

Lecture Notes in Artificial Intelligence

2168

Subseries of Lecture Notes in Computer Science

Edited by J. G. Carbonell and J. Siekmann

Lecture Notes in Computer Science

Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

Springer

Berlin

Heidelberg

New York

Barcelona

Hong Kong

London

Milan

Paris

Tokyo

Luc De Raedt Arno Siebes (Eds.)

Principles of Data Mining and Knowledge Discovery

5th European Conference, PKDD 2001
Freiburg, Germany, September 3-5, 2001
Proceedings

Dedicated to Jan Żytkow



Springer

Series Editors

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

Volume Editors

Luc De Raedt

Albert-Ludwigs University Freiburg

Department of Computer Science

Georges Köhler-Allee, Geb. 079, 79110 Freiburg, Germany

E-mail: deraedt@informatik.uni-freiburg.de

Arno Siebes

University of Utrecht, Inst. of Information and Computing Sciences

Dept. of Mathematics and Computer Science

Padualaan 14, de Uithof, 3508 TB Utrecht, The Netherlands

E-mail: arno@cs.uu.nl

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Principles of data mining and knowledge discovery : 5th European conference proceedings / PKDD 2001, Freiburg, Germany, September 3 - 5, 2001. Luc de Raedt ; Arno Siebes (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Tokyo : Springer, 2001

(Lecture notes in computer science ; Vol. 2168 : Lecture notes in artificial intelligence)

ISBN 3-540-42534-9

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

ISBN 3-540-42534-9 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 2001

Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP Berlin, Stefan Sossna

Printed on acid-free paper SPIN 10840371 06/3142 5 4 3 2 1 0

In Memoriam

The PKDD community has learned with great sadness that Jan Zytkow passed away on Jan. 16, 2001. Jan was a well-known researcher in the area of scientific discovery, a pioneer in machine learning, an author of many publications and books, an organizer of many conferences and meetings, the driving force behind the PKDD conferences, a wonderful person, and a friend. Those who knew him will miss him. These proceedings are dedicated to Jan.

Preface

It is our pleasure to present the proceedings of the 12th European Conference on Machine Learning (*Lecture Notes in Artificial Intelligence 2167*) and the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases (this volume). These two conferences were held from September 3–7, 2001 in Freiburg, Germany, marking the first time – world-wide – that a data mining conference has been co-located with a machine learning conference.

As Program Committee co-chairs of the two conferences, our goal was to co-ordinate the submission and reviewing process as much as possible. Here are some statistics: a total of 117 papers was submitted to ECML 2001, 78 papers were submitted to PKDD 2001, and 45 papers were submitted as joint papers. Each paper was carefully reviewed by 3 (in exceptional circumstances 2 or 4) members of the Program Committees. Out of the 240 submitted papers, 40 were accepted after the first reviewing round, and 54 were accepted on the condition that the final paper would meet the requirements of the reviewers. In the end, 90 papers were accepted for the proceedings (50 for ECML 2001 and 40 for PKDD 2001).

We were also aiming at putting together a 5-day program that would be attractive, in its entirety, to both communities. This would encourage participants to stay the whole week and thus foster interaction and cross-fertilization. The PKDD 2001 conference ran from Monday to Wednesday, and the ECML 2001 conference from Wednesday to Friday, with the Wednesday program carefully selected to be of interest to a mixed audience. On each day there was an invited talk by an internationally renowned scientist. Tom Dietterich spoke on *Support Vector Machines for Reinforcement Learning*; Heikki Mannila on *Combining Discrete Algorithmic and Probabilistic Approaches in Data Mining*; Antony Unwin on *Statistification or Mystification, the Need for Statistical Thought in Visual Data Mining*; Gerhard Widmer on *The Musical Expression Project: A Challenge for Machine Learning and Knowledge Discovery*; and Stefan Wrobel on *Scalability, Search and Sampling: From Smart Algorithms to Active Discovery*. In addition, there was an extensive parallel program of 11 workshops and 8 tutorials. Two workshops were devoted to results achieved by the participants in the two learning and mining challenges that were set prior to the conferences.

It has been a great pleasure for us to prepare and organize such a prestigious event, but of course we could not have done it without the help of many colleagues. We would like to thank all the authors who submitted papers to ECML 2001 and PKDD 2001, the program committee members of both conferences, the other reviewers, the invited speakers, the workshop organizers, and the tutorial speakers. We are particularly grateful to the workshop chairs Johannes Fürnkranz and Stefan Wrobel; the tutorial chairs Michèle Sebag and Hannu Toivonen; and the challenge chairs Petr Berka and Christoph Helma for their assistance in putting together an exciting scientific program. Many

thanks to Michael Keser for his technical support in setting up the CyberChair website, to Richard van de Stadt for developing CyberChair, and to the local team at Freiburg for the organizational support provided. We would also like to thank Alfred Hofmann of Springer-Verlag for his co-operation in publishing these proceedings. Finally, we gratefully acknowledge the financial support provided by the sponsors; EU Network of Excellence MLnet II, National Institute of Environmental Health Sciences (US), SICK AG, the city of Freiburg, and the Albert-Ludwigs University Freiburg and its Lab for Machine Learning.

Although at the time of writing the event is yet to take place, we are confident that history will cast a favorable eye, and we are looking forward to continued and intensified integration of the European machine learning and data mining communities that we hope has been set in motion with this event.

July 2001

Luc De Raedt
Peter Flach
Arno Siebes

Executive Committees

- Program Chairs ECML-2001: Peter Flach (University of Bristol)
 Luc De Raedt (Albert-Ludwigs University)
- Program Chairs PKDD-2001: Arno Siebes (Utrecht University)
 Luc De Raedt (Albert-Ludwigs University)
- Tutorial Chairs: Michèle Sebag (Ecole Polytechnique)
 Hannu Toivonen (Nokia)
- Workshop Chairs: Johannes Fürnkranz (Austrian Research Institute
 for Artificial Intelligence)
 Stefan Wrobel (University of Magdeburg)
- Challenge Chairs: Petr Berka (University of Economics Prague)
 Cristoph Helma (Albert-Ludwigs University)
- Local Chairs: Luc De Raedt (Albert-Ludwigs University)
 Stefan Kramer (Albert-Ludwigs University)
- Local Organization: Catherine Blocher (K&K Freiburg)
 Susanne Bourjaillat (Albert-Ludwigs University)
 Dirk Hähnel (Albert-Ludwigs University)
 Cristoph Helma (Albert-Ludwigs University)
 Michael Keser (Albert-Ludwigs University)
 Kristian Kersting (Albert-Ludwigs University)
 Walter Koch (K&K Freiburg)
 Andrea Kricheldorf (K&K Freiburg)
- Advisory Committee : Ramon Lopez de Mantaras (IIIA-CSIC Artificial
 Intelligence Research Institute)
 Heikki Mannila (Nokia and Helsinki University of
 Technology)
 Jan Rauch (University of Economics Prague)
 Maarten van Someren (University of Amsterdam)
 Stefan Wrobel (University of Magdeburg)
 Djamel Zighed (University of Lyon 2)
- PKDD Steering Committee : Jan Komorowski (Norwegian University of Sci-
 ence and Technology, Trondheim)
 Jan Rauch (University of Economics Prague)
 Djamel Zighed (University of Lyon 2)

ECML Program Committee

- A. Aamodt (Norway)
H. Blockeel (Belgium)
H. Bostrom (Sweden)
I. Bratko (Slovenia)
P. Brazdil (Portugal)
W. Burgard (Germany)
J. Cussens (UK)
W. Daelemans (Belgium)
L. De Raedt (Germany)
M. Dorigo (Belgium)
S. Dzeroski (Slovenia)
F. Esposito (Italy)
D. Fisher (USA)
P. Flach (UK)
J. Fürnkranz (Austria)
J.-G. Ganascia (France)
Y. Kodradoff (France)
S. Kramer (Germany)
N. Lavrac (Slovenia)
R. Lopez de Mantaras (Spain)
D. Malerba (Italy)
B. Manderick (Belgium)
S. Matwin (Canada)
K. Morik (Germany)
H. Motoda (Japan)
V. Moustakis (Greece)
S. Muggleton (UK)
C. Nedelec (France)
E. Plaza (Spain)
G. Paliouras (Greece)
C. Rouveirol (France)
L. Saitta (Italy)
M. Sebag (France)
A. Siebes (The Netherlands)
M. van Someren (The Netherlands)
D. Sleeman (UK)
P. Stone (USA)
P. Turney (Canada)
P. Vitanyi (The Netherlands)
G. Widmer (Austria)
R. Wirth (Germany)
S. Wrobel (Germany)

PKDD Program Committee

- E. Baralis (Italy)
P. Berka (Czech Republic)
J.-F. Boulicaut (France)
H. Briand (France)
L. Carbonara (UK)
L. De Raedt (Germany)
L. Dehaspe (Belgium)
S. Dzeroski (Slovenia)
A. Feelders (The Netherlands)
P. Flach (UK)
A. Flexer (Austria)
E. Frank (New Zealand)
A. Freitas (Brasil)
J. Fürnkranz (Austria)
P. Gallinari (France)
M.-S. Hacid (France)
H.J. Hamilton (Canada)
J. Han (Canada)
D.J. Hand (UK)
J. Keller (Germany)
J.-U. Kietz (Switzerland)
J. Kok (The Netherlands)
R. King (UK)
M. Klemettinen (Finland)
W. Klösgen (Germany)
Y. Kodratoff (France)
J. Koronacki (Poland)
S. Kramer (Germany)
G. Lausen (Germany)
H. Mannila (Finland)
S. Matwin (Canada)
H. Motoda (Japan)
W. Pedrycz (Canada)
R. Rakotomalala (France)
Z.W. Ras (USA)
J. Rauch (Czech Republic)
G. Ritschard (Switzerland)
C. Rouveirol (France)
S. Schulze-Kremer (Germany)
M. Sebag (France)
A. Siebes (The Netherlands)
A. Skowron (Poland)
M. van Someren (The Netherlands)
M. Spiliopoulou (Germany)
N. Spyratos (France)
E. Suzuki (Japan)
L. Todorovski (Slovenia)
S. Tsumoto (Japan)
A. Unwin (Germany)
J. Van den Bussche (Belgium)
G. Venturini (France)
L. Wehenkel (Belgium)
D. Wettschereck (Germany)
G. Widmer (Austria)
W. Winiwarter (Austria)
R. Wirth (Germany)
S. Wrobel (Germany)
D.A. Zighed (France)

Referees

F. Abbattista	F. Guillet	J-M. Petit
E. Alphonse	S. Haustein	B. Pfahringer
I. Androutsopoulos	S. Hoche	O. Pivert
J. Azé	T. Hrycej	K. Pohle
B. Bakker	M. Imhoff	T. Poibeau
M. Bernadet	N. Jacobs	S. Rüping
A. Bianchetti-Jacques	A. Jakulin	J. Ramon
G. Bisson	B. Jeudy	O. Ritthof
K. Bo	A. Jorge	M. Robnik-Sikouja
O. Bousquet	P. Juvan	T. Scheffer
O. Brevik	N. Kabachi	A. Seewald
P. Brockhausen	D. Kalles	C. Soares
J. Cassens	V. Karkaletsis	F. Sormo
G. Castellano	K. Kersting	D. Suc
L-P. Castillo	F. Kirchner	R. Sutton
J-H. Chauchat	M. Kirsten	I. Tellier
R. Cicchetti	R. Kosala	S. ten Hagen
J. Costa	W. Kosters	A. Termier
A. Cournuejols	M-A. Krogel	G. Theocharous
E. de Jong	P. Kuntz	N. Theocharous
M. de Jong	A. Lüthje	M. Tommasi
J. Demsar	N. Labroche	L. Torgo
M. deRougemont	L. Lakhal	F. Torre
K. Driessens	C. Lanquillon	F. Toumani
C. Drummond	A. Lanza	Y. Tzitzikas
N. Fanizzi	D. Laurent	V. Ventos
S. Ferrili	F. Lisi	M. Vetah
R. Fried	P. Marcel	D. Vladusic
J. Gama	D. McAllester	M. Wiering
W. Gersten	N. Meuleau	K. Winkler
P. Geurts	R. Munos	J. Wyatt
A. Giacometti	C. Olaru	T. Yoshida
B. Goethals	A. Oliver	R. Zucker
C. Green	J. Pei	B. Zupan
M. Grobelnik	G. Petasis	

Tutorials

Development and Applications of Ontologies

Alexander Maedche

Bayesian Intelligent Decision Support

Kevin Korb

Support Vector Machines

Thorsten Joachims

Co-training and Learning from Labeled Data and Unlabeled Data

Francois Denis and Rémi Gilleron

KDD for Personalization

Bamshad Mobasher, Bettina Berendt, and Myra Spiliopoulou

Web Mining for E-Commerce: Concepts and Case Studies

Jaideep Srivastava and Robert Cooley

Text Mining: What If Your Data Is Made of Words?

Dunja Mladenic and Marko Grobelnik

Visual Data Mining and Exploration of Large Databases

Daniel A. Keim and Mihael Ankerst

Workshops

Semantic Web Mining

Gerd Stumme, Andreas Hotho and Bettina Berendt

Machine Learning as Experimental Philosophy of Science

Kevin Korb and Hilan Bensusan

Visual Data Mining

Simeon J. Simoff, Monique Noirhomme-Fraiture and Michael H. Böhlen

Integrating Aspects of Data Mining, Decision Support, and Meta-Learning

Christophe Giraud-Carrier, Nada Lavrac, and Stephen Moyle

Active Learning, Database Sampling, Experiment Design: Views on Instance Selection

Tobias Scheffer and Stefan Wrobel

Multi-Relational Data Mining

Arno Knobbe and Daniël van der Wallen

The Discovery Challenge on Thrombosis Data

Petr Berka, Shusaku Tsumoto, Katsuhiko Takabayashi and Shishir Gupta

The Predictive Toxicology Challenge for 2000-2001

Christoph Helma, Ross D. King, Stefan Kramer, and Ashwin Srinivasan

Database Support for KDD

Gunter Saake, Daniel Keim, Kai-Uwe Sattler, and Alexander Hinneburg

Data Mining for Marketing Applications

Wendy Gersten and Koen Vanhoof

Ubiquitous Data Mining: Technology for Mobile and Distributed KDD

Hillol Kargupta, Krishnamoorthy Sivakumar, and Ruediger Wirth

Contents

Regular Papers

Self Similar Layered Hidden Markov Models	1
<i>Jafar Adibi and Wei-Min Shen</i>	
Automatic Text Summarization Using Unsupervised and Semi-supervised Learning.....	16
<i>Massih-Reza Amini and Patrick Gallinari</i>	
Detecting Temporal Changes in Event Sequences: An Application to Demographic Data.....	29
<i>Hendrik Blockeel, Johannes Fürnkranz, Alexia Prskawetz, and Francesco C. Billari</i>	
Knowledge Discovery in Multi-label Phenotype Data.....	42
<i>Amanda Clare and Ross D. King</i>	
Computing Association Rules Using Partial Totals.....	54
<i>Frans Coenen, Graham Goulbourne, and Paul Leng</i>	
Gaphyl: A Genetic Algorithms Approach to Cladistics	67
<i>Clare Bates Congdon</i>	
Parametric Approximation Algorithms for High-Dimensional Euclidean Similarity.....	79
<i>Ömer Eğecioğlu</i>	
Data Structures for Minimization of Total Within-Group Distance for Spatio-Temporal Clustering.....	91
<i>Vladimir Estivill-Castro and Michael E. Houle</i>	
Non-crisp Clustering by Fast, Convergent, and Robust Algorithms.....	103
<i>Vladimir Estivill-Castro and Jianhua Yang</i>	
Pattern Extraction for Time Series Classification.....	115
<i>Pierre Geurts</i>	
Specifying Mining Algorithms with Iterative User-Defined Aggregates: A Case Study	128
<i>Fosca Giannotti, Giuseppe Manco, and Franco Turini</i>	
Interesting Fuzzy Association Rules in Quantitative Databases.....	140
<i>Jeannette M. de Graaf, Walter A. Kosters, and Jeroen J.W. Witteman</i>	

Interestingness Measures for Fuzzy Association Rules	152
<i>Attila Gyenesei and Jukka Teuhola</i>	
A Data Set Oriented Approach for Clustering Algorithm Selection	165
<i>Maria Halkidi and Michalis Vazirgiannis</i>	
Fusion of Meta-knowledge and Meta-data for Case-Based Model Selection.	180
<i>Melanie Hilario and Alexandros Kalousis</i>	
Discovery of Temporal Patterns. Learning Rules about the Qualitative Behaviour of Time Series	192
<i>Frank Höppner</i>	
Temporal Rule Discovery for Time-Series Satellite Images and Integration with RDB	204
<i>Rie Honda and Osamu Konishi</i>	
Using Grammatical Inference to Automate Information Extraction from the Web.....	216
<i>Theodore W. Hong and Keith L. Clark</i>	
Biological Sequence Data Mining	228
<i>Yuh-Jyh Hu</i>	
Implication-Based Fuzzy Association Rules	241
<i>Eyke Hüllermeier</i>	
A General Measure of Rule Interestingness.....	253
<i>Szymon Jaroszewicz and Dan A. Simovici</i>	
Error Correcting Codes with Optimized Kullback-Leibler Distances for Text Categorization.....	266
<i>Jörg Kindermann, Gerhard Paass, and Edda Leopold</i>	
Propositionalisation and Aggregates	277
<i>Arno J. Knobbe, Marc de Haas, and Arno Siebes</i>	
Algorithms for the Construction of Concept Lattices and Their Diagram Graphs	289
<i>Sergei Kuznetsov and Sergei Obiedkov</i>	
Data Reduction Using Multiple Models Integration	301
<i>Aleksandar O. Lazarevic and Zoran A. Obradovic</i>	
Discovering Fuzzy Classification Rules with Genetic Programming and Co-evolution.....	314
<i>Roberto R.F. Mendes, Fabricio de B. Voznika, Alex A. Freitas, and Julio C. Nievola</i>	

Sentence Filtering for Information Extraction in Genomics, a Classification Problem.....	326
<i>Claire Nédellec, Mohamed Ould Abdel Vetah, and Philippe Bessières</i>	
Text Categorization and Semantic Browsing with Self-Organizing Maps on Non-euclidean Spaces.....	338
<i>Jörg Ontrup and Helge Ritter</i>	
A Study on the Hierarchical Data Clustering Algorithm Based on Gravity Theory.....	350
<i>Yen-Jen Oyang, Chien-Yu Chen, and Tsui-Wei Yang</i>	
Internet Document Filtering Using Fourier Domain Scoring.....	362
<i>Laurence A.F. Park, Marimuthu Palaniswami, and Ramamohanarao Kotagiri</i>	
Distinguishing Natural Language Processes on the Basis of fMRI-Measured Brain Activation.....	374
<i>Francisco Pereira, Marcel Just, and Tom Mitchell</i>	
Automatic Construction and Refinement of a Class Hierarchy over Multi-valued Data.....	386
<i>Nathalie Pernelle, Marie-Christine Rousset, and Véronique Ventos</i>	
Comparison of Three Objective Functions for Conceptual Clustering.....	399
<i>Céline Robardet and Fabien Feschet</i>	
Identification of ECG Arrhythmias Using Phase Space Reconstruction...	411
<i>Felice M. Roberts, Richard J. Povinelli, and Kristina M. Ropella</i>	
Finding Association Rules That Trade Support Optimally against Confidence.....	424
<i>Tobias Scheffer</i>	
Bloomy Decision Tree for Multi-objective Classification.....	436
<i>Einoshin Suzuki, Masafumi Gotoh, and Yuta Choki</i>	
Discovery of Temporal Knowledge in Medical Time-Series Databases Using Moving Average, Multiscale Matching, and Rule Induction.....	448
<i>Shusaku Tsumoto</i>	
Mining Positive and Negative Knowledge in Clinical Databases Based on Rough Set Model.....	460
<i>Shusaku Tsumoto</i>	
The TwoKey Plot for Multiple Association Rules Control.....	472
<i>Antony Unwin, Heike Hofmann, and Klaus Bernt</i>	

XVIII Table of Contents

Lightweight Collaborative Filtering Method for Binary-Encoded Data 484
Sholom M. Weiss and Nitin Indurkha

Invited Papers

Support Vectors for Reinforcement Learning 492
Thomas G. Dietterich and Xin Wang

Combining Discrete Algorithmic and Probabilistic Approaches
in Data Mining 493
Heikki Mannila

Statistification or Mystification? The Need for Statistical Thought
in Visual Data Mining 494
Antony Unwin

The Musical Expression Project: A Challenge for Machine Learning and
Knowledge Discovery 495
Gerhard Widmer

Scalability, Search, and Sampling: From Smart Algorithms
to Active Discovery 507
Stefan Wrobel

Author Index 509