

Lecture Notes in Artificial Intelligence

1263

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

Jan Komorowski Jan Zytkow (Eds.)

Principles of Data Mining and Knowledge Discovery

First European Symposium, PKDD '97
Trondheim, Norway, June 24-27, 1997
Proceedings



Springer

Series Editors

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

Volume Editors

Jan Komorowski

Norwegian University of Science and Technology
Department of Computer and Information Science
O.S. Bragstads plass 2E, N-7034 Trondheim, Norway
E-mail: jan.komorowski@idt.ntnu.no

Jan Zytkow

Wichita State University, Department of Computer Science
Wichita, KS 67260-0083, USA
E-mail: zytkow@wise.cs.twsu.edu

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Principles of data mining and knowledge discovery : first European symposium ; proceedings / PKDD '97, Trondheim, Norway, June 24 - 27, 1997. Jan Komorowski ; Jan Zytkow (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Budapest ; Hong Kong ; London ; Milan ; Paris ; Santa Clara ; Singapore ; Tokyo : Springer, 1997

(Lecture notes in computer science ; Vol. 1263 : Lecture notes in artificial intelligence)
ISBN 3-540-63223-9

CR Subject Classification (1991): I.2, H.3, H.5, G.3, J.1

ISBN 3-540-63223-9 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 1997
Printed in Germany

Typesetting: Camera ready by author
SPIN 10548929 06/3142 - 5 4 3 2 1 0 Printed on acid-free paper

Preface

This volume contains papers that were selected for presentation at the First European Symposium on Principles of Data Mining and Knowledge Discovery from Databases – PKDD’97, held in Trondheim, Norway, June 24-27, 1997. The Norwegian University of Science and Technology (NTNU) hosted the symposium. The Department of Computer and Information Science, the Norwegian Research Council (NFR), and the Norwegian Artificial Intelligence Society (NAIS) sponsored the symposium.

PKDD’97 was the first symposium in an intended series of meetings of the data mining and knowledge discovery from databases (KDD) community in Europe. The goal of the PKDD series is to provide a European-based forum for interaction among all theoreticians and practitioners interested in data mining and knowledge discovery. Fostering an interdisciplinary collaboration is one desired outcome, but the main long-term focus is on theoretical principles for the emerging discipline of KDD, especially those new principles that go beyond each of the contributing areas.

The Programming Committee decided to select the following major areas for PKDD’97:

- Data and knowledge representation for data mining
- Statistics and probability in data mining
- Logic-based perspective on data mining
- Man-machine interaction in data mining
- Artificial intelligence contributions to KDD
- High performance computing for data mining
- From machine learning to KDD
- From automated scientific discovery to KDD
- Quality assessment of data mining results
- Applications of data mining and knowledge discovery

The contributed papers were selected from 50 full draft papers by the following program committee:

Andrzej Skowron (U. Warsaw, Poland), Arno Siebes (CWI, The Netherlands), Attilio Giordana (U. Torino, Italy), Bob Henery (U. Strathclyde, UK), David Hand (Open U. UK), Derek Sleeman (U. Aberdeen, UK), Erik Sandewall (Linköping U., Sweden), Gregory Piatetsky-Shapiro (GTE Lab. USA), Heikki Manilla (U. Helsinki, Finland), Jan Komorowski (NTNU, Norway), Jan Zytkow (Wichita State, USA), Lorenza Saitta (U. Torino, Italy), Marjorie Moulet (LRI, U. Paris XI, France), Mikhail Kiselev (Nat. Research Center of Surgery, Russia), Pieter Adriaans (Syllogic, The Netherlands), Raul Valdes-Perez (CMU, USA), Rudiger Wirth (Daimler-Benz, Germany), Shusaku Tsumoto (Tokyo Medical &

Dental U. Japan), Stefan Wrobel (GTE, Germany), Steve Muggleton (Oxford U. UK), Wei-Min Shen (U. South. Calif. USA), Willi Kloesgen (GMD, Germany), Wojtek Ziarko (U. Regina, Canada), Yves Kodratoff (U. Paris VI, France), Zbigniew Ras (UNC Charlotte, USA), Zdzislaw Pawlak (Warsaw Technical U. Poland).

The papers were divided into three categories: 14 plenary papers, 13 parallel session papers, and 11 poster papers that included spot-light presentations in the plenary sessions. In addition, four tutorials were selected: Rough Sets for Data Mining and Knowledge Discovery, Techniques and Applications of KDD, High Performance Data Mining, and Data Mining in the Telecommunications Industry.

The Knowledge Systems Group of the Department of Computer and Information Science of the Norwegian University of Science and Technology together with Jan Zytkow, Wichita State University, organized the symposium. The Congress Department of NTNU provided the secretariat for the symposium.

We wish to express our thanks to the sponsors of the symposium and to all who submitted papers for presentation and publication in the proceedings. Special thanks are due to Alfred Hofmann of Springer-Verlag for his help and support.

Finally, we would like to thank Staal Vinterbo whose quiet work "behind the scenes" largely contributed to the success of this meeting.

May 1997

J. Komorowski and Jan Zytkow

Table of Contents

Invited Talk

B. A. Foss	
<i>Knowledge Discovery - A Control Theory Perspective</i>	1

Plenary Session 1

W. Kowalczyk, F. Slisser	
<i>Modeling Customer Retention with Rough Data Models</i>	4
C. L. Carter, H. J. Hamilton, N. Cercone	
<i>Share Based Measures for Itemsets</i>	14
R. J. Hilderman, H. J. Hamilton, R. J. Kowalchuk, N. Cercone	
<i>Parallel Knowledge Discovery Using Domain Generalization Graphs</i>	25
J. Stefanowski, K. Słowiński	
<i>Rough Set Theory and Rule Induction Techniques for Discovery of Attribute Dependencies in Medical Information Systems</i>	36
J. Rauch	
<i>Logical Calculi for Knowledge Discovery in Databases</i>	47

Parallel Session 2a

S. Tsumoto	
<i>Extraction of Experts' Decision Process from Clinical Databases Using Rough Set Model</i>	58
M. Bull, G. Kundt, L. Gierl	
<i>Discovering of Health Risks and Case-Based Forecasting of Epidemics in a Health Surveillance System</i>	68
S. Wrobel	
<i>An Algorithm for Multi-relational Discovery of Subgroups</i>	78

Parallel Session 2b

G. Das, D. Gunopulos, H. Mannila	
<i>Finding Similar Time Series</i>	88
D. Merkl	
<i>Exploration of Document Collections with Self-Organizing Maps: A Novel Approach to Similarity Representation</i>	101
R. Feldman, W. Klösgen, Y. Ben-Yehuda, G. Kedar, V. Reznikov	
<i>Pattern Based Browsing in Document Collections</i>	112

Parallel Session 3a

D. Rasmussen, R. R. Yager	
<i>Induction of Fuzzy Characteristic Rules</i>	123

M. V. Kiselev, S. M. Ananyan, S. B. Arseniev	
<i>Regression-Based Classification Methods and Their Comparison with Decision Tree Algorithms</i>	134
J. Stepaniuk	
<i>Attribute Discovery and Rough Sets</i>	145
M. Kryszkiewicz	
<i>Generation of Rules from Incomplete Information Systems</i>	156

Parallel Session 3b

G. Ruhe	
<i>Knowledge Discovery from Software Engineering Data: Rough Set Analysis and Its Interaction with Goal-Oriented Measurement</i>	167
T. Elomaa, J. Rousu	
<i>Efficient Multisplitting on Numerical Data</i>	178
P. D. Scott, A. P. M. Coxon, M. H. Hobbs, R. J. Williams	
<i>SNOUT: An Intelligent Assistant for Exploratory Data Analysis . . .</i>	189

Plenary Session 4

R. Simutis	
<i>Exploratory Analysis of Biochemical Processes Using Hybrid Modeling Methods</i>	200
H. André-Jönsson, D. Z. Badal	
<i>Using Signature Files for Querying Time-Series Data</i>	211
A. Amir, R. Feldman, R. Kashi	
<i>A New and Versatile Method for Association Generation</i>	221
J. C. Bioch, O. van der Meer, R. Potharst	
<i>Bivariate Decision Trees</i>	232
R. Wirth, C. Shearer, U. Grimmer, T. Reinartz, J. Schlösser, C. Breitner, R. Engels, G. Lindner	
<i>Towards Process-Oriented Tool Support for KDD</i>	243

Plenary Session 5

K. Schädler, F. Wysotzki	
<i>A Connectionist Approach to Structural Similarity Determination as a Basis of Clustering, Classification and Feature Detection</i>	254
S. H. Nguyen, A. Skowron	
<i>Searching for Relational Patterns in Data</i>	265
F. Gebhardt	
<i>Finding Spatial Clusters</i>	277
E. Boudaillier, G. Hébrail	
<i>Interactive Interpretation of Hierarchical Clustering</i>	288

Poster Session 6

A. A. Freitas <i>The Principle of Transformation between Efficiency and Effectiveness: Towards a Fair Evaluation of the Cost-Effectiveness of KDD Techniques</i>	299
P. Berka <i>Recognizing Reliability of Discovered Knowledge</i>	307
A. M. Manning, A. Brass, C. A. Goble, J. A. Keane <i>Clustering Techniques in Biological Sequence Analysis</i>	315
R. J. Watts, A. L. Porter, S. Cunningham, D. Zhu <i>TOAS Intelligence Mining; Analysis of NLP and Computational Linguistics</i>	323
M. Moshkov <i>Algorithms for Constructing of Decision Trees</i>	335
H. Ahonen, O. Heinonen, M. Klemettinen, A. Inkeri Verkamo <i>Mining in the Phrasal Frontier</i>	343
A. T. Bjorvand <i>Mining Time Series Using Rough Sets - A Case Study</i>	351
H. S. Nguyen, M. S. Szczuka, D. Słezak <i>Neural Networks Design: Rough Set Approach to Continuous Data</i>	359
N. Zhong, S. Ohsuga, C. Liu, Y. Kakemoto, X. Zhang <i>On Meta Levels of an Organized Society of KDD Agents</i>	367
Y. Zhou, Y. Lu, C. Shi <i>Using Neural Network to Extract Knowledge from Database</i>	376
M. Quafafou, M. Boussouf <i>Induction of Strong Feature Subsets</i>	384

Tutorials

J. Komorowski, L. Polkowski, A. Skowron <i>Rough Sets for Data Mining and Knowledge Discovery</i>	393
W. Klösgen, J. Zytkow <i>Techniques and Applications of KDD</i>	394
R. Grossman <i>A Tutorial Introduction to High Performance Data Mining</i>	395
L. Carbonara, H. Roberts, B. Egan <i>Data Mining in the Telecommunications Industry</i>	396

Author Index	397
------------------------	-----