Lecture Notes in Artificial Intelligence 1810

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
Singapore
Tokyo

Ramon López de Mántaras Enric Plaza (Eds.)

Machine Learning: ECML 2000

11th European Conference on Machine Learning
Barcelona, Catalonia, Spain, May 31 – June 2, 2000
Proceedings

Springer

Series Editors

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

Volume Editors

Ramon López de Mántaras
Enric Plaza
Institut d'Investigació en Intelligència Artificial, IIIA
Spanish Council for Scientific Research, CSIC
Campus, U.A.B.
08193 Bellaterra, Catalonia, Spain
E-mail:{mantaras/enric}@iiia.csic.es

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Machine learning : proceedings / ECML 2000, 11th European Conference on Machine Learning, Barcelona, Catalonia, Spain, May 31 - June 2, 2000. Ramón López de Mántaras ; Enric Plaza (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2000
(Lecture notes in computer science ; Vol. 1810 : Lecture notes in artificial intelligence)
ISBN 3-540-67602-3

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

ISSN 0302-9743
ISBN 3-540-67602-3 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 is a company in the BertelsmannSpringer publishing group.
© Springer-Verlag Berlin Heidelberg 2000
Printed in Germany

Typesetting: Camera-ready by author, data conversion by DA-TeX Gerd Blumenstein
Printed on acid-free paper SPIN 10720254 06/3142 5 4 3 2 1 0

Foreword

The biennial European Conference on Machine Learning (ECML) series is intended to provide an international forum for the discussion of the latest high quality research results in machine learning and is the major European scientific event in the field. The eleventh conference (ECML 2000) held in Barcelona, Catalonia, Spain from May 31 to June 2, 2000, has continued this tradition by attracting high quality papers from around the world.

Scientists from 21 countries submitted 100 papers to ECML 2000, from which 20 were selected for long oral presentations and 23 for short oral presentations. This selection was based on the recommendations of at least two reviewers for each submitted paper. It is worth noticing that the number of papers reporting applications of machine learning has increased in comparison to past ECML conferences. We believe this fact shows the growing maturity of the field.

This volume contains the 43 accepted papers as well as the invited talks by Katharina Morik from the University of Dortmund and Pedro Domingos from the University of Washington at Seattle. In addition, three workshops were jointly organized by ECML 2000 and the European Network of Excellence ML-net: "Dealing with Structured Data in Machine Learning and Statistics Web-stites", "Machine Learning in the New Information Age" , and "Meta-Learning: Building Automatic Advice Strategies for Model Selection and Method Combination". Finally, a special workshop on "Learning Agents" was jointly organized by ECML 2000 and the co-located International Conference on Autonomous Agents. Information on the workshops can be found on the ECML 2000 web page: `http://www.iiia.csic.es/ecml2000/`.

We gratefully acknowledge the work of the invited speakers and the authors of the submitted papers that made this conference possible. We also thank the program committee and the additional reviewers for their effort during the paper selection process. Our gratitude also goes to the ECML 2000 sponsors: MLnet, CSIC, CICYT, and ACIA, as well as to Gemma Sales for her assistance in the organization of the conference.

March 2000

Ramon López de Mántaras
Enric Plaza

Organization

The 11th European Conference on Machine Learning took place on May 30th to June 2nd 2000 in Barcelona, Catalonia (Spain).

Program Chairs

Ramon López de Mántaras (mantaras@iiia.csic.es), Enric Plaza (enric@iiia.csic.es), Institut d'Investigació en Intel.ligència Artificial (IIIA-CSIC), Bellaterra, Catalonia (Spain)

Conference Support

ECML 2000 was organized by the Institut d'Investigació en Intel.ligència Artificial (IIIA), Consejo Superior de Investigaciones Científicas (CSIC), Bellaterra, Catalonia (Spain). Institutions that provided support for ECML 2000 are ACIA (Catalan Association for Artificial Intelligence), CICYT (Spanish Science and Technology Council), CSIC (Spanish Council for Scientific Research), and ML-net (European Network of Excellence on Machine Learning IST-29288)

Program Committee

A. Aamodt (Norway)
D. Aha (USA)
K. D. Althoff (Germany)
F. Bergadano (Italy)
D. Borrajo (Spain)
I. Bratko (Slovenia)
P. Brazdil (Portugal)
E. Costa (Portugal)
K. DeJong (USA)
L. De Raedt (Germany)
T. G. Dietterich (USA)
P. Domingos (USA)
M. Dorigo (Belgium)
S. Dzeroski (Slovenia)
F. Esposito (Italy)
D. H. Fisher (USA)
P. Flach (UK)
J. Fuerkranz (Austria)
J. G. Ganascia (France)
O. Gascuel (France)
M. Goldszmidth (USA)
Y. Kodratoff (France)
I. Kononenko (Slovenia)
N. Lavrac (Slovenia)
R. López de Mántaras (Spain)
S. Matwin (Canada)
K. Morik (Germany)
V. Moustakis (Greece)
S. Muggleton (UK)
G. Nakhaeizadeh (Germany)
C. Nedellec (France)
E. Plaza (Spain)
C. Rouveirol (France)
S. Russell
L. Saitta (Italy)
M. van Someren (Netherlands)
D. Sleeman (UK)
S. Thrun (USA)
C. Torras (Spain)
S. Weiss (Germany)
G. Widmer (Austria)
S. Wrobel (Germany)

Additional Reviewers

J. Aguilar Martin (France)
K. Barker (Canada)
K. Bartlmae (Germany)
M. Birattari (Italy)
M. Borth (Germany)
O. A. Brevik (Norway)
P. Brockhausen (Germany)
T. Brychcy (Germany)
A. Cornuejols (France)
D. Faure (France)
R. Gavaldà (Spain)
M. Grobelnik (Slovenia)
D. Gunetti (Italy)
T. Hrycej (Germany)
P. Isasi (Spain)
A. Jorge (Portugal)
T. Joachims (Germany)
H. Kauderer (Germany)
R. Klinkenberg (Germany)
M. Kukar (Slovenia)
C. Lanquillon (Germany)
M. Liquiere (France)
P. Machado (Portugal)
D. Malerba (Italy)
J. M. Molina (Spain)
F. B. Pereira (Portugal)
F. A. Rauscher (Germany)
I. Renz (Germany)
M. Robnik-Sikonja (Slovenia)
G. Ruffo (Italy)
V. Ruiz de Angulo (Spain)
M. Saerens (Belgium)
T. Scheffer (Germany)
R. Schneider (Germany)
G. Semeraro (Italy)
L. Todorovski (Slovenia)
L. Torgo (Portugal)
F. Torre (France)
C. Ungerer (Germany)
M. Wiering (Italy)
D. Zupanic (Slovenia)

Table of Contents

Invited Papers

Beyond Occam's Razor: Process-Oriented Evaluation 3
Pedro Domingos

The Representation Race – Preprocessing for Handling Time Phenomena 4
Katharina Morik

Contributed Papers

Short-Term Profiling for a Case-Based Reasoning Recommendation System 23
Esma Aimeur and Mathieu Vézeau

K-SVCR. A Multi-class Support Vector Machine 31
Cecilio Angulo and Andreu Català

Learning Trading Rules with Inductive Logic Programming 39
Liviu Badea

Improving Knowledge Discovery Using Domain Knowledge in Unsupervised Learning 47
Javier Béjar

Exploiting Classifier Combination for Early Melanoma Diagnosis Support .. 55
E. Blanzieri, C. Eccher, S. Forti and A. Sboner

A Comparison of Ranking Methods for Classification Algorithm Selection .. 63
Pavel B. Brazdil and Carlos Soares

Hidden Markov Models with Patterns and Their Application to Integrated Circuit Testing 75
Laurent Bréhélin, Olivier Gascuel and Gilles Caraux

Comparing Complete and Partial Classification for Identifying Latently Dissatisfied Customers 88
Tom Brijs, Gilbert Swinnen, Koen Vanhoof and Geert Wets

Wrapper Generation via Grammar Induction 96
Boris Chidlovskii, Jon Ragetli and Maarten de Rijke

Diversity versus Quality in Classification Ensembles Based on Feature Selection 109
Pádraig Cunningham and John Carney

Minimax TD-Learning with Neural Nets in a Markov Game 117
Fredrik A. Dahl and Ole Martin Halck

Boosting Applied to Word Sense Disambiguation 129
Gerard Escudero, Lluís Màrquez and German Rigau

A Multiple Model Cost-Sensitive Approach for Intrusion Detection 142
Wei Fan, Wenke Lee, Salvatore J. Stolfo and Matthew Miller

Value Miner: A Data Mining Environment for the Calculation of the Customer Lifetime Value with Application to the Automotive Industry 154
Katja Gelbrich and Reza Nakhaeizadeh

Investigation and Reduction of Discretization Variance in Decision Tree Induction 162
Pierre Geurts and Louis Wehenkel

Asymmetric Co-evolution for Imperfect-Information Zero-Sum Games 171
Ole Martin Halck and Fredrik A. Dahl

A Machine Learning Approach to Workflow Management 183
Joachim Herbst

The Utilization of Context Signals in the Analysis of ABR Potentials by Application of Neural Networks 195
Andrzej Izworski, Ryszard Tadeusiewicz and Andrzej Paslawski

Complexity Approximation Principle and Rissanen's Approach to Real-Valued Parameters 203
Yuri Kalnishkan

Handling Continuous-Valued Attributes in Decision Tree with Neural Network Modeling 211
DaeEun Kim and Jaeho Lee

Learning Context-Free Grammars with a Simplicity Bias 220
Pat Langley and Sean Stromsten

Partially Supervised Text Classification: Combining Labeled and Unlabeled Documents Using an EM-like Scheme 229
Carsten Lanquillon

Toward an Explanatory Similarity Measure for Nearest-Neighbor Classification 238
Mathieu Latourrette

Relative Unsupervised Discretization for Regression Problems 246
Marcus-Christopher Ludl and Gerhard Widmer

Metric-Based Inductive Learning Using Semantic Height Functions 254
Zdravko Markov and Ivo Marinchev

Error Analysis of Automatic Speech Recognition Using Principal Direction Divisive Partitioning 263
David McKoskey and Daniel Boley

A Study on the Performance of Large Bayes Classifier 271
Dimitris Meretakis, Hongjun Lu and Beat Wüthrich

Dynamic Discretization of Continuous Values from Time Series 280
Llanos Mora López, Inmaculada Fortes Ruiz, Rafael Morales Bueno and Francisco Triguero Ruiz

Using a Symbolic Machine Learning Tool to Refine Lexico-syntactic Patterns 292
Emmanuel Morin and Emmanuelle Martienne

Measuring Performance when Positives Are Rare: Relative Advantage versus Predictive Accuracy – A Biological Case-Study 300
Stephen H. Muggleton, Christopher H. Bryant and Ashwin Srinivasan

Mining TCP/IP Traffic for Network Intrusion Detection by Using a Distributed Genetic Algorithm 313
Filippo Neri

Learning Patterns of Behavior by Observing System Events 323
Marlon Núñez

Dimensionality Reduction through Sub-space Mapping for Nearest Neighbour Algorithms 331
Terry R. Payne and Peter Edwards

Nonparametric Regularization of Decision Trees 344
Tobias Scheffer

An Efficient and Effective Procedure for Updating a Competence Model for Case-Based Reasoners 357
Barry Smyth and Elizabeth McKenna

Layered Learning 369
Peter Stone and Manuela Veloso

Problem Decomposition for Behavioural Cloning 382
Dorian Šuc and Ivan Bratko

Dynamic Feature Selection in Incremental Hierarchical Clustering 392
Luis Talavera

On the Boosting Pruning Problem 404
Christino Tamon and Jie Xiang

An Empirical Study of MetaCost Using Boosting Algorithms 413
Kai Ming Ting

Clustered Partial Linear Regression 426
Luís Torgo and Joaquim Pinto da Costa

Knowledge Discovery from Very Large Databases Using Frequent Concept Lattices 437
Kitsana Waiyamai and Lotfi Lakhal

Some Improvements on Event-Sequence Temporal Region Methods 446
Wei Zhang

Author Index 459