

Lecture Notes in Artificial Intelligence 2430

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

Tapio Elomaa Heikki Mannila
Hannu Toivonen (Eds.)

Machine Learning: ECML 2002

13th European Conference on Machine Learning
Helsinki, Finland, August 19-23, 2002
Proceedings



Springer

Series Editors

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

Volume Editors

Tapio Elomaa
Heikki Mannila
Hannu Toivonen
University of Helsinki, Department of Computer Science
P.O. Box 26, 00014 Helsinki, Finland
E-mail: {elomaa, heikki.mannila, hannu.toivonen}@cs.helsinki.fi

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Machine learning : proceedings / ECML 2002, 13th European Conference on
Machine Learning, Helsinki, Finland, August 19 - 23, 2002. Tapio Elomaa ...
(ed.) - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ;
Milan ; Paris ; Tokyo : Springer, 2002
(Lecture notes in computer science ; Vol. 2430 : Lecture notes in
artificial intelligence)
ISBN 3-540-44036-4

CR Subject Classification (1998): I.2, F.2.2, F.4.1

ISSN 0302-9743

ISBN 3-540-44036-4 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 2002
Printed in Germany

Typesetting: Camera-ready by author, data conversion by DA-TeX Gerd Blumenstein
Printed on acid-free paper SPIN: 10873900 06/3142 5 4 3 2 1 0

Preface

We are pleased to present the proceedings of the *13th European Conference on Machine Learning* (LNAI 2430) and the *6th European Conference on Principles and Practice of Knowledge Discovery in Databases* (LNAI 2431). These two conferences were colocated in Helsinki, Finland during August 19–23, 2002. ECML and PKDD were held together for the second year in a row, following the success of the colocation in Freiburg in 2001. Machine learning and knowledge discovery are two highly related fields and ECML/PKDD is a unique forum to foster their collaboration.

The benefit of colocation to both the machine learning and data mining communities is most clearly displayed in the common workshop, tutorial, and invited speaker program. Altogether six workshops and six tutorials were organized on Monday and Tuesday. As invited speakers we had the pleasure to have Erkki Oja (Helsinki Univ. of Technology), Dan Roth (Univ. of Illinois, Urbana-Champaign), Bernhard Schölkopf (Max Planck Inst. for Biological Cybernetics, Tübingen), and Padhraic Smyth (Univ. of California, Irvine).

The main events ran from Tuesday until Friday, comprising 41 ECML technical papers and 39 PKDD papers. In total, 218 manuscripts were submitted to these two conferences: 95 to ECML, 70 to PKDD, and 53 as joint submissions. All papers were assigned at least three reviewers from our international program committees. Out of the 80 accepted papers 31 were first accepted conditionally; the revised manuscripts were accepted only after the conditions set by the reviewers had been met.

Our special thanks go to the tutorial chairs Johannes Fürnkranz and Myra Spiliopoulou and the workshop chairs Hendrik Blockeel and Jean-François Boulicaut for putting together an exiting combined tutorial and workshop program. Also the challenge chair Petr Berka deserves our sincerest gratitude. All the members of both program committees are thanked for devoting their expertise to the continued success of ECML and PKDD. The organizing committee chaired by Helena Ahonen-Myka worked hard to make the conferences possible. A special mention has to be given to Oskari Heinonen for designing and maintaining the web pages and Ilkka Koskenniemi for maintaining CyberChair, which was developed by Richard van de Stadt. We thank Alfred Hofmann of Springer-Verlag for cooperation in publishing these proceedings. We gratefully acknowledge the financial support of the Academy of Finland and KDNet.

We thank all the authors for contributing to what in our mind is a most interesting technical program for ECML and PKDD. We trust that the week in late August was most enjoyable for all members of both research communities.

June 2002

Tapio Elomaa
Heikki Mannila
Hannu Toivonen

ECML/PKDD-2002 Organization

Executive Committee

Program Chairs: Tapio Elomaa (Univ. of Helsinki)
Heikki Mannila (Helsinki Inst. for Information
Technology and Helsinki Univ. of Technology)
Hannu Toivonen (Nokia Research Center and
Univ. of Helsinki)

Tutorial Chairs: Johannes Fürnkranz (Austrian Research Inst.
for Artificial Intelligence)
Myra Spiliopoulou
(Leipzig Graduate School of Management)

Workshop Chairs: Hendrik Blockeel
(Katholieke Universiteit Leuven)
Jean-François Boulicaut (INSA Lyon)

Challenge Chair: Petr Berka (University of Economics, Prague)

Organizing Chair: Helena Ahonen-Myka (Univ. of Helsinki)

Organizing Committee: Oskari Heinonen, Ilkka Koskenniemi, Greger
Lindén, Pirjo Moen, Matti Nykänen, Anna
Pienimäki, Ari Rantanen, Juho Rousu, Marko
Salmenkivi (Univ. of Helsinki)

ECML Program Committee

H. Blockeel, Belgium	A. Hyvärinen, Finland
I. Bratko, Slovenia	T. Joachims, USA
P. Brazdil, Portugal	Y. Kodratoff, France
H. Boström, Sweden	I. Kononenko, Slovenia
W. Burgard, Germany	S. Kramer, Germany
N. Cristianini, USA	M. Kubat, USA
J. Cussens, UK	N. Lavrač, Slovenia
L. De Raedt, Germany	C. X. Ling, Canada
M. Dorigo, Belgium	R. López de Mántaras, Spain
S. Džeroski, Slovenia	D. Malerba, Italy
F. Esposito, Italy	S. Matwin, Canada
P. Flach, UK	R. Meir, Israel
J. Fürnkranz, Austria	J. del R. Millán, Switzerland
J. Gama, Portugal	K. Morik, Germany
J.-G. Ganascia, France	H. Motoda, Japan
T. Hofmann, USA	R. Nock, France
L. Holmström, Finland	E. Plaza, Spain

G. Palioras, Greece	H. Tirri, Finland
J. Rousu, Finland	P. Turney, Canada
L. Saitta, Italy	R. Vilalta, USA
T. Scheffer, Germany	P. Vitányi, The Netherlands
M. Sebag, France	S. Weiss, USA
J. Shawe-Taylor, UK	G. Widmer, Austria
A. Siebes, The Netherlands	R. Wirth, Germany
D. Sleeman, UK	S. Wrobel, Germany
M. van Someren, The Netherlands	Y. Yang, USA
P. Stone, USA	

PKDD Program Committee

H. Ahonen-Myka, Finland	S. Morishita, Japan
E. Baralis, Italy	H. Motoda, Japan
J.-F. Boulicaut, France	G. Nakhaeizadeh, Germany
N. Cercone, Canada	Z.W. Raś, USA
B. Crémilleux, France	J. Rauch, Czech Republic
L. De Raedt, Germany	G. Ritschard, Switzerland
L. Dehaspe, Belgium	M. Sebag, France
S. Džeroski, Slovenia	F. Sebastiani, Italy
M. Ester, Canada	M. Sebban, France
R. Feldman, Israel	B. Seeger, Germany
P. Flach, UK	A. Siebes, The Netherlands
E. Frank, New Zealand	A. Skowron, Poland
A. Freitas, Brazil	M. van Someren, The Netherlands
J. Fürnkranz, Austria	M. Spiliopoulou, Germany
H.J. Hamilton, Canada	N. Spyros, France
J. Han, Canada	E. Suzuki, Japan
R. Hilderman, Canada	A.-H. Tan, Singapore
S.J. Hong, USA	S. Tsumoto, Japan
S. Kaski, Finland	A. Unwin, Germany
D. Keim, USA	J. Wang, USA
J.-U. Kietz, Switzerland	K. Wang, Canada
R. King, UK	L. Wehenkel, Belgium
M. Klemettinen, Finland	D. Wettschereck, Germany
W. Klösgen, Germany	G. Widmer, Austria
Y. Kodratoff, France	R. Wirth, Germany
J.N. Kok, The Netherlands	S. Wrobel, Germany
S. Kramer, Germany	M. Zaki, USA
S. Matwin, Canada	

VIII Organization

Additional Reviewers

N. Abe	S. Haustein	L. Peña
F. Aiolfi	J. He	Y. Peng
Y. Altun	K.G. Herbert	J. Petrak
S. de Amo	J. Himberg	V. Phan Luong
A. Appice	J. Hipp	K. Rajaraman
E. Armengol	S. Hoche	T. Reinartz
T.G. Ault	J. Hosking	I. Renz
J. Azé	E. Hüllermeier	C. Rigotti
M.T. Basile	P. Juvan	F. Rioult
A. Bonarini	M. Kääriäinen	M. Robnik-Šikonja
R. Bouckaert	D. Kalles	M. Roche
P. Brockhausen	V. Karkaletsis	B. Rosenfeld
M. Brodie	A. Karwath	S. Rüping
W. Buntine	K. Kersting	M. Salmenkivi
J. Carbonell	J. Kindermann	A.K. Seewald
M. Ceci	R. Klinkenberg	H. Shan
S. Chikkanna-Naik	P. Koistinen	J. Sinkkonen
S. Chiusano	C. Köpf	J. Struyf
R. Cicchetti	R. Kosala	R. Taouil
A. Clare	W. Kosters	J. Taylor
M. Degemmis	M.-A. Krogel	L. Todorovski
J. Demsar	M. Kukar	T. Urbancic
F. De Rosis	L. Lakhal	K. Vasko
N. Di Mauro	G. Lebanon	H. Wang
G. Dorffner	S.D. Lee	Y. Wang
G. Dounias	F. Li	M. Wiering
N. Durand	J.T. Lindgren	S. Wu
P. Erästö	J. Liu	M.M. Yin
T. Erjavec	Y. Liu	F. Zambetta
J. Farrand	M.-C. Ludl	B. Ženko
S. Ferilli	S. Mannor	J. Zhang
P. Floréen	R. Meo	S. Zhang
J. Franke	N. Meuleau	T. Zhang
T. Gaertner	H. Mogg-Schneider	M. Zlochin
P. Gallinari	R. Natarajan	B. Zupan
P. Garza	S. Nijssen	
A. Giacometti	G. Paaß	

Tutorials

Text Mining and Internet Content Filtering
José María Gómez Hidalgo

Formal Concept Analysis
Gerd Stumme

Web Usage Mining for E-business Applications
Myra Spiliopoulou, Bamshad Mobasher, and Bettina Berendt

Inductive Databases and Constraint-Based Mining
Jean-François Boulicaut and Luc De Raedt

An Introduction to Quality Assessment in Data Mining
Michalis Vazirgiannis and M. Halkidi

Privacy, Security, and Data Mining
Chris Clifton

Workshops

Integration and Collaboration Aspects of Data Mining, Decision Support and Meta-Learning
Marko Bohanec, Dunja Mladenić, and Nada Lavrač

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

Semantic Web Mining
Bettina Berendt, Andreas Hotho, and Gerd Stumme

Mining Official Data
Paula Brito and Donato Malerba

Knowledge Discovery in Inductive Databases
Mika Klemettinen, Rosa Meo, Fosca Giannotti, and Luc De Raedt

Discovery Challenge Workshop
Petr Berka, Jan Rauch, and Shusaku Tsumoto

Table of Contents

Contributed Papers

Convergent Gradient Ascent in General-Sum Games	1
<i>Bikramjit Banerjee and Jing Peng</i>	
Revising Engineering Models: Combining Computational Discovery with Knowledge	10
<i>Stephen D. Bay, Daniel G. Shapiro, and Pat Langley</i>	
Variational Extensions to EM and Multinomial PCA	23
<i>Wray Buntine</i>	
Learning and Inference for Clause Identification	35
<i>Xavier Carreras, Lluís Márquez, Vasin Punyakanok, and Dan Roth</i>	
An Empirical Study of Encoding Schemes and Search Strategies in Discovering Causal Networks	48
<i>Honghua Dai, Gang Li, and Yiqing Tu</i>	
Variance Optimized Bagging	60
<i>Philip Derbeko, Ran El-Yaniv, and Ron Meir</i>	
How to Make AdaBoost.M1 Work for Weak Base Classifiers by Changing Only One Line of the Code	72
<i>Günther Eibl and Karl Peter Pfeiffer</i>	
Sparse Online Greedy Support Vector Regression	84
<i>Yaakov Engel, Shie Mannor, and Ron Meir</i>	
Pairwise Classification as an Ensemble Technique	97
<i>Johannes Fürnkranz</i>	
RIONA: A Classifier Combining Rule Induction and k-NN Method with Automated Selection of Optimal Neighbourhood	111
<i>Grzegorz Góra and Arkadiusz Wojna</i>	
Using Hard Classifiers to Estimate Conditional Class Probabilities	124
<i>Ole Martin Halck</i>	
Evidence that Incremental Delta-Bar-Delta Is an Attribute-Efficient Linear Learner	135
<i>Harlan D. Harris</i>	
Scaling Boosting by Margin-Based Inclusion of Features and Relations	148
<i>Susanne Hoche and Stefan Wrobel</i>	
Multiclass Alternating Decision Trees	161
<i>Geoffrey Holmes, Bernhard Pfahringer, Richard Kirkby, Eibe Frank, and Mark Hall</i>	

XII Table of Contents

Possibilistic Induction in Decision-Tree Learning	173
<i>Eyke Hüllermeier</i>	
Improved Smoothing for Probabilistic Suffix Trees Seen as Variable Order Markov Chains	185
<i>Christopher Kermorvant and Pierre Dupont</i>	
Collaborative Learning of Term-Based Concepts for Automatic Query Expansion	195
<i>Stefan Klink, Armin Hust, Markus Junker, and Andreas Dengel</i>	
Learning to Play a Highly Complex Game from Human Expert Games	207
<i>Tony Kråkenes and Ole Martin Halck</i>	
Reliable Classifications with Machine Learning	219
<i>Matjaž Kukar and Igor Kononenko</i>	
Robustness Analyses of Instance-Based Collaborative Recommendation ...	232
<i>Nicholas Kushmerick</i>	
<i>i</i> Boost: Boosting Using an instance-Based Exponential Weighting Scheme	245
<i>Stephen Kwek and Chau Nguyen</i>	
Towards a Simple Clustering Criterion Based on Minimum Length Encoding	258
<i>Marcus-Christopher Ludl and Gerhard Widmer</i>	
Class Probability Estimation and Cost-Sensitive Classification Decisions ..	270
<i>Dragos D. Margineantu</i>	
On-Line Support Vector Machine Regression	282
<i>Mario Martin</i>	
Q-Cut – Dynamic Discovery of Sub-goals in Reinforcement Learning	295
<i>Ishai Menache, Shie Mannor, and Nahum Shimkin</i>	
A Multistategy Approach to the Classification of Phases in Business Cycles	307
<i>Katharina Morik and Stefan Rüping</i>	
A Robust Boosting Algorithm	319
<i>Richard Nock and Patrice Lefaucheur</i>	
Case Exchange Strategies in Multiagent Learning	331
<i>Santiago Ontañón and Enric Plaza</i>	
Inductive Confidence Machines for Regression	345
<i>Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman</i>	
Macro-Operators in Multirelational Learning: A Search-Space Reduction Technique	357
<i>Lourdes Peña Castillo and Stefan Wrobel</i>	

Propagation of Q-values in Tabular TD(λ).....	369
<i>Philippe Preux</i>	
Transductive Confidence Machines for Pattern Recognition	381
<i>Kostas Proedrou, Ilia Nouretdinov, Volodya Vovk, and Alex Gammerman</i>	
Characterizing Markov Decision Processes	391
<i>Bohdana Ratitch and Doina Precup</i>	
Phase Transitions and Stochastic Local Search in k-Term DNF Learning	405
<i>Ulrich Rückert, Stefan Kramer, and Luc De Raedt</i>	
Discriminative Clustering: Optimal Contingency Tables by Learning Metrics	418
<i>Janne Sinkkonen, Samuel Kaski, and Janne Nikkilä</i>	
Boosting Density Function Estimators	431
<i>Franck Thollard, Marc Sebban, and Philippe Ezequel</i>	
Ranking with Predictive Clustering Trees	444
<i>Ljupčo Todorovski, Hendrik Blockeel, and Sašo Džeroski</i>	
Support Vector Machines for Polycategorical Classification	456
<i>Ioannis Tsacharidis and Thomas Hofmann</i>	
Learning Classification with Both Labeled and Unlabeled Data	468
<i>Jean-Noël Vittaute, Massih-Reza Amini, and Patrick Gallinari</i>	
An Information Geometric Perspective on Active Learning	480
<i>Chen-Hsiang Yeang</i>	
Stacking with an Extended Set of Meta-level Attributes and MLR	493
<i>Bernard Ženko and Sašo Džeroski</i>	

Invited Papers

Finding Hidden Factors Using Independent Component Analysis	505
<i>Erkki Oja</i>	
Reasoning with Classifiers	506
<i>Dan Roth</i>	
A Kernel Approach for Learning from almost Orthogonal Patterns	511
<i>Bernhard Schölkopf, Jason Weston, Eleazar Eskin, Christina Leslie, and William Stafford Noble</i>	
Learning with Mixture Models: Concepts and Applications	529
<i>Padhraic Smyth</i>	
Author Index	531