

LNAI Series Editors

Randy Goebel

University of Alberta, Edmonton, Canada

Yuzuru Tanaka

Hokkaido University, Sapporo, Japan

Wolfgang Wahlster

DFKI and Saarland University, Saarbrücken, Germany

LNAI Founding Series Editor

Joerg Siekmann

DFKI and Saarland University, Saarbrücken, Germany

More information about this series at <http://www.springer.com/series/1244>

Stefan Michaelis · Nico Piatkowski
Marco Stolpe (Eds.)

Solving Large Scale Learning Tasks

Challenges and Algorithms

Essays Dedicated to Katharina Morik
on the Occasion of Her 60th Birthday

Editors
Stefan Michaelis
TU Dortmund
Dortmund
Germany

Marco Stolpe
TU Dortmund
Dortmund
Germany

Nico Piatkowski
TU Dortmund
Dortmund
Germany

Cover illustration: The illustration appearing on the cover belongs to Katharina Morik. Used with permission.

Photograph on p. V: The photograph of the honoree was taken by Jürgen Huhn. Used with permission.

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Artificial Intelligence
ISBN 978-3-319-41705-9 ISBN 978-3-319-41706-6 (eBook)
DOI 10.1007/978-3-319-41706-6

Library of Congress Control Number: 2016942885

LNCS Sublibrary: SL7 – Artificial Intelligence

© Springer International Publishing Switzerland 2016

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 International Publishing AG Switzerland



Foreword

The German word “Festschrift” has made it into international dictionaries because it very succinctly denotes a volume of writings by different researchers that originate from an event and constitutes a tribute to a scholar of extraordinary reputation. As such, a Festschrift offers a unique approach toward a field of science, since at its center, instead of an a priori-defined topical focus, there are the works and scientific vision of an outstanding individual as reflected in the works of collaborators and contributors to the volume.

While this nature of a Festschrift makes it an interesting approach independent of what the field of science is, in the field of machine learning this way of accessing science is of particular interest, since in a certain sense it reflects the very nature of the field itself. In the present volume, which originated at the scientific symposium in honor of Katharina Morik’s 60th birthday, you will see that the individual contributions of her colleagues offer an implicit view of her strategic vision of what machine learning should be and how research in machine learning should be conducted, as reflected in her choice of collaborators. You are thus invited to do what any good machine-learning algorithm would do when presented with examples: use the research presented in this book to induce for yourselves the implicit vision that lies at their heart.

In this foreword, I certainly do not want to take away from the pleasure of drawing these conclusions yourselves, so let me just say that in my view, the papers clearly reflect Katharina Morik’s commitment and conviction that machine learning should be firmly rooted in fundamental research with all its rigor, while at the same time being turned into software and engineering results and demonstrating its usefulness by applications in various disciplines. As you will see, this vision is clearly shared by the excellent researchers who have contributed to this volume.

Enjoy the book!

December 2015

Stefan Wrobel

Preface

In celebration of Prof. Morik's 60th birthday, this Festschrift covers research areas that Prof. Morik worked in and presents various researchers with whom she collaborated. Articles in this Festschrift volume provide challenges and solutions from theoreticians and practitioners on data preprocessing, modeling, learning, and evaluation. Topics include data-mining and machine-learning algorithms, feature selection, optimization as well as efficiency of energy and communication.

March 2016

Stefan Michaelis
Nico Piatkowski
Marco Stolpe

Biographical Details

Katharina Morik was born in 1954. She earned her PhD (1981) at the University of Hamburg and her habilitation (1988) at the TU Berlin. In 1991, Katharina became a full professor of computer science at the TU Dortmund University (former Universität Dortmund), Germany. Starting with natural language processing, her interest moved to machine learning ranging from inductive logic programming to statistical learning, then to the analysis of very large data collections, high-dimensional data, and resource awareness. She is a member of the National Academy of Science and Engineering and the North Rhine-Westphalia Academy of Science and Art. She is the author of more than 200 papers in acknowledged conferences and journals. Her latest results include spatio-temporal random fields and integer Markov random fields, both allowing for complex probabilistic graphical models under resource constraints.

Throughout her career, Katharina has been passionate about teaching. She has often taught more courses than required, and inspired students with her passion for artificial intelligence and computer science in general.

Her aim to share scientific results strongly supports open source developments. For instance, the first efficient implementation of the support vector machine, SVM_{light}, was developed at her lab by Thorsten Joachims. The leading data-mining platform RapidMiner also started out at her lab, which continues to contribute to it. Currently, the Java streams framework is being developed, which abstracts processes on distributed data streams.

Since 2011, she has been leading the collaborative research center SFB876 on resource-aware data analysis, an interdisciplinary center comprising 14 projects, 20 professors, and about 50 PhD students or postdocs.

Katharina was and is strongly engaged in the data mining and machine learning community. She was one of the founders of the IEEE International Conference on Data Mining together with Xindong Wu, and she chaired the program of this conference in 2004. She was the program chair of the European Conference on Machine Learning (ECML) in 1989 and one of the program chairs of ECML PKDD 2008. Katharina is on the editorial boards of the international journals *Knowledge and Information Systems* and *Data Mining and Knowledge Discovery*.

Contents

Online Social Networks Event Detection: A Survey.	1
<i>Mário Cordeiro and João Gama</i>	
Detecting Events in Online Social Networks: Definitions, Trends and Challenges	42
<i>Nikolaos Panagiotou, Ioannis Katakis, and Dimitrios Gunopulos</i>	
Why Do We Need Data Privacy?	85
<i>Volker Klingspor</i>	
Sharing Data with Guaranteed Privacy.	96
<i>Arno Siebes</i>	
Distributed Support Vector Machines: An Overview	109
<i>Marco Stolpe, Kanishka Bhaduri, and Kamalika Das</i>	
Big Data Classification – Aspects on Many Features	139
<i>Claus Weihs</i>	
Knowledge Discovery from Complex High Dimensional Data	148
<i>Sangkyun Lee and Andreas Holzinger</i>	
Local Pattern Detection in Attributed Graphs	168
<i>Jean-François Boullicaut, Marc Plantevit, and Céline Robardet</i>	
Advances in Exploratory Pattern Analytics on Ubiquitous Data and Social Media	184
<i>Martin Atzmüller</i>	
Understanding Human Mobility with Big Data	208
<i>Fosca Giannotti, Lorenzo Gabrielli, Dino Pedreschi, and Salvatore Rinzivillo</i>	
On Event Detection from Spatial Time Series for Urban Traffic Applications.	221
<i>Gustavo Souto and Thomas Liebig</i>	
Compressible Reparametrization of Time-Variant Linear Dynamical Systems	234
<i>Nico Piatkowski and François Schnitzler</i>	

Detection of Local Intensity Changes in Grayscale Images with Robust Methods for Time-Series Analysis.	251
<i>Sermad Abbas, Roland Fried, and Ursula Gather</i>	
SCHEP — A Geometric Quality Measure for Regression Rule Sets, Gauging Ranking Consistency Throughout the Real-Valued Target Space . . .	272
<i>Wouter Duivesteijn and Marvin Meeng</i>	
Bayesian Ordinal Aggregation of Peer Assessments: A Case Study on KDD 2015.	286
<i>Thorsten Joachims and Karthik Raman</i>	
Collaborative Online Learning of an Action Model	300
<i>Christophe Rodrigues, Henry Soldano, Gauvain Bourgne, and Céline Rouveirol</i>	
Ontology-Based Classification – Application of Machine Learning Concepts Without Learning	320
<i>Thomas Hoppe</i>	
Deep Distant Supervision: Learning Statistical Relational Models for Weak Supervision in Natural Language Extraction.	331
<i>Sriraam Natarajan, Ameet Soni, Anurag Wazalwar, Dileep Viswanathan, and Kristian Kersting</i>	
Supervised Extraction of Usage Patterns in Different Document Representations.	346
<i>Christian Pölitz</i>	
Data-Driven Analyses of Electronic Text Books	362
<i>Ahcène Boubekki, Ulf Kröhne, Frank Goldhammer, Waltraud Schreiber, and Ulf Brefeld</i>	
k-Morik: Mining Patterns to Classify Cartified Images of Katharina	377
<i>Elisa Fromont and Bart Goethals</i>	
Author Index	387