

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

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

Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos

New York University, NY, USA

Doug Tygar

University of California, Berkeley, CA, USA

Moshe Y. Vardi

Rice University, Houston, TX, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

A. Fazel Famili Joost N. Kok
José M. Peña Arno Siebes
Ad Feelders (Eds.)

Advances in Intelligent Data Analysis VI

6th International Symposium on
Intelligent Data Analysis, IDA 2005
Madrid, Spain, September 8-10, 2005
Proceedings

Volume Editors

A. Fazel Famili

IIT/ITI - National Research Council Canada, Ottawa University
School of Information Technology and Engineering
1200 Montreal Rd, M-50, Ottawa, ON K1A 0R6, Canada
E-mail: fazel.famili@nrc-cnrc.gc.ca

Joost N. Kok

Leiden University, Leiden Institute of Advanced Computer Science
Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
E-mail: joost@liacs.nl

José M. Peña

Universidad Politécnica de Madrid, DATSI - Facultad de Informática
Campus de Montegancedo S/N, Boadilla del Monte, 28660 Madrid, Spain
E-mail: jmpena@fi.upm.es

Arno Siebes

Ad Feelders

Utrecht University, Department of Information and Computing Sciences
PO Box 80.089, 3508 TB Utrecht, The Netherlands
E-mail: {arno, ad}@cs.uu.nl

Library of Congress Control Number: 2005931595

CR Subject Classification (1998): H.3, I.2, G.3, I.5.1, I.4.5, J.2, J.1, J.3

ISSN 0302-9743

ISBN-10 3-540-28795-7 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-28795-7 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 is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2005

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11552253 06/3142 5 4 3 2 1 0

Preface

One of the superb characteristics of Intelligent Data Analysis (IDA) is that it is an interdisciplinary field in which researchers and practitioners from a number of areas are involved in a typical project. This also creates a challenge in which the success of a team depends on the participation of users and domain experts who need to interact with researchers and developers of any IDA system. All this is usually reflected in successful projects and of course on the papers that were evaluated by this year's program committee from which the final program has been developed.

In our call for papers, we solicited papers on (i) applications and tools, (ii) theory and general principles, and (iii) algorithms and techniques. We received a total of 184 papers, reviewing these was a major challenge. Each paper was assigned to three reviewers. In the end 46 papers were accepted, which are all included in the proceedings and presented at the conference.

This year's papers reflect the results of applied and theoretical research from a number of disciplines all of which are related to the field of Intelligent Data Analysis. To have the best combination of theoretical and applied research and also provide the best focus, we have divided this year's IDA program into tutorials, invited talks, panel discussions and technical sessions.

We have managed to organize two excellent tutorials on the first day by Luc De Raedt and Kristian Kersting, entitled *Probabilistic Inductive Logic Programming*, and by Bruno Apolloni, Dario Malchiodi and Sabrina Gaito, entitled *Statistical Bases of Machine Learning*. Our invited speakers are Prof Ivan Bratko from the Jozef Stefan Institute in Slovenia, and Prof Alex Freitas from the University of Kent.

We wish to express our sincere thanks to many people who have worked hard for the IDA conference to happen in Madrid. Special thanks to tutorial, publicity, local organization, and panels chairs who have been in charge of a large portion of our responsibilities. We would also like to thank Xiaohui Liu and Michael Berthold who worked as advisors to this conference and members of the local organizing committee for their hard work. Finally, we are grateful to the members of our program committee; without their help it would have been impossible to put together such a valuable program.

September 2005

With our best wishes,

A. Fazel Famili,
José Maria S. Peña,
Joost Kok,
Arno Siebes,
Ad Feelders

Organization

Conference Organization

General Chair

A. Fazel Famili
National Research Council
Ottawa, Canada

Program Chairs

José M. Peña
Universidad Politécnica de Madrid
Madrid, Spain

Arno Siebes
Utrecht University
Utrecht, The Netherlands

Joost Kok
Leiden University
Leiden, The Netherlands

Tutorial Chair

Pedro Larrañaga
EHU-Universidad del País Vasco
San Sebastián, Spain

Publication Chair

Ad Feelders
Utrecht University
Utrecht, The Netherlands

Publicity Chairs

Jorge Muruzábal
Universidad Rey Juan Carlos
Madrid, Spain

Julián Sánchez
Quinao S.L.
Madrid, Spain

Local Organization Chair

Víctor Robles
Universidad Politécnica de Madrid
Madrid, Spain

Panel Chair

Sofian Maabout
LaBRI-Université Bordeaux
Bordeaux, France

Local Committee

Universidad Politécnica de Madrid Madrid, Spain

María S. Pérez
Vanessa Herves
Francisco Rosales
Antonio García
Óscar Cubo
Pilar Herrero
Antonio LaTorre
Alberto Sánchez

Universidad Rey Juan Carlos Madrid, Spain

Susana Vegas
Andrés L. Martinez

Program Committee

Niall Adams, Imperial College, UK
Riccardo Bellazzi, University of Pavia, Italy
Bettina Berendt, Humboldt University of Berlin, Germany
Michael Berthold, University of Konstanz, Germany
Hans-Georg Beyer, Vorarlberg University of Applied Sciences, Austria
Jean-François Boulicaut, INSA Lyon, France
Christian Borgelt, Otto-von-Guericke-University of Magdeburg, Germany
Hans-Dieter Burkhard, Humboldt Universität Berlin, Germany
Luis M. de Campos, Universidad de Granada, Spain
Fazel Famili, Institute for Information Technology; NRC, Canada
Giuseppe Di Fatta, University of Konstanz, Germany
Fridtjof Feldbusch, University of Karlsruhe, Germany
Ingrid Fischer, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
Douglas Fisher, Univ. Vanderbilt, USA
Peter Flach, University of Bristol, UK
Eibe Frank, University of Waikato, New Zealand
Karl A. Fröschl, ec3 – eCommerce Competence Center; Vienna DC, Austria
Gabriela Guimaraes, CENTRIA UNL, Portugal
Lawrence O. Hall, University of South Florida, USA
Pilar Herrero, Universidad Politécnica de Madrid, Spain
Tom Heskes, Radboud University Nijmegen, The Netherlands
Alexander Hinneburg, University of Halle, Germany
Frank Hoepfner, University of Wolfenbuettel, Germany
Adele Howe, Colorado State University, USA
Klaus-Peter Huber, SAS Institute, Germany
Anthony Hunter, University College London, UK
Alfred Inselberg, Tel Aviv University, Israel
Bert Kappen, Radboud University Nijmegen, The Netherlands
Frank Klawonn, University of Wolfenbuettel, Germany
Joost N. Kok, Leiden University, The Netherlands
Walter Kosters, Leiden University, The Netherlands
Rudolf Kruse, Otto-von-Guericke-University of Magdeburg, Germany
Pedro Larrañaga, Universidad del País Vasco, Spain
Hans-Joachim Lenz, Freie Universität Berlin, Germany
Xiaohui Liu, Brunel University, UK
Sofian Maabout, LaBRI-Université Bordeaux, France
Rainer Malaka, European Media Laboratory Heidelberg, Germany
Jorge Muruzábal, Universidad Rey Juan Carlos, Spain
Susana Nascimento, CENTRIA-Universidade Nova de Lisboa, Portugal
Detlef Nauck, BTextact Technologies, UK
Tim Oates, University of Maryland Baltimore County, USA
Simon Parsons, Brooklyn College City University of New York, USA
José M. Peña, Universidad Politécnica de Madrid, Spain

María S. Pérez, Universidad Politécnica de Madrid, Spain
 Bhanu Prasad, Florida A&M University, USA
 Víctor Robles, Universidad Politécnica de Madrid, Spain
 Lorenza Saitta, Università del Piemonte Orientale, Italy
 Paola Sebastiani, Boston University School of Public Health, USA
 Arno Siebes, Universiteit Utrecht, The Netherlands
 Maarten van Someren, University of Amsterdam, The Netherlands
 Myra Spiliopoulou, Otto-von-Guericke-Universität Magdeburg, Germany
 Martin Spott, BTextact Technologies, UK
 Reinhard Viertl, Vienna University of Technology, Austria
 Richard Weber, University of Chile, Chile
 Stefan Wrobel, Fraunhofer AIS & Univ. Bonn, Germany
 Mohammed Zaki, Rensselaer Polytechnic Institute, USA

Referees

Silvia Acid	Fabien Jourdan
David Auber	Florian Kaiser
Roland Barriot	Joerg Kindermann
Concha Bielza	Christine Koerner
Bouchra Bouqata	Antonio LaTorre
Kai Broszat	Marie-Jeanne Lesot
Andres Cano	Andres L. Martinez
Javier G. Castellano	Michael Mayo
Nicolas Cebron	Thorsten Meinl
Víctor Uc Cetina	Ernestina Menasalvas
T.K. Cocx	Dagmar Monett
Nuno Correia	Serafin Moral
Óscar Cubo	Siegfried Nijssen
Santiago Eibe	Juan A. Fernández del Pozo
Lukas C. Faulstich	Simon Price
Juan M. Fernández-Luna	Jose M. Puerta
Fulvia Ferrazzi	Simon Rawles
Manuel Gómez	Frank Rügheimer
Daniel Goehring	Lucia Sacchi
Edgar de Graaf	Alberto Sánchez
J.M. de Graaf	Karlton Sequeira
Jose A. Gámez	Zujun Shentu
Mark Hall	David James Sherman
Alexander Hinneburg	Hendrik Stange
Susanne Hoche	Micheal Syrjakow
Geoff Holmes	Xiaomeng Wang
Rainer Holve	Bernd Wiswedel
Tamás Horváth	Marta Elena Zorrilla
Juan F. Huete	

Table of Contents

Probabilistic Latent Clustering of Device Usage <i>Jean-Marc Andreoli, Guillaume Bouchard</i>	1
Condensed Nearest Neighbor Data Domain Description <i>Fabrizio Angiulli</i>	12
Balancing Strategies and Class Overlapping <i>Gustavo E.A.P.A. Batista, Ronaldo C. Prati, Maria C. Monard</i>	24
Modeling Conditional Distributions of Continuous Variables in Bayesian Networks <i>Barry R. Cobb, Rafael Rumí, Antonio Salmerón</i>	36
Kernel K-Means for Categorical Data <i>Julia Couto</i>	46
Using Genetic Algorithms to Improve Accuracy of Economical Indexes Prediction <i>Óscar Cubo, Víctor Robles, Javier Segovia, Ernestina Menasalvas</i>	57
A Distance-Based Method for Preference Information Retrieval in Paired Comparisons <i>Esther Dopazo, Jacinto González-Pachón, Juan Robles</i>	66
Knowledge Discovery in the Identification of Differentially Expressed Genes in Tumoridal Macrophage <i>A. Fazel Famili, Ziyang Liu, Pedro Carmona-Saez, Alaka Mullick</i> ...	74
Searching for Meaningful Feature Interactions with Backward-Chaining Rule Induction <i>Doug Fisher, Mary Edgerton, Lianhong Tang, Lewis Frey, Zhihua Chen</i>	86
Exploring Hierarchical Rule Systems in Parallel Coordinates <i>Thomas R. Gabriel, A. Simona Pintilie, Michael R. Berthold</i>	97
Bayesian Networks Learning for Gene Expression Datasets <i>Giacomo Gamberoni, Evelina Lamma, Fabrizio Riguzzi, Sergio Storari, Stefano Volinia</i>	109

Pulse: Mining Customer Opinions from Free Text <i>Michael Gamon, Anthony Aue, Simon Corston-Oliver, Eric Ringger</i>	121
Keystroke Analysis of Different Languages: A Case Study <i>Daniele Gunetti, Claudia Picardi, Giancarlo Ruffo</i>	133
Combining Bayesian Networks with Higher-Order Data Representations <i>Elias Gyftodimos, Peter A. Flach</i>	145
Removing Statistical Biases in Unsupervised Sequence Learning <i>Yoav Horman, Gal A. Kaminka</i>	157
Learning from Ambiguously Labeled Examples <i>Eyke Hüllermeier, Jürgen Beringer</i>	168
Learning Label Preferences: Ranking Error Versus Position Error <i>Eyke Hüllermeier, Johannes Fürnkranz</i>	180
FCLib: A Library for Building Data Analysis and Data Discovery Tools <i>Wendy S. Koegler, W. Philip Kegelmeyer</i>	192
A Knowledge-Based Model for Analyzing GSM Network Performance <i>Pasi Lehtimäki, Kimmo Raivio</i>	204
Sentiment Classification Using Information Extraction Technique <i>Jian Liu, Jianxin Yao, Gengfeng Wu</i>	216
Extending the SOM Algorithm to Visualize Word Relationships <i>Manuel Martín-Merino, Alberto Muñoz</i>	228
Towards Automatic and Optimal Filtering Levels for Feature Selection in Text Categorization <i>E. Montañés, E.F. Combarro, I. Díaz, J. Ranilla</i>	239
Block Clustering of Contingency Table and Mixture Model <i>Mohamed Nadif, Gérard Govaert</i>	249
Adaptive Classifier Combination for Visual Information Processing Using Data Context-Awareness <i>Mi Young Nam, Phill Kyu Rhee</i>	260
Self-poised Ensemble Learning <i>Ricardo Nanculef, Carlos Valle, Héctor Allende, Claudio Moraga</i>	272

Discriminative Remote Homology Detection Using Maximal Unique Sequence Matches <i>Hasan Oğul, Ü. Erkan Mumcuoğlu</i>	283
From Local Pattern Mining to Relevant Bi-cluster Characterization <i>Ruggero G. Pensa, Jean-François Boulicaut</i>	293
Machine-Learning with Cellular Automata <i>Petra Povalej, Peter Kokol, Tatjana Welzer Družovec, Bruno Stiglic</i> ..	305
MDS _{polar} : A New Approach for Dimension Reduction to Visualize High Dimensional Data <i>Frank Rehm, Frank Klawonn, Rudolf Kruse</i>	316
Miner Ants Colony: A New Approach to Solve a Mine Planning Problem <i>María-Cristina Riff, Michael Moossen, Xavier Bonnaire</i>	328
Extending the GA-EDA Hybrid Algorithm to Study Diversification and Intensification in GAs and EDAs <i>V. Robles, J.M. Peña, M.S. Pérez, P. Herrero, O. Cubo</i>	339
Spatial Approach to Pose Variations in Face Verification <i>Licesio J. Rodríguez-Aragón, Ángel Serrano, Cristina Conde, Enrique Cabello</i>	351
Analysis of Feature Rankings for Classification <i>Roberto Ruiz, Jesús S. Aguilar-Ruiz, José C. Riquelme, Norberto Díaz-Díaz</i>	362
A Mixture Model-Based On-line CEM Algorithm <i>Allou Samé, Gérard Govaert, Christophe Ambroise</i>	373
Reliable Hierarchical Clustering with the Self-organizing Map <i>Elena V. Samsonova, Thomas Bäck, Joost N. Kok, Ad P. IJzerman</i>	385
Statistical Recognition of Noun Phrases in Unrestricted Text <i>José I. Serrano, Lourdes Araujo</i>	397
Successive Restrictions Algorithm in Bayesian Networks <i>Linda Smail, Jean Pierre Raoult</i>	409
Modelling the Relationship Between Streamflow and Electrical Conductivity in Hollin Creek, Southeastern Australia <i>Jess Spate</i>	419

Biological Cluster Validity Indices Based on the Gene Ontology <i>Nora Speer, Christian Spieth, Andreas Zell</i>	429
An Evaluation of Filter and Wrapper Methods for Feature Selection in Categorical Clustering <i>Luis Talavera</i>	440
Dealing with Data Corruption in Remote Sensing <i>Choh Man Teng</i>	452
Regularized Least-Squares for Parse Ranking <i>Eugeni Tsvitsivadze, Tapio Pahikkala, Sampo Pyysalo, Jorma Boberg, Aleksandr Mylläri, Tapio Salakoski</i>	464
Bayesian Network Classifiers for Time-Series Microarray Data <i>Allan Tucker, Veronica Vinciotti, Peter A.C. 't Hoen, Xiaohui Liu</i>	475
Feature Discovery in Classification Problems <i>Manuel del Valle, Beatriz Sánchez, Luis F. Lago-Fernández, Fernando J. Corbacho</i>	486
A New Hybrid NM Method and Particle Swarm Algorithm for Multimodal Function Optimization <i>Fang Wang, Yuhui Qiu, Yun Bai</i>	497
Detecting Groups of Anomalously Similar Objects in Large Data Sets <i>Zhicheng Zhang, David J. Hand</i>	509
Author Index	521