

*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*

João Gama Elizabeth Bradley  
Jaakko Hollmén (Eds.)

# Advances in Intelligent Data Analysis X

10th International Symposium, IDA 2011  
Porto, Portugal, October 29-31, 2011  
Proceedings

## Volume Editors

João Gama  
University of Porto, Faculty of Economics  
LIAAD-INESC Porto, L.A.  
Rua de Ceuta, 118, 6°, 4050-190 Porto, Portugal  
E-mail: [jgama@fep.up.pt](mailto:jgama@fep.up.pt)

Elizabeth Bradley  
University of Colorado, Department of Computer Science  
Boulder, CO 80309-0430, USA  
E-mail: [lizb@cs.colorado.edu](mailto:lizb@cs.colorado.edu)

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](mailto:jaakko.hollmen@aalto.fi)

ISSN 0302-9743  
ISBN 978-3-642-24799-6  
DOI 10.1007/978-3-642-24800-9  
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349  
e-ISBN 978-3-642-24800-9

Library of Congress Control Number: 2011938470

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

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

© Springer-Verlag Berlin Heidelberg 2011

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](http://www.springer.com))

# Preface

We are proud to present the proceedings of the 10th International Symposium on Intelligent Data Analysis, which was held during October 29-31, in Porto, Portugal. The IDA series started in 1995 and was held biennially until 2009. The year 2010 marked a special year: the IDA symposium refocused on the original goals of the series. This focus was continued and consolidated this year, as outlined in the Call for Papers:

“IDA 2011 solicits papers on all aspects of intelligent data analysis, particularly papers on intelligent support for modeling and analyzing complex, dynamical systems. Many important scientific and social systems are poorly understood, particularly their evolving nature. Intelligent support for understanding these kinds of systems goes beyond the usual algorithmic offerings in the data mining literature and includes data collection, novel modes of data acquisition, such as crowd sourcing and other *citizen science* mechanisms; data cleaning, semantics and markup; searching for data and assembling datasets from multiple sources; data processing, including workflows, mixed-initiative data analysis, and planning; data and information fusion; incremental, mixed-initiative model development, testing and revision; and visualization and dissemination of results...”

The IDA community responded enthusiastically to this call and the symposium program was accordingly very strong. We received 73 submissions from people in 29 countries. Each submission was evaluated by at least three members of the Program Committee; acceptance was based on these experts’ assessment of the quality and novelty of the work. Members of the Senior Program Committee were asked to solicit and advocate for papers that reported on preliminary but highly novel work. Overall, 73 papers were submitted; 19 and 16 were accepted for oral and poster presentation, respectively.

The meeting was hosted by the Department of Computer Sciences, Faculty of Science, University of Porto in Portugal. Marking the IDA 10th anniversary, we were honored by the participation of three distinguished invited speakers. Xiaohui Liu from Brunel University in the United Kingdom – one of the founders of the IDA symposium series – presented a talk on “Intelligent Data Analysis: Keeping Pace with Technological Advance”. Michael R. Berthold from the University of Konstanz in Germany presented “From Pattern Discovery to Discovery Support: Creativity and Heterogeneous Information Networks”. Carla Gomes from Cornell University in the United States presented “Computational Sustainability”. These invited talks were complemented by a special General Chair’s Lecture presented by Ricardo Baeza-Yates from Yahoo! Research Barcelona in Spain entitled “Web Mining or the Wisdom of the Crowds”.

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. We gratefully acknowledge those who were involved in the local organization of the symposium: Pedro Pereira Rodrigues (Chair), Carlos Ferreira, Raquel Sebastião, Márcia Oliveira, and Petr Kosina.

The 2011 IDA symposium would not have been possible without the support of the following organizations:

- Associação Portuguesa para a Inteligência Artificial
- Centro Internacional de Matemática
- Fundação para a Ciência e a Tecnologia
- INESC Porto
- KNIME — Konstanz Information Miner
- LIAAD — The Laboratory of Artificial Intelligence and Decision Support
- University of Porto
- Yahoo! Research Center, Barcelona

We especially thank Ericsson for funding the IDA Frontier Prize, a newly established award for a novel, visionary contribution to data analysis in the understanding of complex systems.

August 2011

João Gama  
Elizabeth Bradley  
Jaakko Hollmén

# Organization

## Conference Chair

João Gama	University of Porto, Portugal
-----------	-------------------------------

## Program Chairs

Elizabeth Bradley	University of Colorado, USA
Jaakko Hollmén	Aalto University, Finland

## Poster Chairs

Fazel Famili	CNRC, Canada
Alípio Jorge	University of Porto, Portugal

## Local Organization Chair

Pedro Pereira Rodrigues	University of Porto, Portugal
-------------------------	-------------------------------

## Publicity Chairs

David Weston	Imperial College, UK
Carlos Ferreira	University of Porto, Portugal

## Frontier Prize Chairs

Michael R. Berthold	University of Konstanz, Germany
Niall M. Adams	Imperial College, UK

## Webmasters

Luis Matias	University of Porto, Portugal
Petr Kosina	University of Porto, Portugal

## Local Organization

Carlos Ferreira	LIAAD-Inesc Porto, University of Porto
Raquel Sebastião	LIAAD-Inesc Porto, University of Porto
Márcia Oliveira	LIAAD-Inesc Porto, University of Porto
Petr Kosina	LIAAD-Inesc Porto, University of Porto

## Senior Program Committee Members

Niall M. Adams	Imperial College, UK
Michael R. Berthold	University of Konstanz, Germany
Elizabeth Bradley	University of Colorado, USA
Paul Cohen	University of Arizona, USA
João Gama	University of Porto, Portugal
Jaakko Hollmén	Aalto University, Finland
Frank Klawonn	Ostfalia University of Applied Sciences, Germany
Joost Kok	Leiden University, The Netherlands
Xioahui Liu	Brunel University West London, UK
Hannu Toivonen	University of Helsinki, Finland
David Weston	Imperial College, UK

## Program Committee Members

Henrik Boström	Stockholm University, Sweden
Jean-François Boulicaut	INSA Lyon, France
Andre de Carvalho	University of Sao Paulo, Brazil
Jose del Campo	University of Malaga, Spain
Bruno Crémilleux	University of Caen, France
Werner Dubitzky	University of Ulster, UK
Sašo Džeroski	Jožef Stefan Institute, Ljubljana, Slovenia
Fazel Famili	IIT - National Research Council Canada, Canada
Ad Feelders	Utrecht University, The Netherlands
Ingrid Fischer	University of Konstanz, Germany
Johannes Fürnkranz	Technische Universität Darmstadt, Germany
Gemma Garriga	INRIA, France
Ricard Gavaldà	Technical University of Catalonia (UPC), Spain
Gerard Govaert	University of Technology of Compiègne, France
Kenny Gruchalla	National Renewable Energy Lab, USA
Lawrence Hall	University of South Florida, USA
Howard Hamilton	University of Regina, Canada
Eyke Hüllermeier	University of Marburg, Germany
Alípio Jorge	University of Porto, Portugal
Eammon Keogh	University of California Riverside, USA
Wesley Kerr	University of Arizona, USA
Rudolf Kruse	Otto von Guericke University Magdeburg, Germany

Pedro Larranaga	Universidad Politecnica de Madrid, Spain
Manuel Martin-Merino	Pontifical University of Salamanca, Spain
Ernestina Menasalvas	Universidad Politecnica de Madrid, Spain
Maria-Carolina Monard	University of Sao Paulo, Brazil
Giovanni Montana	Imperial College, UK
Miguel Prada	University of Leon, Spain
J. Sunil Rao	The University of Miami, USA
Victor Robles	Technical University of Madrid, Spain
Vítor Santos Costa	University of Porto, Portugal
Roberta Siciliano	University of Naples Federico II, Italy
Myra Spiliopoulou	Otto von Guericke University Magdeburg, Germany
Stephen Swift	Brunel University, UK
Evimaria Terzi	Boston University, USA
Allan Tucker	Brunel University, UK
Juha Vesanto	Xtract Ltd., Finland
Ricardo Vigário	Aalto University, Finland

## Additional Referees

Satrajit Basu	Jefrey Lijffijt
Igor Braga	Christian Moewes
Ivica Dimitrovski	Bruno Nogueira
Dora Erdos	Márcia Oliveira
Carlos Ferreira	Jonathon Parker
Valentin Gjorgjioski	Georg Ruß
Pascal Held Frederik Janssen	Leander Schietgat
Dragi Koccev	Raquel Sebastião
John N. Korecki	Newton Spolaôr
Mikko Korpela	Matthias Steinbrecher
Petr Kosina	Janne Toivola
Mikko Kurimo	Lucas Vendramin



# Table of Contents

## Invited Papers

Bisociative Knowledge Discovery .....	1
<i>Michael R. Berthold</i>	
Computational Sustainability (Abstract) .....	8
<i>Carla P. Gomes</i>	
Intelligent Data Analysis: Keeping Pace with Technological Advances (Abstract) .....	9
<i>Xiaohui Liu</i>	

## Selected Contributions

Comparative Analysis of Power Consumption in University Buildings Using envSOM .....	10
<i>Serafín Alonso, Manuel Domínguez, Miguel Angel Prada, Mika Sulkava, and Jaakko Hollmén</i>	
Context-Aware Collaborative Data Stream Mining in Ubiquitous Devices .....	22
<i>João Bártolo Gomes, Mohamed Medhat Gaber, Pedro A.C. Sousa, and Ernestina Menasalvas</i>	
Intra-firm Information Flow: A Content-Structure Perspective .....	34
<i>Yakir Berchenko, Or Daliot, and Nir N. Brueller</i>	
Mining Fault-Tolerant Item Sets Using Subset Size Occurrence Distributions .....	43
<i>Christian Borgelt and Tobias Kötter</i>	
Finding Ensembles of Neurons in Spike Trains by Non-linear Mapping and Statistical Testing .....	55
<i>Christian Braune, Christian Borgelt, and Sonja Grün</i>	
Towards Automatic Pathway Generation from Biological Full-Text Publications .....	67
<i>Ekaterina Buyko, Jörg Linde, Steffen Priebe, and Udo Hahn</i>	
Online Writing Data Representation: A Graph Theory Approach .....	80
<i>Gilles Caporossi and Christophe Leblay</i>	

Online Evaluation of Email Streaming Classifiers Using GNUsmail . . . . .	90
<i>José M. Carmona-Cejudo, Manuel Baena-García, José del Campo-Ávila, Albert Bifet, João Gama, and Rafael Morales-Bueno</i>	
The Dynamic Stage Bayesian Network: Identifying and Modelling Key Stages in a Temporal Process . . . . .	101
<i>Stefano Ceccon, David Garway-Heath, David Crabb, and Allan Tucker</i>	
Mining Train Delays . . . . .	113
<i>Boris Cule, Bart Goethals, Sven Tassenoy, and Sabine Verboven</i>	
Robustness of Change Detection Algorithms . . . . .	125
<i>Tamraparni Dasu, Shankar Krishnan, and Gina Maria Pomann</i>	
GaMuSo: Graph Base Music Recommendation in a Social Bookmarking Service . . . . .	138
<i>Jeroen De Knijf, Anthony Liekens, and Bart Goethals</i>	
Resampling-Based Change Point Estimation . . . . .	150
<i>Jelena Fiosina and Maksims Fiosins</i>	
Learning about the Learning Process . . . . .	162
<i>João Gama and Petr Kosina</i>	
Predicting Computer Performance Dynamics . . . . .	173
<i>Joshua Garland and Elizabeth Bradley</i>	
Prototype-Based Classification of Dissimilarity Data . . . . .	185
<i>Barbara Hammer, Bassam Mokbel, Frank-Michael Schleif, and Xibin Zhu</i>	
Automatic Layout Design Solution . . . . .	198
<i>Fadratul Hafinaz Hassan and Allan Tucker</i>	
An Alternative to ROC and AUC Analysis of Classifiers . . . . .	210
<i>Frank Klawonn, Frank Höppner, and Sigrun May</i>	
The Algorithm APT to Classify in Concurrence of Latency and Drift . . .	222
<i>Georg Matthias Kreml</i>	
Identification of Nuclear Magnetic Resonance Signals via Gaussian Mixture Decomposition . . . . .	234
<i>Martin Krone, Frank Klawonn, Thorsten Lühns, and Christiane Ritter</i>	
Graphical Feature Selection for Multilabel Classification Tasks . . . . .	246
<i>Gerardo Lastra, Oscar Luaces, Jose Ramon Quevedo, and Antonio Bahamonde</i>	

A Web2.0 Strategy for the Collaborative Analysis of Complex Bioimages .....	258
<i>Christian Loyek, Jan Kölling, Daniel Langenkämper, Karsten Niehaus, and Tim W. Nattkemper</i>	
Data Quality through Model Checking Techniques .....	270
<i>Mario Mezzanzanica, Roberto Boselli, Mirko Cesarini, and Fabio Mercorio</i>	
Generating Automated News to Explain the Meaning of Sensor Data ...	282
<i>Martin Molina, Amanda Stent, and Enrique Parodi</i>	
Binding Statistical and Machine Learning Models for Short-Term Forecasting of Global Solar Radiation .....	294
<i>Llanos Mora-López, Ildefonso Martínez-Marchena, Michel Piliouguine, and Mariano Sidrach-de-Cardona</i>	
Bisociative Discovery of Interesting Relations between Domains .....	306
<i>Uwe Nagel, Kilian Thiel, Tobias Kötter, Dawid Piątek, and Michael R. Berthold</i>	
Collaboration-Based Function Prediction in Protein-Protein Interaction Networks .....	318
<i>Hossein Rahmani, Hendrik Blockeel, and Andreas Bender</i>	
Mining Sentiments from Songs Using Latent Dirichlet Allocation .....	328
<i>Govind Sharma and M. Narasimha Murty</i>	
Analyzing Parliamentary Elections Based on Voting Advice Application Data .....	340
<i>Jaakko Talonen and Mika Sulkava</i>	
Integrating Marine Species Biomass Data by Modelling Functional Knowledge .....	352
<i>Allan Tucker and Daniel Duplisea</i>	
A Stylometric Study and Assessment of Machine Translators .....	364
<i>V. Suresh, Avanthi Krishnamurthy, Rama Badrinath, and C.E. Veni Madhavan</i>	
Traffic Events Modeling for Structural Health Monitoring .....	376
<i>Ugo Vespier, Arno Knobbe, Joaquin Vanschoren, Shengfa Miao, Arne Koopman, Bas Obladen, and Carlos Bosma</i>	
Supervised Learning in Parallel Universes Using Neighborgrams .....	388
<i>Bernd Wiswedel and Michael R. Berthold</i>	

<i>i</i> MMPC: A Local Search Approach for Incremental Bayesian Network Structure Learning .....	401
<i>Amanullah Yasin and Philippe Leray</i>	
Analyzing Emotional Semantics of Abstract Art Using Low-Level Image Features .....	413
<i>He Zhang, Eimontas Augilius, Timo Honkela, Jorma Laaksonen, Hannes Gamper, and Henok Alene</i>	
<b>Author Index</b> .....	425