Takashi Washio Zhi-Hua Zhou
Joshua Zhexue Huang Xiaohua Hu
Jinyan Li Chao Xie
Jieyue He Deqing Zou
Kuan-Ching Li Mário M. Freire (Eds.)

LNAI 4819

Emerging Technologies in Knowledge Discovery and Data Mining

PAKDD 2007 International Workshops
Nanjing, China, May 2007
Revised Selected Papers

Springer

TP18-53
K73
2007-2

Takashi Washio Zhi-Hua Zhou
Joshua Zhexue Huang Xiaohua Hu
Jinyan Li Chao Xie Jieyue He
Deqing Zou Kuan-Ching Li
Mário M. Freire (Eds.)

Emerging Technologies in Knowledge Discovery and Data Mining

PAKDD 2007 International Workshops
Nanjing, China, May 22, 2007
Revised Selected Papers

Springer

Series Editors

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

Volume Editors

Takashi Washio, Osaka University, Japan (washio@ar.sanken.osaka-u.ac.jp)

Zhi-Hua Zhou, Nanjing University, China (zhouzh@nju.edu.cn)

Joshua Zhexue Huang, The University of Hong Kong, China (jhuang@eti.hku.hk)

Xiaohua Hu, Drexel University, USA (thu@cis.drexel.edu)

Jinyan Li, Nanyang Technological University, Singapore (jyli@ntu.edu.sg)

Chao Xie, Georgia State University, USA (cxie@cs.gsu.edu)

Jieyue He, Southeast University, China (jieyuehe@seu.edu.cn)

Deqing Zou, Huazhong University of Science and Technology, China
(deqingzou@hust.edu.cn)

Kuan-Ching Li, Providence University, Taiwan (kuancli@pu.edu.tw)

Mário M. Freire, University of Beira Interior, Portugal (mario@di.ubi.pt)

Library of Congress Control Number: 2007941074

CR Subject Classification (1998): I.2, H.2.7-8, H.3, H.5.1, G.3, J.3

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743
ISBN-10 3-540-77016-X Springer Berlin Heidelberg New York
ISBN-13 978-3-540-77016-9 Springer 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. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media

springer.com

© Springer-Verlag Berlin Heidelberg 2007
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12197576 06/3180 5 4 3 2 1 0

Lecture Notes in Artificial Intelligence 4819

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Lecture Notes in Artificial Intelligence (LNAI)

Vol. 4874: J. Neves, M.F. Santos, J.M. Machado (Eds.), Progress in Artificial Intelligence. XVIII, 704 pages. 2007.

Vol. 4845: N. Zhong, J. Liu, Y. Yao, J. Wu, S. Lu, K. Li (Eds.), Web Intelligence Meets Brain Informatics. XI, 516 pages. 2007.

Vol. 4830: M.A. Orgun, J. Thornton (Eds.), AI 2007: Advances in Artificial Intelligence. XIX, 841 pages. 2007.

Vol. 4828: M. Randall, H.A. Abbass, J. Wiles (Eds.), Progress in Artificial Life. XII, 402 pages. 2007.

Vol. 4827: A. Gelbukh, Á.F. Kuri Morales (Eds.), MICAI 2007: Advances in Artificial Intelligence. XXIV, 1234 pages. 2007.

Vol. 4826: P. Perner, O. Salvetti (Eds.), Advances in Mass Data Analysis of Signals and Images in Medicine, Biotechnology and Chemistry. X, 183 pages. 2007.

Vol. 4819: T. Washio, Z.-H. Zhou, J.Z. Huang, X. Hu, J. Li, C. Xie, J. He, D. Zou, K.-C. Li, M.M. Freire (Eds.), Emerging Technologies in Knowledge Discovery and Data Mining. XIV, 675 pages. 2007.

Vol. 4798: Z. Zhang, J.H. Siekmann (Eds.), Knowledge Science and Engineering and Management. XVI, 669 pages. 2007.

Vol. 4795: F. Schilder, G. Katz, J. Pustejovsky (Eds.), Annotating, Extracting and Reasoning about Time and Events. VII, 141 pages. 2007.

Vol. 4790: N. Dershowitz, A. Voronkov (Eds.), Logic for Programming, Artificial Intelligence, and Reasoning. XIII, 562 pages. 2007.

Vol. 4788: D. Borrajo, L. Castillo, J.M. Corchado (Eds.), Current Topics in Artificial Intelligence. XI, 280 pages. 2007.

Vol. 4775: A. Esposito, M. Faundez-Zanuy, E. Keller, M. Marinaro (Eds.), Verbal and Nonverbal Communication Behaviours. XII, 325 pages. 2007.

Vol. 4772: H. Prade, V.S. Subrahmanian (Eds.), Scalable Uncertainty Management. X, 277 pages. 2007.

Vol. 4766: N. Maudet, S. Parsons, I. Rahwan (Eds.), Argumentation in Multi-Agent Systems. XII, 211 pages. 2007.

Vol. 4755: V. Corruble, M. Takeda, E. Suzuki (Eds.), Discovery Science. XI, 298 pages. 2007.

Vol. 4754: M. Hutter, R.A. Servedio, E. Takimoto (Eds.), Algorithmic Learning Theory. XI, 403 pages. 2007.

Vol. 4737: B. Berendt, A. Hotho, D. Mladenic, G. Semeraro (Eds.), From Web to Social Web: Discovering and Deploying User and Content Profiles. XI, 161 pages. 2007.

Vol. 4733: R. Basili, M.T. Pazienza (Eds.), AI*IA 2007: Artificial Intelligence and Human-Oriented Computing. XVII, 858 pages. 2007.

Vol. 4724: K. Mellouli (Ed.), Symbolic and Quantitative Approaches to Reasoning with Uncertainty. XV, 914 pages. 2007.

Vol. 4722: C. Pelachaud, J.-C. Martin, E. André, G. Chollet, K. Karpouzis, D. Pelé (Eds.), Intelligent Virtual Agents. XV, 425 pages. 2007.

Vol. 4720: B. Konev, F. Wolter (Eds.), Frontiers of Combining Systems. X, 283 pages. 2007.

Vol. 4702: J.N. Kok, J. Koronacki, R. Lopez de Mantaras, S. Matwin, D. Mladenič, A. Skowron (Eds.), Knowledge Discovery in Databases: PKDD 2007. XXIV, 640 pages. 2007.

Vol. 4701: J.N. Kok, J. Koronacki, R. Lopez de Mantaras, S. Matwin, D. Mladenič, A. Skowron (Eds.), Machine Learning: ECML 2007. XXII, 809 pages. 2007.

Vol. 4696: H.-D. Burkhard, G. Lindemann, R. Verbrugge, L.Z. Varga (Eds.), Multi-Agent Systems and Applications V. XIII, 350 pages. 2007.

Vol. 4694: B. Apolloni, R.J. Howlett, L. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part III. XXIX, 1126 pages. 2007.

Vol. 4693: B. Apolloni, R.J. Howlett, L. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part II. XXXII, 1380 pages. 2007.

Vol. 4692: B. Apolloni, R.J. Howlett, L. Jain (Eds.), Knowledge-Based Intelligent Information and Engineering Systems, Part I. LV, 882 pages. 2007.

Vol. 4687: P. Petta, J.P. Müller, M. Klusch, M. Georgeff (Eds.), Multiagent System Technologies. X, 207 pages. 2007.

Vol. 4682: D.-S. Huang, L. Heutte, M. Loog (Eds.), Advanced Intelligent Computing Theories and Applications. XXVII, 1373 pages. 2007.

Vol. 4676: M. Klusch, K.V. Hindriks, M.P. Papazoglou, L. Sterling (Eds.), Cooperative Information Agents XI. XI, 361 pages. 2007.

Vol. 4667: J. Hertzberg, M. Beetz, R. Englert (Eds.), KI 2007: Advances in Artificial Intelligence. IX, 516 pages. 2007.

Vol. 4660: S. Džeroski, L. Todorovski (Eds.), Computational Discovery of Scientific Knowledge. X, 327 pages. 2007.

Vol. 4659: V. Mařík, V. Vyatkin, A.W. Colombo (Eds.), Holonic and Multi-Agent Systems for Manufacturing. VIII, 456 pages. 2007.

此为试读,需要完整PDF请访问: www.ertongbook.com

Vol. 4651: F. Azevedo, P. Barahona, F. Fages, F. Rossi (Eds.), Recent Advances in Constraints. VIII, 185 pages. 2007.

Vol. 4648: F. Almeida e Costa, L.M. Rocha, E. Costa, I. Harvey, A. Coutinho (Eds.), Advances in Artificial Life. XVIII, 1215 pages. 2007.

Vol. 4635: B. Kokinov, D.C. Richardson, T.R. Roth-Berghofer, L. Vieu (Eds.), Modeling and Using Context. XIV, 574 pages. 2007.

Vol. 4632: R. Alhajj, H. Gao, X. Li, J. Li, O.R. Zaïane (Eds.), Advanced Data Mining and Applications. XV, 634 pages. 2007.

Vol. 4629: V. Matoušek, P. Mautner (Eds.), Text, Speech and Dialogue. XVII, 663 pages. 2007.

Vol. 4626: R.O. Weber, M.M. Richter (Eds.), Case-Based Reasoning Research and Development. XIII, 534 pages. 2007.

Vol. 4617: V. Torra, Y. Narukawa, Y. Yoshida (Eds.), Modeling Decisions for Artificial Intelligence. XII, 502 pages. 2007.

Vol. 4612: I. Miguel, W. Ruml (Eds.), Abstraction, Reformulation, and Approximation. XI, 418 pages. 2007.

Vol. 4604: U. Priss, S. Polovina, R. Hill (Eds.), Conceptual Structures: Knowledge Architectures for Smart Applications. XII, 514 pages. 2007.

Vol. 4603: F. Pfenning (Ed.), Automated Deduction – CADE-21. XII, 522 pages. 2007.

Vol. 4597: P. Perner (Ed.), Advances in Data Mining. XI, 353 pages. 2007.

Vol. 4594: R. Bellazzi, A. Abu-Hanna, J. Hunter (Eds.), Artificial Intelligence in Medicine. XVI, 509 pages. 2007.

Vol. 4585: M. Kryszkiewicz, J.F. Peters, H. Rybinski, A. Skowron (Eds.), Rough Sets and Intelligent Systems Paradigms. XIX, 836 pages. 2007.

Vol. 4578: F. Masulli, S. Mitra, G. Pasi (Eds.), Applications of Fuzzy Sets Theory. XVIII, 693 pages. 2007.

Vol. 4573: M. Kauers, M. Kerber, R. Miner, W. Windsteiger (Eds.), Towards Mechanized Mathematical Assistants. XIII, 407 pages. 2007.

Vol. 4571: P. Perner (Ed.), Machine Learning and Data Mining in Pattern Recognition. XIV, 913 pages. 2007.

Vol. 4570: H.G. Okuno, M. Ali (Eds.), New Trends in Applied Artificial Intelligence. XXI, 1194 pages. 2007.

Vol. 4565: D.D. Schmorrow, L.M. Reeves (Eds.), Foundations of Augmented Cognition. XIX, 450 pages. 2007.

Vol. 4562: D. Harris (Ed.), Engineering Psychology and Cognitive Ergonomics. XXIII, 879 pages. 2007.

Vol. 4548: N. Olivetti (Ed.), Automated Reasoning with Analytic Tableaux and Related Methods. X, 245 pages. 2007.

Vol. 4539: N.H. Bshouty, C. Gentile (Eds.), Learning Theory. XII, 634 pages. 2007.

Vol. 4529: P. Melin, O. Castillo, L.T. Aguilar, J. Kacprzyk, W. Pedrycz (Eds.), Foundations of Fuzzy Logic and Soft Computing. XIX, 830 pages. 2007.

Vol. 4520: M.V. Butz, O. Sigaud, G. Pezzulo, G. Baldassarre (Eds.), Anticipatory Behavior in Adaptive Learning Systems. X, 379 pages. 2007.

Vol. 4511: C. Conati, K. McCoy, G. Paliouras (Eds.), User Modeling 2007. XVI, 487 pages. 2007.

Vol. 4509: Z. Kobti, D. Wu (Eds.), Advances in Artificial Intelligence. XII, 552 pages. 2007.

Vol. 4496: N.T. Nguyen, A. Grzech, R.J. Howlett, L.C. Jain (Eds.), Agent and Multi-Agent Systems: Technologies and Applications. XXI, 1046 pages. 2007.

Vol. 4483: C. Baral, G. Brewka, J. Schlipf (Eds.), Logic Programming and Nonmonotonic Reasoning. IX, 327 pages. 2007.

Vol. 4482: A. An, J. Stefanowski, S. Ramanna, C.J. Butz, W. Pedrycz, G. Wang (Eds.), Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. XIV, 585 pages. 2007.

Vol. 4481: J. Yao, P. Lingras, W.-Z. Wu, M. Szczuka, N.J. Cercone, D. Ślęzak (Eds.), Rough Sets and Knowledge Technology. XIV, 576 pages. 2007.

Vol. 4476: V. Gorodetsky, C. Zhang, V.A. Skormin, L. Cao (Eds.), Autonomous Intelligent Systems: Multi-Agents and Data Mining. XIII, 323 pages. 2007.

Vol. 4460: S. Aguzzoli, A. Ciabattoni, B. Gerla, C. Manara, V. Marra (Eds.), Algebraic and Proof-theoretic Aspects of Non-classical Logics. VIII, 309 pages. 2007.

Vol. 4457: G.M.P. O'Hare, A. Ricci, M.J. O'Grady, O. Dikenelli (Eds.), Engineering Societies in the Agents World VII. XI, 401 pages. 2007.

Vol. 4456: Y. Wang, Y.-m. Cheung, H. Liu (Eds.), Computational Intelligence and Security. XXIII, 1118 pages. 2007.

Vol. 4455: S. Muggleton, R. Otero, A. Tamaddoni-Nezhad (Eds.), Inductive Logic Programming. XII, 456 pages. 2007.

Vol. 4452: M. Fasli, O. Shehory (Eds.), Agent-Mediated Electronic Commerce. VIII, 249 pages. 2007.

Vol. 4451: T.S. Huang, A. Nijholt, M. Pantic, A. Pentland (Eds.), Artifical Intelligence for Human Computing. XVI, 359 pages. 2007.

Vol. 4442: L. Antunes, K. Takadama (Eds.), Multi-Agent-Based Simulation VII. X, 189 pages. 2007.

Vol. 4441: C. Müller (Ed.), Speaker Classification II. X, 309 pages. 2007.

Vol. 4438: L. Maicher, A. Sigel, L.M. Garshol (Eds.), Leveraging the Semantics of Topic Maps. X, 257 pages. 2007.

Vol. 4434: G. Lakemeyer, E. Sklar, D.G. Sorrenti, T. Takahashi (Eds.), RoboCup 2006: Robot Soccer World Cup X. XIII, 566 pages. 2007.

Vol. 4429: R. Lu, J.H. Siekmann, C. Ullrich (Eds.), Cognitive Systems. X, 161 pages. 2007.

Vol. 4428: S. Edelkamp, A. Lomuscio (Eds.), Model Checking and Artificial Intelligence. IX, 185 pages. 2007.

Vol. 4426: Z.-H. Zhou, H. Li, Q. Yang (Eds.), Advances in Knowledge Discovery and Data Mining. XXV, 1161 pages. 2007.

¥671.00元

Preface

The techniques of knowledge discovery and data mining (KDD) have rapidly developed along the significant progress of the computer and its network technologies in the last two decades. The attention and the number of researchers in this domain continue to grow in both international academia and industry. The Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD) is a worldwide representative international conference on the research areas of KDD. Under the current spread of KDD techniques in our society, PAKDD 2007 invited the organizers to an industrial track on KDD techniques and, moreover, called for enterprising proposals of workshops on novel and emerging topics in KDD studies. The PAKDD industrial track and the PAKDD workshops both attracted more than 350 paper submissions from 22 countries in total, of which only 81 high-quality papers were accepted after strict reviews. The conference provided many informal and vibrant opportunities for researchers and industry practitioners to share their research positions, original research results and practical development experiences on specific new challenges, emerging issues, new technology trends and solutions to real-world problems.

The objective of this volume is to offer the excellent presentations to the public, and to promote the study exchange among researchers worldwide. Sixty-two outstanding papers in the industrial track and the workshops were further selected, among the 81 paper presentations, under even more rigorous reviews by the organizers, and these papers were carefully brushed up and included in this post-proceedings volume.

The first part of this volume contains ten outstanding papers presented in the industrial track. This track was organized to attract papers on new technology trends and real-world solutions in different industry sectors. The succeeding chapters include papers selected from the three workshops of BioDM 2007, HPDMA 2007 and SSDU 2007. The 2nd BioDM Workshop on Data Mining for Biomedical Applications (BioDM 2007) aimed at attracting top researchers, practitioners and students from around the world to discuss data mining applications in the field of bioinformatics. There are many computationally challenging problems in the analysis of the diverse and voluminous data provided in the field, and the readers will find 11 challenging papers on these problems. The 2007 International Workshop on High-Performance Data Mining and Applications (HPMDA 2007) addressed high-performance data mining methods and applications from both algorithmic and system perspectives. It includes 25 papers on how to achieve efficient mining of useful information from the data available, as well as on the topics of parallel hardware platforms, clusters and large-scale distributed computing infrastructures. The 2007 International Workshop on Service, Security and Its Data Management for Ubiquitous Computing (SSDU 2007) was held to foster research in the areas of security and intelligence integration

into ubiquitous computing and data management technology. Sixteen papers discussing the many security risks and problems in the ubiquitous computing environment are included in the volume.

We hope this book contributes to the growth of the worldwide community of KDD researchers.

September 2007

Takashi Washio
Zhi-Hua Zhou

Organization and Editorial Board

The paper selection of the industrial track and the workshops was made by the Program Committee of each respective organization. After the paper selection, the book was edited and managed by the volume editors.

Volume Editors

Takashi Washio (Osaka University, Japan)
Zhi-Hua Zhou (Nanjing University, China)
Joshua Z. Huang (The University of Hong Kong, Hong Kong)
Xiaohua Hu (Drexel University, USA)
Jinyan Li (Institute for Infocomm Research, Singapore)
Chao Xie (Georgia State University, USA)
Jieyue He (Southeast University, China)
Deqing Zou (Huazhong University of Science and Technology, China)
Kuan-Ching Li (Providence University, Taiwan)
Mario Freire (University of Beira Interior, Portugal)

Table of Contents

PAKDD Industrial Track and Workshops 2007

Industrial Track

PAKDD 2007 Industrial Track Workshop 1
Joshua Zhexue Huang and Yunming Ye

A Survey of Open Source Data Mining Systems 3
Xiaojun Chen, Yunming Ye, Graham Williams, and Xiaofei Xu

Predicting the Short-Term Market Reaction to Asset Specific News: Is Time Against Us? 15
Calum Robertson, Shlomo Geva, and Rodney Wolff

Frequency-Weighted Fuzzy Time-Series Based on Fibonacci Sequence for TAIEX Forecasting 27
Hia Jong Teoh, Tai-Liang Chen, and Ching-Hsue Cheng

Probabilistic Techniques for Corporate Blog Mining 35
Flora S. Tsai, Yun Chen, and Kap Luk Chan

Mining Chat Conversations for Sex Identification 45
Cemal Köse, Özcan Özyurt, and Guychmyrat Amanmyradov

Mining High Impact Exceptional Behavior Patterns 56
Longbing Cao, Yanchang Zhao, Fernando Figueiredo, Yuming Ou, and Dan Luo

Practical Issues on Privacy-Preserving Health Data Mining 64
Huidong (Warren) Jin

Data Mining for Intelligent Structure Form Selection Based on Association Rules from a High Rise Case Base 76
Shihai Zhang, Shujun Liu, and Jinping Ou

CommonKADS Methodology for Developing Power Grid Switching Orders Systems 87
Ming Zhou, Jianwen Ren, Jianxun Qi, Dongxiao Niu, and Gengyin Li

Discovering Prediction Model for Environmental Distribution Maps 99
Ke Zhang, Huidong Jin, Nianjun Liu, Rob Lesslie, Lei Wang, Zhouyu Fu, and Terry Caelli

Workshop of BioDM 2007

Workshop BioDM'07—An Overview 110
Jinyan Li and Xiaohua Hu

Extracting Features from Gene Ontology for the Identification of Protein Subcellular Location by Semantic Similarity Measurement 112
Guoqi Li and Huanye Sheng

Detecting Community Structure in Complex Networks by Optimal Rearrangement Clustering 119
Rui-Sheng Wang, Yong Wang, Xiang-Sun Zhang, and Luonan Chen

The HIV Data Mining Tool for Government Decision-Making Support 131
Huijun Liu, Qunying Xiao, and Zhengwei Zhu

Negative Localized Relationship Among p70S6 with Smad1, 2, 3 and p38 in Three Treated Human Cancer Cell Lines 142
Lin Wang, Minghu Jiang, Stefan Wolfl, and Yinghua Lu

Cancer Identification Based on DNA Microarray Data 153
Yihui Liu

Incorporating Dictionary Features into Conditional Random Fields for Gene/Protein Named Entity Recognition 162
Hongfei Lin, Yanpeng Li, and Zhihao Yang

Translation and Rotation Invariant Mining of Frequent Trajectories: Application to Protein Unfolding Pathways 174
Alexander Andreopoulos, Bill Andreopoulos, Aijun An, and Xiaogang Wang

Genetic-Annealing Algorithm for 3D Off-lattice Protein Folding Model 186
Xiaolong Zhang, Xiaoli Lin, Chengpeng Wan, and Tingting Li

Biclustering of Microarray Data Based on Singular Value Decomposition 194
Wen-Hui Yang, Dao-Qing Dai, and Hong Yan

On the Number of Partial Least Squares Components in Dimension Reduction for Tumor Classification 206
Xue-Qiang Zeng, Guo-Zheng Li, Geng-Feng Wu, and Hua-Xing Zou

Mining Biosignal Data: Coronary Artery Disease Diagnosis Using Linear and Nonlinear Features of HRV 218
Heon Gyu Lee, Ki Yong Noh, and Keun Ho Ryu

此为试读,需要完整PDF请访问: www.ertongbook.com

Workshop of HPDMA 2007

High Performance Data Mining and Applications Overview 229
Chao Xie and Jieyue He

Approximately Mining Recently Representative Patterns on Data Streams 231
Jia-Ling Koh and Yuan-Bin Don

Finding Frequent Items in Data Streams Using ESBF 244
ShuYun Wang, XiuLan Hao, HeXiang Xu, and YunFa Hu

A New Decision Tree Classification Method for Mining High-Speed Data Streams Based on Threaded Binary Search Trees 256
Tao Wang, Zhoujun Li, Xiaohua Hu, Yuejin Yan, and Huowang Chen

Progressive Subspace Skyline Clusters Mining on High Dimensional Data 268
Rong Hu, Yansheng Lu, Lei Zou, and Chong Zhou

Efficient Privacy Preserving Distributed Clustering Based on Secret Sharing 280
Selim V. Kaya, Thomas B. Pedersen, Erkay Savaş, and Yücel Saygın

SePMa: An Algorithm That Mining Sequential Processes from Hybrid Log 292
Xiaoyu Huang, Huiling Zhong, and Wenxue Cai

Evaluate Structure Similarity in XML Documents with Merge-Edit-Distance 301
Chong Zhou, Yansheng Lu, Lei Zou, and Rong Hu

Ensemble Learning Based Distributed Clustering 312
Genlin Ji and Xiaohan Ling

Deploying Mobile Agents in Distributed Data Mining 322
Xining Li and JingBo Ni

ODDC: Outlier Detection Using Distance Distribution Clustering 332
Kun Niu, Chong Huang, Shubo Zhang, and Junliang Chen

Spatial Clustering with Obstacles Constraints Using Ant Colony and Particle Swarm Optimization 344
Xueping Zhang, Jiayao Wang, Zhongshan Fan, and Bin Li

A High Performance Hierarchical Cubing Algorithm and Efficient OLAP in High-Dimensional Data Warehouse 357
Kongfa Hu, Zhenzhi Gong, Qingli Da, and Ling Chen

Grid-Based Clustering Algorithm Based on Intersecting Partition and Density Estimation 368
Bao-Zhi Qiu, Xiang-Li Li, and Jun-Yi Shen

Depth First Generation of Frequent Patterns Without Candidate Generation 378
Qunxiong Zhu and Xiaoyong Lin

Efficient Time Series Data Classification and Compression in Distributed Monitoring 389
Sheng Di, Hai Jin, Shengli Li, Jing Tie, and Ling Chen

Best-Match Method Used in Co-training Algorithm 401
Hui Wang, Liping Ji, and Wanli Zuo

A General Method of Mining Chinese Web Documents Based on GA&SA and Position-Factors 410
Xi Bai, Jigui Sun, Haiyan Che, and Jin Wang

Data Management Services in ChinaGrid for Data Mining Applications 421
Song Wu, Wei Wang, Muzhou Xiong, and Hai Jin

Two-Phase Algorithms for a Novel Utility-Frequent Mining Model 433
Jieh-Shan Yeh, Yu-Chiang Li, and Chin-Chen Chang

Top-Down and Bottom-Up Strategies for Incremental Maintenance of Frequent Patterns 445
Qunxiong Zhu and Xiaoyong Lin

GC-Tree: A Fast Online Algorithm for Mining Frequent Closed Itemsets 457
Junbo Chen and ShanPing Li

Integration of Distributed Biological Data Using Modified K-Means Algorithm 469
Jongil Jeong, Byunggul Ryu, Dongil Shin, and Dongkyoo Shin

A Parallel Algorithm for Enumerating All the Maximal k-Plexes 476
Bin Wu and Xin Pei

A Multi-dependency Language Modeling Approach to Information Retrieval 484
Keke Cai, Chun Chen, Jiajun Bu, Guang Qiu, and Peng Huang

Factoid Mining Based Content Trust Model for Information Retrieval ... 492
Wei Wang, Guosun Zeng, Mingjun Sun, Huanan Gu, and Quan Zhang

Workshop of SSDU 2007

Service, Security and Its Data Management for Ubiquitous Computing – Overview 500
Jong Hyuk Park and Deqing Zou

Study on Trust Inference and Emergence of Economical Small-World Phenomena in P2P Environment 502
Yufeng Wang, Yoshiaki Hori, and Kouichi Sakurai

A Secure Time Synchronization Protocol for Sensor Network 515
Hui Li, Yanfei Zheng, Mi Wen, and Kefei Chen

On Grid-Based Key Pre-distribution: Toward a Better Connectivity in Wireless Sensor Network 527
Abedelaziz Mohaisen, YoungJae Maeng, and DaeHun Nyang

A Distributed and Cooperative Black Hole Node Detection and Elimination Mechanism for Ad Hoc Networks 538
Chang Wu Yu, Tung-Kuang Wu, Rei Heng Cheng, and Shun Chao Chang

A Novel Adaptive and Safe Framework for Ubicomp 550
Xuanhua Shi and Jean-Louis Pazat

Reducing Inter-piconet Delay for Large-Scale Bluetooth Scatternets 562
Chang Wu Yu, Kun-Ming Yu, and Shu Ling Lin

Security Analysis and Enhancement of One-Way Hash Based Low-Cost Authentication Protocol (OHLCAP) 574
JeaCheol Ha, SangJae Moon, Juan Manuel Gonzalez Nieto, and Colin Boyd

An Effective Design of an Active RFID Reader Using a Cache of Tag Memory Data 584
Seok-Young Jang, Sang-Hwa Chung, Won-Ju Yoon, and Seong-Joon Lee

Privacy Protection Scheme of RFID Using Random Number 596
Soo-Young Kang and Im-Yeong Lee

A Hierarchical Composition of LU Matrix-Based Key Distribution Scheme for Sensor Networks 608
Mi Wen, Yanfei Zheng, Hui Li, and Kefei Chen

Security Framework for Home Network: Authentication, Authorization, and Security Policy 621
Geon Woo Kim, Deok Gyu Lee, Jong Wook Han, Sang Choon Kim, and Sang Wook Kim

Bogus Data Filtering in Sensor Networks 629
Yong Ho Kim, Jong Hwan Park, Dong Hoon Lee, and Jongin Lim

Streaming Media Securely over Multipath Multihop Wireless Network 636
Binod Vaidya, SangDuck Lee, Eung-Kon Kim, and SeungJo Han

Flexible Selection of Wavelet Coefficients Based on the Estimation Error of Predefined Queries 644
Jaehoon Kim and Seog Park

Secured Web Services Based on Extended Usage Control 656
Woochul Shin and Sang Bong Yoo

A Digital Rights Management Architecture for Multimedia in P2P 664
Cheng Yang, Jianbo Liu, Aina Sui, and Yongbin Wang

Author Index 673

PAKDD 2007 Industrial Track Workshop

Joshua Zhexue Huang[1] and Yunming Ye[2]

[1] E-Business Technology Institute
The University of Hong Kong, Hong Kong, China
jhuang@eti.hku.hk
[2] E-Business & Intelligent Enterprise Computing Research Center
Shenzhen Graduate School of Harbin Institute of Technology, Shenzhen 518055, China
yeyunming@hit.edu.cn

1 Workshop Overview

The PAKDD 2007 Industrial Track Workshop was held on the 22nd of May 2007 in conjunction with The 11th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2007) at the Mandarin Garden Hotel in Nanjing, China. Following the PAKDD conference tradition to promote industry applications of new data mining techniques, methodologies and systems, the theme of this workshop was defined as "Data Mining for Practitioners". Industrial oriented papers on innovative applications of data mining technology to solve real world problems were solicited with emphasis on the following areas:

1. New data mining methodologies
2. Data mining application systems
3. Data cleansing and transformation in data mining
4. New application problems and data mining solutions
5. Data mining in government applications
6. Innovative data mining case studies

The workshop received 82 submissions from 9 countries and regions in Asia, Europe and North America. The topics of these submissions covered new techniques, algorithms, systems, traditional applications, such as banking and finance, and new applications in manufacturing, facility management, environment and government services. After a rigorous review by the committee members, 13 papers were selected for presentation at the workshop. The acceptance rate was about 16%. For the industrial track papers in PAKDD series conferences, this low acceptance rate was unprecedented.

After reevaluation of the 13 presentations, we selected 10 papers for publication in this proceedings. The first paper by (Xiaojun Chen et al.) gives a survey and evaluation on the current open source data mining systems. The rest 9 papers present real world data mining applications in five areas: (1) Financial Market by (Calum Robertson et al. and Hia-Jong Teoh et al.), (2) new Web Applications by (Flora Tsai et al. and Cemal Köse et al.), (3) Social Services by (Longbin Cao et al. and Huidong Jin), (4) Construction and Power Industry by (Shihai Zhang et al. and Ming Zhou et al.), and (5) Environment by

T. Washio et al. (Eds.): PAKDD 2007 Workshops, LNAI 4819, pp. 1–2, 2007.
© Springer-Verlag Berlin Heidelberg 2007

(Ke Zhang et al.). These applications demonstrate that innovative use of data mining technology has a lot of potentials in solving various real world problems.

As Co-chairs of this workshop committee, we wish to place on record our sincere thanks to many people who have contributed to the Industrial Track Workshop. Firstly, we wish to thank all the authors of the 82 submissions. Without them, the workshop would not be able to take place. Secondly, we must express our appreciations to the committee members who did an excellent job in completing the quality paper review work in a very short time. Finally, we would like to thank the PAKDD 2007 organizing committee and local organizers for their assistance in organizing this workshop.

2 Program Committee

Workshop Chairs

Workshop Chair	Joshua Zhexue Huang (The University of Hong Kong, China)
Workshop Co-chair	Yunming Ye (Harbin Institute of Technology, China)

Program Committee

Graham Williams	ATO, Australia
Ke Wang	Simon Fraser University, Canada
William Song	University of Durham, UK
Xuelong Li	University of London, UK
Warren Jin	CSIRO, Australia
Simeon J. Simoff	University of Technology, Sydney, Australia
Tao Lin	SAP Research, USA
Xiongfei Li	Jilin University, China
Shuicheng Yan	University of Illinois at Urbana-Champaign, USA
Dacheng Tao	University of London, UK
Wen Jun Yin	IBM China
Yuan Yuan	Aston University, UK
Dong Xu	Columbia University, USA
Yoon-Joon Lee	KAIST, Korea
Longbing Cao	University of Technology in Sydney, Australia
Liping Jing	University of Hong Kong, China
Jun Xu	University of Hong Kong, China
Chunkai Zhang	HIT, China
Shijie Zhou	UESTC, China
Feilong Tang	SJTU, China
Hongbing Wang	SEU, China
Xiaoyuan Jin	NJUPT, China

此为试读,需要完整PDF请访问: www.ertongbook.com