

Lecture Notes in Artificial Intelligence 3202

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science



Jean-François Boulicaut  
Floriana Esposito Fosca Giannotti  
Dino Pedreschi (Eds.)

# Knowledge Discovery in Databases: PKDD 2004

8th European Conference on Principles and Practice  
of Knowledge Discovery in Databases  
Pisa, Italy, September 20-24, 2004  
Proceedings

 Springer

### Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA  
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

### Volume Editors

Jean-François Boulicaut  
INSA Lyon  
LIRIS CNRS FRE 2672, 69621 Villeurbanne Cedex, France  
E-mail: jean-francois.boulicaut@insa-lyon.fr

Floriana Esposito  
University of Bari  
Department of Computer Science  
Via Orabona 4, 70126 Bari, Italy  
E-mail: esposito@di.uniba.it

Fosca Giannotti  
Science and Technology Institute  
Knowledge Discovery and Delivery (KDD)  
Via G. Moruzzi 1, 56124 Pisa, Italy  
E-mail: Fosca.Giannotti@isti.cnr.it

Dino Pedreschi  
University of Pisa  
Department of Computer Science  
Via F. Buonarroti 2, 56125 Pisa, Italy  
E-mail: pedre@di.unipi.it

Library of Congress Control Number: 2004111516

CR Subject Classification (1998): I.2, H.2, J.1, H.3, G.3, I.7, F.4.1

ISSN 0302-9743

ISBN 3-540-23108-0 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 is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2004  
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik  
Printed on acid-free paper SPIN: 11322801 06/3142 5 4 3 2 1 0

# Preface

The proceedings of ECML/PKDD 2004 are published in two separate, albeit intertwined, volumes: the Proceedings of the 15th European Conference on Machine Learning (LNAI 3201) and the Proceedings of the 8th European Conferences on Principles and Practice of Knowledge Discovery in Databases (LNAI 3202). The two conferences were co-located in Pisa, Tuscany, Italy during September 20–24, 2004.

It was the fourth time in a row that ECML and PKDD were co-located. After the successful co-locations in Freiburg (2001), Helsinki (2002), and Cavtat-Dubrovnik (2003), it became clear that researchers strongly supported the organization of a major scientific event about machine learning and data mining in Europe.

We are happy to provide some statistics about the conferences. 581 different papers were submitted to ECML/PKDD (about a 75% increase over 2003); 280 were submitted to ECML 2004 only, 194 were submitted to PKDD 2004 only, and 107 were submitted to both. Around half of the authors for submitted papers are from outside Europe, which is a clear indicator of the increasing attractiveness of ECML/PKDD.

The Program Committee members were deeply involved in what turned out to be a highly competitive selection process. We assigned each paper to 3 reviewers, deciding on the appropriate PC for papers submitted to both ECML and PKDD. As a result, ECML PC members reviewed 312 papers and PKDD PC members reviewed 269 papers. We accepted for publication regular papers (45 for ECML 2004 and 39 for PKDD 2004) and short papers that were associated with poster presentations (6 for ECML 2004 and 9 for PKDD 2004). The global acceptance rate was 14.5% for regular papers (17% if we include the short papers).

The scientific program of ECML/PKDD 2004 also included 5 invited talks, a wide workshop and tutorial program (10 workshops plus a Discovery Challenge workshop, and seven tutorials) and a demo session.

We wish to express our gratitude to:

- the authors of the submitted papers;
- the program committee members and the additional referees for their exceptional contribution to a tough but crucial selection process;
- the invited speakers: Dimitris Achlioptas (Microsoft Research, Redmond), Rakesh Agrawal (IBM Almaden Research Center), Soumen Chakrabarti (Indian Institute of Technology, Bombay), Pedro Domingos (University of Washington, Seattle), and David J. Hand (Imperial College, London);
- the workshop chairs Donato Malerba and Mohammed J. Zaki;
- the tutorial chairs Katharina Morik and Franco Turini;
- the discovery challenge chairs Petr Berka and Bruno Crémilleux;

- the publicity chair Salvatore Ruggieri;
- the demonstration chairs Rosa Meo, Elena Baralis, and Codrina Lauth;
- the members of the ECML/PKDD Steering Committee Peter Flach, Luc De Raedt, Arno Siebes, Nada Lavrač, Dragan Gamberger, Ljupčo Todorovski, Hendrik Blockeel, Tapio Elomaa, Heikki Mannila, and Hannu T.T. Toivonen;
- the members of the Award Committee, Michael May and Foster Provost;
- the workshops organizers and the tutorialists;
- the extremely efficient Organization Committee members, Maurizio Atzori, Miriam Baglioni, Sergio Barsocchi, J r my Besson, Francesco Bonchi, Stefano Ferilli, Tiziana Mazzone, Mirco Nanni, Ruggero Pensa, Simone Puntoni, Chiara Renso, Salvatore Rinzivillo, as well as all the other members of the KDD Lab in Pisa, Laura Balbarini and Cristina Rosamilia of L&B Studio, Elena Perini and Elena Tonsini of the University of Pisa;
- the great Web masters Mirco Nanni, Chiara Renso and Salvatore Rinzivillo;
- the directors of the two research institutions in Pisa that jointly made this event possible, Piero Maestrini (ISTI-CNR) and Ugo Montanari (Dipartimento di Informatica);
- the administration staff of the two research institutions in Pisa, in particular Massimiliano Farnesi (ISTI-CNR), Paola Fabiani and Letizia Petrellese (Dipartimento di Informatica);
- Richard van de Stadt ([www.borbala.com](http://www.borbala.com)) for his efficient support to the management of the whole submission and evaluation process by means of the CyberChairPRO software;
- Alfred Hofmann of Springer for co-operation in publishing the proceedings.

We gratefully acknowledge the financial support of KDNet, the Pascal Network, Kluwer and the Machine Learning journal, Springer, the Province of Lucca, the Province of Pisa, the Municipality of Pisa, Microsoft Research, COOP, Exeura, Intel, Talent, INSA-Lyon, ISTI-CNR Pisa, the University of Pisa, the University of Bari, and the patronage of Regione Toscana.

There is no doubt that the impressive scientific activities in machine learning and data mining world-wide were well demonstrated in Pisa. We had an exciting week in Tuscany, enhancing further co-operations between the many researchers who are pushing knowledge discovery into becoming a mature scientific discipline.

July 2004

Jean-Fran ois Boulicaut,  
 Floriana Esposito,  
 Fosca Giannotti,  
 and Dino Pedreschi

# ECML/PKDD 2004 Organization

## Executive Committee

Program Chairs	Jean-François Boulicaut (INSA Lyon) Floriana Esposito (Università di Bari) Fosca Giannotti (ISTI-CNR) Dino Pedreschi (Università di Pisa)
Workshop Chairs	Donato Malerba (University of Bari) Mohammed J. Zaki (Rensselaer Polytechnic Institute)
Tutorial Chairs	Katharina Morik (University of Dortmund) Franco Turini (University of Pisa)
Discovery Challenge Chairs	Petr Berka (University of Economics, Prague) Bruno Crémilleux (University of Caen)
Publicity Chair	Salvatore Ruggieri (University of Pisa)
Demonstration Chairs	Rosa Meo (University of Turin) Elena Baralis (Politecnico of Turin) Ina Lauth (Fraunhofer Institute for Autonomous Intelligent Systems)
Steering Committee	Peter Flach (University of Bristol) Luc De Raedt (Albert-Ludwigs University, Freiburg) Arno Siebes (Utrecht University) Nada Lavrač (Jozef Stefan Institute) Dragan Gamberger (Rudjer Boskovic Institute) Ljupčo Todorovski (Jozef Stefan Institute) Hendrik Blockeel (Katholieke Universiteit Leuven) Tapio Elomaa (Tampere University of Technology) Heikki Mannila (Helsinki Institute for Information Technology) Hannu T.T. Toivonen (University of Helsinki)
Awards Committee	Michael May (Fraunhofer Institute for Autonomous Intelligent Systems, KNet representative) Floriana Esposito (PC representative) Foster Provost (Editor-in-Chief of Machine Learning Journal, Kluwer)
Organizing Committee	Maurizio Atzori (KDDLab, ISTI-CNR) Miriam Baglioni (KDDLab, University of Pisa) Sergio Barsocchi (KDDLab, ISTI-CNR) Jérémy Besson (INSA Lyon) Francesco Bonchi (KDDLab, ISTI-CNR) Stefano Ferilli (University of Bari) Tiziana Mazzone (KDDLab) Mirco Nanni (KDDLab, ISTI-CNR) Ruggero Pensa (INSA Lyon) Chiara Renso (KDDLab, ISTI-CNR) Salvatore Rinzivillo (KDDLab, University of Pisa)

## ECML 2004 Program Committee

Hendrik Blockeel, Belgium  
Marco Botta, Italy  
Henrik Boström, Sweden  
Jean-François Boulicaut, France  
Ivan Bratko, Slovenia  
Pavel Brazdil, Portugal  
Nello Cristianini, USA  
James Cussens, UK  
Ramon Lopes de Mantaras, Spain  
Luc De Raedt, Germany  
Luc Dehaspe, Belgium  
José del R. Millan, Switzerland  
Sašo Džeroski, Slovenia  
Tapio Elomaa, Finland  
Floriana Esposito, Italy  
Peter Flach, UK  
Johannes Fürnkranz, Germany  
Joao Gama, Portugal  
Dragan Gamberger, Croatia  
Jean-Gabriel Ganascia, France  
Fosca Giannotti, Italy  
Attilio Giordana, Italy  
Haym Hirsh, USA  
Thomas Hofmann, USA  
Tamas Horvath, Germany  
Thorsten Joachims, USA  
Dimitar Kazakov, UK  
Roni Khardon, USA  
Joerg Kindermann, Germany  
Yves Kodratoff, France  
Igor Kononenko, Slovenia  
Stefan Kramer, Germany  
Miroslav Kubat, USA  
Stephen Kwek, USA  
Nada Lavrač, Slovenia  
Charles Ling, Canada  
Donato Malerba, Italy  
Heikki Mannila, Finland  
Stan Matwin, Canada  
Dunja Mladenic, Slovenia  
Katharina Morik, Germany  
Hiroshi Motoda, Japan  
Remi Munos, France  
Richard Nock, France  
David Page, USA  
Georgios Paliouras, Greece  
Dino Pedreschi, Italy  
Bernhard Pfahringer, New Zealand  
Enric Plaza, Spain  
Juho Rousu, UK  
Celine Rouveirol, France  
Tobias Scheffer, Germany  
Michele Sebag, France  
Giovanni Semeraro, Italy  
Arno Siebes, The Netherlands  
Robert Sloan, USA  
Gerd Stumme, Germany  
Henry Tirri, Finland  
Ljupčo Todorovski, Slovenia  
Luis Torgo, Portugal  
Peter Turney, Canada  
Maarten van Someren, The Netherlands  
Paul Vitanyi, The Netherlands  
Sholom Weiss, USA  
Dietrich Wettschereck, UK  
Gerhard Widmer, Austria  
Marco Wiering, The Netherlands  
Ruediger Wirth, Germany  
Stefan Wrobel, Germany  
Thomas Zeugmann, Germany  
Tong Zhang, USA  
Blaž Zupan, Slovenia

## PKDD 2004 Program Committee

Elena Baralis, Italy  
Michael Berthold, Germany  
Elisa Bertino, USA  
Hendrik Blockeel, Belgium  
Jean-François Boulicaut, France  
Christopher W. Clifton, USA  
Bruno Cremilleux, France  
Luc De Raedt, Germany  
Luc Dehaspe, Belgium  
Sašo Džeroski, Slovenia  
Tapio Elomaa, Finland  
Floriana Esposito, Italy  
Martin Ester, Canada  
Ad Feelders, The Netherlands  
Ronen Feldman, IL  
Peter Flach, UK  
Eibe Frank, New Zealand  
Alex Freitas, UK  
Johannes Fürnkranz, Germany  
Dragan Gamberger, Croatia  
Minos Garofalakis, USA  
Fosca Giannotti, Italy  
Christophe Giraud-CARRIER, Switzerland  
Bart Goethals, Finland  
Howard Hamilton, Canada  
Robert Hilderman, Canada  
Haym Hirsh, USA  
Frank Hoepfner, Germany  
Se Hong, USA  
Samuel Kaski, Finland  
Daniel Keim, Germany  
Jorg-Uwe Kietz, Switzerland  
Ross King, UK  
Yves Kodratoff, France  
Joost Kok, The Netherlands  
Stefan Kramer, Germany  
Laks Lakshmanan, Canada  
Nada Lavrač, Slovenia  
Donato Malerba, Italy  
Giuseppe Manco, Italy  
Heikki Mannila, Finland  
Stan Matwin, Canada  
Michael May, Germany  
Rosa Meo, Italy  
Dunja Mladenic, Slovenia  
Katharina Morik, Germany  
Shinichi Morishita, Japan  
Hiroshi Motoda, Japan  
Gholamreza Nakhaeizadeh, Germany  
Claire Nedellec, France  
David Page, USA  
Dino Pedreschi, Italy  
Zbigniew Ras, USA  
Jan Rauch, Czech Republic  
Christophe Rigotti, France  
Gilbert Ritschard, Switzerland  
John Roddick, Australia  
Yucel Saygin, Turkey  
Michele Sebag, France  
Marc Sebban, France  
Arno Siebes, The Netherlands  
Andrzej Skowron, Poland  
Myra Spiliopoulou, Germany  
Nicolas Spyrtatos, France  
Reinhard Stolle, USA  
Gerd Stumme, Germany  
Einoshin Suzuki, Japan  
Ah-Hwee Tan, Singapore  
Ljupčo Todorovski, Slovenia  
Hannu Toivonen, Finland  
Luis Torgo, Portugal  
Shusaku Tsumoto, Japan  
Franco Turini, Italy  
Maarten van Someren, The Netherlands  
Ke Wang, Canada  
Louis Wehenkel, Belgium  
Dietrich Wettschereck, UK  
Gerhard Widmer, Austria  
Ruediger Wirth, Germany  
Stefan Wrobel, Germany  
Osmar R. Zaiane, Canada  
Mohammed Zaki, USA  
Carlo Zaniolo, USA  
Djamel Zighed, France  
Blaž Zupan, Slovenia

**ECML/PKDD 2004 Additional Reviewers**

Fabio Abbattista	Martin R.J. Carpenter	François Fleuret
Markus Ackermann	Costantina Caruso	Francesco Folino
Erick Alphonse	Ciro Castiello	Francesco Fornasari
Oronzo Altamura	Barbara Catania	Blaz Fortuna
Massih Amini	Davide Cavagnino	Andrew Foss
Ahmed Amrani	Michelangelo Ceci	Keith Frikken
Anastasia Analiti	Alessio Ceroni	Barbara Furletti
Nicos Angelopoulos	Jesús Cerquides	Thomas Gärtner
Fabrizio Angiulli	Eugenio Cesario	Ugo Galassi
Luiza Antonie	Silvia Chiusano	Arianna Gallo
Annalisa Appice	Fang Chu	Byron Gao
Josep-Lluis Arcos	Antoine Cornuéjols	Paolo Garza
Eva Armengol	Fabrizio Costa	Liqiang Geng
Thierry Artieres	Gianni Costa	Claudio Gentile
Maurizio Atzori	Tom Croonenborghs	Pierre Geurts
Anne Auger	Tomaz Curk	Zoubin Ghahramani
Ilkka Autio	Maria Damiani	Arnaud Giacometti
Jérôme Azé	Agnieszka Dardzinska	Emiliano Giovannetti
Vincenzo Bacarella	Tijl De Bie	Piotr Gmytrasiewicz
Miriam Baglioni	Edwin D. De Jong	Judy Goldsmith
Yijian Bai	Kurt De Grave	Anna Gomolinska
Cristina Baroglio	Marco Degemmis	Udo Grimmer
Teresa Basile	Janez Demšar	Matthew Grounds
Ganesan Bathumalai	Damjan Demšar	Antonella Guzzo
Fadila Bentayeb	Michel de Rougemont	Amaury Habrard
Margherita Berardi	Nicola Di Mauro	Stephan ten Hagen
Bettina Berendt	Christos Dimitrakakis	Jörg Hakenberg
Petr Berka	Simon Dixon	Mark Hall
Guillaume Beslon	Kurt Driessens	Greg Hamerly
Philippe Bessières	Isabel Drost	Ji He
Matjaz Bevk	Chris Drummond	Jaana Heino
Steffen Bickel	Wenliang Du	Thomas Heitz
Gilles Bisson	Nicolas Durand	Frank Herrmann
Avrim Blum	Michael Egmont-Petersen	Haitham Hindi
Axel Blumenstock	Craig Eldershaw	Ayca Azgin Hintoglu
Damjan Bojadžiev	Mohammed El-Hajj	Joachim Hipp
Francesco Bonchi	Roberto Esposito	Susanne Hoche
Toufik Boudellal	Timm Euler	Pieter Jan 't Hoen
Omar Boussaid	Theodoros Evgeniou	Andreas Hotho
Janez Brank	Anna Maria Fanelli	Tomas Hrycej
Nicolas Bredeche	Nicola Fanizzi	Luigi Iannone
Ulf Brefeld	Ayman Farahat	Inaki Inza
Wray Buntine	Sebastien Ferre	François Jacquenet
Christoph Büscher	Stefano Ferilli	Aleks Jakulin
Benjamin Bustos	Daan Fierens	Jean-Christophe Janodet
Niccolo Capanni	Thomas Finley	Nathalie Japkowicz
Amedeo Cappelli	Sergio Flesca	Tony Jebara

Tao-Yuan Jen	Marco Locatelli	Johann Petrak
Tao Jiang	Huma Lodhi	Sergios Petridis
Xing Jiang	Ricardo Lopes	Viet Phan-Luong
Yuelong Jiang	Pasquale Lops	Dimitris Pierrakos
Alípio Jorge	Robert Lothian	Joël Plisson
Pierre-Emmanuel Jouve	Claudio Lucchese	Neoklis Polyzotis
Matti Kääriäinen	Jack Lutz	Luboš Popelínský
Spiros Kapetanakis	Tuomo Malinen	Roland Priemer
Vangelis Karkaletsis	Michael Maltrud	Kai Puolamäki
Andreas Karwath	Suresh Manandhar	Sabine Rabaseda
Branko Kavšek	Alain-Pierre Manine	Filip Radlinski
Steffen Kempe	Raphael Marée	Mika Raento
Kristian Kersting	Berardi Margherita	Jan Ramon
Jahwan Kim	Elio Masciari	Ari Rantanen
Minsoo Kim	Cyrille Masson	Pierre Renaux
Svetlana Kiritchenko	Nicolas Méger	Chiara Renso
Richard Kirkby	Carlo Meghini	Rita Ribeiro
Jyrki Kivinen	Corrado Mencar	Lothar Richter
Willi Kloesgen	Amar-Djalil Mezaour	Salvatore Rinzivillo
Gabriella Kókai	Tatiana Miazhynskaia	François Rioult
Petri Kontkanen	Alessio Micheli	Stefano Rizzi
Dimitrios Kosmopoulos	Taneli Mielikäinen	Céline Robardet
Mark-A. Krogel	Ingo Mierswa	Mathieu Roche
Jussi Kujala	Tommi Mononen	Pedro Rodrigues
Matjaž Kukar	Martin Možina	Teemu Roos
Kari Laasonen	Thierry Murgue	Benjamin Rosenfeld
Krista Lagus	Mirco Nanni	Roman Rosipal
Lotfi Lakhil	Phu Chien Nguyen	Fabrice Rossi
Stéphane Lallich	Tuan Trung Nguyen	Olga Roudenko
Gert Lanckriet	Alexandru Niculescu-Mizil	Antonin Rozsypal
John Langford	Siegfried Nijssen	Ulrich Rückert
Carsten Lanquillon	Janne Nikkilä	Salvatore Ruggieri
Antonietta Lanza	Blaž Novak	Stefan Rüping
Michele Lapi	Alexandros Ntoulas	Nicolas Sabouret
Dominique Laurent	William O'Neill	Aleksander Sadikov
Yan-Nei Law	Kouzou Ohara	Taro L. Saito
Neil Lawrence	Arlindo L. Oliveira	Lorenza Saitta
Gregor Leban	Santiago Ontañón	Luka Šajn
Sau Dan Lee	Riccardo Ortale	Apkar Salatian
Gaëlle Legrand	Martijn van Otterlo	Marko Salmenkivi
Edda Leopold	Gerhard Paass	Craig Saunders
Claire Leschi	Ignazio Palmisano	Alexandr Savinov
Guichong Li	Christian Panse	Jelber Sayyad Shirabad
Oriana Licchelli	Andrea Passerini	Francesco Scarcello
Per Lidén	Jaakko Peltonen	Christoph Schmitz
Jussi T. Lindgren	Lourdes Pena	Joern Schneidewind
Francesca A. Lisi	Raffaele Perego	Martin Scholz
Bing Liu	José Ramón Quevedo Pérez	Tobias Schreck
Zhenyu Liu	Fernando Perez-Cruz	Ingo Schwab
Peter Ljubič	Georgios Petasis	Mihaela Scuturici

Vasile-Marian Scuturici	Prasad Tadepalli	Alessandro Vullo
Alexander K. Seewald	Andrea Tagarelli	Bernard Ženko
Jouni K. Seppänen	Julien Tane	Martin Žnidaršič
Jun Sese	Alexandre Termier	Haixun Wang
Georgios Sigletos	Evimaria Terzi	Xin Wang
Marko Robnik-Šikonja	Franck Thollard	Yizhou Wang
Fabrizio Silvestri	Andrea Torsello	Hannes Wettig
Janne Sinkkonen	Alain Trubuil	Nirmalie Wiratunga
Mike Sips	Athanasios Tsakonas	Jakub Wroblewski
Dominik Slezak	Chrisa Tsinaraki	Michael Wurst
Giovanni Soda	Ville Tuulos	Dan Xiao
Larisa Soldatova	Yannis Tzitzikas	Tomoyuki Yamada
Arnaud Soulet	Jaideep S. Vaidya	Robert J. Yan
Alessandro Sperduti	Pascal Vaillant	Hong Yao
Jaroslav Stepaniuk	Alexandros Valarakos	Ghim-Eng Yap
Olga Stepankova	Anneleen Van Assche	Kihoon Yoon
Umberto Straccia	Antonio Varlaro	Bianca Zadrozny
Alexander L. Strehl	Guillaume Vauvert	Fabio Zambetta
Thomas Strohmman	Julien Velcin	Farida Zehraoui
Jan Struyf	Celine Vens	Bernard Zenko
Dorian Šuc	Naval K. Verma	Xiang Zhang
Henri-Maxime Suchier	Ricardo Vilalta	Alexander Zien
Johan Suykens	Alexei Vinokourov	Albrecht Zimmermann
Piotr Synak	Daniel Vladusic	
Marcin Szczuka	Nikos Vlassis	

## ECML/PKDD 2004 Tutorials

Evaluation in Web Mining

*Bettina Berendt, Ernestina Menasalvas, Myra Spiliopoulou*

Symbolic Data Analysis

*Edwin Diday, Carlos Marcelo*

Radial Basis Functions: An Algebraic Approach (with Data Mining Applications)

*Amrit L. Goel, Miyoung Shin*

Mining Unstructured Data

*Ronen Feldman*

Statistical Approaches Used in Machine Learning

*Bruno Apolloni, Dario Malchiodi*

Rule-Based Data Mining Methods for Classification Problems in the Biomedical Domain

*Jinyan Li, Limsoon Wong*

Distributed Data Mining for Sensor Networks

*Hillol Kargupta*

## ECML/PKDD 2004 Workshops

Statistical Approaches for Web Mining (SAWM)

*Marco Gori, Michelangelo Ceci, Mirco Nanni*

Symbolic and Spatial Data Analysis: Mining Complex Data Structures

*Paula Brito, Monique Noirhomme*

Third International Workshop on Knowledge Discovery in Inductive Databases (KDID 2004)

*Bart Goethals, Arno Siebes*

Data Mining and Adaptive Modelling Methods for Economics and Management (IWAMEM 2004)

*Pavel Brazdil, Fernando S. Oliveira, Giulio Bottazzi*

Privacy and Security Issues in Data Mining

*Yücel Saygin*

Knowledge Discovery and Ontologies

*Paul Buitelaar, Jürgen Franke, Marko Grobelnik, Gerhard Paaß, Vojtech Svátek*

Mining Graphs, Trees and Sequences (MGTS 2004)

*Joost Kok, Takashi Washio*

Advances in Inductive Rule Learning

*Johannes Fürnkranz*

Data Mining and Text Mining for Bioinformatics

*Tobias Scheffer*

Knowledge Discovery in Data Streams

*Jesus Aguilar-Ruiz, Joao Gama*

# Table of Contents

## Invited Papers

Random Matrices in Data Analysis . . . . .	1
<i>Dimitris Achlioptas</i>	
Data Privacy . . . . .	8
<i>Rakesh Agrawal</i>	
Breaking Through the Syntax Barrier: Searching with Entities and Relations . . . . .	9
<i>Soumen Chakrabarti</i>	
Real-World Learning with Markov Logic Networks . . . . .	17
<i>Pedro Domingos</i>	
Strength in Diversity: The Advance of Data Analysis . . . . .	18
<i>David J. Hand</i>	

## Contributed Papers

Mining Positive and Negative Association Rules: An Approach for Confined Rules . . . . .	27
<i>Maria-Luiza Antonie and Osmar R. Zaiane</i>	
An Experiment on Knowledge Discovery in Chemical Databases . . . . .	39
<i>Sandra Berasaluce, Claude Laurenço, Amedeo Napoli, and Gilles Niel</i>	
Shape and Size Regularization in Expectation Maximization and Fuzzy Clustering . . . . .	52
<i>Christian Borgelt and Rudolf Kruse</i>	
Combining Multiple Clustering Systems . . . . .	63
<i>Constantinos Boulis and Mari Ostendorf</i>	
Reducing Data Stream Sliding Windows by Cyclic Tree-Like Histograms . . . . .	75
<i>Francesco Buccafurri and Gianluca Lax</i>	
A Framework for Data Mining Pattern Management . . . . .	87
<i>Barbara Catania, Anna Maddalena, Maurizio Mazza, Elisa Bertino, and Stefano Rizzi</i>	
Spatial Associative Classification at Different Levels of Granularity: A Probabilistic Approach . . . . .	99
<i>Michelangelo Ceci, Annalisa Appice, and Donato Malerba</i>	

AutoPart: Parameter-Free Graph Partitioning and Outlier Detection . . . . .	112
<i>Deepayan Chakrabarti</i>	
Properties and Benefits of Calibrated Classifiers . . . . .	125
<i>Ira Cohen and Moises Goldszmidt</i>	
A Tree-Based Approach to Clustering XML Documents by Structure . . . . .	137
<i>Gianni Costa, Giuseppe Manco, Riccardo Ortale, and Andrea Tagarelli</i>	
Discovery of Regulatory Connections in Microarray Data . . . . .	149
<i>Michael Egmont-Petersen, Wim de Jonge, and Arno Siebes</i>	
Learning from Little: Comparison of Classifiers Given Little Training . . . . .	161
<i>George Forman and Ira Cohen</i>	
Geometric and Combinatorial Tiles in 0–1 Data . . . . .	173
<i>Aristides Gionis, Heikki Mannila, and Jouni K. Seppänen</i>	
Document Classification Through Interactive Supervision of Document and Term Labels . . . . .	185
<i>Shantanu Godbole, Abhay Harpale, Sunita Sarawagi, and Soumen Chakrabarti</i>	
Classifying Protein Fingerprints . . . . .	197
<i>Melanie Hilario, Alex Mitchell, Jee-Hyub Kim, Paul Bradley, and Terri Attwood</i>	
Finding Interesting Pass Patterns from Soccer Game Records . . . . .	209
<i>Shoji Hirano and Shusaku Tsumoto</i>	
Discovering Unexpected Information for Technology Watch . . . . .	219
<i>François Jacquenet and Christine Largeron</i>	
Scalable Density-Based Distributed Clustering . . . . .	231
<i>Eshref Januzaj, Hans-Peter Kriegel, and Martin Pfeifle</i>	
Summarization of Dynamic Content in Web Collections . . . . .	245
<i>Adam Jatowt and Mitsuru Ishizuka</i>	
Mining Thick Skylines over Large Databases . . . . .	255
<i>Wen Jin, Jiawei Han, and Martin Ester</i>	
Ensemble Feature Ranking . . . . .	267
<i>Kees Jong, Jérémie Mary, Antoine Cornuéjols, Elena Marchiori, and Michèle Sebag</i>	
Privately Computing a Distributed $k$ -nn Classifier . . . . .	279
<i>Murat Kantarcioğlu and Chris Clifton</i>	
Incremental Nonlinear PCA for Classification . . . . .	291
<i>Byung Joo Kim and Il Kon Kim</i>	

A Spectroscopy of Texts for Effective Clustering . . . . .	301
<i>Wenyuan Li, Wee-Keong Ng, Kok-Leong Ong, and Ee-Peng Lim</i>	
Constraint-Based Mining of Episode Rules and Optimal Window Sizes . . . . .	313
<i>Nicolas Méger and Christophe Rigotti</i>	
Analysing Customer Churn in Insurance Data – A Case Study . . . . .	325
<i>Katharina Morik and Hanna Köpcke</i>	
Nomograms for Visualization of Naive Bayesian Classifier . . . . .	337
<i>Martin Možina, Janez Demšar, Michael Kattan, and Blaž Zupan</i>	
Using a Hash-Based Method for Apriori-Based Graph Mining . . . . .	349
<i>Phu Chien Nguyen, Takashi Washio, Kouzou Ohara, and Hiroshi Motoda</i>	
Evaluation of Rule Interestingness Measures with a Clinical Dataset on Hepatitis . . . . .	362
<i>Miho Ohsaki, Shinya Kitaguchi, Kazuya Okamoto, Hideto Yokoi, and Takahira Yamaguchi</i>	
Classification in Geographical Information Systems . . . . .	374
<i>Salvatore Rinzivillo and Franco Turini</i>	
Digging into Acceptor Splice Site Prediction: An Iterative Feature Selection Approach . . . . .	386
<i>Yvan Saeys, Sven Degroeve, and Yves Van de Peer</i>	
Itemset Classified Clustering . . . . .	398
<i>Jun Sese and Shinichi Morishita</i>	
Combining Winnow and Orthogonal Sparse Bigrams for Incremental Spam Filtering . . . . .	410
<i>Christian Siefkes, Fidelis Assis, Shalendra Chhabra, and William S. Yerazunis</i>	
Asynchronous and Anticipatory Filter-Stream Based Parallel Algorithm for Frequent Itemset Mining . . . . .	422
<i>Adriano Veloso, Wagner Meira Jr., Renato Ferreira, Dorgival Guedes Neto, and Srinivasan Parthasarathy</i>	
A Quantification of Cluster Novelty with an Application to Martian Topography . . . . .	434
<i>Ricardo Vilalta, Tom Stepinski, Muralikrishna Achari, and Francisco Ocegueda-Hernandez</i>	
Density-Based Spatial Clustering in the Presence of Obstacles and Facilitators . . . . .	446
<i>Xin Wang, Camilo Rostoker, and Howard J. Hamilton</i>	

Text Mining for Finding Functional Community of Related Genes  
Using TCM Knowledge ..... 459  
*Zhaohui Wu, Xuezhong Zhou, Baoyan Liu, and Junli Chen*

Dealing with Predictive-but-Unpredictable Attributes  
in Noisy Data Sources ..... 471  
*Ying Yang, Xindong Wu, and Xingquan Zhu*

A New Scheme on Privacy Preserving Association Rule Mining ..... 484  
*Nan Zhang, Shengquan Wang, and Wei Zhao*

**Posters**

A Unified and Flexible Framework for Comparing Simple  
and Complex Patterns ..... 496  
*Ilaria Bartolini, Paolo Ciaccia, Irene Ntoutsis, Marco Patella,  
and Yannis Theodoridis*

Constructing (Almost) Phylogenetic Trees  
from Developmental Sequences Data ..... 500  
*Ronnie Bathoorn and Arno Siebes*

Learning from Multi-source Data ..... 503  
*Élisa Fromont, Marie-Odile Cordier, and René Quiniou*

The Anatomy of SnakeT: A Hierarchical Clustering Engine  
for Web-Page Snippets ..... 506  
*Paolo Ferragina and Antonio Gulli*

COCOA: Compressed Continuity Analysis for Temporal Databases ..... 509  
*Kuo-Yu Huang, Chia-Hui Chang, and Kuo-Zui Lin*

Discovering Interpretable Muscle Activation Patterns  
with the Temporal Data Mining Method ..... 512  
*Fabian Mörchen, Alfred Ultsch, and Olaf Hoos*

A Tolerance Rough Set Approach to Clustering Web Search Results ..... 515  
*Chi Lang Ngo and Hung Son Nguyen*

Improving the Performance of the RISE Algorithm ..... 518  
*Aloisio Carlos de Pina and Gerson Zaverucha*

Mining History of Changes to Web Access Patterns ..... 521  
*Qiankun Zhao and Sourav S. Bhowmick*

**Demonstration Papers**

Visual Mining of Spatial Time Series Data ..... 524  
*Gennady Andrienko, Natalia Andrienko, and Peter Gatalsky*

Detecting Driving Awareness . . . . .	528
<i>Bruno Apolloni, Andrea Brega, Dario Malchiodi, and Cristian Mesiano</i>	
An Effective Recommender System for Highly Dynamic and Large Web Sites . . . . .	531
<i>Ranieri Baraglia, Francesco Merlo, and Fabrizio Silvestri</i>	
SemanticTalk: Software for Visualizing Brainstorming Sessions and Thematic Concept Trails on Document Collections . . . . .	534
<i>Chris Biemann,, Karsten Böhm, Gerhard Heyer, and Ronny Melz</i>	
Orange: From Experimental Machine Learning to Interactive Data Mining . . . . .	537
<i>Janez Demšar, Blaž Zupan, Gregor Leban, and Tomaz Curk</i>	
Terrorist Detection System . . . . .	540
<i>Yuval Elovici, Abraham Kandel, Mark Last, Bracha Shapira, Omer Zaafrany, Moti Schneider, and Menahem Friedman</i>	
Experimenting SnakeT: A Hierarchical Clustering Engine for Web-Page Snippets . . . . .	543
<i>Paolo Ferragina and Antonio Gulli</i>	
HIClass: Hyper-interactive Text Classification by Interactive Supervision of Document and Term Labels . . . . .	546
<i>Shantanu Godbole, Abhay Harpale, Sunita Sarawagi, and Soumen Chakrabarti</i>	
BALIOS – The Engine for Bayesian Logic Programs . . . . .	549
<i>Kristian Kersting and Uwe Dick</i>	
SEWeP: A Web Mining System Supporting Semantic Personalization . . . . .	552
<i>Stratos Paulakis, Charalampos Lampos, Magdalini Eirinaki, and Michalis Vazirgiannis</i>	
SPIN! Data Mining System Based on Component Architecture . . . . .	555
<i>Alexandr Savinov</i>	
<b>Author Index . . . . .</b>	<b>559</b>