

# Lecture Notes in Artificial Intelligence

1805

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*

Takao Terano   Huan Liu  
Arbee L.P. Chen   (Eds.)

# Knowledge Discovery and Data Mining

Current Issues and New Applications

4th Pacific-Asia Conference, PAKDD 2000  
Kyoto, Japan, April 18-20, 2000  
Proceedings



Springer

## Series Editors

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

## Volume Editors

Takao Terano  
University of Tsukuba  
Graduate School of Systems Management  
3-29-1 Otsuka, Bunkyo-ku, Tokyo 112-0012, Japan  
E-mail: terano@gssm.otsuka.tsukuba.ac.jp

Huan Liu  
Arizona State University  
Department of Computer Science and Engineering  
P.O. Box 875 406, Tempe, AZ, 85287-5406  
E-mail: hliu@asu.edu

Arbee L.P. Chen  
National Tsing Hua University  
Department of Computer Science  
Hsinchu, Taiwan 300, ROC  
E-mail: alpchen@cs.nthu.edu.tw

## Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Knowledge discovery and data mining : current issues and new applications ; 4th Pacific Asia conference ; proceedings / PAKDD 2000, Kyoto, Japan, April 18 - 20, 2000. Takao Terano ... (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2000  
(Lecture notes in computer science ; Vol. 1805 : Lecture notes in artificial intelligence)  
ISBN 3-540-67382-2

CR Subject Classification (1991): I.2, H.3, H.5.1, G.3, J.1

ISBN 3-540-67382-2 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 Steingraber Satztechnik GmbH, Heidelberg  
Printed on acid-free paper      SPIN 10720199      06/3142      5 4 3 2 1 0

# Preface

The Fourth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2000) was held at the Keihanna-Plaza, Kyoto, Japan, April 18 – 20, 2000. PAKDD 2000 provided an international forum for researchers and application developers to share their original research results and practical development experiences. A wide range of current KDD topics were covered including machine learning, databases, statistics, knowledge acquisition, data visualization, knowledge-based systems, soft computing, and high performance computing. It followed the success of PAKDD 97 in Singapore, PAKDD 98 in Australia, and PAKDD 99 in China by bringing together participants from universities, industry, and government from all over the world to exchange problems and challenges and to disseminate the recently developed KDD techniques.

This PAKDD 2000 proceedings volume addresses both current issues and novel approaches in regards to theory, methodology, and real world application. The technical sessions were organized according to subtopics such as Data Mining Theory, Feature Selection and Transformation, Clustering, Application of Data Mining, Association Rules, Induction, Text Mining, Web and Graph Mining. Of the 116 worldwide submissions, 33 regular papers and 16 short papers were accepted for presentation at the conference and included in this volume. Each submission was critically reviewed by two to four program committee members based on their relevance, originality, quality, and clarity.

The PAKDD 2000 program was enhanced by two keynote speeches and one invited talk: Yoshiharu Sato of Hokkaido University, Japan (Statistics), Ryszard S. Michalski of George Mason University, USA (Machine Learning), and Andrew Tomkins of IBM Almaden Research Center (Databases). The PAKDD 2000 program was further complemented by five tutorials: Enterprise Data Mining with Case Studies (Zhexue Huang and Graham J Williams), Data Mining with Decision Trees (Johannes Gehrke), Knowledge Extraction from Texts - Application to Human Resources in Industry (Yves Kodratoff), Rough Sets in KDD: A Tutorial (Andrzej Skowron and Ning Zhong), and Data Mining on the World Wide Web (Wee-Keong NG). Two international workshops were co-hosted with PAKDD 2000 focusing on two KDD frontiers: the International Workshop on Web Knowledge Discovery and Data Mining (WKDDM 2000) and the International Workshop of KDD Challenge on Real-World Data (KDD Challenge 2000).

The success of PAKDD 2000 would not have been possible without the generous help rendered to us. We would like to extend our heartfelt gratitude to the program committee members and the steering committee members for their invaluable contributions. Special thanks go to the conference chairs: Hiroshi Motoda and Masaru Kitsuregawa for their leadership and involvement in making the conference run smoothly. We would like to express our immense gratitude to all the contributors to the conference for submitting and presenting papers, offering tutorials, giving talks, and organizing workshops. Special thanks are

due to Yuko Ichiki of Keihanna-Plaza Co. for her excellent secretarial work. The conference was sponsored by various academic societies in Japan. These are the Japanese Society of Artificial Intelligence (JSAI), SIG-DE (Data Engineering) and SIG-AI (Artificial Intelligence) of The Institute of Electronics, Information and Communication Engineers, SIG-DM (Data Mining) of the Japan Society for Software Science and Technology, SIG-DB (Data Base) and SIG-ICS (Intelligent and Complex Systems) of the Information Processing Society of Japan, and ACM SIG-MOD Japan. PAKDD 2000 was also generously supported by the SAS Institute, Japan and the Telecommunication Advancement Foundation (TAF), Japan.

We hope all participants had a pleasant stay at PAKDD 2000 as well as in Japan, exchanged refreshing views, and we wish them great success in their KDD endeavors.

April 2000

Takao Terano, Huan Liu, and Arbee L. P. Chen

# PAKDD 2000 Conference Committee

## *Conference Chairs:*

Masaru Kitsuregawa, University of Tokyo, Japan  
Hiroshi Motoda, Osaka University, Japan

## *Program Chairs:*

Takao Terano, Tsukuba University, Japan (Chair)  
Huan Liu, Arizona State University, USA (Co-chair)  
Arbee L. P. Chen, National Tsing Hua University, Taiwan (Co-chair)

## *Publicity Chair:*

Shinichi Morishita, University of Tokyo, Japan

## *Workshop Chair:*

Takahira Yamaguchi, Shizuoka University, Japan

## *Tutorial Chair:*

Shusaku Tsumoto, Shimane Medical University, Japan

## *Local Organizing Committee Chair:*

Takashi Washio, Osaka University, Japan

## *PAKDD Steering Committee:*

Xindong Wu,	Colorado School of Mines, USA (Chair)
Hongjun Lu,	National University of Singapore (Co-chair)
Ramamohanarao Kotagiri,	University of Melbourne, Australia
Huan Liu,	Arizona State University, USA
Hiroshi Motoda,	Osaka University, Japan
Lizhu Zhou,	Tsinghua University, China
Ning Zhong,	Yamaguchi University, Japan

## PAKDD 2000 Program Committee

Akinori Abe	NTT Communication Science Laboratories, Japan
Tatsuya Akutsu	University of Tokyo, Japan
Hiroki Arimura	Kyushuu University, Japan
Ming-Syan Chen	National Taiwan University, Taiwan
David Cheung	Hong Kong University, Hong Kong
Vic Ciesielski	RMIT, Australia
Honghua Dai	Deakin University, Australia
Usama Fayyad	Microsoft Research, USA
Takeshi Fukuda	IBM, Tokyo Research Laboratory, Japan
Yike Guo	Imperial College, UK
Jiawei Han	Simon Fraser University, Canada
Tomoyuki Higuchi	Institute of Statistical Mathematics, Japan
Tu Bao Ho	Japan Advanced Institute of Science and Technolo
Howard Ho	IBM, Almaden Research Center, USA
Xiaohua Hu	Knowledge Stream, USA
Zhexue Huang	MIP, Australia
Seiji Isobe	NTT Software Corporation, Japan
Manabu Iwasaki	Seikei University, Japan
Won-Chul Jhee	Hong-Ik University, Korea
Hiroyuki Kawano	Kyoto University, Japan
Kyung-Chang Kim	Hong-Ik University, Korea
Kevin Korb	Monash University, Australia
Ramamohanarao Kotagiri	University of Melbourne, Australia
Deyi Li	Beijing System Engineering Institute, China
Jianmin Li	University of Illinois at Chicago, USA
T.Y. Lin	San Jose State University, USA
Charles X. Ling	Univ. Western Ontario, Canada
Bing Liu	National University of Singapore
Chunnian Liu	Beijing Polytechnic University, China
Jiming Liu	Hong Kong Baptist University
Hongjun Lu	National University of Singapore
Yasuhiko Morimoto	IBM, Tokyo Research Laboratory, Japan
Shinichi Morishita	University of Tokyo, Japan
Masayuki Numao	Tokyo Institute of Technology, Japan
Tadashi Ohmori	University of Electro-Communications, Japan
Yukio Ohsawa	Tsukuba University, Japan
Gregory Piatetsky-Shapiro	Knowledge Stream, USA
Mohamed Quafafou	University of Nantes, France
Zbigniew W. Ras	University of North Carolina, USA
Keun Ho Ryu	Chungbuk National University, Korea



Peter Scheuermann	Northwestern University, USA
John C. Shafer	IBM, Almaden Research Center, USA
Arun Sharma	University of New South Wales, Australia
Zhongzhi Shi	Chinese Academy of Sciences, China
Arul Siromoney	Anna University, India
Andrzej Skowron	Warsaw University, Poland
Wataru Sunayama	Osaka University, Japan
Einoshin Suzuki	Yokohama National University, Japan
Atsuhiko Takasu	NACSIS, Japan
Shiro Takata	ATR, Japan
Bhavani Thuraisingham	MITRE Corporation, USA
Kai Ming Ting	Deakin University, Australia
Hiroshi Tsukimoto	Toshiba Corporation, Japan
Shusaku Tsumoto	Shimane Medical University
Jeffrey D. Ullman	Stanford University
Lipo Wang	Nanyang Technical University, Singapore
Takashi Washio	Osaka University, Japan
Graham Williams	CSIRO, Australia
Xindong Wu	Colorado School of Mines, USA
Beat Wuthrich	Hong Kong University of Science & Technology
Takahira Yamaguchi	Shizuoka University, Japan
Yiyu Yao	University of Regina, Canada
Suk-Chung Yoon	Widener University, USA
Tetsuya Yoshida	Osaka University, Japan
Meide Zhao	University of Illinois at Chicago, USA
Zijian Zheng	Blue Martini Software, USA
Aoying Zhou	Fudan University, China
Ning Zhong	Yamaguchi University, Japan
Lizhu Zhou	Tsinghua University, China
Jan Zytkow	University of North Carolina, USA

## PAKDD 2000 Reviewers

Akinori Abe	Akiko Aizawa	Tatsuya Akutsu
Hiroki Arimura	Ming-Syan Chen	David Cheung
Vic Ciesielski	Honghua Dai	Manoranjan Dash
Yi Du	Usama Fayyad	Takeshi Fukuda
Xiujun Gong	Yike Guo	Jiawei Han
Hisaaki Hatano	Lawrence. J. Henschen	Tomoyuki Higuchi
Tu Bao Ho	Howard Ho	Charles Hu
Xiaohua Hu	Zhexhe Huang	Md. Farhad Hussain
Tetsuya Iizuka	Seiji Isobe	Manabu Iwasaki
Won-Chul Jhee	Rong Jiang	Hiroyuki Kawano
Kyung-Chang Kim	Masaru Kitsuregawa	Kevin Korb
Ramamohanarao Kotagiri	Doheon Lee	Deyi Li
Jianmin Li	Xiaoli Li	T.Y. Lin
Charles X. Ling	Bing Liu	Chunnian Liu
Huan Liu	Jimin Liu	Jiming Liu
Oscar Ortega Lobo	Hongjun Lu	Yasuhiko Morimoto
Shinichi Morishita	Chie Morita	Koichi Moriyama
Hiroshi Motoda	Svetlozar Nestorov	N. Binh Nguyen
T. N. Nguyen	Masayuki Numao	Kensuke Ohmori
Tadashi Ohmori	Yukio Ohsawa	Wen Pen
Gregory Piatetsky-Shapiro	Mohamed Quafafou	Zbigniew W. Ras
Keun Ho Ryu	Makoto Sato	Peter Scheuermann
John C. Shafer	Arun Sharma	John Shepherd
Zhongzhi Shi	Hisako Shiohara	Arul Siromoney
Andrzej Skowron	Wataru Sunayama	Einoshin Suzuki
Atsuhiko Takasu	Takao Terano	Bhavani Thuraisingham
Kai Ming Ting	Hiroshi Tsukimoto	Shusaku Tsumoto
Wisut Sae-Tung	Jeffrey D. Ullman	Lipo Wang
Takashi Washio	Alicja Wieczorkowska	Graham Williams
J. Wroblewski	Xindong Wu	Beat Wuthrich
Takahira Yamaguchi	Yiyu Yao	Shiren Ye
Suk-Chung Yoon	Mariko Yoshida	Tetsuya Yoshida
Xiangtao You	Meide Zhao	Zijian Zheng
Ning Zhong	Aoying Zhou	Lizhu Zhou
Jan Zytkow		

# Table of Contents

## Keynote Speeches and Invited Talk

Perspective on Data Mining from Statistical Viewpoints . . . . .	1
<i>Yoshiharu Sato</i>	
Inductive Databases and Knowledge Scouts . . . . .	2
<i>Ryszard S. Michalski</i>	
Hyperlink-Aware Mining and Analysis of the Web . . . . .	4
<i>Andrew Tomkins</i>	

## Data Mining Theory

Polynomial Time Matching Algorithms for Tree-Like Structured Patterns in Knowledge Discovery . . . . .	5
<i>Tetsuhiro Miyahara, Takayoshi Shoudai, Tomoyuki Uchida, Kenichi Takahashi, Hiroaki Ueda</i>	
Fast Discovery of Interesting Rules . . . . .	17
<i>Nobuhiro Yugami, Yuiko Ohta, Seishi Okamoto</i>	
Performance Controlled Data Reduction for Knowledge Discovery in Distributed Databases . . . . .	29
<i>Slobodan Vucetic, Zoran Obradovic</i>	
Minimum Message Length Criterion for Second-Order Polynomial Model Discovery . . . . .	40
<i>Grace W. Rumantir</i>	
Frequent Itemset Counting Across Multiple Tables . . . . .	49
<i>Viviane Crestana Jensen, Nandit Soparkar</i>	
Frequent Closures as a Concise Representation for Binary Data Mining . . .	62
<i>Jean-François Boulicaut, Artur Bykowski</i>	
An Optimization Problem in Data Cube System Design . . . . .	74
<i>Edward Hung, David W. Cheung, Ben Kao, Yilong Liang</i>	
Exception Rule Mining with a Relative Interestingness Measure . . . . .	86
<i>Farhad Hussain, Huan Liu, Einoshin Suzuki, Hongjun Lu</i>	

## Feature Selection and Transformation

Consistency Based Feature Selection . . . . .	98
<i>Manoranjan Dash, Huan Liu, Hiroshi Motoda</i>	

Feature Selection for Clustering .....	110
<i>Manoranjan Dash, Huan Liu</i>	
A Simple Dimensionality Reduction Technique for Fast Similarity Search in Large Time Series Databases .....	122
<i>Eamonn J. Keogh, Michael J. Pazzani</i>	
Missing Value Estimation Based on Dynamic Attribute Selection .....	134
<i>K. C. Lee, J. S. Park, Y. S. Kim, Y. T. Byun</i>	
On Association, Similarity and Dependency of Attributes .....	138
<i>Yi Yu Yao, Ning Zhong</i>	
<b>Clustering</b>	
Prototype Generation Based on Instance Filtering and Averaging .....	142
<i>Chi-Kin Keung, Wai Lam</i>	
A Visual Method of Cluster Validation with Fastmap .....	153
<i>Zhexue Huang, Tao Lin</i>	
COE: Clustering with Obstacles Entities. A Preliminary Study .....	165
<i>Anthony K.H. Tung, Jean Hou, Jiawei Han</i>	
Combining Sampling Technique with DBSCAN Algorithm for Clustering Large Spatial Databases .....	169
<i>Shuigeng Zhou, Aoying Zhou, Jing Cao, Jin Wen, Ye Fan, Yunfa Hu</i>	
Predictive Adaptive Resonance Theory and Knowledge Discovery in Databases .....	173
<i>Ah-Hwee Tan, Hui-Shin Vivien Soon</i>	
Improving Generalization Ability of Self-Generating Neural Networks Through Ensemble Averaging .....	177
<i>Hiroataka Inoue, Hiroyuki Narihisa</i>	
<b>Application of Data Mining</b>	
Attribute Transformations on Numerical Databases .....	181
<i>Tsau Young Lin, Joseph Tremba</i>	
Efficient Detection of Local Interactions in the Cascade Model .....	193
<i>Takashi Okada</i>	
Extracting Predictors of Corporate Bankruptcy: Empirical Study on Data Mining Methods .....	204
<i>Cindy Yoshiko Shirata, Takao Terano</i>	

Evaluating Hypothesis-Driven Exception-Rule Discovery with Medical Data Sets .....	208
<i>Einoshin Suzuki, Shusaku Tsumoto</i>	
Discovering Protein Functional Models Using Inductive Logic Programming .....	212
<i>Takashi Ishikawa, Masayuki Numao, Takao Terano</i>	
Mining Web Transaction Patterns in an Electronic Commerce Environment .....	216
<i>Ching-Huang Yun, Ming-Syan Chen</i>	
<b>Association Rules and Related Topics</b>	
Making Use of the Most Expressive Jumping Emerging Patterns for Classification .....	220
<i>Jinyan Li, Guozhu Dong, Ramamohanarao Kotagiri</i>	
Mining Structured Association Patterns from Databases .....	233
<i>Hirofumi Matsuzawa, Takeshi Fukuda</i>	
Association Rules .....	245
<i>Tao Zhang</i>	
Density-Based Mining of Quantitative Association Rules .....	257
<i>David W. Cheung, Lian Wang, S.M. Yiu, Bo Zhou</i>	
AViz: A Visualization System for Discovering Numeric Association Rules ..	269
<i>Jiancho Han, Nick Cercone</i>	
Discovering Unordered and Ordered Phrase Association Patterns for Text Mining .....	281
<i>Ryoichi Fujino, Hiroki Arimura, Setsuo Arikawa</i>	
Using <i>Random Walks</i> for Mining Web Document Associations .....	294
<i>K. Selçuk Candan, Wen-Syan Li</i>	

## Induction

A Concurrent Approach to the Key-Preserving Attribute-Oriented Induction Method .....	306
<i>Maybin K. Muyebe, John A. Keane</i>	
Scaling Up a Boosting-Based Learner via Adaptive Sampling .....	317
<i>Carlos Domingo, Osamu Watanabe</i>	
Adaptive Boosting for Spatial Functions with Unstable Driving Attributes	329
<i>Aleksandar Lazarevic, Tim Fiez, Zoran Obradovic</i>	

Robust Ensemble Learning for Data Mining .....	341
<i>Gunnar Rätsch, Bernhard Schölkopf, Alexander Johannes Smola, Sebastian Mika, Takashi Onoda, Klaus-Robert Müller</i>	
Interactive Visualization in Mining Large Decision Trees .....	345
<i>Trong Dung Nguyen, Tu Bao Ho, Hiroshi Shimodaira</i>	
VQTree: Vector Quantization for Decision Tree Induction .....	349
<i>Shlomo Geva, Lawrence Buckingham</i>	
Making Knowledge Extraction and Reasoning Closer .....	360
<i>Fosca Giannotti, Giuseppe Manco</i>	
Discovery of Relevant Weights by Minimizing Cross-Validation Error .....	372
<i>Kazumi Saito, Ryohei Nakano</i>	
Efficient and Comprehensible Local Regression .....	376
<i>Luís Torgo</i>	
Information Granules for Spatial Reasoning .....	380
<i>Andrzej Skowron, Jaroslaw Stepaniuk, Shusaku Tsumoto</i>	
<b>Text, Web, and Graph Mining</b>	
Uncovering the Hierarchical Structure of Text Archives by Using an Unsupervised Neural Network with Adaptive Architecture .....	384
<i>Dieter Merkl, Andreas Rauber</i>	
Mining Access Patterns Efficiently from Web Logs .....	396
<i>Jian Pei, Jiawei Han, Behzad Mortazavi-ast, Hua Zhu</i>	
A Comparative Study of Classification Based Personal E-mail Filtering ...	408
<i>Yanlei Diao, Hongjun Lu, Dekai Wu</i>	
Extension of Graph-Based Induction for General Graph Structured Data ..	420
<i>Takashi Matsuda, Tadashi Horiuchi, Hiroshi Motoda, Takashi Washio</i>	
Text-Source Discovery and GLOSS Update in a Dynamic Web .....	432
<i>Chi-Yuen Ng, Ben Kao, David Cheung</i>	
Extraction of Fuzzy Clusters from Weighted Graphs .....	442
<i>Seiji Hotta, Kohei Inoue, Kiichi Urahama</i>	
Text Summarization by Sentence Segment Extraction Using Machine Learning Algorithms .....	454
<i>Wesley T. Chuang, Jihoon Yang</i>	
<b>Author Index .....</b>	<b>459</b>