

# **Lecture Notes in Artificial Intelligence**

**1510**

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*

*Budapest*

*Hong Kong*

*London*

*Milan*

*Paris*

*Singapore*

*Tokyo*

Jan M. Żytkow   Mohamed Quafafou (Eds.)

# Principles of Data Mining and Knowledge Discovery

Second European Symposium, PKDD '98  
Nantes, France, September 23-26, 1998  
Proceedings



Springer

## Series Editors

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

## Volume Editors

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

Mohamed Quafafou  
Université de Nantes, IRIN  
2, rue de la Houssinière, F-44322 Nantes Cedex 3, France  
E-mail: quafafou@irin.univ-nantes.fr

Cataloging-in-Publication Data applied for

## Die Deutsche Bibliothek - CIP-Einheitsaufnahme

**Principles of data mining and knowledge discovery : second European symposium ; proceedings / PKDD '98, Nantes, France, September 23 - 26, 1998. Jan M. Żytkow ; Mohamed Quafafou (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Budapest ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 1998 (Lecture notes in computer science ; Vol. 1510 : Lecture notes in artificial intelligence)**  
ISBN 3-540-65068-7

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

ISBN 3-540-65068-7 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 1998  
Printed in Germany

Typesetting: Camera ready by author  
SPIN 10692663 06/3142 - 5 4 3 2 1 0 Printed on acid-free paper

# Preface

Knowledge Discovery in Databases (KDD), also known as Data Mining, has emerged in the last decade in response to the challenge of turning large and ubiquitous databases into knowledge that can be used in practice.

KDD has been able to grow very rapidly by drawing its techniques and data mining experiences from a combination of many existing research areas: databases, statistics, machine learning, automated scientific discovery, inductive logic programming, artificial intelligence, visualization, decision science, and high performance computing. While each of these areas can contribute in specific ways, the strength of KDD comes from the value that is added by creative combination of techniques from the contributing areas.

The practical successes of KDD in a broad range of application domains have led to very high expectations. In the long run, KDD can meet or fail those expectations. In order to maintain its status of interdisciplinary research of great practical payoff KDD has to establish its own theoretical principles that go beyond each of the contributing areas, and demonstrate how they jointly create a broad and exciting area of research. Such principles will be instrumental in maintaining the identity of KDD research, in effective communication and in guiding the practitioners.

Seeking the principles has always been a part of the European research tradition. Thus “Principles of KDD” (PKDD) make a suitable focus for the annual meetings of the KDD community in Europe. The main long-term interest is in theoretical principles for the emerging discipline of KDD. Another 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 as well as fostering the interdisciplinary collaboration. The first meeting was held in Trondheim, Norway, in June 1997.

This volume contains papers selected for presentation at the Second European Symposium on Principles of Data Mining and Knowledge Discovery in Databases – PKDD’98, held in Nantes, France, September 23-26, 1998. The University of Nantes hosted the symposium. The symposium was sponsored by the Ville de Nantes, Université de Nantes, MENRT, Conseil Régional des Pays de Loire, Centre de Recherche et Développement de France Télécom (CNET). We wish to express our thanks to the sponsors of the symposium for their generous support. The contributed papers were selected from 73 full draft papers by the following program committee:

**Pieter Adriaans** (Syllogic, The Netherlands), **Pawel Brazdil** (U. Porto, Portugal), **Henri Briand** (U. Nantes, France), **Leo Carbonara** (British Telecom, UK), **A. Fazl Famili** (IIT-NRC, Canada), **Ronen Feldman** (Bar Ilan, U. Israel), **Patrick Gallinari** (U. Paris VI, France), **Jean Gabriel Ganascia** (U. Paris VI, France), **Attilio Giordana** (U. Torino, Italy), **David Hand** (Open U. UK), **Bob Henery** (U. Strathclyde, UK), **Mikhail Kiselev**

(Megaputer Intelligence, Russia), **Willi Kloesgen** (GMD, Germany), **Yves Kodratoff** (U. Paris VI, France), **Jan Komorowski** (NTNU, Norway), **Nada Lavrac** (Josef Stefan Inst. Slovenia), **Heikki Manilla** (U. Helsinki, Finland), **Steve Muggleton** (Oxford U. UK), **Zdzislaw Pawlak** (Warsaw Technical U. Poland), **Gregory Piatetsky-Shapiro** (Knowledge Stream, Boston, USA), **Lech Polkowski** (U. Warsaw, Poland), **Mohamed Quafafou** (U. Nantes, France), **Zbigniew Ras** (UNC Charlotte, USA), **Lorenza Saitta** (U. Torino, Italy), **Wei-Min Shen** (U. South Calif. USA), **Arno Siebes** (CWI, Netherlands), **Andrzej Skowron** (U. Warsaw, Poland), **Derek Sleeman** (U. Aberdeen, UK), **Nicolas Spyrtos** (U. Paris XI, France), **Shusaku Tsumoto** (Tokyo Medical & Dental U. Japan), **Raul Valdes-Perez** (CMU, USA), **Thierry Van de Merckt** (CSC, Belgium), **Rudiger Wirth** (Daimler-Benz, Germany), **Stefan Wrobel** (GTE, Germany), **Ning Zhong** (Yamaguchi U. Japan), **Wojtek Ziarko** (U. Regina, Canada), **Djamel A. Zighed** (U. Lyon II, France), **Jan M. Żytkow** (UNC Charlotte, USA).

PKDD was truly international: papers came from 24 countries, including Australia (2), Belgium (1), Brazil (1), Bulgaria (1), Canada (1), China (1), Cuba (1), Czech Republic (3), Finland (2), France (18), Germany (6), Israel (3), Italy (2), Japan (5), Norway (1), Poland (4), Portugal (2), Russia (1), Singapore (2), Spain (5), Sweden (1), Switzerland (1), United Kingdom (6), United States of America (3). Many thanks to all who submitted papers for review and for publication in the proceedings. The accepted papers were divided into two categories: 26 oral presentations and 30 poster presentations. In addition to poster sessions each poster paper was allocated 4 minutes highlight presentation. All papers were allocated the same number of 9 pages in the proceedings.

Three tutorials were offered to all symposium participants on September 23rd: Scalable, High-Performance Data Mining with Parallel Processing by Alex Alves Freitas, Industrial Applications of Data Mining by Gholamreza Nakhaeizadeh, and Practical Text Mining by Ronen Feldman.

The members of PKDD'98 local organizing committee: Emmanuelle Martienne, Laurent Ughetto, and Abdellatif Saoudi, all of the University of Nantes, did an enormous amount of work and deserve the special gratitude of all the participants. We wish to express our thanks to the following colleagues for their reviewing participation: M.R. Amini, J-F. Boulicaut, V. Corruble, A. Giacometti, A. Jorge, A. Leger, D. Laurent, F. Mitchell, H.S. Nguyen, D. Slezak, N. Valette, H. Zaragoza, J-D. Zucker.

Special thanks are due to Alfred Hofmann of Springer-Verlag for his help and support.

# Table of Contents

## Communications

### Session 1. Rule evaluation

On Objective Measures of Rule Suprisingness .....	1
<i>A.A. Freitas</i>	
Discovery of Surprising Exception Rules Based on Intensity of Implication .....	10
<i>E. Suzuki, Y. Kodratoff</i>	
A Metric for Selection of the Most Promising Rules .....	19
<i>P. Gago, C. Bento</i>	

### Session 2. Visualization

For Visualization-Based Analysis Tools in Knowledge Discovery Process: A Multilayer Perceptron versus Principal Components Analysis - A Comparative Study .....	28
<i>X. Polanco, C. François, M.A. Ould Louly</i>	
Trend Graphs: Visualizing the Evolution of Concept Relationships in Large Document Collections .....	38
<i>R. Feldman, Y. Aumann, A. Zilberstein, Y. Ben-Yehuda</i>	
Ranked Rules and Data Visualization .....	47
<i>L. Bobrowski, T. Sowiński</i>	

### Session 3. Association rules and text mining

TextVis: An Integrated Visual Environment for Text Mining .....	56
<i>D. Landau, R. Feldman, O. Zamir, Y. Aumann, , M. Fresko, Y. Lindell, O. Lipshtat</i>	
Text Mining at the Term Level .....	65
<i>R. Feldman, M. Fresko, Y. Kinar, Y. Lindell, O. Lipshtat, M. Rajman, Y. Schler, O. Zamir</i>	
A New Algorithm for Faster Mining of Generalized Association Rules .....	74
<i>J. Hipp, A. Myka, R. Wirth, U. Güntzer</i>	

### Session 4. Clustering and discretization

Knowledge Discovery with Clustering Based on Rules. Interpreting Results .....	83
<i>K. Gibert, T. Aluja, U. Cortés</i>	

Efficient Construction of Comprehensible Hierarchical Clusterings . . . . .	93
<i>L. Talavera, J. Béjar</i>	

Cost Sensitive Discretization of Numeric Attributes . . . . .	102
<i>T. Brijs, K. Vanhoof</i>	

## Session 5. KDD process and software

Handling KDD Process Changes by Incremental Replanning . . . . .	111
<i>N. Zhong, C. Liu, Y. Kakemoto, S. Ohsuga</i>	

Object Mining: A Practical Application of Data Mining for the Construction and Maintenance of Software Components . . . . .	121
<i>A. T. Bjorvand</i>	

A Relational Data Mining Tool Based on Genetic Programming . . . . .	130
<i>L. Martin, F. Moal, C. Vrain</i>	

## Session 6. Tree construction

Inducing Cost-Sensitive Trees via Instance Weighting . . . . .	139
<i>K.M. Ting</i>	

Model Switching for Bayesian Classification Trees with Soft Splits . . . . .	148
<i>J. Kindermann, G. Paass</i>	

Interactive Visualization for Predictive Modelling with Decision Tree Induction . . . . .	158
<i>T.B. Ho, T.D. Nguyen</i>	

## Session 7. Sequential and spatial data mining

Discovery of Diagnostic Patterns from Protein Sequence Databases . . . . .	167
<i>B. Olsson, K. Laurio</i>	

The PSP Approach for Mining Sequential Patterns . . . . .	176
<i>F. Masseglia, F. Cathala, P. Poncelet</i>	

Knowledge Discovery in Spatial Data by Means of ILP . . . . .	185
<i>L. Popelínský</i>	

Querying Inductive Databases: A Case Study on the MINE RULE Operator	194
<i>J.F. Boulicaut, M. Klemettinen, H. Mannila</i>	

## Session 8. Attribute selection

Classes of Four-Fold Table Quantifiers . . . . .	203
<i>J. Rauch</i>	



Detection of Interdependences in Attribute Selection .....	212
<i>J. Lorenzo, M. Hernández, J. Méndez</i>	
Postponing the Evaluation of Attributes with a High Number of Boundary Points .....	221
<i>T. Elomaa, J. Rousu</i>	
A Hybrid Approach to Feature Selection .....	230
<i>M. Boussouf</i>	

## Posters

Discretization and Grouping: Preprocessing Steps for Data Mining .....	239
<i>P. Berka, I. Bruha</i>	
Fuzzy Spacial OQL for Fuzzy Knowledge Discovery in Databases .....	246
<i>N.M. Bigolin, C. Marsala</i>	
Extended Functional Dependencies as a Basis for Linguistic Summaries ...	255
<i>P. Bosc, L. Liétard, O. Pivert</i>	
A Comparison of Batch and Incremental Supervised Learning Algorithms .	264
<i>L. Carbonara, A. Borrowman</i>	
Knowledge Discovery with Qualitative Influences and Synergies .....	273
<i>J. Cerquides, R. López de Màntaras</i>	
Language Support for Temporal Data Mining .....	282
<i>X. Chen, I. Petrounias</i>	
Resampling in an Indefinite Database to Approximate Functional Dependencies .....	291
<i>E. Collopy, M. Levene</i>	
Knowledge Discovery from Client-Server Databases .....	300
<i>N. Dewhurst, S. Lavington</i>	
Discovery of Common Subsequences in Cognitive Evoked Potentials .....	309
<i>A. Flexer, H. Bauer</i>	
Improving the Discovery of Association Rules with Intensity of Implication	318
<i>S. Guillaume, F. Guillet, J. Philippé</i>	
Generalization Lattices .....	328
<i>H.J. Hamilton, R.J. Hilderman, L. Li, D.J. Randall</i>	

Overcoming Fragmentation in Decision Trees Through Attribute Value Grouping .....	337
<i>K.M. Ho, P.D. Scott</i>	
Data Mining at a Major Bank: Lessons from a Large Marketing Application	345
<i>P. Hunziker, A. Maier, A. Nippe, M. Tresch, D. Weers, P. Zemp</i>	
PolyAnalyst Data Analysis Technique and Its Specialization for Processing Data Organized as a Set of Attribute Values .....	352
<i>M.V. Kiselev, S.M. Ananyan, S.B. Arseniev</i>	
Representative Association Rules and Minimum Condition Maximum Consequence Association Rules .....	361
<i>M. Kryszkiewicz</i>	
Discovery of Decision Rules from Databases: An Evolutionary Approach ..	370
<i>W. Kwedlo, M. Krętowski</i>	
Using Loglinear Clustering for Subcategorization Identification .....	379
<i>N.M. Marques, G.P. Lopes, C.A. Coelho</i>	
Exploratory Attributes Search in Times-Series Data: An Experimental System for Agricultural Application .....	388
<i>K. Matsumoto</i>	
A Procedure to Compute Prototypes for Data Mining in Non-structured Domains .....	396
<i>J. Méndez, M. Hernández, J. Lorenzo</i>	
From the Data Mine to the Knowledge Mill: Applying the Principles of Lexical Analysis to the Data Mining and Knowledge Discovery Process ...	405
<i>J. Moscarola, R. Bolden</i>	
Preprocessing of Missing Values Using Robust Associations Rules .....	414
<i>A. Ragel</i>	
Similarity-Driven Sampling for Data Mining .....	423
<i>T. Reinartz</i>	
Modeling the Business Process by Mining Multiple Databases .....	432
<i>A.P. Sanjeev, J.M. Żytkow</i>	
Data Transformation and Rough Sets .....	441
<i>J. Stepaniuk, M. Maj</i>	
Conceptual Knowledge Discovery in Databases Using Formal Concept Analysis Methods .....	450
<i>G. Stumme, R. Wille, U. Wille</i>	

CLASITEX+: A Tool for Knowledge Discovery from Texts . . . . .	459
<i>J.F. Martínez Trinidad, B. Beltrán Martínez, A. Guzmán Arenas,</i>	
<i>J. Ruiz Shulcloper</i>	
Discovery of Approximate Medical Knowledge Based on Rough set Model .	468
<i>S. Tsumoto</i>	

## Tutorials

Scalable, High-Performance Data Mining with Parallel Processing . . . . .	477
<i>Alex Alves Freitas, CEFET-PR DAINF, Brazil</i>	
Practical Text Mining . . . . .	478
<i>Ronen Feldman, Bar-Ilan University</i>	
Industrial Applications of Data Mining . . . . .	479
<i>Gholamreza Nakhaeizadeh, Daimler-Benz, Germany</i>	
Author Index . . . . .	481