

Lecture Notes in Artificial Intelligence 6119

Edited by R. Goebel, J. Siekmann, and W. Wahlster

Subseries of Lecture Notes in Computer Science

Mohammed J. Zaki Jeffrey Xu Yu
B. Ravindran Vikram Pudi (Eds.)

Advances in Knowledge Discovery and Data Mining

14th Pacific-Asia Conference, PAKDD 2010
Hyderabad, India, June 21-24, 2010
Proceedings
Part II

Series Editors

Randy Goebel, University of Alberta, Edmonton, Canada

Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Wolfgang Wahlster, DFKI and University of Saarland, Saarbrücken, Germany

Volume Editors

Mohammed J. Zaki

Rensselaer Polytechnic Institute

Troy, NY, USA

E-mail: zaki@cs.rpi.edu

Jeffrey Xu Yu

The Chinese University of Hong Kong

Hong Kong, China

E-mail: yu@se.cuhk.edu.hk

B. Ravindran

IIT Madras, Chennai, India

E-mail: ravi@cse.iitm.ac.in

Vikram Pudi

IIIT, Hyderabad, India

E-mail: vikram@iiit.ac.in

Library of Congress Control Number: 2010928262

CR Subject Classification (1998): I.2, H.3, H.4, H.2.8, I.4, C.2

LNCS Sublibrary: SL 7 – Artificial Intelligence

ISSN 0302-9743

ISBN-10 3-642-13671-0 Springer Berlin Heidelberg New York

ISBN-13 978-3-642-13671-9 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.com

© Springer-Verlag Berlin Heidelberg 2010

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper 06/3180

Preface

The 14th Pacific-Asia Conference on Knowledge Discovery and Data Mining was held in Hyderabad, India during June 21–24, 2010; this was the first time the conference was held in India.

PAKDD is a major international conference in the areas of data mining (DM) and knowledge discovery in databases (KDD). It provides an international forum for researchers and industry practitioners to share their new ideas, original research results and practical development experiences from all KDD-related areas including data mining, data warehousing, machine learning, databases, statistics, knowledge acquisition and automatic scientific discovery, data visualization, causal induction and knowledge-based systems.

PAKDD-2010 received 412 research papers from over 34 countries including: Australia, Austria, Belgium, Canada, China, Cuba, Egypt, Finland, France, Germany, Greece, Hong Kong, India, Iran, Italy, Japan, S. Korea, Malaysia, Mexico, The Netherlands, New Caledonia, New Zealand, San Marino, Singapore, Slovenia, Spain, Switzerland, Taiwan, Thailand, Tunisia, Turkey, UK, USA, and Vietnam. This clearly reflects the truly international stature of the PAKDD conference.

After an initial screening of the papers by the Program Committee Chairs, for papers that did not conform to the submission guidelines or that were deemed not worthy of further reviews, 60 papers were rejected with a brief explanation for the decision. The remaining 352 papers were rigorously reviewed by at least three reviewers. The initial results were discussed among the reviewers and finally judged by the Program Committee Chairs. In some cases of conflict additional reviews were sought. As a result of the deliberation process, only 42 papers (10.2%) were accepted as long presentations (25 mins), and an additional 55 papers (13.3%) were accepted as short presentations (15 mins). The total acceptance rate was thus about 23.5% across both categories.

The PAKDD 2010 conference program also included seven workshops: Workshop on Data Mining for Healthcare Management (DMHM 2010), Pacific Asia Workshop on Intelligence and Security Informatics (PAISI 2010), Workshop on Feature Selection in Data Mining (FSDM 2010), Workshop on Emerging Research Trends in Vehicle Health Management (VHM 2010), Workshop on Behavior Informatics (BI 2010), Workshop on Data Mining and Knowledge Discovery for e-Governance (DMEG 2010), Workshop on Knowledge Discovery for Rural Systems (KDRS 2010).

The conference would not have been successful without the support of the Program Committee members (164), external reviewers (195), Conference Organizing Committee members, invited speakers, authors, tutorial presenters, workshop organizers, reviewers, authors and the conference attendees. We highly appreciate the conscientious reviews provided by the Program Committee

members, and external reviewers. The Program Committee members were matched with the papers using the SubSift system (<http://subsift.ilrt.bris.ac.uk/>) for bid matching; we thank Simon Price and Peter Flach, of Bristol University, for developing this wonderful system. Thanks also to Andrei Voronkov for hosting the entire PAKDD reviewing process on the easychair.org site.

We are indebted to the members of the PAKDD Steering Committee for their invaluable suggestions and support throughout the organization process. We thank Vikram Pudi (Publication Chair), Pabitra Mitra (Workshops Chair), Kamal Karlapalem (Tutorials Chair), and Arnab Bhattacharya (Publicity Chair). Special thanks to the Local Arrangements Committee and Chair R.K. Bagga, and the General Chairs: Jaideep Srivastava, Masaru Kitsuregawa, and P. Krishna Reddy. We would also like to thank all those who contributed to the success of PAKDD 2010 but whose names may not be listed.

We greatly appreciate the support from various institutions. The conference was organized by IIIT Hyderabad. It was sponsored by the Office of Naval Research Global (ONRG) and the Air Force Office of Scientific Research/Asian Office of Aerospace Research and Development (AFOSR/AOARD).

We hope you enjoy the proceedings of the PAKDD conference, which presents cutting edge research in data mining and knowledge discovery. We also hope all participants took this opportunity to share and exchange ideas with each other and enjoyed the cultural and social attractions of the wonderful city of Hyderabad!

June 2010

Mohammed J. Zaki
Jeffrey Xu Yu
B. Ravindran

PAKDD 2010 Conference Organization

Honorary Chair

Rajeev Sangal	IIIT Hyderabad, India
---------------	-----------------------

General Chairs

Jaideep Srivastava	University of Minnesota, USA
Masaru Kitsuregawa	University of Tokyo, Japan
P. Krishna Reddy	IIIT Hyderabad, India

Program Committee Chairs

Mohammed J. Zaki	Rensselaer Polytechnic Institute, USA
Jeffrey Xu Yu	The Chinese University of Hong Kong
B. Ravindran	IIT Madras, India

Workshop Chair

Pabitra Mitra	IIT Kharagpur, India
---------------	----------------------

Tutorial Chairs

Kamal Karlapalem	IIIT Hyderabad, India
------------------	-----------------------

Publicity Chairs

Arnab Bhattacharya	IIT Kanpur, India
--------------------	-------------------

Publication Chair

Vikram Pudi	IIIT Hyderabad, India
-------------	-----------------------

Local Arrangements Committee

Raji Bagga (Chair)	IIIT Hyderabad, India
K.S. Vijaya Sekhar	IIIT Hyderabad, India
T. Ragunathan	IIIT Hyderabad, India
P. Radhakrishna	Infosys SET Labs, Hyderabad, India
A. Govardhan	JNTU, Hyderabad, India
R.B.V. Subramanyam	NIT, Warangal, India

Program Committee

Osman Abul
Muhammad Abulaish
Arun Agarwal
Hisham Al-Mubaid
Reda Alhajj
Hiroki Arimura
Hideo Bannai
Jayanta Basak
M.M. Sufyan Beg
Bettina Berendt
Fernando Berzal
Raj Bhatnagar
Vasudha Bhatnagar
Arnab Bhattacharya
Pushpak Bhattacharyya
Chiranjib Bhattacharyya
Vivek Borkar
Keith C.C. Chan
Longbing Cao
Doina Caragea
Venkatesan Chakaravarthy
Vineet Chaoji
Sanjay Chawla
Arbee Chen
Phoebe Chen
Jake Yue Chen
Zheng Chen
Hong Cheng
James Cheng
Alok Choudhary
Diane Cook
Alfredo Cuzzocrea
Sanmay Das
Anne Denton
Lipika Dey
Guozhu Dong
Petros Drineas
Tina Eliassi-Rad
Wei Fan
Eibe Frank
Benjamin C.M. Fung
Sachin Garg
Mohamed Gaber

Joao Gama
Jean-Gabriel Ganascia
Gemma Garriga
Ravi Gupta
Mohammad Hasan
Tu Bao Ho
Vasant Honavar
Wynne Hsu
Xiaohua Hu
Akihiro Inokuchi
Shen Jialie
Hasan Jamil
Daxin Jiang
Ruoming Jin
Bo Jin
Hiroyuki Kawano
Tamer Kahveci
Toshihiro Kamishima
Ben Kao
Panagiotis Karras
Hisashi Kashima
Yiping Ke
Latifur Khan
Hiroyuki Kitagawa
Ravi Kothari
Mehmet Koyuturk
Bharadwaja Kumar
Wai Lam
Chung-Hong Lee
Xue Li
Jinyan Li
Tao Li
Chun-hung Li
Ee-Peng Lim
Chih-Jen Lin
Guimei Liu
Tie-Yan Liu
Changtien Lu
Vasilis Megalooikonomou
Tao Mei
Wagner Meira Jr.
Rosa Meo
Dunja Mladenec

Yang-Sae Moon
 Yasuhiko Morimoto
 Tsuyoshi Murata
 Atsuyoshi Nakamura
 J. Saketha Nath
 Juggapong Natwichai
 Richi Nayak
 Wilfred Ng
 Mitsunori Ogihara
 Salvatore Orlando
 Satoshi Oyama
 Prabin Panigrahi
 Spiros Papadimitriou
 Srinivasan Parthasarathy
 Wen-Chih Peng
 Bernhard Pfahringer
 Srinivasa Raghavan
 R. Rajesh
 Naren Ramakrishnan
 Ganesh Ramakrishnan
 Jan Ramon
 Sanjay Ranka
 Rajeev Rastogi
 Chandan Reddy
 Patricia Riddle
 S. Raju Bapi
 Saeed Salem
 Sudeshna Sarkar
 Tamas Sarlos
 C. Chandra Sekhar
 Srinivasan Sengamedu
 Shirish Shevade
 M. Shimbo

Ambuj K. Singh
 Mingli Song
 P. Sreenivasa Kumar
 Aixin Sun
 S. Sundararajan
 Ashish Sureka
 Domenico Talia
 Kay Chen Tan
 Ah-Hwee Tan
 Pang-Ning Tan
 David Taniar
 Ashish Tendulkar
 Thanaruk Theeramunkong
 W. Ivor Tsang
 Vincent S. Tseng
 Tomoyuki Uchida
 Lipo Wang
 Jason Wang
 Jianyong Wang
 Graham Williams
 Xintao Wu
 Xindong Wu
 Meng Xiaofeng
 Hui Xiong
 Seiji Yamada
 Jiong Yang
 Dit-Yan Yeung
 Tetsuya Yoshida
 Aidong Zhang
 Zhongfei (Mark) Zhang
 Zhi-Hua Zhou
 Chengqi Zhang

External Reviewers

Abdul Nizar
 Abhinav Mishra
 Alessandra Raffaeta
 Aminul Islam
 Andrea Tagarelli
 Anitha Varghese
 Ankit Agrawal
 Anuj Mahajan
 Anupam Bhattacharjee

Atul Saroop
 Blaz Novak
 Brian Rutenber
 Bum-Soo Kim
 Carlo Mastroianni
 Carlos Ferreira
 Carmela Comito
 Cha Lun Li
 Chandra Sekhar Chellu

Chao Luo
 Chen-Yi Lin
 Chih jui Lin Wang
 Chuancong Gao
 Chun Kit Chui
 Chun Wei Seah
 Chun-Hao Chen
 Chung An Yeh
 Claudio Silvestri
 Da Jun Li
 David Uthus
 De-Chuan Zhan
 Delia Rusu
 Dhruv Mahajan
 Di Wang
 Dinesh Garg
 Elena Ikonmovska
 En Tzu Wang
 Eugenio Cesario
 Feilong Chen
 Feng Chen
 Ferhat Ay
 Gokhan Yavas
 Gongqing Wu
 Hea-Suk Kim
 Hideyuki Kawashima
 Hui Zhu Su
 Ilija Subasic
 Indranil Palit
 Jan Rupnik
 Janez Brank
 Jeyashanker Ramamirtham
 Jitendra Ajmera
 Junjie Wu
 Kathy Macropol
 Khalil Al-Hussaeni
 Kong Wah Wan
 Krishna Prasad Chitrapura
 Kungpeng Zhang
 Kyle Chipman
 L. Venkata Subramaniam
 Lang Huo
 Lei Liu
 Lei Shi
 Lei Yang

Leting Wu
 Li Zheng
 Li An
 Li Da Jun
 Lili Jiang
 Ling Guo
 Linhong Zhu
 Lixin Duan
 Lorand Dali
 Luka Bradesko
 Luming Zhang
 Manas Somaiya
 Mark Beltramo
 Markus Ojala
 Masayuki Okabe
 Miao Qiao
 Michael Barnathan
 Min-Ling Zhang
 Mingkui Tan
 Mitja Trampus
 Mohammed Aziz
 Mohammed Imran
 Muad Abu-Ata
 Nagaraj Kota
 Nagarajan Krishnamurthy
 Narasimha Murty Musti
 Narayan Bhamidipati
 Ngoc Khanh Pham
 Ngoc Tu Le
 Nicholas Larusso
 Nidhi Raj
 Nikolaj Tatti
 Ning Ruan
 Nirmalya Bandyopadhyay
 Nithi Gupta
 Noman Mohammed
 Palvali Teja
 Pannagadatta Shivaswamy
 Paolo Trunfio
 Parthasarathy Krishnaswamy
 Pedro P. Rodrigues
 Peipei Li
 Prahladavaradan Sampath
 Prakash Mandayam Comare
 Prasad Deshpande

Prithviraj Sen
Pruet Boonma
Qi Mao
Qiang Wang
Qingyan Yang
Quan Yuan
Quang Khoat Than
Rahul Chougule
Ramanathan Narayanan
Raquel Sebastiao
Rashid Ali
Rui Chen
S. Sathiya Keerthi
Shailesh Kumar
Sai Sundarakrishna
Saikat Dey
J. Saketha Nath
SakethaNath Jagarlapudi
Salim Akhter Chowdhury
Samah Fodeh
Sami Hanhijärvi
Satnam Singh
Sau Dan Lee
Sayan Ranu
Sergio Flesca
Shafkat Amin
Shailesh Kumar
Shalabh Bhatnagar
Shantanu Godbole
Sharanjit Kaur
Shazzad Hosain
Shenghua Gao
Shirish Shevade
Shirish Tatikonda
Shirong Li
Shumo Chu
Shuo Miao
Sinan Erten
Sk. Mirajul Haque
Soumen De
Sourangshu Bhattacharya
Sourav Dutta
Srinivasan Sengamedu
Srujana Merugu

Subhajit Sanyal
Sufyan Beg
Sugato Chakrabarty
Sundararajan Sellamanickam
Tadej Štajner
Takehiko Sakamoto
Thi Nhan Le
Tianrui Li
Timothy DeVries
Toshiyuki Amagasa
Venu Satuluri
Victor Lee
Vikram Pudi
Vishwakarma Singh
Viswanath G
Wang Wen-Chi
Wei Chu
Wei Jin
Wei Peng
Wei Su
Wendell Jordan-Brangman
Wenjun Zhou
Wenting Liu
Xiaofeng Zhu
Xiaogang Wu
Xin Liu
Xing Jiang
Xintian Yang
Xutong Liu
Yan Zhang
Yanchang Zhao
Yang Xiang
Yang Zhou
Yasufumi Takama
Yezhou Yang
Yilin Kang
Yin Zhang
Yong Ge
Yuan Liu
Yukai He
Yuko Itokawa
Zakia Sultana
Zubin Abraham

Organized by

IIIT Hyderabad, India



Sponsoring Institutions

AFOSR, USA



AOARD, Tokyo, Japan



ONRG, USA



Table of Contents – Part II

Session 4B. Dimensionality Reduction/Parallelism

Subclass-Oriented Dimension Reduction with Constraint Transformation and Manifold Regularization	1
<i>Bin Tong and Einoshin Suzuki</i>	
Distributed Knowledge Discovery with Non Linear Dimensionality Reduction	14
<i>Panagis Magdalinos, Michalis Vazirgiannis, and Dialecti Valsamou</i>	
DPSP: Distributed Progressive Sequential Pattern Mining on the Cloud	27
<i>Jen-Wei Huang, Su-Chen Lin, and Ming-Syan Chen</i>	
An Approach for Fast Hierarchical Agglomerative Clustering Using Graphics Processors with CUDA	35
<i>S.A. Arul Shalom, Manoranjan Dash, and Minh Tue</i>	

Session 5A. Novel Applications

Ontology-Based Mining of Brainwaves: A Sequence Similarity Technique for Mapping Alternative Features in Event-Related Potentials (ERP) Data	43
<i>Haishan Liu, Gwen Frishkoff, Robert Frank, and Dejing Dou</i>	
Combining Support Vector Machines and the t -statistic for Gene Selection in DNA Microarray Data Analysis	55
<i>Tao Yang, Vojislave Kecman, Longbing Cao, and Chengqi Zhang</i>	
Satrap: Data and Network Heterogeneity Aware P2P Data-Mining	63
<i>Hock Hee Ang, Vivekanand Gopalkrishnan, Anwitaman Datta, Wee Keong Ng, and Steven C.H. Hoi</i>	
Player Performance Prediction in Massively Multiplayer Online Role-Playing Games (MMORPGs)	71
<i>Kyong Jin Shim, Richa Sharan, and Jaideep Srivastava</i>	
Relevant Gene Selection Using Normalized Cut Clustering with Maximal Compression Similarity Measure	81
<i>Rajni Bala, R.K. Agrawal, and Manju Sardana</i>	

Session 5B. Feature Selection/Visualization

A Novel Prototype Reduction Method for the K -Nearest Neighbor Algorithm with $K \geq 1$	89
<i>Tao Yang, Longbing Cao, and Chengqi Zhang</i>	
Generalized Two-Dimensional FLD Method for Feature Extraction: An Application to Face Recognition	101
<i>Shiladitya Chowdhury, Jamuna Kanta Sing, Dipak Kumar Basu, and Mita Nasipuri</i>	
Learning Gradients with Gaussian Processes	113
<i>Xinwei Jiang, Junbin Gao, Tianjiang Wang, and Paul W. Kwan</i>	
Analyzing the Role of Dimension Arrangement for Data Visualization in Radviz	125
<i>Luigi Di Caro, Vanessa Frias-Martinez, and Enrique Frias-Martinez</i>	

Session 6A. Graph Mining

Subgraph Mining on Directed and Weighted Graphs	133
<i>Stephan Günnemann and Thomas Seidl</i>	
Finding Itemset-Sharing Patterns in a Large Itemset-Associated Graph	147
<i>Mutsumi Fukuzaki, Mio Seki, Hisashi Kashima, and Jun Sese</i>	
A Framework for SQL-Based Mining of Large Graphs on Relational Databases	160
<i>Sriganesh Srihari, Shruti Chandrashekar, and Srinivasan Parthasarathy</i>	
Fast Discovery of Reliable k -terminal Subgraphs	168
<i>Melissa Kasari, Hannu Toivonen, and Petteri Hintsanen</i>	
GTRACE2: Improving Performance Using Labeled Union Graphs	178
<i>Akihiro Inokuchi and Takashi Washio</i>	

Session 6B. Clustering II

Orthogonal Nonnegative Matrix Tri-factorization for Semi-supervised Document Co-clustering	189
<i>Huifang Ma, Weizhong Zhao, Qing Tan, and Zhongzhi Shi</i>	
Rule Synthesizing from Multiple Related Databases	201
<i>Dan He, Xindong Wu, and Xingquan Zhu</i>	

Fast Orthogonal Nonnegative Matrix Tri-Factorization for Simultaneous Clustering	214
<i>Zhao Li, Xindong Wu, and Zhenyu Lu</i>	

Hierarchical Web-Page Clustering via In-Page and Cross-Page Link Structures	222
<i>Cindy Xide Lin, Yintao Yu, Jiawei Han, and Bing Liu</i>	

Mining Numbers in Text Using Suffix Arrays and Clustering Based on Dirichlet Process Mixture Models	230
<i>Minoru Yoshida, Issei Sato, Hiroshi Nakagawa, and Akira Terada</i>	

Session 7A. Opinion/Sentiment Mining

Opinion-Based Imprecise Query Answering	238
<i>Muhammad Abulaish, Tanvir Ahmad, Jahiruddin, and Mohammad Najmud Doja</i>	

Blog Opinion Retrieval Based on Topic-Opinion Mixture Model	249
<i>Peng Jiang, Chunxia Zhang, Qing Yang, and Zhendong Niu</i>	

Feature Subsumption for Sentiment Classification in Multiple Languages	261
<i>Zhongwu Zhai, Hua Xu, Jun Li, and Peifa Jia</i>	

Decentralisation of ScoreFinder: A Framework for Credibility Management on User-Generated Contents	272
<i>Yang Liao, Aaron Harwood, and Kotagiri Ramamohanarao</i>	

Classification and Pattern Discovery of Mood in Weblogs	283
<i>Thin Nguyen, Dinh Phung, Brett Adams, Truyen Tran, and Svetha Venkatesh</i>	

Capture of Evidence for Summarization: An Application of Enhanced Subjective Logic	291
<i>Sukanya Manna, B. Sumudu U. Mendis, and Tom Gedeon</i>	

Session 7B. Stream Mining

Fast Perceptron Decision Tree Learning from Evolving Data Streams ...	299
<i>Albert Bifet, Geoff Holmes, Bernhard Pfahringer, and Eibe Frank</i>	

Classification and Novel Class Detection in Data Streams with Active Mining	311
<i>Mohammad M. Masud, Jing Gao, Latifur Khan, Jiawei Han, and Bhavani Thuraisingham</i>	

Bulk Loading Hierarchical Mixture Models for Efficient Stream Classification	325
<i>Philipp Kranen, Ralph Krieger, Stefan Denker, and Thomas Seidl</i>	
Summarizing Multidimensional Data Streams: A Hierarchy-Graph-Based Approach	335
<i>Yoann Pitarch, Anne Laurent, and Pascal Poncelet</i>	
Efficient Trade-Off between Speed Processing and Accuracy in Summarizing Data Streams	343
<i>Nesrine Gabsi, Fabrice Clérot, and Georges Hébrail</i>	
Subsequence Matching of Stream Synopses under the Time Warping Distance	354
<i>Su-Chen Lin, Mi-Yen Yeh, and Ming-Syan Chen</i>	

Session 8A. Similarity and Kernels

Normalized Kernels as Similarity Indices	362
<i>Julien Ah-Pine</i>	
Adaptive Matching Based Kernels for Labelled Graphs	374
<i>Adam Woźnica, Alexandros Kalousis, and Melanie Hilario</i>	
A New Framework for Dissimilarity and Similarity Learning	386
<i>Adam Woźnica and Alexandros Kalousis</i>	
Semantic-Distance Based Clustering for XML Keyword Search	398
<i>Weidong Yang and Hao Zhu</i>	

Session 8B. Graph Analysis

OddBall: Spotting Anomalies in Weighted Graphs	410
<i>Leman Akoglu, Mary McGlohon, and Christos Faloutsos</i>	
Robust Outlier Detection Using Commute Time and Eigenspace Embedding	422
<i>Nguyen Lu Dang Khoa and Sanjay Chawla</i>	
EigenSpokes: Surprising Patterns and Scalable Community Chipping in Large Graphs	435
<i>B. Aditya Prakash, Ashwin Sridharan, Mukund Seshadri, Sridhar Machiraju, and Christos Faloutsos</i>	
BASSET: Scalable Gateway Finder in Large Graphs	449
<i>Hanghang Tong, Spiros Papadimitriou, Christos Faloutsos, Philip S. Yu, and Tina Eliassi-Rad</i>	

Session 8C. Classification II

Ensemble Learning Based on Multi-Task Class Labels	464
<i>Qing Wang and Liang Zhang</i>	
Supervised Learning with Minimal Effort	476
<i>Eileen A. Ni and Charles X. Ling</i>	
Generating Diverse Ensembles to Counter the Problem of Class Imbalance.....	488
<i>T. Ryan Hoens and Nitesh V. Chawla</i>	
Relationship between Diversity and Correlation in Multi-Classifier Systems	500
<i>Kuo-Wei Hsu and Jaideep Srivastava</i>	
Compact Margin Machine.....	507
<i>Bo Dai and Gang Niu</i>	
Author Index	515

Table of Contents – Part I

Keynote Speeches

Empower People with Knowledge: The Next Frontier for Web Search . . . <i>Wei-Ying Ma</i>	1
Discovery of Patterns in Global Earth Science Data Using Data Mining <i>Vipin Kumar</i>	2
Game Theoretic Approaches to Knowledge Discovery and Data Mining <i>Y. Narahari</i>	3

Session 1A. Clustering I

A Set Correlation Model for Partitional Clustering <i>Nguyen Xuan Vinh and Michael E. Houle</i>	4
iVAT and aVAT: Enhanced Visual Analysis for Cluster Tendency Assessment <i>Liang Wang, Uyen T.V. Nguyen, James C. Bezdek, Christopher A. Leckie, and Kotagiri Ramamohanarao</i>	16
A Robust Seedless Algorithm for Correlation Clustering <i>Mohammad S. Aziz and Chandan K. Reddy</i>	28
Integrative Parameter-Free Clustering of Data with Mixed Type Attributes <i>Christian Böhm, Sebastian Goebel, Annahita Oswald, Claudia Plant, Michael Plavinski, and Bianca Wackersreuther</i>	38
Data Transformation for Sum Squared Residue <i>Hyuk Cho</i>	48

Session 1B. Social Networks

A Better Strategy of Discovering Link-Pattern Based Communities by Classical Clustering Methods <i>Chen-Yi Lin, Jia-Ling Koh, and Arbee L.P. Chen</i>	56
Mining Antagonistic Communities from Social Networks <i>Kuan Zhang, David Lo, and Ee-Peng Lim</i>	68

As Time Goes by: Discovering Eras in Evolving Social Networks	81
<i>Michele Berlingerio, Michele Coscia, Fosca Giannotti, Anna Monreale, and Dino Pedreschi</i>	
Online Sampling of High Centrality Individuals in Social Networks	91
<i>Arun S. Maiya and Tanya Y. Berger-Wolf</i>	
Estimate on Expectation for Influence Maximization in Social Networks	99
<i>Yao Zhang, Qing Gu, Jun Zheng, and Daoxu Chen</i>	

Session 1C. Classification I

A Novel Scalable Multi-class ROC for Effective Visualization and Computation	107
<i>Md. Rafiul Hassan, Kotagiri Ramamohanarao, Chandan Karmakar, M. Maruf Hossain, and James Bailey</i>	
Efficiently Finding the Best Parameter for the Emerging Pattern-Based Classifier PCL	121
<i>Thanh-Son Ngo, Mengling Feng, Guimei Liu, and Limsoon Wong</i>	
Rough Margin Based Core Vector Machine	134
<i>Gang Niu, Bo Dai, Lin Shang, and Yangsheng Ji</i>	
BoostML: An Adaptive Metric Learning for Nearest Neighbor Classification	142
<i>Nayyar Abbas Zaidi, David McG. Squire, and David Suter</i>	
A New Emerging Pattern Mining Algorithm and Its Application in Supervised Classification	150
<i>Milton García-Borroto, José Francisco Martínez-Trinidad, and Jesús Ariel Carrasco-Ochoa</i>	

Session 2A. Privacy

Hiding Emerging Patterns with Local Recoding Generalization	158
<i>Michael W.K. Cheng, Byron Koon Kau Choi, and William Kwok Wai Cheung</i>	
Anonymizing Transaction Data by Integrating Suppression and Generalization	171
<i>Junqiang Liu and Ke Wang</i>	
Satisfying Privacy Requirements: One Step before Anonymization	181
<i>Xiaorun Sun, Hua Wang, and Jiuyong Li</i>	

Computation of Ratios of Secure Summations in Multi-party Privacy-Preserving Latent Dirichlet Allocation	189
<i>Bin Yang and Hiroshi Nakagawa</i>	
Privacy-Preserving Network Aggregation	198
<i>Troy Raeder, Marina Blanton, Nitesh V. Chawla, and Keith Frikken</i>	
Multivariate Equi-width Data Swapping for Private Data Publication . . .	208
<i>Yidong Li and Hong Shen</i>	

Session 2B. Spatio-Temporal Mining

Correspondence Clustering: An Approach to Cluster Multiple Related Spatial Datasets	216
<i>Vadeerat Rinsurongkawong and Christoph F. Eick</i>	
Mining Trajectory Corridors Using Fréchet Distance and Meshing Grids	228
<i>Haohan Zhu, Jun Luo, Hang Yin, Xiaotao Zhou, Joshua Zhexue Huang, and F. Benjamin Zhan</i>	
Subseries Join: A Similarity-Based Time Series Match Approach	238
<i>Yi Lin and Michael D. McCool</i>	
TWave: High-Order Analysis of Spatiotemporal Data	246
<i>Michael Barnathan, Vasileios Megalooikonomou, Christos Faloutsos, Feroze B. Mohamed, and Scott Faro</i>	
Spatial Clustering with Obstacles Constraints by Dynamic Piecewise-Mapped and Nonlinear Inertia Weights PSO	254
<i>Xueping Zhang, Haohua Du, and Jiayao Wang</i>	

Session 3A. Pattern Mining

An Efficient GA-Based Algorithm for Mining Negative Sequential Patterns	262
<i>Zhigang Zheng, Yanchang Zhao, Ziye Zuo, and Longbing Cao</i>	
Valency Based Weighted Association Rule Mining	274
<i>Yun Sing Koh, Russel Pears, and Wai Yeap</i>	
Ranking Sequential Patterns with Respect to Significance	286
<i>Robert Gwadera and Fabio Crestani</i>	
Mining Association Rules in Long Sequences	300
<i>Boris Cule and Bart Goethals</i>	
Mining Closed Episodes from Event Sequences Efficiently	310
<i>Wenzhi Zhou, Hongyan Liu, and Hong Cheng</i>	

Most Significant Substring Mining Based on Chi-square Measure	319
<i>Sourav Dutta and Arnab Bhattacharya</i>	

Session 3B. Recommendations/Answers

Probabilistic User Modeling in the Presence of Drifting Concepts	328
<i>Vikas Bhardwaj and Ramaswamy Devarajan</i>	
Using Association Rules to Solve the Cold-Start Problem in Recommender Systems	340
<i>Gavin Shaw, Yue Xu, and Shlomo Geva</i>	
Semi-supervised Tag Recommendation - Using Untagged Resources to Mitigate Cold-Start Problems	348
<i>Christine Preisach, Leandro Balby Marinho, and Lars Schmidt-Thieme</i>	
Cost-Sensitive Listwise Ranking Approach	358
<i>Min Lu, MaoQiang Xie, Yang Wang, Jie Liu, and YaLou Huang</i>	
Mining Wikipedia and Yahoo! Answers for Question Expansion in Opinion QA	367
<i>Yajie Miao and Chunping Li</i>	
Answer Diversification for Complex Question Answering on the Web ...	375
<i>Palakorn Achananuparp, Xiaohua Hu, Tingting He, Christopher C. Yang, Yuan An, and Lifan Guo</i>	
Vocabulary Filtering for Term Weighting in Archived Question Search	383
<i>Zhao-Yan Ming, Kai Wang, and Tat-Seng Chua</i>	

Session 3C. Topic Modeling/Information Extraction

On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations	391
<i>R. Arun, V. Suresh, C.E. Veni Madhavan, and M.N. Narasimha Murthy</i>	
Supervising Latent Topic Model for Maximum-Margin Text Classification and Regression	403
<i>Wanhong Xu</i>	
Resource-bounded Information Extraction: Acquiring Missing Feature Values On Demand	415
<i>Pallika Kanani, Andrew McCallum, and Shaohan Hu</i>	
Efficient Deep Web Crawling Using Reinforcement Learning	428
<i>Lu Jiang, Zhaohui Wu, Qian Feng, Jun Liu, and Qinghua Zheng</i>	

Topic Decomposition and Summarization	440
<i>Wei Chen, Can Wang, Chun Chen, Lijun Zhang, and Jiajun Bu</i>	
Session 4A. Skylines/Uncertainty	
UNN: A Neural Network for Uncertain Data Classification	449
<i>Jiaqi Ge, Yuni Xia, and Chandima Nadungodage</i>	
SkyDist: Data Mining on Skyline Objects	461
<i>Christian Böhm, Annahita Oswald, Claudia Plant, Michael Plavinski, and Bianca Wackersreuther</i>	
Multi-Source Skyline Queries Processing in Multi-Dimensional Space ...	471
<i>Cuiping Li, Wenlin He, and Hong Chen</i>	
Efficient Pattern Mining of Uncertain Data with Sampling	480
<i>Toon Calders, Calin Garboni, and Bart Goethals</i>	
Classifier Ensemble for Uncertain Data Stream Classification	488
<i>Shirui Pan, Kuan Wu, Yang Zhang, and Xue Li</i>	
Author Index	497