

Lecture Notes in Artificial Intelligence 1910

Subseries of Lecture Notes in Computer Science

Djamel A. Zighed Jan Komorowski
Jan Żytkow (Eds.)

Principles of Data Mining and Knowledge Discovery

4th European Conference, PKDD 2000

Lyon, France, September 2000

Proceedings



Springer

Djamel A. Zighed Jan Komorowski
Jan Żytkow (Eds.)

Principles of Data Mining and Knowledge Discovery

4th European Conference, PKDD 2000
Lyon, France, September 13-16, 2000
Proceedings



Springer

Series Editors

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

Volume Editors

Djamel A. Zighed
Université Lyon 2, Laboratoire ERIC
5 avenue Pierre Mendès-France, 69676 Bron, France
E-mail: zighed@univ-lyon2.fr

Jan Komorowski
Norwegian University of Science and Technology
Department of Computer and Information Science
O.S. Bragstads plass 2E, 7491 Trondheim, Norway
E-mail: janko@idi.ntnu.no

Jan Żytkow
University of North Carolina, Department of Computer Science
Charlotte, NC 28223, USA
E-mail: zytkow@uncc.edu

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Principles of data mining and knowledge discovery : 4th European conference ; proceedings / PKDD 2000, Lyon, France, September 13 - 16, 2000. Djamel A. Zighed ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2000
(Lecture notes in computer science ; Vol. 1910 : Lecture notes in artificial intelligence)
ISBN 3-540-41066-X

CR Subject Classification (1998): I.2, H.3, H.5, G.3, J.1, F.4.1

ISBN 3-540-41066-X 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
© Springer-Verlag Berlin Heidelberg 2000
Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Stefan Sossna
Printed on acid-free paper SPIN: 10722719 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence

1910

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

Preface

This volume contains papers selected for presentation at PKDD'2000, the Fourth European Conference on Principles and Practice of Knowledge Discovery in Databases. The first meeting was held in Trondheim, Norway, in June 1997, the second in Nantes, France, in September 1998, and the third in Prague, Czech Republic, in September 1999.

PKDD 2000 was organized in Lyon, France, on 13–16 September 2000. The conference was hosted by the Equipe de Recherche en Ingénierie des Connaissances at the Université Lumière Lyon 2. We wish to express our thanks to the sponsors of the Conference, to the University Claude Bernard Lyon 1, the INSA of Lyon, the Conseil général of the Rhône, the Région Rhône Alpes, SPSS France, AFIA, and the University of Lyon 2, for their generous support.

Knowledge discovery in databases (KDD), also known as data mining, provides tools for turning large databases into knowledge that can be used in practice. KDD has been able to grow very rapidly since its emergence a decade ago by drawing its techniques and data mining experiences from a combination of many existing research areas: databases, statistics, mathematical logic, machine learning, automated scientific discovery, inductive logic programming, artificial intelligence, visualization, decision science, knowledge management, and high performance computing. The strength of KDD came initially from the value added by the creative combination of techniques from the contributing areas. But in order to establish its identity, KDD has to create its own theoretical principles and to demonstrate how they stimulate KDD research, facilitate communication, and guide practitioners towards successful applications.

Seeking the principles that can guide and strengthen practical applications has been always a part of the European research tradition. Thus “Principles and Practice of KDD” (PKDD) make a suitable focus for annual meetings of the KDD community in Europe. The main long-term interest is in theoretical principles for the emerging discipline of KDD and in practical applications that demonstrate utility of those principles. Other goals of the PKDD series are to provide a European-based forum for interaction among all theoreticians and practitioners interested in data mining and knowledge discovery as well as to foster new directions in interdisciplinary collaboration.

A Discovery Challenge hosted at PKDD 2000 was conducted for the second year in a row. It followed a successful and very broadly attended Discovery Challenge in Prague. Discovery challenge is a new initiative that promotes cooperative research on original and practically important real-world databases. The Challenge requires a broad and unified view of knowledge and methods of discovery, and emphasizes business problems that require an open-minded search for knowledge in data. Two multi-relational databases, in banking and in medicine, were available to all conference participants. The Challenge was born out of the conviction that knowledge discovery in real-world databases requires

an open-minded discovery process rather than application of one or another tool limited to one form of knowledge. A discoverer should consider a broad scope of techniques that can reach many forms of knowledge. The discovery process cannot be rigid and selection of techniques must be driven by knowledge hidden in the data, so that the most and the best of knowledge can be reached.

The contributed papers were selected from 157 full papers (48% growth over PKDD'99) by the following program committee:

Magnus L. Andersson (AstraZeneca, Sweden),
Petr Berka (U. Economics, Czech Republic),
Jean-Francois Boulicaut (INSA Lyon, France),
Henri Briand (U. Nantes, France),
Leo Carbonara (American Management System, UK),
Luc De Raedt (U. Freiburg, Germany),
Ronen Feldman (Bar Ilan U., Israel),
Arthur Flexer (Austrian Research Inst. for AI, Austria),
Alex Freitas (PUC-PR, Brazil),
Patrick Gallinari (U. Paris 6, France),
Jean Gabriel Ganascia (U. Paris 6, France),
Mohand-Said Hacid (U. Lyon I, INSA Lyon, France),
Petr Hájek (Acad. Science, Czech Republic),
Howard Hamilton (U. Regina, Canada),
David Hand (Imperial Col. London, UK),
Mika Klemettinen (U. Helsinki, Finland),
Willi Kloesgen (GMD, Germany),
Yves Kodratoff (U. Paris-Sud, France),
Jan Komorowski (Norwegian U. Sci. & Tech.),
Jacek Koronacki (Acad. Science, Poland),
Michel Lamure (U. Lyon I, France),
Heikki Manilla (Nokia, Finland),
Stan Matwin (U. Ottawa, Canada),
Hiroshi Motoda (Osaka U., Japan),
Jan Mrazek (Bank of Montreal, Canada),
Witold Pedrycz (U. Alberta, Canada),
Mohamed Quafafou (U. Nantes, France),
Ricco Rakotomalala (F. Lyon II, France),
Jan Rauch (U. Economics, Czech Republic),
Zbigniew Ras (UNC Charlotte, USA),
Gilbert Ritschard (U. Geneva, Switzerland),
Michele Sebag (Ecole Polytech. Paris, France),
Arno Siebes (CWI, Netherlands),
Andrzej Skowron (U. Warsaw, Poland),
Myra Spiliopoulou (Humboldt U. Berlin, Germany),
Nicolas Spyratos (U. Paris-Sud, France),
Olga Štěpánková (Czech Tech. U., Czech Republic),
Einoshin Suzuki (Yokohama Natl. U., Japan),

Ljupco Todorovski (Josef Stefan Inst., Slovenia),
Shusaku Tsumoto (Shimane Medical U., Japan),
Gilles Venturini (U. Tours, France),
Inkeri Verkamo (U. Helsinki, Finland),
Louis Wehenkel (U. Liege, Belgium),
Gerhard Widmer (U. Vienna, Austria),
Rudiger Wirth (DaimlerChrysler, Germany),
Stefan Wrobel (U. Magdeburg, Germany),
Ning Zhong (Maebashi Inst. Techn., Japan),
Djamel A. Zighed (U. Lyon 2, France), and
Jan Żytkow (UNC Charlotte, USA).

The following colleagues also reviewed papers for the conference and are due our special thanks:

Florence d'Alche buc (U. Paris 6),
Jérôme Darmont (U. Lyon 2),
Stéphane Lallich (U. Lyon 2), and
Christel Vrain (U. Orleans).

Classified according to the first author's nationality, papers submitted to PKDD'2000 came from 34 countries on 5 continents (Europe: 102 papers; Asia: 21; North America: 17; Australia: 3; and South America: 4), including Argentina (1 paper), Australia (3), Austria (3), Belgium (5), Brazil (3), Canada (4), Croatia (1), Czech Republic (4), Finland (3), France (38), Germany (7), Greece (5), Hong Kong (1), Italy (3), Japan (9), Mexico (2), Netherlands (1), Norway (1), Poland (5), Portugal (2), Romania (1), Russia (1), Singapore (4), Slovakia (1), Slovenia (1), South Korea (3), Spain (5), Sweden (1), Switzerland (3), Taiwan (4), Turkey (1), United Kingdom (10), and USA (20).

Many thanks to all who submitted papers for review and for publication in the proceedings. The accepted papers were divided into two categories: 29 oral presentations and 57 poster presentations. In addition to poster sessions each poster paper was available for the web-based discussion forum in the months preceding the conference.

There were three invited presentations:

- Sep.14, **Willi Kloesgen** (GMD, Germany), Multi-relational, statistical, and visualization approaches for spatial knowledge discovery
- Sep.15 **Luc De Raedt** (U.Freiburg, Germany), Data mining in multi-relational databases
- Sep.16 **Arno Siebes** (Utrecht U.), Developing KDD Systems

Six workshops were affiliated with the conference. All were held on 12 September.

- **Workshop 1** Data Mining, Decision Support, Meta-learning, and ILP. Chairs: **Pavel Brazdil** and **Alipio Jorge**, U. Porto, Portugal;
- **Workshop 2** Temporal, Spatial, and Spatio-temporal Data Mining (TSDM 2000). Chairs: **John F. Roddick**, Flinders U., Australia & **Kathleen Hornsby**, U. Maine, USA;

- **Workshop 3** Knowledge Discovery in Biology. Chair: **Jan Komorowski**, Norwegian U. Sci. & Techn, Norway;
- **Workshop 4** Machine Learning and Textual Information Access. Chairs: **Hugo Zaragoza**, LIP6, U. Paris 6, France **Patrick Gallinari**, LIP6, U. Paris 6, France **Martin Rajman**, EPFL, Switzerland;
- **Workshop 5** Knowledge Management: Theory and Applications. Chair: **Jean-Louis Ermine**, CEA Paris, France;
- **Workshop 6** Symbolic Data Analysis: Theory, Software, and Applications for Knowledge Mining. Chair: **Edwin Diday**, U. Paris Dauphine, France.

Five tutorials were offered to all conference participants on 13 September:

- **Tutorial 1** An Introduction to Distributed Data Mining, presented by **H. Kargupta**, Washington State U., USA;
- **Tutorial 2** Clustering Techniques for Large Data Sets: From the Past to the Future, presented by **A. Hinneburg** and **D.A. Keim**, U. Halle, Germany;
- **Tutorial 3** Data Analysis for Web Marketing and Merchandizing Applications, presented by **M. Spiliopoulou**, Humboldt U. Berlin, Germany;
- **Tutorial 4** Database Support for Business Intelligence Applications, presented by **W. Lehner**, U. Erlangen-Nürnberg, Germany;
- **Tutorial 5** Text Mining, presented by **Yves Kodratoff**, **Djamel Zighed**, and **Serge Di Palma** (France).

Members of the PKDD'2000 organizing committee did an enormous amount of work and deserve the special gratitude of all participants: **Willi Kloesgen** – Tutorials Chair, **Jan Rauch** – Workshops Chair, **Arno Siebes** and **Petr Berka** – Discovery Challenge Chairs, **Leonardo Carbonara** – Industrial Program Chair, **Djamel Zighed** – Local Arrangement Chair, and **Jean Hugues Chauchat**, **Fabrice Muhlenbach**, **Laure Tougne**, **Fadila Bentayeb**, **Omar Boussaid**, **Salima Hassas**, **Stéphane Lallich**, **Céline Agier**, **David Coeurjolly**, **Alfredos Mar**, **Luminita Firanescu** **Astrid Varaine**, **Sabine Rabaséda**, **Fabien Feschet**, and **Nicolas Nicoloyannis**.

Special thanks go to **Alfred Hofmann** of Springer-Verlag for his continuous help and support and to **Karin Henzold** for the preparation of the Proceedings.

July 2000

Djamel Zighed and Jan Komorowski
Jan Żytkow

Lecture Notes in Artificial Intelligence (LNAI)

- Vol. 1761: R. Caferra, G. Salzer (Eds.), Automated Deduction in Classical and Non-Classical Logics. Proceedings. VIII, 299 pages. 2000.
- Vol. 1771: P. Lambrix, Part-Whole Reasoning in an Object-Centered Framework. XII, 195 pages. 2000.
- Vol. 1772: M. Beetz, Concurrent Reactive Plans. XVI, 213 pages. 2000.
- Vol. 1775: M. Thielscher, Challenges for Action Theories. XIII, 138 pages. 2000.
- Vol. 1778: S. Wermter, R. Sun (Eds.), Hybrid Neural Systems. IX, 403 pages. 2000.
- Vol. 1788: A. Moukas, C. Sierra, F. Ygge (Eds.), Agent Mediated Electronic Commerce II. IX, 239 pages. 2000.
- Vol. 1792: E. Lamma, P. Mello (Eds.), AI*IA 99: Advances in Artificial Intelligence. Proceedings, 1999. XI, 392 pages. 2000.
- Vol. 1793: O. Cairo, L.E. Sucar, F.J. Cantu (Eds.), MICA 2000: Advances in Artificial Intelligence. Proceedings, 2000. XIV, 750 pages. 2000.
- Vol. 1794: H. Kirchner, C. Ringeissen (Eds.), Frontiers of Combining Systems. Proceedings, 2000. X, 291 pages. 2000.
- Vol. 1804: B. Azvine, N. Azarmi, D.D. Nauck (Eds.), Intelligent Systems and Soft Computing. XVII, 359 pages. 2000.
- Vol. 1805: T. Terano, H. Liu, A.L.P. Chen (Eds.), Knowledge Discovery and Data Mining. Proceedings, 2000. XIV, 460 pages. 2000.
- Vol. 1809: S. Biundo, M. Fox (Eds.), Recent Advances in AI Planning. Proceedings, 1999. VIII, 373 pages. 2000.
- Vol. 1810: R. López de Mántaras, E. Plaza (Eds.), Machine Learning: ECML 2000. Proceedings, 2000. XII, 460 pages. 2000.
- Vol. 1813: P.L. Lanzi, W. Stolzmann, S.W. Wilson (Eds.), Learning Classifier Systems. X, 349 pages. 2000.
- Vol. 1821: R. Loganathanaraj, G. Palm, M. Ali (Eds.), Intelligent Problem Solving. Proceedings, 2000. XVII, 751 pages. 2000.
- Vol. 1822: H.H. Hamilton, Advances in Artificial Intelligence. Proceedings, 2000. XII, 450 pages. 2000.
- Vol. 1831: D. McAllester (Ed.), Automated Deduction – CADE-17. Proceedings, 2000. XIII, 519 pages. 2000.
- Vol. 1834: J.-C. Heudin (Ed.), Virtual Worlds. Proceedings, 2000. XI, 314 pages. 2000.
- Vol. 1835: D. N. Christodoulakis (Ed.), Natural Language Processing – NLP 2000. Proceedings, 2000. XII, 438 pages. 2000.
- Vol. 1836: B. Masand, M. Spiliopoulou (Eds.), Web Usage Analysis and User Profiling. Proceedings, 2000. V, 183 pages. 2000.
- Vol. 1847: R. Dyckhoff (Ed.), Automated Reasoning with Analytic Tableaux and Related Methods. Proceedings, 2000. X, 441 pages. 2000.
- Vol. 1849: C. Freksa, W. Brauer, C. Habel, K.F. Wender (Eds.), Spatial Cognition II. XI, 420 pages. 2000.
- Vol. 1856: M. Veloso, E. Pagello, H. Kitano (Eds.), RoboCup-99: Robot Soccer World Cup III. XIV, 802 pages. 2000.
- Vol. 1860: M. Klusch, L. Kerschberg (Eds.), Cooperative Information Agents IV. Proceedings, 2000. XI, 285 pages. 2000.
- Vol. 1861: J. Lloyd, V. Dahl, U. Furbach, M. Kerber, K.-K. Lau, C. Palamidessi, L. Moniz Pereira, Y. Sagiv, P.J. Stuckey (Eds.), Computational Logic – CL 2000. Proceedings, 2000. XIX, 1379 pages.
- Vol. 1864: B. Y. Choueiry, T. Walsh (Eds.), Abstraction, Reformulation, and Approximation. Proceedings, 2000. XI, 333 pages. 2000.
- Vol. 1865: K.R. Apt, A.C. Kakas, E. Monfroy, F. Rossi (Eds.), New Trends Constraints. Proceedings, 1999. X, 339 pages. 2000.
- Vol. 1866: J. Cussens, A. Frisch (Eds.), Inductive Logic Programming. Proceedings, 2000. X, 265 pages. 2000.
- Vol. 1867: B. Ganter, G.W. Mineau (Eds.), Conceptual Structures: Logical, Linguistic, and Computational Issues. Proceedings, 2000. XI, 569 pages. 2000.
- Vol. 1881: C. Zhang, V.-W. So (Eds.), Design and Applications of Intelligent Agents. Proceedings, 2000. X, 183 pages. 2000.
- Vol. 1886: R. Mizoguchi, J. Slaney (Eds.), PRICAI 2000: Topics in Artificial Intelligence. Proceedings, 2000. XX, 835 pages. 2000.
- Vol. 1898: E. Blanzieri, L. Portinale (Eds.), Advances in Case-Based Reasoning. Proceedings, 2000. XII, 530 pages. 2000.
- Vol. 1889: M. Anderson, P. Cheng, V. Haarslev (Eds.), Theory and Application of Diagrams. Proceedings, 2000. XII, 504 pages. 2000.
- Vol. 1891: A.L. Oliveira (Ed.), Grammatical Inference: Algorithms and Applications. Proceedings, 2000. VIII, 313 pages. 2000.
- Vol. 1902: P. Sojka, I. Kopeček, K. Pala (Eds.), Text, Speech and Dialogue. Proceedings, 2000. XIII, 463 pages. 2000.
- Vol. 1904: S.A. Cerri, D. Dochev (Eds.), Artificial Intelligence: Methodology, Systems, and Applications. Proceedings, 2000. XII, 366 pages. 2000.
- Vol. 1910: D.A. Zighed, J. Komorowski, J. Żytkow (Eds.), Principles of Data Mining and Knowledge Discovery. Proceedings, 2000. XV, 701 pages. 2000.

Lecture Notes in Computer Science

- Vol. 1882: D. Kotz, F. Mattern (Eds.), Agent Systems, Mobile Agents, and Applications. Proceedings, 2000. XII, 275 pages. 2000.
- Vol. 1883: B. Triggs, A. Zisserman, R. Szeliski (Eds.), Vision Algorithms: Theory and Practice. Proceedings, 1999. X, 383 pages. 2000.
- Vol. 1884: J. Štuller, J. Pokorný, B. Thalheim, Y. Masunaga (Eds.), Current Issues in Databases and Information Systems. Proceedings, 2000. XIII, 396 pages. 2000.
- Vol. 1885: K. Havelund, J. Penix, W. Visser (Eds.), SPIN Model Checking and Software Verification. Proceedings, 2000. X, 343 pages. 2000.
- Vol. 1886: R. Mizoguchi, J. Slaney (Eds.), PRICAI 2000: Topics in Artificial Intelligence. Proceedings, 2000. XX, 835 pages. 2000. (Subseries LNAI).
- Vol. 1888: G. Sommer, Y.Y. Zeevi (Eds.), Algebraic Frames for the Perception-Action Cycle. Proceedings, 2000. X, 349 pages. 2000.
- Vol. 1889: M. Anderson, P. Cheng, V. Haarslev (Eds.), Theory and Application of Diagrams. Proceedings, 2000. XII, 504 pages. 2000. (Subseries LNAI).
- Vol. 1890: C. Linnhoff-Popien, H.-G. Hegering (Eds.), Trends in Distributed Systems: Towards a Universal Service Market. Proceedings, 2000. XI, 341 pages. 2000.
- Vol. 1891: A.L. Oliveira (Ed.), Grammatical Inference: Algorithms and Applications. Proceedings, 2000. VIII, 313 pages. 2000. (Subseries LNAI).
- Vol. 1892: P. Brusilovsky, O. Stock, C. Strapparava (Eds.), Adaptive Hypermedia and Adaptive Web-Based Systems. Proceedings, 2000. XIII, 422 pages. 2000.
- Vol. 1893: M. Nielsen, B. Rován (Eds.), Mathematical Foundations of Computer Science 2000. Proceedings, 2000. XIII, 710 pages. 2000.
- Vol. 1894: R. Dechter (Ed.), Principles and Practice of Constraint Programming – CP 2000. Proceedings, 2000. XII, 556 pages. 2000.
- Vol. 1895: F. Cuppens, Y. Deswarte, D. Gollmann, M. Waidner (Eds.), Computer Security – ESORICS 2000. Proceedings, 2000. X, 325 pages. 2000.
- Vol. 1896: R. W. Hartenstein, H. Grünbacher (Eds.), Field-Programmable Logic and Applications. Proceedings, 2000. XVII, 856 pages. 2000.
- Vol. 1897: J. Gutknecht, W. Weck (Eds.), Modular Programming Languages. Proceedings, 2000. XII, 299 pages. 2000.
- Vol. 1898: E. Blanzieri, L. Portinale (Eds.), Advances in Case-Based Reasoning. Proceedings, 2000. XII, 530 pages. 2000. (Subseries LNAI).
- Vol. 1899: H.-H. Nagel, F.J. Perales López (Eds.), Articulated Motion and Deformable Objects. Proceedings, 2000. X, 183 pages. 2000.
- Vol. 1900: A. Bode, T. Ludwig, W. Karl, R. Wismüller (Eds.), Euro-Par 2000 Parallel Processing. Proceedings, 2000. XXXV, 1368 pages. 2000.
- Vol. 1901: O. Etzion, P. Scheuermann (Eds.), Cooperative Information Systems. Proceedings, 2000. XI, 336 pages. 2000.
- Vol. 1902: P. Sojka, I. Kopeček, K. Pala (Eds.), Text, Speech and Dialogue. Proceedings, 2000. XIII, 463 pages. 2000. (Subseries LNAI).
- Vol. 1904: S.A. Cerri, D. Dochev (Eds.), Artificial Intelligence: Methodology, Systems, and Applications. Proceedings, 2000. XII, 366 pages. 2000. (Subseries LNAI).
- Vol. 1906: A. Porto, G.-C. Roman (Eds.), Coordination Languages and Models. Proceedings, 2000. IX, 353 pages. 2000.
- Vol. 1908: J. Dongarra, P. Kacsuk, N. Podhorszki (Eds.), Recent Advances in Parallel Virtual Machine and Message Passing Interface. Proceedings, 2000. XV, 364 pages. 2000.
- Vol. 1910: D.A. Zighed, J. Komorowski, J. Żytkow (Eds.), Principles of Data Mining and Knowledge Discovery. Proceedings, 2000. XV, 701 pages. 2000. (Subseries LNAI).
- Vol. 1912: Y. Gurevich, P.W. Kutter, M. Odersky, L. Thiele (Eds.), Abstract State Machines. Proceedings, 2000. X, 381 pages. 2000.
- Vol. 1913: K. Jansen, S. Khuller (Eds.), Approximation Algorithms for Combinatorial Optimization. Proceedings, 2000. IX, 275 pages. 2000.
- Vol. 1917: M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J.J. Merelo, H.-P. Schwefel (Eds.), Parallel Problem Solving from Nature – PPSN VI. Proceedings, 2000. XXI, 915 pages. 2000.
- Vol. 1918: D. Soudris, P. Pirsch, E. Barke (Eds.), Integrated Circuit Design. Proceedings, 2000. XII, 338 pages. 2000.
- Vol. 1923: J. Borbinha, T. Baker (Eds.), Research and Advanced Technology for Digital Libraries. Proceedings, 2000. XVII, 513 pages. 2000.
- Vol. 1924: W. Taha (Ed.), Semantics, Applications, and Implementation of Program Generation. Proceedings, 2000. VIII, 231 pages. 2000.
- Vol. 1926: M. Joseph (Ed.), Formal Techniques in Real-Time and Fault-Tolerant Systems. Proceedings, 2000. X, 305 pages. 2000.
- Vol. 1931: E. Horlait (Ed.), Mobile Agents for Telecommunication Applications. Proceedings, 2000. IX, 271 pages. 2000.

Table of Contents

Session 1A – Towards Broader Foundations

Multi-relational Data Mining, Using UML for ILP	1
<i>Arno J. Knobbe, Arno Siebes, Hendrik Blockeel, and Daniël Van Der Wallen</i>	
An Apriori-Based Algorithm for Mining Frequent Substructures from Graph Data	13
<i>Akihiro Inokuchi, Takashi Washio, and Hiroshi Motoda</i>	
Basis of a Fuzzy Knowledge Discovery System	24
<i>Maurice Bernadet</i>	

Session 1B – Rules and Trees

Confirmation Rule Sets	34
<i>Dragan Gamberger and Nada Lavrač</i>	
Contribution of Dataset Reduction Techniques to Tree-Simplification and Knowledge Discovery	44
<i>Marc Sebban and Richard Nock</i>	
Combining Multiple Models with Meta Decision Trees	54
<i>Ljupčo Todorovski and Sašo Džeroski</i>	

Session 2A – Databases and Reward-Based Learning

Materialized Data Mining Views	65
<i>Tadeusz Morzy, Marek Wojciechowski, and Maciej Zakrzewicz</i>	
Approximation of Frequency Queries by Means of Free-Sets	75
<i>Jean-François Boulicaut, Arthur Bykowski, and Christophe Rigotti</i>	
Application of Reinforcement Learning to Electrical Power System Closed-Loop Emergency Control	86
<i>Christophe Druet, Damien Ernst, and Louis Wehenkel</i>	
Efficient Score-Based Learning of Equivalence Classes of Bayesian Networks	96
<i>Paul Munteanu and Denis Cau</i>	

Session 2B – Classification

Quantifying the Resilience of Inductive Classification Algorithms	106
<i>Melanie Hilario and Alexandros Kalousis</i>	
Bagging and Boosting with Dynamic Integration of Classifiers	116
<i>Alexey Tsymbal and Seppo Puuronen</i>	
Zoomed Ranking: Selection of Classification Algorithms Based on Relevant Performance Information	126
<i>Carlos Soares and Pavel B. Brazdil</i>	
Some Enhancements of Decision Tree Bagging	136
<i>Pierre Geurts</i>	

Session 3A – Association Rules and Exceptions

Relative Unsupervised Discretization for Association Rule Mining	148
<i>Marcus-Christopher Ludl and Gerhard Widmer</i>	
Mining Association Rules: Deriving a Superior Algorithm by Analyzing Today’s Approaches	159
<i>Jochen Hipp, Ulrich Güntzer, and Gholamreza Nakhaeizadeh</i>	
Unified Algorithm for Undirected Discovery of Exception Rules	169
<i>Einoshin Suzuki and Jan Żytkow</i>	
Sampling Strategies for Targeting Rare Groups from a Bank Customer Database	181
<i>Jean-Hugues Chauchat, Ricco Rakotomalala, and Didier Robert</i>	

Session 3B – Instance-Based Discovery

Instance-Based Classification by Emerging Patterns	191
<i>Jinyan Li, Guozhu Dong, and Kotagiri Ramamohanarao</i>	
Context-Based Similarity Measures for Categorical Databases	201
<i>Gautam Das and Heikki Mannila</i>	
A Mixed Similarity Measure in Near-Linear Computational Complexity for Distance-Based Methods	211
<i>Ngoc Binh Nguyen and Tu Bao Ho</i>	
Fast Feature Selection Using Partial Correlation for Multi-valued Attributes	221
<i>Stéphane Lallich and Ricco Rakotomalala</i>	

Session 4A – Clustering and Classification

Fast Hierarchical Clustering Based on Compressed Data and OPTICS	232
<i>Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander</i>	
Accurate Recasting of Parameter Estimation Algorithms Using Sufficient Statistics for Efficient Parallel Speed-Up: Demonstrated for Center-Based Data Clustering Algorithms	243
<i>Bin Zhang, Meichun Hsu, and George Forman</i>	
Predictive Performance of Weighted Relative Accuracy	255
<i>Ljupčo Todorovski, Peter Flach, and Nada Lavrač</i>	
Quality Scheme Assessment in the Clustering Process	265
<i>Maria Halkidi, M. Vazirgiannis, and Y. Batistakis</i>	

Session 5A – Time Series

Algorithm for Matching Sets of Time Series	277
<i>Iztok Sarnik, Georg Lausen, Hans-Peter Kahle, Heinrich Spiecker, and Sebastian Hein</i>	
MSTS: A System for Mining Sets of Time Series	289
<i>Georg Lausen, Iztok Sarnik, and Aldar Dougarjapov</i>	
Learning First Order Logic Time Series Classifiers: Rules and Boosting . . .	299
<i>Juan J. Rodríguez, Carlos J. Alonso, and Henrik Boström</i>	

Posters

Learning Right Sized Belief Networks by Means of a Hybrid Methodology .	309
<i>Sylvia Acid and Luis M. De Campos</i>	
Algorithms for Mining Share Frequent Itemsets Containing Infrequent Subsets	316
<i>Brock Barber and Howard J. Hamilton</i>	
Discovering Task Neighbourhoods through Landmark Learning Performances	325
<i>Hilan Bensusan and Christophe Giraud-Carrier</i>	
Induction of Multivariate Decision Trees by Using Dipolar Criteria	331
<i>Leon Bobrowski and Marek Krętowski</i>	
Inductive Logic Programming in Clementine	337
<i>Sam Brewer and Tom Khabaza</i>	
A Genetic Algorithm-Based Solution for the Problem of Small Disjuncts . .	345
<i>Deborah R. Carvalho and Alex A. Freitas</i>	

Clustering Large, Multi-level Data Sets: An Approach Based on Kohonen Self Organizing Maps	353
<i>Antonio Ciampi and Yves Lechevallier</i>	
Trees and Induction Graphs for Multivariate Response	359
<i>Antonio Ciampi, Djamel A. Zighed, and J��r��my Clech</i>	
CEM - Visualisation and Discovery in Email	367
<i>Richard Cole, Peter Eklund and Gerd Stumme</i>	
Image Access and Data Mining: An Approach	375
<i>Chabane Djeraba</i>	
Decision Tree Toolkit: A Component-Based Library of Decision Tree Algorithms	381
<i>Nikos Drossos, Athanasios Papagelis, and Dimitris Kalles</i>	
Determination of Screening Descriptors for Chemical Reaction Databases .	388
<i>Laurent Dury, Laurence Leherter, and Daniel P. Vercauteren</i>	
Prior Knowledge in Economic Applications of Data Mining	395
<i>A.J. Feelders</i>	
Temporal Machine Learning for Switching Control	401
<i>Pierre Geurts and Louis Wehenkel</i>	
Improving Dissimilarity Functions with Domain Knowledge, Applications with IKBS System	409
<i>David Grosser, Jean Diatta, and No��l Conruyt</i>	
Mining Weighted Association Rules for Fuzzy Quantitative Items	416
<i>Attila Gyenesei</i>	
Centroid-Based Document Classification: Analysis and Experimental Results	424
<i>Eui-Hong (Sam) Han and George Karypis</i>	
Applying Objective Interestingness Measures in Data Mining Systems	432
<i>Robert J. Hilderman and Howard J. Hamilton</i>	
Observational Logic Integrates Data Mining Based on Statistics and Neural Networks	440
<i>Martin Hole��a</i>	
Supporting Discovery in Medicine by Association Rule Mining of Bibliographic Databases	446
<i>Dimitar Hristovski, Sa��o D��zeroski, Borut Peterlin, and Anamarija Rozic-Hristovski</i>	

Collective Principal Component Analysis from Distributed, Heterogeneous Data	452
<i>Hillol Kargupta, Weiyun Huang, Krishnamoorthy Sivakumar, Byung-Hoon Park, and Shuren Wang</i>	
Hierarchical Document Clustering Based on Tolerance Rough Set Model ..	458
<i>Saori Kawasaki, Ngoc Binh Nguyen, and Tu Bao Ho</i>	
Application of Data-Mining and Knowledge Discovery in Automotive Data Engineering	464
<i>Jörg Keller, Valerij Bauer, and Wojciech Kwedlo</i>	
Towards Knowledge Discovery from cDNA Microarray Gene Expression Data	470
<i>Jan Komorowski, Torgeir R. Hvidsten, Tor-Kristian Jenssen, Dyre Tjeldvoll, Eivind Hovig, Arne K. Sanvik, and Astrid Lægreid</i>	
Mining with Cover and Extension Operators	476
<i>Marzena Kryszkiewicz</i>	
A User-Driven Process for Mining Association Rules	483
<i>Pascale Kuntz, Fabrice Guillet, Rémi Lehn, and Henri Briand</i>	
Learning from Labeled and Unlabeled Documents: A Comparative Study on Semi-Supervised Text Classification	490
<i>Carsten Lanquillon</i>	
Schema Mining: Finding Structural Regularity among Semistructured Data	498
<i>P.A. Laur, F. Masseglia, and P. Poncelet</i>	
Improving an Association Rule Based Classifier	504
<i>Bing Liu, Yiming Ma, and Ching Kian Wong</i>	
Discovery of Generalized Association Rules with Multiple Minimum Supports	510
<i>Chung-Leung Lui and Fu-Lai Chung</i>	
Learning Dynamic Bayesian Belief Networks Using Conditional Phase-Type Distributions	516
<i>Adele Marshall, Sally McClean, Mary Shapcott, and Peter Millard</i>	
Discovering Differences in Patients with Uveitis through Typical Testors by Class	524
<i>José F. Martínez-Trinidad, Miriam Velasco-Sánchez, and Edgar E. Contreras-Aravelo</i>	

Web Usage Mining: How to Efficiently Manage New Transactions and New Clients	530
<i>F. Masseglia, P. Poncelet, and M. Teisseire</i>	
Mining Relational Databases	536
<i>Frédéric Moal, Teddy Turmeaux, and Christel Vrain</i>	
Interestingness in Attribute-Oriented Induction (AOI): Multiple-Level Rule Generation	542
<i>Maybin K. Muyeba and John A. Keane</i>	
Discovery of Characteristic Subgraph Patterns Using Relative Indexing and the Cascade Model	550
<i>Takashi Okada and Mayumi Oyama</i>	
Transparency and Predictive Power: Explaining Complex Classification Models	558
<i>Gerhard Paass and Jörg Kindermann</i>	
Clustering Distributed Homogeneous Datasets	566
<i>Srinivasan Parthasarathy and Mitsunori Ogihara</i>	
Empirical Evaluation of Feature Subset Selection Based on a Real-World Data Set	575
<i>Petra Perner and Chid Apte</i>	
Discovery of Ambiguous Patterns in Sequences: Application to Bioinformatics	581
<i>Gerard Ramstein, Pascal Bunelle, and Yannick Jacques</i>	
Action-Rules: How to Increase Profit of a Company	587
<i>Zbigniew W. Ras and Alicja Wiczorkowska</i>	
Aggregation and Association in Cross Tables	593
<i>Gilbert Ritschard and Nicolas Nicoloyannis</i>	
An Experimental Study of Partition Quality Indices in Clustering	599
<i>Céline Robardet, Fabien Feschet, and Nicolas Nicoloyannis</i>	
Expert Constrained Clustering: A Symbolic Approach	605
<i>Fabrice Rossi and Frédérick Vautrain</i>	
An Application of Association Rules Discovery to Geographic Information Systems	613
<i>Ansaf Salleb and Christel Vrain</i>	
Generalized Entropy and Projection Clustering of Categorical Data	619
<i>Dan A. Simovici, Dana Cristofor, and Laurentiu Cristofor</i>	