

Commenced Publication in 1973

Founding and Former Series Editors:

Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison

Lancaster University, UK

Takeo Kanade

Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler

University of Surrey, Guildford, UK

Jon M. Kleinberg

Cornell University, Ithaca, NY, USA

Alfred Kobsa

University of California, Irvine, CA, USA

Friedemann Mattern

ETH Zurich, Switzerland

John C. Mitchell

Stanford University, CA, USA

Moni Naor

Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz

University of Bern, Switzerland

C. Pandu Rangan

Indian Institute of Technology, Madras, India

Bernhard Steffen

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Jaakko Hollmén Frank Klawonn
Allan Tucker (Eds.)

Advances in Intelligent Data Analysis XI

11th International Symposium, IDA 2012
Helsinki, Finland, October 25-27, 2012
Proceedings

Volume Editors

Jaakko Hollmén
Aalto University School of Science
Department of Information and Computer Science
P.O. Box 15400, 00076 Aalto, Finland
E-mail: jaakko.hollmen@aalto.fi

Frank Klawonn
Ostfalia University of Applied Sciences
Department of Computer Science
Salzdahlumer Straße 46/48, 38302 Wolfenbüttel, Germany
E-mail: f.klawonn@ostfalia.de

Allan Tucker
Brunel University, School of Information Systems
Computing and Mathematics
Uxbridge, Middlesex UB8 3PH, UK
E-mail: allan.tucker@brunel.ac.uk

ISSN 0302-9743
ISBN 978-3-642-34155-7
DOI 10.1007/978-3-642-34156-4
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349
e-ISBN 978-3-642-34156-4

Library of Congress Control Number: 2012949063

CR Subject Classification (1998): H.3, H.4, I.2, F.1, H.2.8, J.3, I.4-5

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

© Springer-Verlag Berlin Heidelberg 2012

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

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

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

We are proud to present the proceedings of the 11th International Symposium on Intelligent Data Analysis, which was held during October 25–27 in Helsinki, Finland. The series started in 1995 and was held biennially until 2009. In 2010, the symposium re-focussed 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 2012 continued this approach and included an important and still emerging class of problems: the analysis of data from networked digital information systems such as mobile devices, remote sensors, and streaming applications.

The symposium seeks first-look papers that might elsewhere be considered preliminary, but contain potentially high-impact research. The IDA Symposium, which is A-ranked according to ERA, 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. IDA 2012 particularly encouraged papers about:

Novel applications of IDA techniques to, for example:

- Networked digital information systems
- Novel modes of data acquisition and the associated issues
- Robustness and scalability issues of intelligent data analysis techniques
- Visualization and dissemination of results

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 IDA 2012 has published. 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 Chair advisor wrote an informed, thoughtful, positive review was accepted, irrespective of other reviews.

We were pleased to have a very strong program. We received 88 submissions from people in 24 countries on five continents.

As in IDA 2011, we included a poster session for PhD students to promote their work and for the first time we also introduced the use of a 2-minute video slot for all PhD posters and standard posters.

We were honored to have distinguished invited speakers at IDA 2012: Arno Siebes from the University of Utrecht in The Netherlands, who talked about ways of getting multiple good models that complement each other in the insight

they give. Paola Sebastiani from Boston University in the USA, who talked about the intelligent data analysis of human genetic data. Gavin Cawley from the University of East Anglia in the UK, who talked about avoiding over-fitting in model selection.

The conference was held at Finlandia Hall in Helsinki. We wish to express our gratitude to all authors of submitted papers for their intellectual contributions; to the Program Committee members and the additional reviewers for their effort in reviewing, discussing, and commenting on the submitted papers; to the members of the IDA Steering Committee for their ongoing guidance and support; and to the Senior Program Committee for their active involvement. We thank Richard van de Stadt for running the submission website and handling the production of the proceedings. Special thanks go to the Poster Chair, Frank Höppner, and the Frontier Prize Chairs, Elizabeth Bradley and João Gama. We gratefully acknowledge those who were involved in the local organization of the symposium: Anu Juslin, Outi Elina Kansanen, Mikko Korpela, Janne Toivola, Prem Raj Adhikari, and Olli-Pekka Rinta-Koski. The help of volunteers during the IDA 2012 symposium is also thankfully acknowledged.

We are grateful for our sponsors: Aalto University, the Helsinki Institute for Information technology (HIIT), and Algodan — Finnish Centre of Excellence for Algorithmic Data Analysis Research. 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.

August 2012

Jaakko Hollmén
Frank Klawonn
Allan Tucker

Organization

General Chair

Jaakko Hollmén	Aalto University, Finland
----------------	---------------------------

Program Chairs

Frank Klawonn	Ostfalia University of Applied Sciences, Germany
Allan Tucker	Brunel University, UK

Poster Chair

Frank Höppner	Ostfalia University of Applied Sciences, Germany
---------------	---

Frontier Prize Chairs

Elizabeth Bradley	University of Colorado, USA
João Gama	University of Porto, Portugal

Local Organizing Committee

Anu Juslin	Aalto University, Finland
Outi Elina Kansanen	Aalto University, Finland
Mikko Korpela	Aalto University, Finland
Janne Toivola	Aalto University, Finland
Prem Raj Adhikari	Aalto University, Finland
Olli-Pekka Rinta-Koski	Aalto University, Finland

Program Chair Advisors

Michael Berthold	University of Konstanz, Germany
Liz Bradley	University of Colorado, USA
João Gama	University of Porto, Portugal
Jaakko Hollmén	Aalto University, Finland
Frank Höppner	Ostfalia University of Applied Sciences, Germany
Joost Kok	Leiden University, The Netherlands
Xiaohui Liu	Brunel University, UK

Arno Siebes
Hannu Toivonen
David Weston

Universiteit Utrecht, The Netherlands
University of Helsinki, Finland
Imperial College, UK

Program Committee

Fabrizio Angiulli
Alexandre Aussem
Christian Borgelt
Henrik Boström
Andre de Carvalho
Jose del Campo
Bruno Crémilleux
Werner Dubitzky
Sašo Džeroski
Fazel Famili

Ad Feelders
Ingrid Fischer
Élisa Fromont
Johannes Fürnkranz
Alex Gammernan
Ricard Gavalda
Kenny Gruchalla
Lawrence Hall
Patrik Hoyer
Eyke Hüllermeier
Wesley Kerr
Rudolf Kruse

Nada Lavrač
Jose Lozano
George Magoulas
Maria-Carolina Monard
Giovanni Montana
Mohamed Nadif
Detlef Nauck
Andreas Nürnberger

Panagiotis Papapetrou
Mykola Pechenizkiy

José-Maria Peña
Ruggero Pensa
Alexandra Poulouvassilis

University of Calabria, Italy
University of Lyon 1, France
European Centre for Soft Computing, Spain
Stockholm University, Sweden
University of Sao Paulo, Brazil
University of Malaga, Spain
University of Caen, France
University of Ulster, UK
Jožef Stefan Institute, Slovenia
IIT - National Research Council Canada,
Canada
Utrecht University, The Netherlands
University of Konstanz, Germany
University of Jean Monnet, France
Technische Universität Darmstadt, Germany
University of London, UK
Technical University of Catalonia, Spain
National Renewable Energy Lab, USA
University of South Florida, USA
University of Helsinki, Finland
University of Marburg, Germany
Adknowledge Inc., USA
Otto von Guericke University Magdeburg,
Germany
Jožef Stefan Institute, Slovenia
The University of the Basque Country, Spain
Birkbeck College, University of London, UK
University of Sao Paulo, Brazil
Imperial College, UK
Descartes University of Paris, France
British Telecom, UK
Otto von Guericke University Magdeburg,
Germany
Aalto University, Finland
Eindhoven University of Technology,
The Netherlands
Polytechnic University of Madrid, Spain
University of Turin, Italy
University of London, UK

Miguel A. Prada	University of Leon, Spain
Ronaldo Prati	Universidade Federal do ABC, Brazil
Fabrizio Riguzzi	ENDIF - University of Ferrara, Italy
Stefan Rüping	Fraunhofer IAIS, Germany
Antonio Salmerón Cerdán	University of Almeria, Spain
Vítor Santos Costa	University of Porto, Portugal
Roberta Siciliano	University of Naples Federico II, Italy
Maarten van Someren	Universiteit van Amsterdam, The Netherlands
Myra Spiliopoulou	Otto von Guericke University Magdeburg, Germany
Stephen Swift	Brunel University, UK
Maguelonne Teisseire	TETIS - Irstea, France
Evimaria Terzi	Boston University, USA
Panayiotis Tsaparas	University of Ioannina, Greece
Antti Ukkonen	Yahoo! Research Barcelona, Spain
Antony Unwin	University of Augsburg, Germany
Veronica Vinciotti	Brunel University, UK
Zidong Wang	Brunel University, UK

Additional Referees

Darko Aleksovski	Dragi Koccev
Satrajit Basu	John N. Korecki
Igor Braga	Mikko Korpela
Christian Braune	Antonio LaTorre
Bertrand Cuissart	Thomas Low
Jeroen De Knijf	Gjorgji Madjarov
Ivica Dimitrovski	Christian Moewes
Fabio Fasseti	Jesus Montes
Tatiana Gossen	Santiago Muelas
Pascal Held	Ilia Nouretdinov
Dino Ienco	Luigi Palopoli
R.P. Jagadeesh Chandra Bose	Hai Nhat Phan
Baptiste Jeudy	Simona E. Rombo
Julia Kiseleva	Bernard Zenko
Arto Klami	

Table of Contents

Invited Papers

Over-Fitting in Model Selection and Its Avoidance (Abstract)	1
<i>Gavin C. Cawley</i>	
Intelligent Data Analysis of Human Genetic Data	2
<i>Paola Sebastiani</i>	
Queries for Data Analysis	7
<i>Arno Siebes</i>	

Selected Contributions

Parallel Data Mining Revisited. Better, Not Faster	23
<i>Zaenal Akbar, Violeta N. Ivanova, and Michael R. Berthold</i>	
Weighting Features for Partition around Medoids Using the Minkowski Metric	35
<i>Renato Cordeiro de Amorim and Trevor Fenner</i>	
On Initializations for the Minkowski Weighted K-Means	45
<i>Renato Cordeiro de Amorim and Peter Komisarczuk</i>	
A Skew- t -Normal Multi-level Reduced-Rank Functional PCA Model for the Analysis of Replicated Genomics Time Course Data	56
<i>Maurice Berk and Giovanni Montana</i>	
Methodological Foundation for Sign Language 3D Motion Trajectory Analysis	67
<i>Mehrez Boulares and Mohamed Jemni</i>	
Assembly Detection in Continuous Neural Spike Train Data	78
<i>Christian Braune, Christian Borgelt, and Sonja Grün</i>	
Online Techniques for Dealing with Concept Drift in Process Mining . . .	90
<i>Josep Carmona and Ricard Gavalda</i>	
An Evolutionary Based Clustering Algorithm Applied to Dada Compression for Industrial Systems	103
<i>Jun Chen, Mahdi Mahfouf, Chris Bingham, Yu Zhang, Zhijing Yang, and Michael Gallimore</i>	

Multi-label LeGo — Enhancing Multi-label Classifiers with Local Patterns	114
<i>Wouter Duivesteijn, Eneldo Loza Mencía, Johannes Fürnkranz, and Arno Knobbe</i>	
Discriminative Dimensionality Reduction Mappings	126
<i>Andrej Gisbrecht, Daniela Hofmann, and Barbara Hammer</i>	
Finding Interesting Contexts for Explaining Deviations in Bus Trip Duration Using Distribution Rules	139
<i>Alípio M. Jorge, João Mendes-Moreira, Jorge Freire de Sousa, Carlos Soares, and Paulo J. Azevedo</i>	
Curve Fitting for Short Time Series Data from High Throughput Experiments with Correction for Biological Variation	150
<i>Frank Klawonn, Nada Abidi, Evelin Berger, and Lothar Jänsch</i>	
Formalizing Complex Prior Information to Quantify Subjective Interestingness of Frequent Pattern Sets	161
<i>Kleanthis-Nikolaos Kontonassios and Tijl DeBie</i>	
MCut: A Thresholding Strategy for Multi-label Classification	172
<i>Christine Langeron, Christophe Moulin, and Mathias Géry</i>	
Improving Tag Recommendation Using Few Associations	184
<i>Matthijs van Leeuwen and Diyah Puspitaningrum</i>	
Identifying Anomalous Social Contexts from Mobile Proximity Data Using Binomial Mixture Models	195
<i>Eric Malmi, Juha Raitio, Oskar Kohonen, Krista Lagus, and Timo Honkela</i>	
Constrained Clustering Using SAT	207
<i>Jean-Philippe Métivier, Patrice Boizumault, Bruno Crémilleux, Mehdi Khiari, and Samir Loudni</i>	
Two-Stage Approach for Electricity Consumption Forecasting in Public Buildings	219
<i>Antonio Morán, Miguel A. Prada, Serafín Alonso, Pablo Barrientos, Juan J. Fuertes, Manuel Domínguez, and Ignacio Díaz</i>	
Online Predictive Model for Taxi Services	230
<i>Luís Moreira-Matias, João Gama, Michel Ferreira, João Mendes-Moreira, and Luís Damas</i>	
Analysis of Noisy Biosignals for Musical Performance	241
<i>Joonas Paalasmaa, David J. Murphy, and Ove Holmqvist</i>	

Information Theoretic Approach to Improve Performance of Networked Control Systems	253
<i>Marko Paavola, Mika Ruusunen, Aki Sorsa, and Kauko Leiviskä</i>	
Learning Pattern Graphs for Multivariate Temporal Pattern Retrieval	264
<i>Sebastian Peter, Frank Höppner, and Michael R. Berthold</i>	
GET_MOVE: An Efficient and Unifying Spatio-temporal Pattern Mining Algorithm for Moving Objects	276
<i>Phan Nhat Hai, Pascal Poncelet, and Maguelonne Teisseire</i>	
Fuzzy Frequent Pattern Mining in Spike Trains	289
<i>David Picado Muño, Iván Castro León, and Christian Borgelt</i>	
Mass Scale Modeling and Simulation of the Air-Interface Load in 3G Radio Access Networks	301
<i>Dejan Radosavljevik, Peter van der Putten, and Kim Kylesbech Larsen</i>	
Batch-Incremental versus Instance-Incremental Learning in Dynamic and Evolving Data	313
<i>Jesse Read, Albert Bifet, Bernhard Pfahringer, and Geoff Holmes</i>	
Applying Piecewise Approximation in Perceptron Training of Conditional Random Fields	324
<i>Teemu Ruokolainen</i>	
Patch-Based Data Analysis Using Linear-Projection Diffusion	334
<i>Moshe Salhov, Guy Wolf, Amir Averbuch, and Pekka Neittaanmäki</i>	
Dictionary Construction for Patch-to-Tensor Embedding	346
<i>Moshe Salhov, Guy Wolf, Amit Bermanis, Amir Averbuch, and Pekka Neittaanmäki</i>	
Where Are We Going? Predicting the Evolution of Individuals	357
<i>Zaigham Faraz Siddiqui, Márcia Oliveira, João Gama, and Myra Spiliopoulou</i>	
Use of General Purpose GPU Programming to Enhance the Classification of Leukaemia Blast Cells in Blood Smear Images	369
<i>Stefan Skrobanski, Stelios Pavlidis, Waidah Ismail, Rosline Hassan, Steve Counsell, and Stephen Swift</i>	
Intelligent Data Analysis by a Home-Use Human Monitoring Robot	381
<i>Shinsuke Sugaya, Daisuke Takayama, Asuki Kouno, and Einoshin Suzuki</i>	

Sleep Musicalization: Automatic Music Composition from Sleep
Measurements 392
*Aurora Tulilaulu, Joonas Paalasmaa, Mikko Waris, and
Hannu Toivonen*

Engine Parameter Outlier Detection: Verification by Simulating PID
Controllers Generated by Genetic Algorithm 404
*Joni Vesterback, Vladimir Bochko, Mika Ruohonen, Jarmo Alander,
Andreas Bäck, Martin Nylund, Allan Dal, and Fredrik Östman*

Unit Operational Pattern Analysis and Forecasting Using EMD and
SSA for Industrial Systems 416
*Zhijing Yang, Chris Bingham, Wing-Kuen Ling, Yu Zhang,
Michael Gallimore, and Jill Stewart*

Author Index 425