

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

University of Dortmund, 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 of Computer Science, Saarbruecken, Germany

Niall M. Adams Céline Robardet
Arno Siebes Jean-François Boulicaut (Eds.)

Advances in Intelligent Data Analysis VIII

8th International Symposium
on Intelligent Data Analysis, IDA 2009
Lyon, France, August 31 – September 2, 2009
Proceedings



Springer

Volume Editors

Niall M. Adams
Imperial College London
Department of Mathematics
South Kensington Campus, London SW7 2PG, UK
E-mail: n.adams@imperial.ac.uk

Céline Robardet
Jean-François Boulicaut
University of Lyon, INSA Lyon, LIRIS CNRS UMR 5205
Bâtiment Blaise Pascal, F-69621 Villeurbanne, France
E-mail: {celine.robardet,jean-francois.boulicaut}@insa-lyon.fr

Arno Siebes
University of Utrecht
Department of Information and Computer Science
Utrecht, The Netherlands
E-mail: Arno.Siebes@cs.uu.nl

Library of Congress Control Number: 2009933040

CR Subject Classification (1998): H.3, I.5, G.3, J.1, J.3, H.3.3, I.5.3

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

ISSN 0302-9743
ISBN-10 3-642-03914-6 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-03914-0 Springer Berlin Heidelberg New York

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.

springer.com

© Springer-Verlag Berlin Heidelberg 2009
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12745715 06/3180 5 4 3 2 1 0

Preface

The general theme of the Intelligent Data Analysis (IDA) Symposia is the intelligent use of computers in complex data analysis problems. The field has matured sufficiently that some re-consideration of our objectives was required in order to retain the distinctiveness of IDA. Thus, in addition to the more traditional algorithm- and application-oriented submissions, we sought submissions that specifically focus on aspects of the data analysis process. For example, interactive tools to guide and support data analysis in complex scenarios. With the increasing availability of automatically collected data, tools that intelligently support and assist human analysts are becoming important.

IDA-09, the 8th International Symposium on Intelligent Data Analysis, took place in Lyon from August 31 to September 2, 2009. The invited speakers were Paul Cohen (University of Arizona, USA) and Pablo Jensen (ENS Lyon, France). The meeting received more than 80 submissions. The Programme Committee selected 33 submissions for publication: 18 for full oral presentation, and 15 for poster and short oral presentation. Each contribution was evaluated by three experts and has been allocated 12 pages in the proceedings. The accepted papers cover a broad range of topics and applications, and include contributions on the refined focus of IDA.

The symposium was supported by INSA Lyon (Institut National des Sciences Appliquées de Lyon) and by IXXI (Rhône-Alpes region Complex Systems Institute). We also thank our generous sponsors, La Région Rhône-Alpes and Le Ministère de l'Enseignement Supérieur et de la Recherche. We would like to express our gratitude to the many people involved in the organization of the symposium and the reviewing of submissions. The local Organizing Committee was co-chaired by Guillaume Beslon and Céline Robardet, and the committee members were Loïc Cerf, Serge Fenet, Pierre-Nicolas Mougel and David Parsons (all at INSA Lyon, France). We thank them for making the conference an unforgettable event. We are grateful to Joaquina Labro from Insavator s.a., who managed the registration process. We are thankful for the support of the IDA council, especially for the advice of Michael Berthold, Joost Kok, Xiaohui Liu and José-Maria Peña. Finally we would like to thank Richard van de Stadt and Springer for preparing the proceedings.

August 2009

Niall Adams
Céline Robardet
Arno Siebes
Jean-François Boulicaut

Conference Organization

Conference Chair

Jean-François Boulicaut University of Lyon, France

Programme Committee

Niall Adams	
(PC Co-chair)	Imperial College London, UK
Fabrizio Angiulli	University of Calabria, Italy
Alexandre Aussem	University of Lyon, France
Tony Bagnall	University of East Anglia, UK
Riccardo Bellazzi	University of Pavia, Italy
Bettina Berendt	KU Leuven, Belgium
Daniel Berrar	Systems Biology Institute, Tokyo, Japan
Michael Berthold	University of Konstanz, Germany
Klemens Böhm	University of Karlsruhe, Germany
Christian Borgelt	European Centre for Soft Computing, Spain
Elizabeth Bradley	University of Colorado, USA
Pavel Brazdil	University of Porto, Portugal
Bruno Crémilleux	University of Caen, France
Werner Dubitzky	University of Ulster, UK
Sašo Džeroski	Jozef Stefan Institute, Slovenia
Fazel Famili	IIT - NRC, Canada
Jason Farquhar	University of Nijmegen, The Netherlands
Ad Feelders	University of Utrecht, The Netherlands
Ingrid Fischer	University of Konstanz, Germany
Eibe Frank	University of Waikato, New Zealand
Elisa Fromont	University of Saint-Etienne, France
Johannes Fürnkranz	TU Darmstadt, Germany
Alex Gammerman	University of London, UK
Gérard Govaert	TU Compiègne, France
Pilar Herrero	Polytechnic University of Madrid, Spain
Alexander Hinneburg	University of Halle, Germany
Frank Höppner	University of Applied Sciences, Germany
Jaakko Hollmén	Helsinki University of Technology, Finland
Eyke Hüllermeier	University of Marburg, Germany
Daniel Keim	University of Konstanz, Germany
Frank Klawonn	University of Wolfenbüttel, Germany
Jiri Klema	Czech Technical University, Czech Republic
Arno Knobbe	Leiden University, The Netherlands

VIII Organization

Joost Kok	Leiden University, The Netherlands
Peter Kokol	University of Maribor, Slovenia
Walter Kosters	Leiden University, The Netherlands
Paul Krause	University of Surrey, UK
Rudolf Kruse	University of Magdeburg, Germany
Evelina Lamma	University of Ferrara, Italy
Pedro Larranaga	TU Madrid, Spain
Mark Last	Ben-Gurion University of the Negev, Israel
Nada Lavrac	Jozef Stefan Institute, Slovenia
Steven Lemm	Fraunhofer FIRST Institute, Germany
Hans-J. Lenz	Free University Berlin, Germany
Xiaohui Liu	Brunel University, UK
Rainer Malaka	University of Bremen, Germany
Trevor Martin	University of Bristol, UK
Dunja Mladenč	Jozef Stefan Institute, Slovenia
Maria-Carolina Monard	University of Sao Paulo, Brasil
Clayton Morrison	University of Arizona, USA
Alberto Munoz Garcia	Carlos III University, Spain
Mohamed Nadif	Paris Descartes University, France
Detlef Nauck	British Telecom, UK
Andreas Nürnberg	University of Magdeburg, Germany
Tim Oates	University of Maryland Baltimore, USA
Nicos Pavlidis	Imperial College London, UK
Mykola Pechenizkiy	TU Eindhoven, The Netherlands
José-Maria Peña	TU Madrid, Spain
Ruggero Pensa	University of Turin, Italy
Adriana Prado	University of Antwerp, Belgium
Bhanu Prasad	Florida A&M University, USA
Ronaldo Prati	Universidade Federal do ABC, Brazil
Christophe Rigotti	University of Lyon, France
Fabrizio Riguzzi	University of Ferrara, Italy
Céline Robardet (PC Co-chair)	University of Lyon, France
Céline Rouveiro	University Paris-Nord, France
Stefan Rüping	TU Dortmund, Germany
Antonio Salmeron	University of Almeria, Spain
Paola Sebastiani	Boston University, USA
Arno Siebes (PC Co-chair)	Utrecht University, The Netherlands
Maarten van Someren	University of Amsterdam, The Netherlands
Myra Spiliopoulou	Magdeburg University, Germany
Martin Spott	British Telecom, UK
Stephen Swift	Brunel University, UK
Dimitris Tasoulis	Imperial College London, UK
Maguelonne Teisseire	University of Montpellier, France

Hannu Toivonen	University of Helsinki, Finland
Koji Tsuda	Max Planck Institute, Germany
Allan Tucker	Brunel University, UK
Antony Unwin	University of Augsburg, Germany
Richard Weber	University of Chile, Chile
Stefan Wrobel	Fraunhofer IAIS, Germany

Local Arrangements Chairs

Guillaume Beslon	University of Lyon, France
Céline Robardet	University of Lyon, France

Local Organizing Committee

Loïc Cerf	University of Lyon, France
Serge Fenet	University of Lyon, France
Pierre-Nicolas Mougel	University of Lyon, France
David Parsons	University of Lyon, France

Additional Referees

Hidir Aras	Inmaculada López
Tassadit Bouadi	Antonio LaTorre
Sandra Bringay	Dongning Luo
Laurent Boyer	Florian Mansmann
Matthias Bracht	Christian Moewes
Massimiliano Cattafi	Santiago Muelas
Bertrand Cuissart	Pance Panov
Frank Eichinger	Robert Porzel
Mohamed Elati	Natalja Punko
Fabio Fassetti	Iead Rezek
Denis Ferraretti	Georg Ruß
Maxim Foursa	Matthias Steinbrecher
Valentin Gjorgioski	Mika Sulkava
Henrik Grosskreutz	Gerben de Vries
Ajay Jasra	David Weston
Lisa Di Jorio	Christian von der Weth
Joerg Kindermann	Bernard Zenko

Table of Contents

Invited Papers

- Intelligent Data Analysis in the 21st Century 1
Paul Cohen and Niall Adams

- Analyzing the Localization of Retail Stores with Complex Systems Tools 10
Pablo Jensen

Selected Contributions 1 (Long Talks)

- Change (Detection) You Can Believe in: Finding Distributional Shifts in Data Streams 21

*Tamraparni Dasu, Shankar Krishnan, Dongyu Lin,
Suresh Venkatasubramanian, and Kevin Yi*

- Exploiting Data Missingness in Bayesian Network Modeling 35
Sérgio Rodrigues de Moraes and Alex Aussem

- DEMScale: Large Scale MDS Accounting for a Ridge Operator and Demographic Variables 47
Stephen L. France and J. Douglas Carroll

- How to Control Clustering Results? Flexible Clustering Aggregation.... 59
*Martin Hahmann, Peter B. Volk, Frank Rosenthal,
Dirk Habich, and Wolfgang Lehner*

- Compensation of Translational Displacement in Time Series Clustering Using Cross Correlation..... 71
Frank Höppner and Frank Klawonn

- Context-Based Distance Learning for Categorical Data Clustering 83
Dino Ienco, Ruggero G. Pensa, and Rosa Meo

- Semi-supervised Text Classification Using RBF Networks 95
Eric P. Jiang

- Improving k -NN for Human Cancer Classification Using the Gene Expression Profiles 107
Manuel Martín-Merino and Javier De Las Rivas

- Subgroup Discovery for Test Selection: A Novel Approach and Its Application to Breast Cancer Diagnosis 119
*Marianne Mueller, Rómer Rosales, Harald Steck, Sriram Krishnan,
Bharat Rao, and Stefan Kramer*

Trajectory Voting and Classification Based on Spatiotemporal Similarity in Moving Object Databases	131
<i>Costas Panagiotakis, Nikos Pelekis, and Ioannis Kopanakis</i>	
Leveraging Call Center Logs for Customer Behavior Prediction	143
<i>Anju G. Parvathy, Bintu G. Vasudevan, Abhishek Kumar, and Rajesh Balakrishnan</i>	
Condensed Representation of Sequential Patterns According to Frequency-Based Measures	155
<i>Marc Planetevit and Bruno Crémilleux</i>	
ART-Based Neural Networks for Multi-label Classification	167
<i>Elena P. Sapozhnikova</i>	
Two-Way Grouping by One-Way Topic Models	178
<i>Eerika Savia, Kai Puolamäki, and Samuel Kaski</i>	
Selecting and Weighting Data for Building Consensus Gene Regulatory Networks	190
<i>Emma Steele and Allan Tucker</i>	
Incremental Bayesian Network Learning for Scalable Feature Selection	202
<i>Grégory Thibault, Alex Aussem, and Stéphane Bonnevay</i>	
Feature Extraction and Selection from Vibration Measurements for Structural Health Monitoring	213
<i>Janne Toivola and Jaakko Hollmén</i>	
Zero-Inflated Boosted Ensembles for Rare Event Counts	225
<i>Alexander Borisov, George Runger, Eugene Tuv, and Nuttha Lurpongulukana-Strand</i>	
Selected Contributions 2 (Short Talks)	
Mining the Temporal Dimension of the Information Propagation	237
<i>Michele Berlingario, Michele Coscia, and Fosca Giannotti</i>	
Adaptive Learning from Evolving Data Streams	249
<i>Albert Bifet and Ricard Gavaldà</i>	
An Application of Intelligent Data Analysis Techniques to a Large Software Engineering Dataset	261
<i>James Cain, Steve Counsell, Stephen Swift, and Allan Tucker</i>	
Which Distance for the Identification and the Differentiation of Cell-Cycle Expressed Genes?	273
<i>Alpha Diallo, Ahlame Douzal-Chouakria, and Francoise Giroud</i>	

Ontology-Driven KDD Process Composition	285
<i>Claudia Diamantini, Domenico Potena, and Emanuele Storti</i>	
Mining Frequent Gradual Itemsets from Large Databases	297
<i>Lisa Di-Jorio, Anne Laurent, and Maguelonne Teisseire</i>	
Selecting Computer Architectures by Means of Control-Flow-Graph Mining	309
<i>Frank Eichinger and Klemens Böhm</i>	
Visualization-Driven Structural and Statistical Analysis of Turbulent Flows	321
<i>Kenny Gruchalla, Mark Rast, Elizabeth Bradley, John Clyne, and Pablo Mininni</i>	
Distributed Algorithm for Computing Formal Concepts Using Map-Reduce Framework	333
<i>Petr Krajca and Vilem Vychodil</i>	
Multi-Optimisation Consensus Clustering	345
<i>Jian Li, Stephen Swift, and Xiaohui Liu</i>	
Improving Time Series Forecasting by Discovering Frequent Episodes in Sequences	357
<i>Francisco Martínez-Álvarez, Alicia Troncoso, and José C. Riquelme</i>	
Measure of Similarity and Compactness in Competitive Space	369
<i>Nikolay Zagoruiko</i>	
Bayesian Solutions to the Label Switching Problem	381
<i>Kai Puolamäki and Samuel Kaski</i>	
Efficient Vertical Mining of Frequent Closures and Generators	393
<i>Laszlo Szathmary, Petko Valtchev, Amedeo Napoli, and Robert Godin</i>	
Isotonic Classification Trees	405
<i>Rémon van de Kamp, Ad Feelders, and Nicola Barile</i>	
Author Index	417