institute of Automation of the Chinese Academy of Sciences-Handwriting Databases
CAR	Computer-assisted reporting
CART	Classification and regression trees algorithm in MATLAB's classregtree function
CDBN	Convolutional deep belief network
CNN	Convolutional neural network
CONN. GEN. STAT. ANN.	Connecticut general statutes annotated
CS	Compressed sensing
CVPR***	Computer vision and pattern recognition
DAS	Domain awareness system
DBN	Deep belief network
DCNN	Deep convolutional neural network
DHS	US Department of Homeland Security
DList	Decision lists algorithm in SPSS
DNA	Deoxyribonucleic acid
DNNs	Deep neural networks
DTD	Describable Textures Dataset

EU	European Union
EU GSPR	European Union General Data Protection Regulation
EDPS	European Data Protection Supervisor
EFF	Electronic Frontier Foundation
EUCJ	European Union Court of Justice
EULA	End user license agreement
FLIC	Frames Labeled in Cinema
FMD	Flickr Material Database
FTC	Federal Trade Commission
GA. CODE ANN.	Code of Georgia Annotated
GDPR	General Data Protection Regulation
GPS	Global Positioning System
GSM	Global System for Mobile Communications
GSMA	GSM Association
GTSRB	German Traffic Sign Recognition Benchmark
HCI	Human-computer interaction
HDI	Human-data interaction
ICCV***	International Conference on Computer Vision
ICT	Information and communications technology
IEEE***	Institute of Electrical and Electronics Engineers
ILPD	Indian Liver Patient Dataset
Ionos	The Ionosphere Dataset
IP	Integer programming
IRB	Institutional review board
ISLVR	ImageNet Large-Scale Visual Recognition Challenge
kNN	The k-nearest neighbor algorithm in SPSS
Liver	BUPA Liver Disorders Dataset
LFW	Labeled Faces in the Wild
LP	Linear programming
LSP	Leeds Sports Pose
MCDNN	Multicolumn deep neural network
MNIST	Mixed National Institute of Standards and Technology
NP-hard	Non-deterministic Polynomial-time
NSA	National Security Agency
NHS	National Health Service
NIPS***	Neural information processing systems
NPR	National Public Radio
Parkin	Parkinson's Dataset
PETs	Privacy-enhancing technologies
PGP	Pretty Good Privacy
Pima	Pima Indian Diabetes Dataset
PPTCs	Privacy policy terms and conditions

QII	Quantitative input influence
RTDNA	Radio Television Digital News Association
RuB	Boosting approach rule learner
RuSC	Set covering approach rule learner
SCM	Set covering machine
SDNY	United States District Court for the Southern District of New York
Sonar	Connectionist bench sonar dataset
SQGT	Semiquantitative group testing
SRF	Schweizer Radio und Fernsehen
SVM	Support vector machine
SUS	Secondary Uses Service
T3	Tastes, Ties, and Time
t-SNE	Stochastic neighbor embedding
TGT	Threshold group testing
ToS	Terms of service
Trans	Blood Transfusion Service Center Dataset
TrBag	The random forests classifier in MATLAB's Tree-Bagger class
UCI	University of California, Irvine
VOC	Visual object classes
WAF	We Are Family
WDBC	Wisconsin Diagnostic Breast Cancer Dataset
WEF	World Economic Forum
WPF	World Privacy Forum
YTF	YouTube Faces

Note that acronyms marked with \*\*\* are never used without the long form in the text.