

David Cheung
Graham J. Williams
Qing Li (Eds.)

LNAI 2035

Advances in Knowledge Discovery and Data Mining

5th Pacific-Asia Conference, PAKDD 2001
Hong Kong, China, April 2001
Proceedings



Springer

David Cheung Graham J. Williams
Qing Li (Eds.)

Advances in Knowledge Discovery and Data Mining

5th Pacific-Asia Conference, PAKDD 2001
Hong Kong, China, April 16-18, 2001
Proceedings



Springer

Series Editors

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

Volume Editors

David Cheung

The University of Hong Kong, Dept. of Computer Science and Information Systems
Pokfulam, Hong Kong, China

E-mail: dcheung@csis.hku.hk

Graham J. Williams

CSIRO Mathematical and Information Sciences

GPO Box 664, Canberra, ACT 2601, Australia

E-mail: Graham.Williams@cmis.csiro.au

Qing Li

City University of Hong Kong, Department of Computer Science

83 Tat Chee Ave., Kowloon, Hong Kong, China

E-mail: csqli@cityu.edu.hk

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Advances in knowledge discovery and data mining : 5th Pacific Asia
conference ; proceedings / PAKDD 2001, Hong Kong, China, April 16 -
18, 2001. David Cheung ... (ed.). - Berlin ; Heidelberg ; New York ;
Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo :
Springer, 2001

(Lecture notes in computer science ; Vol. 2035 : Lecture notes in
artificial intelligence)

ISBN 3-540-41910-1

CR Subject Classification (1998): I.2, H.3, H.5.1, G.3, J.1, K.4

ISBN 3-540-41910-1 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

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2001

Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Stefan Sossna

Printed on acid-free paper SPIN: 10782484 06/3142 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence

2035

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

Lecture Notes in Artificial Intelligence (LNAI)

- 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.
- Vol. 1916: F. Dignum, M. Greaves (Eds.), *Issues in Agent Communication. X*, 351 pages. 2000.
- Vol. 1919: M. Ojeda-Aciego, I.P. de Guzman, G. Brewka, L. Moniz Pereira (Eds.), *Logics in Artificial Intelligence. Proceedings, 2000. XI*, 407 pages. 2000.
- Vol. 1925: J. Cussens, S. Džeroski (Eds.), *Learning Language in Logic. X*, 301 pages. 2000.
- Vol. 1932: Z.W. Raś, S. Ohsuga (Eds.), *Foundations of Intelligent Systems. Proceedings, 2000. XII*, 646 pages.
- Vol. 1934: J.S. White (Ed.), *Envisioning Machuine Translation in the Information Future. Proceedings, 2000. XV*, 254 pages. 2000.
- Vol. 1937: R. Dieng, O. Corby (Eds.), *Knowledge Engineering and Knowledge Management. Proceedings, 2000. XIII*, 457 pages. 2000.
- Vol. 1952: M.C. Monard, J. Simão Sichman (Eds.), *Advances in Artificial Intelligence. Proceedings, 2000. XV*, 498 pages. 2000.
- Vol. 1955: M. Parigot, A. Voronkov (Eds.), *Logic for Programming and Automated Reasoning. Proceedings, 2000. XIII*, 487 pages. 2000.
- Vol. 1967: S. Arikawa, S. Morishita (Eds.), *Discovery Science. Proceedings, 2000. XII*, 332 pages. 2000.
- Vol. 1968: H. Arimura, S. Jain, A. Sharma (Eds.), *Algorithmic Learning Theory. Proceedings, 2000. XI*, 335 pages. 2000.
- Vol. 1972: A. Omicini, R. Tolksdorf, F. Zambonelli (Eds.), *Engineering Societies in the Agents World. Proceedings, 2000. IX*, 143 pages. 2000.
- Vol. 1979: S. Moss, P. Davidsson (Eds.), *Multi-Agent-Based Simulation. Proceedings, 2000. VIII*, 267 pages. 2001.
- Vol. 1991: F. Dignum, C. Sierra (Eds.), *Agent Mediated Electronic Commerce. VIII*, 241 pages. 2001.
- Vol. 2003: F. Dignum, U. Cortés (Eds.), *Agent-Mediated Electronic Commerce III. XII*, 193 pages. 2001.
- Vol. 2007: J.F. Roddick, K. Hornsby (Eds.), *Temporal, Spatial, and Spatio-Temporal Data Mining. Proceedings, 2000. VII*, 165 pages. 2001.
- Vol. 2035: D. Cheung, G.J. Williams, Q. Li (Eds.), *Advances in Knowledge Discovery and Data Mining – PAKDD 2001. Proceedings, 2001. XVIII*, 596 pages. 2001.

Lecture Notes in Computer Science

- Vol. 1979: S. Moss, P. Davidsson (Eds.), Multi-Agent-Based Simulation. Proceedings, 2000. VIII, 267 pages. 2001. (Subseries LNAI).
- Vol. 1983: K.S. Leung, L.-W. Chan, H. Meng (Eds.), Intelligent Data Engineering and Automated Learning – IDEAL 2000. Proceedings, 2000. XVI, 573 pages. 2000.
- Vol. 1984: J. Marks (Ed.), Graph Drawing. Proceedings, 2000. XII, 419 pages. 2001.
- Vol. 1985: J. Davidson, S.L. Min (Eds.), Languages, Compilers, and Tools for Embedded Systems. Proceedings, 2000. VIII, 221 pages. 2001.
- Vol. 1987: K.-L. Tan, M.J. Franklin, J. C.-S. Lui (Eds.), Mobile Data Management. Proceedings, 2001. XIII, 289 pages. 2001.
- Vol. 1988: L. Vulkov, J. Waśniewski, P. Yalamov (Eds.), Numerical Analysis and Its Applications. Proceedings, 2000. XIII, 782 pages. 2001.
- Vol. 1989: M. Ajmone Marsan, A. Bianco (Eds.), Quality of Service in Multimedia IP Networks. Proceedings, 2001. XII, 440 pages. 2001.
- Vol. 1990: I.V. Ramakrishnan (Ed.), Practical Aspects of Declarative Languages. Proceedings, 2001. VIII, 353 pages. 2001.
- Vol. 1991: F. Dignum, C. Sierra (Eds.), Agent Mediated Electronic Commerce. VIII, 241 pages. 2001. (Subseries LNAI).
- Vol. 1992: K. Kim (Ed.), Public Key Cryptography. Proceedings, 2001. XI, 423 pages. 2001.
- Vol. 1993: E. Zitzler, K. Deb, L. Thiele, C.A. Coello Coello, D. Corne (Eds.), Evolutionary Multi-Criterion Optimization. Proceedings, 2001. XIII, 712 pages. 2001.
- Vol. 1995: M. Sloman, J. Lobo, E.C. Lupu (Eds.), Policies for Distributed Systems and Networks. Proceedings, 2001. X, 263 pages. 2001.
- Vol. 1997: D. Suciu, G. Vossen (Eds.), The World Wide Web and Databases. Proceedings, 2000. XII, 275 pages. 2001.
- Vol. 1998: R. Klette, S. Peleg, G. Sommer (Eds.), Robot Vision. Proceedings, 2001. IX, 285 pages. 2001.
- Vol. 1999: W. Emmerich, S. Tai (Eds.), Engineering Distributed Objects. Proceedings, 2000. VIII, 271 pages. 2001.
- Vol. 2000: R. Wilhelm (Ed.), Informatics: 10 Years Back, 10 Years Ahead. IX, 369 pages. 2001.
- Vol. 2003: F. Dignum, U. Cortés (Eds.), Agent Mediated Electronic Commerce III. XII, 193 pages. 2001. (Subseries LNAI).
- Vol. 2004: A. Gelbukh (Ed.), Computational Linguistics and Intelligent Text Processing. Proceedings, 2001. XII, 528 pages. 2001.
- Vol. 2006: R. Dunke, A. Abran (Eds.), New Approaches in Software Measurement. Proceedings, 2000. VIII, 245 pages. 2001.
- Vol. 2007: J.F. Roddick, K. Hornsby (Eds.), Temporal, Spatial, and Spatio-Temporal Data Mining. Proceedings, 2000. VII, 165 pages. 2001. (Subseries LNAI).
- Vol. 2009: H. Federrath (Ed.), Designing Privacy Enhancing Technologies. Proceedings, 2000. X, 231 pages. 2001.
- Vol. 2010: A. Ferreira, H. Reichel (Eds.), STACS 2001. Proceedings, 2001. XV, 576 pages. 2001.
- Vol. 2013: S. Singh, N. Murshed, W. Kropatsch (Eds.), Advances in Pattern Recognition – ICAPR 2001. Proceedings, 2001. XIV, 476 pages. 2001.
- Vol. 2015: D. Won (Ed.), Information Security and Cryptology – ICISC 2000. Proceedings, 2000. X, 261 pages. 2001.
- Vol. 2018: M. Pollefeys, L. Van Gool, A. Zisserman, A. Fitzgibbon (Eds.), 3D Structure from Images – SMILE 2000. Proceedings, 2000. X, 243 pages. 2001.
- Vol. 2020: D. Naccache (Ed.), Progress in Cryptology – CT-RSA 2001. Proceedings, 2001. XII, 473 pages. 2001.
- Vol. 2021: J. N. Oliveira, P. Zave (Eds.), FME 2001: Formal Methods for Increasing Software Productivity. Proceedings, 2001. XIII, 629 pages. 2001.
- Vol. 2024: H. Kuchen, K. Ueda (Eds.), Functional and Logic Programming. Proceedings, 2001. X, 391 pages. 2001.
- Vol. 2027: R. Wilhelm (Ed.), Compiler Construction. Proceedings, 2001. XI, 371 pages. 2001.
- Vol. 2028: D. Sands (Ed.), Programming Languages and Systems. Proceedings, 2001. XIII, 433 pages. 2001.
- Vol. 2029: H. Hussmann (Ed.), Fundamental Approaches to Software Engineering. Proceedings, 2001. XIII, 349 pages. 2001.
- Vol. 2030: F. Honsell, M. Miculan (Eds.), Foundations of Software Science and Computation Structures. Proceedings, 2001. XII, 413 pages. 2001.
- Vol. 2031: T. Margaria, W. Yi (Eds.), Tools and Algorithms for the Construction and Analysis of Systems. Proceedings, 2001. XIV, 588 pages. 2001.
- Vol. 2034: M.D. Di Benedetto, A. Sangiovanni-Vincentelli (Eds.), Hybrid Systems: Computation and Control. Proceedings, 2001. XIV, 516 pages. 2001.
- Vol. 2035: D. Cheung, G.J. Williams, Q. Li (Eds.), Advances in Knowledge Discovery and Data Mining – PAKDD 2001. Proceedings, 2001. XVIII, 596 pages. 2001. (Subseries LNAI).
- Vol. 2038: J. Miller, M. Tomassini, P.L. Lanzi, C. Ryan, A.G.B. Tettamanzi, W.B. Langdon (Eds.), Genetic Programming. Proceedings, 2001. XI, 384 pages. 2001.

Preface

PAKDD 2001, Hong Kong, 16–18 April, was organized by the E-Business Technology Institute of The University of Hong Kong in cooperation with ACM Hong Kong, IEEE Hong Kong Chapter, and The Hong Kong Web Society. It was the Fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining and the successor of earlier PAKDD conferences held in Singapore (1997), Melbourne, Australia (1998), Beijing, China (1999), and Kyoto, Japan (2000).

PAKDD 2001 brought together participants from universities, industry, and government to present, discuss, and address both current issues and novel approaches in the practise, deployment, theory, and methodology of Knowledge Discovery and Data Mining. The conference provides an international forum for the sharing of original research results and practical development experiences among researchers and application developers from the many KDD related areas including machine learning, databases, statistics, internet, e-commerce, knowledge acquisition, data visualization, knowledge-based systems, soft computing, and high performance computing.

The PAKDD 2001 conference included technical sessions organized around important subtopics such as: Web Mining; Text Mining; Applications and Tools; Interestingness; Feature Selection; Sequence Mining; Spatial and Temporal Mining; Concept Hierarchies; Association Mining; Classification and Rule Induction; Clustering; and Advanced Topics and New Methods.

Following careful review of the 152 submissions by members of the international program committee 38 regular papers and 22 short papers were selected for presentation at the conference and for publication in this volume.

The conference program also included invited keynote presentations from three international researchers and developers in data mining: H. V. Jagadish of the University of Michigan, Ronny Kohavi of Blue Martini, and Hongjun Lu of the University of Science and Technology, Hong Kong. Abstracts of their presentations are included in this volume.

The conference presented six tutorials from experts in their respective disciplines: An Introduction to MARS (Dan Steinberg); Static and Dynamic Data Mining Using Advanced Machine Learning Methods (Ryszard S. Michalski); Sequential Pattern Mining: From Shopping History Analysis to Weblog Mining and DNA Mining (Jiawei Han and Jian Pei); Recent Advances in Data Mining Algorithms for Large Databases (Rajeev Rastogi and Kyuseok Shim); Web Mining for E-Commerce (Jaideep Srivastava); and From Evolving Single Neural Networks to Evolving Ensembles (Xin Yao).

Associated workshops included: Spatial and Temporal Data; Statistical Techniques in Data Mining; and Data Mining and Electronic Business.

A conference such as this can only succeed as a team effort. We would like to thank the program committee members and reviewers for their efforts and the PAKDD steering committee members for their invaluable input and advice. Our sincere gratitude goes to all of the authors who submitted papers. We are grateful to our sponsors for their generous support. Special thanks go to Ms Winnie Yau, E-Business Technology Institute, The University of Hong Kong, for her considerable efforts, seamlessly keeping everything running smoothly and coordinating the many streams of the conference organization.

On behalf of the organizing and program committees of PAKDD 2001 we trust you found the conference a fruitful experience and hope you had an enjoyable stay in Hong Kong.

February 2001

Chung-Jun Tan, Jiawei Han
David Cheung, Graham J. Williams, Qing Li

PAKDD 2001 Organization

PAKDD 2001 was organized by the E-Business Technology Institute of The University of Hong Kong in cooperation with ACM Hong Kong, IEEE Hong Kong Chapter, and The Hong Kong Web Society.

Conference Committee

Conference Chairs:	Chung-Jen Tan (University of Hong Kong and IBM Watson) Jiawei Han (Simon Fraser University, Canada)
Program Chairs:	David Cheung (University of Hong Kong) Graham J. Williams (CSIRO, Australia) Qing Li (City University of Hong Kong)
Tutorial Chair:	Joshua Huang (University of Hong Kong)
Workshop Chair:	Michael Ng (University of Hong Kong)
Industrial Chair:	Joseph Fong (City University of Hong Kong)
Demonstration Chair:	Jiming Liu (Baptist University of Hong Kong)
Local Arrangements Chairs:	Ronnie Cheung (Hong Kong Polytechnic University) Ben Kao (University of Hong Kong)
Publicity Chairs:	Vincent Ng (Hong Kong Polytechnic University) Rohan Baxter (CSIRO, Australia) Hiroyuki Kawano (Kyoto University, Japan)
Treasurer:	Ada Fu (Chinese University of Hong Kong)

PAKDD Steering Committee

Xindong Wu	Colorado School of Mines, USA (Chair)
Hongjun Lu	Hong Kong Univ. of Science & Technology (Co-chair)
Ramamohanarao Kotagiri	University of Melbourne, Australia
Huan Liu	National University of Singapore
Hiroshi Motoda	Osaka University, Japan
Lizhu Zhou	Tsinhua University, China
Ning Zhong	Yamaguchi University, Japan
Masaru Kitsuregawa	University of Tokyo, Japan
Takao Terano	University of Tsukuba, Japan

Program Committee

Rohan Baxter	CSIRO, Australia
Keith Chan	Hong Kong Polytechnic University
Surajit Chaudhuri	Microsoft, USA
Ming-Syan Chen	National Taiwan University
David Cheung	University of Hong Kong
Umeshwar Dayal	Hewlett-Packard Labs, USA
Tharam Dillon	Hong Kong Polytechnic University
Guozhu Dong	Wright State University, USA
David Dowe	Monash University, Australia
Ling Feng	Tilburg University, The Netherlands
Ada Fu	Chinese University of Hong Kong
Chi Wai Fung	VTC-IVE, Hong Kong
Yike Guo	Imperial College, UK
Howard Hamilton	University of Regina, Canada
Jayant Haritysa	Indian Institute of Science, India
Simon Hawkins	CSIRO, Australia
Markus Hegland	Australian National University
Robert Hilderman	University of Regina, Canada
Howard Ho	IBM Almaden, USA
Joshua Z. Huang	University of Hong Kong
Moon Yul Huh	SKK University, Korea
Vijay Iyengar	IBM T.J. Watson Research Center, USA
Ben Kao	University of Hong Kong
Kamal Karlapalem	Hong Kong UST
Hiroyuki Kawano	University of Kyoto, Japan
Jinho Kim	Kangwon National U., Korea
Sang-Wook Kim	Kangwon National University, Korea
Masaru Kitsuregawa	University of Tokyo, Japan
Kevin Korb	Monash University, Australia
Laks V. S. Lakshmanan	Concordia U. and IIT, Bombay
Wai Lam	Chinese University of Hong Kong
Doheon Lee	Chonnam National University, Korea
Hong-Va Leong	Hong Kong Polytechnic University
Deyi Li	Electronic System Eng. Inst., Beijing
Qing Li	City University of Hong Kong
T. Y. Lin	San Jose State University, USA
Charles Ling	University of Western Ontario, Canada
Bing Liu	National University of Singapore
Huan Liu	Arizona State University, USA
Jiming Liu	Hong Kong Baptist University
Hongjun Lu	Hong Kong UST

Hiroshi Motoda
Michael Ng
Raymond Ng
Vincent Ng
Kyuseok Shim
Il-Yeol Song
Ah-Hwee Tan
Changjie Tang
Zhaohui Tang
Takao Terano
Bhavani Thurasingham
Ke Wang
Kyu-Young Whang
Graham J. Williams
Ian Witten
Xindong Wu
Yiyu Yao
Siu Ming Yiu
Clement Yu
Jeffrey Yu
Philip Yu
Osamar R. Zaiane
Mohammed Zaki
Ning Zhong
Aoying Zhou
Lizhu Zhou

Osaka University, Japan
University of Hong Kong
UBC, Canada
Hong Kong Polytechnic University
KAIST, Korea
Drexel University, USA
Kent Ridge Digital Labs, Singapore
Sichuan University, China
Microsoft, USA
University of Tsukuba, Japan
MITRE, USA
Simon Fraser University, Canada
KAIST, Korea
CSIRO, Australia
University of Waikato, New Zealand
Colorado School of Mines, USA
University of Regina, Canada
University of Hong Kong
University of Illinois, USA
Chinese University of Hong Kong
IBM T.J. Watson Research Center, USA
University of Alberta, Canada
Rensselaer Poly. Institute, USA
Maebashi Institute of Technology
Fudan University, China
Tsinghua University, China

Reviewers

David Albrecht	Vijay Iyengar	Weining Qian
Lloyd Allison	He Ji	Kanagasabai Rajaraman
A. Ammoura	Wenyun Ji	Kyuseok Shim
Rohan Baxter	Ben Kao	Ho Wai Shing
Keith Chan	Kamal Karlapalem	Ho Kei Shiu
Surajit Chaudhuri	Hiroyuki Kawano	Il Yeol Song
Ming Syan Chen	Jeong Ja Kim	John Sum
David Cheung	Jinho Kim	Katsumi Takahashi
W. Cheung	Sang Wook Kim	Ah Hwee Tan
Yin Ling Cheung	Masaru Kitsuregawa	Changjie Tang
Vishal Chitkara	Kevin Korb	Zhaohui Tang
Moon Young Chung	Laks V.S. Lakshmanan	Takao Terano
Woong Kyo Chung	Wai Lam	Bhavani Thurasingham
Shi Chunyi	Yip Chi Lap	Masashi Toyoda
Hang Cui	Doheon Lee	Dean Trower
Gautam Das	Raymond Lee	Vassilios Verykios
Manoranjana Dash	Hong Va Leong	Murli Viswanathan
Umeshwar Dayal	Deyi Li	Dianlui Wang
Tharam Dillon	Qing Li	Haiyun Wang
Guozhu Dong	Seung Chun Li	Ke Wang
David Dowe	Jong Hwa Lim	Wei Wang
Leigh Fitzgibbon	T Y Lin	Weinan Wang
Ada Fu	Charles Ling	Kyu Young Whang
Chi Wai Fung	Feng Ling	Graham J. Williams
Venkatesh Ganti	Bing Liu	Ian Witten
Moustafa Ghanem	Huan Liu	Yong Gwan Won
Ishaan Goyal	James Liu	Xindong Wu
Baohua Gu	Jiming Liu	Jim Yang
Yike Guo	Hongjun Lu	Jiong Yang
M. El Hajj	Jinyoung Moon	Yiyu Yao
Howard Hamilton	Hiroshi Motoda	Siu Ming Yiu
Jayant Haritysa	Ka Ka Ng	Ji Hee Yoon
Simon Hawkins	Michael Ng	Clement Yu
Markus Hegland	Raymond Ng	Jeffrey Yu
Chun Hung Heng	Vincent Ng	Philip Yu
Robert Hilderman	Ann Nicholsons	Lui Yuchang
Howard Ho	Masato Oguchi	Osamar R. Zaiane
Kei Shiu Ho	Tadashi Ohmori	Mohammed Zaki
Eui Kyeong Hong	Andrew Paplinski	Shi Hui Zheng
Mintz Hsieh	Jong Soo Park	Ning Zhong
Joshua Z Huang	Sanghyun Park	Aoying Zhou
Moon Yul Huh	Jon Patrick	Lizhu Zhou
Dongjoon Hyun	Iko Pramudiono	

Sponsors

SAS

ETNet

Hong Kong Pei Hua Education Foundation

IEEE Computer Society, Hong Kong Section, Computer Chapter

ACM Hong Kong

Hong Kong Computer Society

Hong Kong Productivity Council

Hong Kong Web Society

E-Business Technology Institute, The University of Hong Kong

Department of Computer Science and Information Systems, The University of Hong Kong

Table of Contents

Keynote Presentations

Incompleteness in Data Mining	1
<i>Hosagrahar Visvesvaraya Jagadish</i>	
Mining E-Commerce Data: The Good, the Bad, and the Ugly	2
<i>Ronny Kohavi</i>	
Seamless Integration of Data Mining with DBMS and Applications	3
<i>Hongjun Lu</i>	

Web Mining

Applying Pattern Mining to Web Information Extraction	4
<i>Chia-Hui Chang, Shao-Chen Lui, Yen-Chin Wu</i>	
Empirical Study of Recommender Systems Using Linear Classifiers	16
<i>Vijay S. Iyengar, Tong Zhang</i>	
iJADE eMiner—A Web-Based Mining Agent Based on Intelligent Java Agent Development Environment (iJADE) on Internet Shopping	28
<i>Raymond S.T. Lee, James N.K. Liu</i>	
A Characterized Rating Recommend System	41
<i>Yao-Tsung Lin, Shian-Shyong Tseng</i>	
Discovery of Frequent Tree Structured Patterns in Semistructured Web Documents	47
<i>Tetsuhiro Miyahara, Takayoshi Shoudai, Tomoyuki Uchida, Kenichi Takahashi, Hiroaki Ueda</i>	

Text Mining

Text Categorization Using Weight Adjusted k -Nearest Neighbor Classification	53
<i>Eui-Hong Han, George Karypis, Vipin Kumar</i>	
Predictive Self-Organizing Networks for Text Categorization	66
<i>Ah-Hwee Tan</i>	
Meta-learning Models for Automatic Textual Document Categorization ...	78
<i>Kwok-Yin Lai, Wai Lam</i>	
Efficient Algorithms for Concept Space Construction	90
<i>Chi Yuen Ng, Joseph Lee, Felix Cheung, Ben Kao, David Cheung</i>	

Topic Detection, Tracking, and Trend Analysis Using Self-Organizing
Neural Networks 102
Kanagasabi Rajaraman, Ah-Hwee Tan

Automatic Hypertext Construction through a Text Mining Approach by
Self-Organizing Maps 108
Hsin-Chang Yang, Chung-Hong Lee

Applications and Tools

Semantic Expectation-Based Causation Knowledge Extraction: A Study
on Hong Kong Stock Movement Analysis..... 114
Boon-Toh Low, Ki Chan, Lei-Lei Choi, Man-Yee Chin, Sin-Ling Lay

A Toolbox Approach to Flexible and Efficient Data Mining..... 124
*Ole M. Nielsen, Peter Christen, Markus Hegland, Tatiana Semenova,
Timothy Hancock*

Determining Progression in Glaucoma Using Visual Fields 136
*Andrew Turpin, Eibe Frank, Mark Hall, Ian H. Witten,
Chris A. Johnson*

Seabreeze Prediction Using Bayesian Networks..... 148
Russell J. Kennett, Kevin B. Korb, Ann E. Nicholson

Semi-supervised Learning in Medical Image Database..... 154
Chun Hung Li, Pong Chi Yuen

On Application of Rough Data Mining Methods to Automatic
Construction of Student Models 161
Feng-Hsu Wang, Shiou-Wen Hung

Concept Hierarchies

Concept Approximation in Concept Lattice 167
Keyun Hu, Yuefei Sui, Yuchang Lu, Ju Wang, Chunyi Shi

Generating Concept Hierarchies/Networks: Mining Additional Semantics
in Relational Data 174
T.Y. Lin

Representing Large Concept Hierarchies Using Lattice Data Structure 186
Yanee Kachai, Kitsana Waiyamai

Feature Selection

Feature Selection for Temporal Health Records 198
Rohan A. Baxter, Graham J. Williams, Hongxing He

Boosting the Performance of Nearest Neighbour Methods with Feature Selection	210
<i>Shlomo Geva</i>	

Feature Selection for Meta-learning	222
<i>Alexandros Kalousis, Melanie Hilario</i>	

Interestingness

Efficient Mining of Niches and Set Routines	234
<i>Guozhu Dong, Kaustubh Deshpande</i>	

Evaluation of Interestingness Measures for Ranking Discovered Knowledge	247
<i>Robert J. Hilderman, Howard J. Hamilton</i>	

Peculiarity Oriented Mining and Its Application for Knowledge Discovery in Amino-Acid Data	260
<i>Ning Zhong, Muneaki Ohshima, Setsuo Ohsuga</i>	

Sequence Mining

Mining Sequence Patterns from Wind Tunnel Experimental Data for Flight Control	270
<i>Zhenyu Liu, Wesley W. Chu, Adam Huang, Chris Folk, Chih-Ming Ho</i>	

Scalable Hierarchical Clustering Method for Sequences of Categorical Values	282
<i>Tadeusz Morzy, Marek Wojciechowski, Maciej Zakrzewicz</i>	

FFS—An I/O-Efficient Algorithm for Mining Frequent Sequences	294
<i>Minghua Zhang, Ben Kao, Chi-Lap Yip, David Cheung</i>	

Sequential Index Structure for Content-Based Retrieval	306
<i>Maciej Zakrzewicz</i>	

Spatial and Temporal Mining

The S^2 -Tree: An Index Structure for Subsequence Matching of Spatial Objects	312
<i>Haixun Wang, Chang-Shin Perng</i>	

Temporal Data Mining Using Hidden Markov-Local Polynomial Models ...	324
<i>Weiqliang Lin, Mehmet A. Orgun, Graham J. Williams</i>	

Patterns Discovery Based on Time-Series Decomposition	336
<i>Jeffrey Xu Yu, Michael K. Ng, Joshua Zhexue Huang</i>	