

Lecture Notes in Artificial Intelligence 2838

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Springer

Berlin

Heidelberg

New York

Hong Kong

London

Milan

Paris

Tokyo

Nada Lavrač Dragan Gamberger
Ljupčo Todorovski Hendrik Blockeel (Eds.)

Knowledge Discovery in Databases: PKDD 2003

7th European Conference on Principles and Practice of
Knowledge Discovery in Databases
Cavtat-Dubrovnik, Croatia, September 22-26, 2003
Proceedings



Springer

Series Editors

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

Volume Editors

Nada Lavrač
Ljupčo Todorovski
Jožef Stefan Institute, Dept. of Intelligent Systems
Jamova 39, 1000 Ljubljana, Slovenia
E-mail: {Nada.Lavrac/Ljupco.Todorovski}@ijs.si

Dragan Gamberger
Rudjer Bošković Institute
Bijenička 54, 10000 Zagreb, Croatia
E-mail: Dragan.Gamberger@irb.hr

Hendrik Blockeel
Katholieke Universiteit Leuven, Dept. of Computer Science
Celestijnenlaan 200A, 3001 Leuven, Belgium
E-mail: Hendrik.Blokeel@cs.kuleuven.ac.be

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress

Bibliographic information published by Die Deutsche Bibliothek
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie;
detailed bibliographic data is available in the Internet at <<http://dnb.ddb.de>>.

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

ISSN 0302-9743

ISBN 3-540-20085-1 Springer-Verlag 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-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer-Verlag Berlin Heidelberg New York,
a member of BertelsmannSpringer Science+Business Media GmbH

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2003
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 10955635 06/3142 5 4 3 2 1 0

Preface

The proceedings of ECML/PKDD 2003 are published in two volumes: the *Proceedings of the 14th European Conference on Machine Learning* (LNAI 2837) and the *Proceedings of the 7th European Conference on Principles and Practice of Knowledge Discovery in Databases* (LNAI 2838). The two conferences were held on September 22–26, 2003 in Cavtat, a small tourist town in the vicinity of Dubrovnik, Croatia.

As machine learning and knowledge discovery are two highly related fields, the co-location of both conferences is beneficial for both research communities. In Cavtat, ECML and PKDD were co-located for the third time in a row, following the successful co-location of the two European conferences in Freiburg (2001) and Helsinki (2002). The co-location of ECML 2003 and PKDD 2003 resulted in a joint program for the two conferences, including paper presentations, invited talks, tutorials, and workshops.

Out of 332 submitted papers, 40 were accepted for publication in the ECML 2003 proceedings, and 40 were accepted for publication in the PKDD 2003 proceedings. All the submitted papers were reviewed by three referees. In addition to submitted papers, the conference program consisted of four invited talks, four tutorials, seven workshops, two tutorials combined with a workshop, and a discovery challenge.

We wish to express our gratitude to

- the authors of submitted papers,
- the program committee members, for thorough and timely paper evaluation,
- invited speakers Pieter Adriaans, Leo Breiman, Christos Faloutsos, and Donald B. Rubin,
- tutorial and workshop chairs Stefan Kramer, Luis Torgo, and Luc Dehaspe,
- local and technical organization committee members,
- advisory board members Luc De Raedt, Tapio Elomaa, Peter Flach, Heikki Mannila, Arno Siebes, and Hannu Toivonen,
- awards and grants committee members Dunja Mladenić, Rob Holte, and Michael May,
- Richard van der Stadt for the development of CyberChair which was used to support the paper submission and evaluation process,
- Alfred Hofmann of Springer-Verlag for co-operation in publishing the proceedings, and finally
- we gratefully acknowledge the financial support of the Croatian Ministry of Science and Technology, Slovenian Ministry of Education, Science, and Sports, and the Knowledge Discovery Network of Excellence (KDNet). KDNet also sponsored the student grants and best paper awards, while Kluwer Academic Publishers (the Machine Learning Journal) awarded a prize for the best student paper.

We hope and trust that the week in Cavtat in late September 2003 will be remembered as a fruitful, challenging, and enjoyable scientific and social event.

June 2003

Nada Lavrač
Dragan Gamberger
Hendrik Blockeel
Ljupčo Todorovski

ECML/PKDD 2003 Organization

Executive Committee

Program Chairs:	Nada Lavrač (Jožef Stefan Institute, Slovenia) ECML and PKDD chair Dragan Gamberger (Rudjer Bošković Institute, Croatia) ECML and PKDD co-chair Hendrik Blockeel (Katholieke Universiteit Leuven, Belgium) ECML co-chair Ljupčo Todorovski (Jožef Stefan Institute, Slovenia) PKDD co-chair
Tutorial and Workshop Chair:	Stefan Kramer (Technische Universität München, Germany)
Workshop Co-chair:	Luis Torgo (University of Porto, Portugal)
Tutorial Co-chair:	Luc Dehaspe (PharmaDM, Belgium)
Challenge Chair:	Petr Berka (University of Economics, Prague, Czech Republic)
Advisory Board:	Luc De Raedt (Albert-Ludwigs University Freiburg, Germany) Tapio Elomaa (University of Helsinki, Finland) Peter Flach (University of Bristol, UK) Heikki Mannila (Helsinki Institute for Information Technology, Finland) Arno Siebes (Utrecht University, The Netherlands) Hannu Toivonen (University of Helsinki, Finland)
Awards and Grants Committee:	Dunja Mladenić (Jožef Stefan Institute, Slovenia) Rob Holte (University of Alberta, Canada) Michael May (Fraunhofer AIS, Germany) Hendrik Blockeel (Katholieke Universiteit Leuven, Belgium)
Local Chairs:	Dragan Gamberger, Tomislav Šmuc (Rudjer Bošković Institute)
Organization Committee:	Darek Krzywania, Celine Vens, Jan Struyf (Katholieke Universiteit Leuven, Belgium), Damjan Demšar, Branko Kavšek, Milica Bauer, Bernard Ženko, Peter Ljubič (Jožef Stefan Institute, Slovenia), Mirna Benat (Rudjer Bošković Institute), Dalibor Ivušić (The Polytechnic of Dubrovnik, Croatia), Zdenko Sonicki (University of Zagreb, Croatia)

ECML 2003 Program Committee

H. Blockeel, Belgium	D. Malerba, Italy
A. van den Bosch, The Netherlands	H. Mannila, Finland
H. Boström, Sweden	S. Matwin, Canada
I. Bratko, Slovenia	J. del R. Millán, Switzerland
P. Brazdil, Portugal	D. Mladenić, Slovenia
W. Buntine, Finland	K. Morik, Germany
M. Craven, USA	H. Motoda, Japan
N. Cristianini, USA	R. Nock, France
J. Cussens, UK	D. Page, USA
W. Daelemans, Belgium	G. Paliouras, Greece
L. Dehaspe, Belgium	B. Pfahringer, New Zealand
L. De Raedt, Germany	E. Plaza, Spain
S. Džeroski, Slovenia	J. Rousu, Finland
T. Elomaa, Finland	C. Rouveirol, France
F. Esposito, Italy	L. Saitta, Italy
B. Filipič, Slovenia	T. Scheffer, Germany
P. Flach, UK	M. Sebag, France
J. Fürnkranz, Austria	J. Shawe-Taylor, UK
J. Gama, Portugal	A. Siebes, The Netherlands
D. Gamberger, Croatia	D. Sleeman, UK
J.-G. Ganascia, France	R.H. Sloan, USA
L. Getoor, USA	M. van Someren, The Netherlands
H. Hirsh, USA	P. Stone, USA
T. Hofmann, USA	J. Suykens, Belgium
T. Horvath, Germany	H. Tirri, Finland
T. Joachims, USA	L. Todorovski, Slovenia
D. Kazakov, UK	L. Torgo, Portugal
R. Khardon, USA	P. Turney, Canada
Y. Kodratoff, France	P. Vitanyi, The Netherlands
I. Kononenko, Slovenia	S.M. Weiss, USA
S. Kramer, Germany	G. Widmer, Austria
M. Kubat, USA	M. Wiering, The Netherlands
S. Kwek, USA	R. Wirth, Germany
N. Lavrač, Slovenia	S. Wrobel, Germany
C. Ling, Canada	T. Zeugmann, Germany
R. López de Màntaras, Spain	B. Zupan, Slovenia

PKDD 2003 Program Committee

H. Ahonen-Myka, Finland	H. Mannila, Finland
E. Baralis, Italy	S. Matwin, Canada
R. Bellazzi, Italy	M. May, Germany
M.R. Berthold, USA	D. Mladenić, Slovenia
H. Blockeel, Belgium	S. Morishita, Japan
M. Bohanec, Slovenia	H. Motoda, Japan
J.F. Boulicaut, France	G. Nakhaeizadeh, Germany
B. Crémilleux, France	C. Nédellec, France
L. Dehaspe, Belgium	D. Page, USA
L. De Raedt, Germany	Z.W. Ras, USA
S. Džeroski, Slovenia	J. Rauch, Czech Republic
T. Elomaa, Finland	G. Ritschard, Switzerland
M. Ester, Canada	M. Sebag, France
A. Feelders, The Netherlands	F. Sebastiani, Italy
R. Feldman, Israel	M. Sebban, France
P. Flach, UK	A. Siebes, The Netherlands
E. Frank, New Zealand	A. Skowron, Poland
A. Freitas, UK	M. van Someren, The Netherlands
J. Fürnkranz, Austria	M. Spiliopoulou, Germany
D. Gamberger, Croatia	N. Spyratos, France
F. Giannotti, Italy	R. Stolle, USA
C. Giraud-Carrier, Switzerland	E. Suzuki, Japan
M. Grobelnik, Slovenia	A. Tan, Singapore
H.J. Hamilton, Canada	L. Todorovski, Slovenia
J. Han, USA	H. Toivonen, Finland
R. Hilderman, Canada	L. Torgo, Portugal
H. Hirsh, USA	S. Tsumoto, Japan
S.J. Hong, USA	A. Unwin, Germany
F. Höppner, Germany	K. Wang, Canada
S. Kaski, Finland	L. Wehenkel, Belgium
J.-U. Kietz, Switzerland	D. Wettschereck, Germany
R.D. King, UK	G. Widmer, Austria
W. Kloesgen, Germany	R. Wirth, Germany
Y. Kodratoff, France	S. Wrobel, Germany
J.N. Kok, The Netherlands	M.J. Zaki, USA
S. Kramer, Germany	D.A. Zighed, France
N. Lavrač, Slovenia	B. Zupan, Slovenia
G. Manco, Italy	

ECML/PKDD 2003 Additional Reviewers

F. Aiolli	L. Geng	H.S. Nguyen
A. Amrani	A. Giacometti	S. Nijssen
A. Appice	T. Giorgino	A. Nowé
E. Armengol	B. Goethals	M. Ohtani
I. Autio	M. Grabert	S. Ontańón
J. Azé	E. Gyftodimos	R. Ortale
I. Azzini	W. Härmäläinen	M. Ould Abdel Vetah
M. Baglioni	A. Habrard	G. Paaß
A. Banerjee	M. Hall	I. Palmisano
T.M.A. Basile	S. Hoche	J. Peltonen
M. Bendou	E. Hüllermeier	L. Peña
M. Berardi	L. Jacobs	D. Pedreschi
G. Beslon	A. Jakulin	G. Petasis
M. Bevk	T.Y. Jen	J. Petrak
A. Blumenstock	B. Jeudy	V. Phan Luong
D. Bojadžiev	A. Jorge	D. Pierrakos
M. Borth	R.J. Jun	U. Rückert
J. Brank	P. Juvan	S. Rüping
P. Brockhausen	M. Kääriäinen	J. Ramon
M. Ceci	K. Karimi	S. Ray
E. Cesario	K. Kersting	C. Rigotti
S. Chiusano	J. Kindermann	F. Rioult
J. Clech	S. Kiritchenko	M. Robnik-Šikonja
A. Cornuéjols	W. Kusters	M. Roche
J. Costa	I. Koychev	B. Rosenfeld
T. Curk	M. Kukar	A. Sadikov
M. Degemmis	S. Lallich	T. Saito
D. Demšar	C. Larizza	E. Savia
J. Demšar	D. Laurent	C. Savu-Krohn
M. Denecker	G. Leban	G. Schmidberger
N. Di Mauro	S.D. Lee	M. Scholz
K. Driessens	G. Legrand	A.K. Seewald
T. Erjavec	E. Leopold	J. Sese
T. Euler	J. Leskovec	G. Sigletos
N. Fanizzi	O. Licchelli	T. Silander
S. Ferilli	J.T. Lindgren	D. Slezak
M. Fernandes	F.A. Lisi	C. Soares
D. Finton	T. Malinen	D. Sonntag
S. Flesca	O. Matte-Tailliez	H.-M. Suchier
J. Franke	A. Mazzanti	B. Sudha
F. Furfaro	P. Medas	P. Synak
T. Gärtner	R. Meo	A. Tagarelli
P. Garza	T. Mielikäinen	Y. Tzitzikas

R. Vilalta
D. Vladušić
X. Wang
A. Wojna
J. Wróblewski

M. Wurst
R.J. Yan
X. Yan
H. Yao
X. Yin

I. Zogalis
W. Zou
M. Žnidaršič
B. Ženko

ECML/PKDD 2003 Tutorials

KD Standards

Sarabjot S. Anand, Marko Grobelnik, and Dietrich Wettschereck

Data Mining and Machine Learning in Time Series Databases

Eamonn Keogh

Exploratory Analysis of Spatial Data and Decision Making Using Interactive Maps and Linked Dynamic Displays

Natalia Andrienko and Gennady Andrienko

Music Data Mining

Darrell Conklin

ECML/PKDD 2003 Workshops

First European Web Mining Forum

Bettina Berendt, Andreas Hotho, Dunja Mladenić, Maarten van Someren, Myra Spiliopoulou, and Gerd Stumme

Multimedia Discovery and Mining

Dunja Mladenić and Gerhard Paaf

Data Mining and Text Mining in Bioinformatics

Tobias Scheffer and Ulf Leser

Knowledge Discovery in Inductive Databases

Jean-François Boulicaut, Sašo Džeroski, Mika Klemettinen, Rosa Meo, and Luc De Raedt

Graph, Tree, and Sequence Mining

Luc De Raedt and Takashi Washio

Probabilistic Graphical Models for Classification

Pedro Larrañaga, Jose A. Lozano, Jose M. Peña, and Iñaki Inza

Parallel and Distributed Computing for Machine Learning

Rui Camacho and Ashwin Srinivasan

Discovery Challenge: A Collaborative Effort in Knowledge Discovery from Databases

Petr Berka, Jan Rauch, and Shusaku Tsumoto

ECML/PKDD 2003 Joint Tutorials-Workshops

Learning Context-Free Grammars

Colin de la Higuera, Jose Oncina, Pieter Adriaans, Menno van Zaanen

Adaptive Text Extraction and Mining

Fabio Ciravegna, Nicholas Kushmerick

Table of Contents

Invited Papers

From Knowledge-Based to Skill-Based Systems: Sailing as a Machine Learning Challenge	1
<i>Pieter Adriaans</i>	
Two-Eyed Algorithms and Problems	9
<i>Leo Breiman</i>	
Next Generation Data Mining Tools: Power Laws and Self-similarity for Graphs, Streams and Traditional Data	10
<i>Christos Faloutsos</i>	
Taking Causality Seriously: Propensity Score Methodology Applied to Estimate the Effects of Marketing Interventions	16
<i>Donald B. Rubin</i>	

Contributed Papers

Efficient Statistical Pruning of Association Rules	23
<i>Alan Ableson and Janice Glasgow</i>	
Majority Classification by Means of Association Rules	35
<i>Elena Baralis and Paolo Garza</i>	
Adaptive Constraint Pushing in Frequent Pattern Mining	47
<i>Francesco Bonchi, Fosca Giannotti, Alessio Mazzanti, and Dino Pedreschi</i>	
ExAnte: Anticipated Data Reduction in Constrained Pattern Mining	59
<i>Francesco Bonchi, Fosca Giannotti, Alessio Mazzanti, and Dino Pedreschi</i>	
Minimal k -Free Representations of Frequent Sets	71
<i>Toon Calders and Bart Goethals</i>	
Discovering Unbounded Episodes in Sequential Data	83
<i>Gemma Casas-Garriga</i>	
Mr-SBC: A Multi-relational Naïve Bayes Classifier	95
<i>Michelangelo Ceci, Annalisa Appice, and Donato Malerba</i>	

SMOTEBoost: Improving Prediction of the Minority Class in Boosting ... 107
*Nitesh V. Chawla, Aleksandar Lazarevic, Lawrence O. Hall,
and Kevin W. Bowyer*

Using Belief Networks and Fisher Kernels
for Structured Document Classification 120
Ludovic Denoyer and Patrick Gallinari

A Skeleton-Based Approach to Learning Bayesian Networks from Data ... 132
Steven van Dijk, Linda C. van der Gaag, and Dirk Thierens

On Decision Boundaries of Naïve Bayes in Continuous Domains 144
Tapio Elomaa and Juho Rousu

Application of Inductive Logic Programming
to Structure-Based Drug Design 156
David P. Enot and Ross D. King

Visualizing Class Probability Estimators 168
Eibe Frank and Mark Hall

Automated Detection of Epidemics from the Usage Logs
of a Physicians' Reference Database 180
Jaana Heino and Hannu Toivonen

An Indiscernibility-Based Clustering Method with Iterative Refinement
of Equivalence Relations 192
Shoji Hirano and Shusaku Tsumoto

Preference Mining: A Novel Approach on Mining User Preferences
for Personalized Applications 204
Stefan Holland, Martin Ester, and Werner Kießling

Explaining Text Clustering Results Using Semantic Structures 217
Andreas Hotho, Steffen Staab, and Gerd Stumme

Analyzing Attribute Dependencies 229
Aleks Jakulin and Ivan Bratko

Ranking Interesting Subspaces for Clustering High Dimensional Data 241
Karin Kailing, Hans-Peter Kriegel, Peer Kröger, and Stefanie Wanka

Efficiently Finding Arbitrarily Scaled Patterns
in Massive Time Series Databases 253
Eamonn Keogh

Using Transduction and Multi-view Learning to Answer Emails 266
Michael Kockelkorn, Andreas Lüneburg, and Tobias Scheffer

Exploring Fringe Settings of SVMs for Classification	278
<i>Adam Kowalczyk and Bhavani Raskutti</i>	
Rule Discovery and Probabilistic Modeling for Onomastic Data	291
<i>Antti Leino, Heikki Mannila, and Ritva Liisa Pitkänen</i>	
Constraint-Based Mining of Sequential Patterns over Datasets with Consecutive Repetitions	303
<i>Marion Leleu, Christophe Rigotti, Jean-François Boulıcaut, and Guillaume Euvrard</i>	
Symbolic Distance Measurements Based on Characteristic Subspaces	315
<i>Marcus-Christopher Ludl</i>	
The Pattern Ordering Problem	327
<i>Taneli Mielikäinen and Heikki Mannila</i>	
Collaborative Filtering Using Restoration Operators	339
<i>Atsuyoshi Nakamura, Mineichi Kudo, and Akira Tanaka</i>	
Efficient Frequent Query Discovery in FARMER	350
<i>Siegfried Nijssen and Joost N. Kok</i>	
Towards Behaviometric Security Systems: Learning to Identify a Typist	363
<i>Mordechai Nisenson, Ido Yariv, Ran El-Yaniv, and Ron Meir</i>	
Efficient Density Clustering Method for Spatial Data	375
<i>Fei Pan, Baoying Wang, Yi Zhang, Dongmei Ren, Xin Hu, and William Perrizo</i>	
Statistical σ -Partition Clustering over Data Streams	387
<i>Nam Hun Park and Won Suk Lee</i>	
Enriching Relational Learning with Fuzzy Predicates	399
<i>Henri Prade, Gilles Richard, and Mathieu Serrurier</i>	
Text Categorisation Using Document Profiling	411
<i>Maximilien Sauban and Bernhard Pfahringer</i>	
A Simple Algorithm for Topic Identification in 0–1 Data	423
<i>Jouni K. Seppänen, Ella Bingham, and Heikki Mannila</i>	
Bottom-Up Learning of Logic Programs for Information Extraction from Hypertext Documents	435
<i>Bernd Thomas</i>	
Predicting Outliers	447
<i>Luis Torgo and Rita Ribeiro</i>	

Mining Rules of Multi-level Diagnostic Procedure from Databases 459
Shusaku Tsumoto

Learning Characteristic Rules Relying on Quantified Paths 471
*Teddy Turmeaux, Ansaf Salleb, Christel Vrain,
and Daniel Cassard*

Topic Learning from Few Examples 483
Huaiyu Zhu, Shivakumar Vaithyanathan, and Mahesh V. Joshi

Arbogodaï, a New Approach for Decision Trees 495
*Djamel A. Zighed, Gilbert Ritschard, Walid Erray,
and Vasile-Marian Scuturici*

Author Index 507