Lecture Notes in Computer Science 9897

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

More information about this series at http://www.springer.com/series/7409

Henrik Boström · Arno Knobbe
Carlos Soares · Panagiotis Papapetrou (Eds.)

Advances in Intelligent Data Analysis XV

15th International Symposium, IDA 2016
Stockholm, Sweden, October 13–15, 2016
Proceedings

Springer

Editors
Henrik Boström
Stockholm University
Stockholm
Sweden

Arno Knobbe
Leiden University
Leiden
The Netherlands

Carlos Soares
University of Porto
Porto
Portugal

Panagiotis Papapetrou
Stockholm University
Stockholm
Sweden

ISSN 0302-9743 ISSN 1611-3349 (electronic)
Lecture Notes in Computer Science
ISBN 978-3-319-46348-3 ISBN 978-3-319-46349-0 (eBook)
DOI 10.1007/978-3-319-46349-0

Library of Congress Control Number: 2016950907

LNCS Sublibrary: SL3 – Information Systems and Applications, incl. Internet/Web, and HCI

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

We are proud to present the proceedings of the 15th International Symposium on Intelligent Data Analysis, which took place during October 13–15 in Stockholm, Sweden. The series started in 1995 and was held biennially until 2009. In 2010, the symposium re-focused to support papers that go beyond established technology and offer genuinely novel and game-changing ideas, while not always being as fully realized as papers submitted to other conferences.

IDA 2016 continued this approach and sought first-look papers that might elsewhere be considered preliminary, but contain potentially high-impact research. In addition, for the first time this year, IDA introduced an industrial challenge track. For the industrial challenge, researchers were invited to participate in a machine learning prediction challenge, where the task was to devise a prediction model for judging whether or not a vehicle faces imminent failure of a specific component, exploiting data collected from heavy Scania trucks in everyday usage.

The IDA symposium is open to all kinds of modelling and analysis methods, irrespective of discipline. It is an interdisciplinary meeting that seeks abstractions that cut across domains. IDA solicits papers on all aspects of intelligent data analysis, including papers on intelligent support for modelling and analyzing data from complex, dynamical systems.

Intelligent support for data analysis goes beyond the usual algorithmic offerings in the literature. Papers about established technology were only accepted if the technology was embedded in intelligent data analysis systems, or was applied in novel ways to analyzing and/or modelling complex systems. The conventional reviewing process, which favors incremental advances on established work, can discourage the kinds of papers that were selected for IDA 2016. The reviewing process addressed this issue explicitly: referees evaluated papers against the stated goals of the symposium, and any paper for which at least one program committee advisor wrote an informed, thoughtful, positive review was accepted, irrespective of other reviews. Indeed, this had a notable impact on what papers were included in the program.

We were pleased to see a very strong program. We received 75 submissions by 198 authors from 30 different countries, out of which 15 were accepted as regular papers, 12 as regular poster papers, and 4 as short papers (industrial challenge papers). All submissions were reviewed by three PC members and one PC advisor.

In addition, we were happy to accept two abstracts to the IDA horizon track:

– “Usable analytics at societal scale”, by Daniel Gillblad
– “Cognitive Computing for the Automated Society”, by Devdatt Dubhashi

We were honored to have the following distinguished invited speakers at IDA 2016:

- Samuel Kaski, Aalto University and University of Helsinki, Finland; on the topic “Bayesian Factorization of Multiple Data Sources”

- Sihem Amer Yahia, CNRS at LIG, Grenoble, France; on the topic "Worker-Centricity Could Be Today's Disruptive Innovation in Crowdsourcing"
- Foster Provost, New York University, USA; on the topic "The Predictive Power of Massive Data about Our Fine-Grained Behavior".

The conference was held at the Department of Computer and Systems Sciences of Stockholm University, Sweden.

We wish to express our gratitude to all authors of submitted papers for their intellectual contributions; to the program committee members and advisors and additional reviewers for their effort in reviewing, discussing, and commenting on the submitted papers, and to the members of the IDA steering committee for their ongoing guidance and support. We thank Isak Karlsson for running the conference website. Special thanks go to the industrial challenge chair, Tony Lindgren, for handling the submission and reviewing process of the industrial challenge papers. We gratefully acknowledge those who were involved in the local organization of the symposium: Lars Asker, Isak Karlsson, Jing Zhao, and Ram Gurung. We are grateful to our sponsors: Stockholm University, Scania AB, Vetenskapsrådet, Springer, The Artificial Intelligence Journal, and SERSC. We are especially indebted to KNIME, who funded the IDA Frontier Prize for the most visionary contribution presenting a novel and surprising approach to data analysis in the understanding of complex systems.

July 2016

Henrik Boström
Arno Knobbe
Carlos Soares
Panagiotis Papapetrou

Organization

General Chair

Panagiotis Papapetrou	Stockholm University, Sweden

Program Committee Chairs

Henrik Boström	Stockholm University, Sweden
Arno Knobbe	Leiden University, The Netherlands
Carlos Soares	University of Porto, Portugal

Local Chair

Lars Asker	Stockholm University, Sweden

Industrial Challenge Chair

Tony Lindgren	Stockholm University, Sweden

Sponsorship and Publicity Chairs

Lars Asker	Stockholm University, Sweden
Panagiotis Papapetrou	Stockholm University, Sweden

Advisory Chairs

Joost Kok	Leiden University, The Netherlands
Jaakko Hollmen	Aalto University, Finland
Matthijs van Leeuween	Leiden University, The Netherlands

Webmaster

Isak Karlsson	Stockholm University, Sweden

Program Committee Advisors

Stephen Swift	Brunel University, UK
Allan Tucker	Brunel University, UK
Michael Berthold	University of Konstanz, Germany
Matthijs van Leeuwen	Leiden University, The Netherlands
Elisa Fromont	University of Saint-Etienne, France

Arno Siebes	Universiteit Utrecht, The Netherlands
Hannu Toivonen	University of Helsinki, Finland
Nada Lavrac	Jozef Stefan Institute, Slovenia
Xiaohui Liu	Brunel University, UK
Elizabeth Bradley	University of Colorado, USA
Hendrik Blockeel	K.U. Leuven, Belgium
Frank Klawonn	Ostfalia University of Applied Sciences, Germany
Jaakko Hollmen	Aalto University School of Science, Finland
Tijl De Bie	Ghent University, Belgium

Program Committee

Wouter Duivesteijn	Ghent University, Belgium
Mykola Pechenizkiy	Eindhoven University of Technology, The Netherlands
Jefrey Lijffijt	Ghent University, Belgium
Lubos Popelinsky	Masaryk University, Czech Republic
Alexandra Poulovassilis	Birkbeck, University of London, UK
Saso Dzeroski	Jozef Stefan Institute, Slovenia
Christine Solnon	LIRIS CNRS UMR 5205/INSA Lyon, France
Nicos Pavlidis	Lancaster University, UK
Marc Plantevit	LIRIS - Université Claude Bernard Lyon 1, France
Maguelonne Teisseir	Cemagref - UMR Tetis, France
Albrecht Zimmermann	University of Normandy, France
George Magoulas	Birkbeck, University of London, UK
Ruggero G. Pensa	University of Turin, Italy
Andre Carvalho	USP, Brazil
Maarten Van Someren	University of Amsterdam, The Netherlands
Frank Takes	Leiden University, The Netherlands
Ricardo Cachucho	Leiden University, The Netherlands
Antonio Salmeron	University of Almeria, Spain
Wannes Meert	KU Leuven, Belgium
Joaquin Vanschoren	TU Eindhoven, The Netherlands
Indre Zliobaite	Aalto University, Finland
Martin Atzmueller	University of Kassel, Germany
Rudolf Kruse	University of Magdeburg, Germany
Paulo Cortez	University of Minho, Portugal
Brett Drury	LIAAD – INESC, Portugal
Jan N. van Rijn	Leiden University, The Netherlands
Vera Oliveira	University of Porto, Portugal
Fabrizio Angiulli	University of Calabria, Italy
Mohamed Nadif	University of Paris Descartes, France
Kaustubh Patil	MIT, USA
Nuno Escudeiro	ISEP - Instituto Superior de Engenharia do Porto, Portugal
Ana Aguiar	FEUP, Portugal
Roberta Siciliano	University of Naples Federico II, Italy

Kenny Gruchalla	NREL/CU-Boulder, USA
Miguel A. Prada	Universidad de Leon, Spain
Myra Spilliopoulou	Otto-von-Guericke-University of Magdeburg, Germany
Paula Brito	University of Porto, Portugal
Andreas Nuernberger	Otto-von-Guericke University of Magdeburg, Germany
Giovanni Montana	Imperial College, UK
Ricard Gavalda	Universitat Politècnica de Catalunya, Spain
Peter van der Putten	Leiden University and Pegasystems, The Netherlands
Loic Cerf	Universidade Federal de Minas Gerais, Brazil
Bernard De Baets	Ghent University, Belgium
Jose-Maria Pena	Universidad Politècnica de Madrid, Spain
Anton Dries	KU Leuven, Belgium
Johannes Furnkranz	TU Darmstadt, Germany
Alipio M. Jorge	University of Porto, Portugal
Antti Ukkonen	Finnish Institute of Occupational Health, Finland
Thibault Sellam	CWI, The Netherlands
Fabrizio Riguzzi	University of Ferrara, Italy
Gustavo Batista	University of Sao Paulo, Brazil
Ulf Brefeld	Leuphana University of Lüneburg, Germany
Niklas Lavesson	Blekinge Institute of Technology, Sweden
Jose A. Lozano	The University of the Basque Country, Spain
Jose Del Campo	Universidad de Málaga, Spain
Frank Hoppner	Ostfalia University of Applied Sciences, Germany
Ingrid Fischer	University of Konstanz, Germany
Ad Feelders	Universiteit Utrecht, The Netherlands
Wojtek Kowalczyk	Leiden University, The Netherlands
Bruno Cremilleux	Université de Caen, France
Maria Bielikova	Slovak University of Technology in Bratislava, Slovakia
Javier Gonzalez	University of Sheffield, UK
Harm de Vries	Leiden University, The Netherlands
Irena Koprinska	The University of Sydney, Australia
Vitor Santos Costa	Universidade do Porto, Portugal
Jesse Read	Aalto University, Espoo, Finland
Cor Veenman	Netherlands Forensic Institute, The Netherlands
Peter Flach	University of Bristol, UK
Adolfo Martinez-Uso	Technical University of Valencia, Spain
Lawrence Hall	University of South Florida, USA
François Portet	University of Grenoble Alpes, France
Jose Balcazar	Universitat Politècnica de Catalunya, Spain
Tias Guns	KU Leuven, Belgium
Douglas Fisher	Vanderbilt University, USA
Norbert Jankowski	Nicolaus Copernicus University, Poland
Eirini Ntoutsi	Leibniz University Hanover, Germany

Sponsors and Supporters

- SCANIA
- The Swedish Research Council
- KNIME
- Springer
- The Artificial Intelligence Journal
- Stockholm University
- SERSC

Contents

DSCo-NG: A Practical Language Modeling Approach for Time Series Classification 1
Daoyuan Li, Tegawendé F. Bissyandé, Jacques Klein, and Yves Le Traon

Ranking Accuracy for Logistic-GEE Models 14
Nasser Davarzani, Ralf Peeters, Evgueni Smirnov, Joël Karel, and Hans-Peter Brunner-La Rocca

The Morality Machine: Tracking Moral Values in Tweets 26
Livia Teernstra, Peter van der Putten, Liesbeth Noordegraaf-Eelens, and Fons Verbeek

A Hybrid Approach for Probabilistic Relational Models Structure Learning . . . 38
Mouna Ben Ishak, Philippe Leray, and Nahla Ben Amor

On the Impact of Data Set Size in Transfer Learning Using Deep Neural Networks 50
Deepak Soekhoe, Peter van der Putten, and Aske Plaat

Obtaining Shape Descriptors from a Concave Hull-Based Clustering Algorithm 61
Christian Braune, Marco Dankel, and Rudolf Kruse

Visual Perception of Discriminative Landmarks in Classified Time Series . . . 73
Tobias Sobek and Frank Höppner

Spotting the Diffusion of New Psychoactive Substances over the Internet . . . 86
Fabio Del Vigna, Marco Avvenuti, Clara Bacciu, Paolo Deluca, Marinella Petrocchi, Andrea Marchetti, and Maurizio Tesconi

Feature Selection Issues in Long-Term Travel Time Prediction 98
Syed Murtaza Hassan, Luis Moreira-Matias, Jihed Khiari, and Oded Cats

A Mean-Field Variational Bayesian Approach to Detecting Overlapping Communities with Inner Roles Using Poisson Link Generation 110
Gianni Costa and Riccardo Ortale

Online Semi-supervised Learning for Multi-target Regression in Data Streams Using AMRules 123
Ricardo Sousa and João Gama

A Toolkit for Analysis of Deep Learning Experiments 134
Jim O'Donoghue and Mark Roantree

The Optimistic Method for Model Estimation 146
James Brofos, Rui Shu, and Frank Zhang

Does Feature Selection Improve Classification? A Large Scale Experiment in OpenML 158
Martijn J. Post, Peter van der Putten, and Jan N. van Rijn

Learning from the News: Predicting Entity Popularity on Twitter 171
Pedro Saleiro and Carlos Soares

Multi-scale Kernel PCA and Its Application to Curvelet-Based Feature Extraction for Mammographic Mass Characterization 183
Sami Dhahbi, Walid Barhoumi, and Ezzeddine Zagrouba

Weakly-Supervised Symptom Recognition for Rare Diseases in Biomedical Text 192
Pierre Holat, Nadi Tomeh, Thierry Charnois, Delphine Battistelli, Marie-Christine Jaulent, and Jean-Philippe Métivier

Estimating Sequence Similarity from Read Sets for Clustering Sequencing Data 204
Petr Ryšavý and Filip Železný

Widened Learning of Bayesian Network Classifiers 215
Oliver R. Sampson and Michael R. Berthold

Vote Buying Detection via Independent Component Analysis 226
Antonio Neme and Omar Neme

Unsupervised Relation Extraction in Specialized Corpora Using Sequence Mining 237
Kata Gábor, Haïfa Zargayouna, Isabelle Tellier, Davide Buscaldi, and Thierry Charnois

A Framework for Interpolating Scattered Data Using Space-Filling Curves . . . 249
David J. Weston

Privacy-Awareness of Distributed Data Clustering Algorithms Revisited 261
Josenildo C. da Silva, Matthias Klusch, and Stefano Lodi

Bi-stochastic Matrix Approximation Framework for Data Co-clustering 273
Lazhar Labiod and Mohamed Nadif

Sequential Cost-Sensitive Feature Acquisition 284
Gabriella Contardo, Ludovic Denoyer, and Thierry Artières

Explainable and Efficient Link Prediction in Real-World Network Data 295
Jesper E. van Engelen, Hanjo D. Boekhout, and Frank W. Takes

DGRMiner: Anomaly Detection and Explanation in Dynamic Graphs 308
Karel Vaculík and Luboš Popelínský

Similarity Based Hierarchical Clustering with an Application to Text Collections 320
Julien Ah-Pine and Xinyu Wang

Determining Data Relevance Using Semantic Types and Graphical Interpretation Cues 332
Eduardo Haruo Kamioka, André Freitas, Frederico Caroli, and Siegfried Handschuh

A First Step Toward Quantifying the Climate's Information Production over the Last 68,000 Years. 343
Joshua Garland, Tyler R. Jones, Elizabeth Bradley, Ryan G. James, and James W.C. White

HAUCA Curves for the Evaluation of Biomarker Pilot Studies with Small Sample Sizes and Large Numbers of Features. 356
Frank Klawonn, Junxi Wang, Ina Koch, Jörg Eberhard, and Mohamed Omar

Stability Evaluation of Event Detection Techniques for Twitter. 368
Andreas Weiler, Joeran Beel, Bela Gipp, and Michael Grossniklaus

IDA 2016 Industrial Challenge: Using Machine Learning for Predicting Failures 381
Camila Ferreira Costa and Mario A. Nascimento

An Optimized k-NN Approach for Classification on Imbalanced Datasets with Missing Data. 387
Ezgi Can Ozan, Ekaterina Riabchenko, Serkan Kiranyaz, and Moncef Gabbouj

Combining Boosted Trees with Metafeature Engineering for Predictive Maintenance. 393
Vítor Cerqueira, Fábio Pinto, Claudio Sá, and Carlos Soares

Prediction of Failures in the Air Pressure System of Scania Trucks Using a Random Forest and Feature Engineering 398
Christopher Gondek, Daniel Hafner, and Oliver R. Sampson

Author Index 403