

# Lecture Notes in Artificial Intelligence 7819

## Subseries of Lecture Notes in Computer Science

### LNAI Series Editors

Randy Goebel

*University of Alberta, Edmonton, Canada*

Yuzuru Tanaka

*Hokkaido University, Sapporo, Japan*

Wolfgang Wahlster

*DFKI and Saarland University, Saarbrücken, Germany*

### LNAI Founding Series Editor

Joerg Siekmann

*DFKI and Saarland University, Saarbrücken, Germany*

Jian Pei Vincent S. Tseng Longbing Cao  
Hiroshi Motoda Guandong Xu (Eds.)

# Advances in Knowledge Discovery and Data Mining

17th Pacific-Asia Conference, PAKDD 2013  
Gold Coast, Australia, April 14-17, 2013  
Proceedings, Part II



Springer

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

Jian Pei  
Simon Fraser University, Burnaby, BC, Canada  
E-mail: [jpei@cs.sfu.ca](mailto:jpei@cs.sfu.ca)

Vincent S. Tseng  
National Cheng Kung University, Tainan, Taiwan  
E-mail: [tsengsm@mail.ncku.edu.tw](mailto:tsengsm@mail.ncku.edu.tw)

Longbing Cao  
Guandong Xu  
University of Technology Sydney, NSW, Australia  
E-mail: {[longbing.cao](mailto:longbing.cao@uts.edu.au), [guandong.xu](mailto:guandong.xu@uts.edu.au)}@uts.edu.au

Hiroshi Motoda  
Osaka University, Japan  
E-mail: [motoda@ar.sanken.osaka-u.ac.jp](mailto:motoda@ar.sanken.osaka-u.ac.jp)

ISSN 0302-9743  
ISBN 978-3-642-37455-5  
DOI 10.1007/978-3-642-37456-2  
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349  
e-ISBN 978-3-642-37456-2

Library of Congress Control Number: 2013934551

CR Subject Classification (1998): H.2.8, I.2, H.3, H.4, J.1, I.5

LNCS Sublibrary: SL 7 – Artificial Intelligence

© Springer-Verlag Berlin Heidelberg 2013

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.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

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

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

# Preface

As the Program Committee Co-chairs, we welcome you to the proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2013), held at Gold Coast, Australia, during April 14-17, 2013.

The PAKDD conference series, since its inception in 1997, has been a leading international conference in the areas of data mining and knowledge discovery (KDD). It provides an inviting and inspiring forum for researchers and practitioners, from both academia and industry, to share new ideas, original research results, and practical experience. The 17th edition continued the great tradition, and had three world-class keynote speeches, a wonderful technical program, a handful of high-quality tutorials and workshops, as well as an interesting invited talk from industry.

The PAKDD 2013 conference received 363 submissions to the technical program, involving more than 1,000 authors in total. In the rigorous double-blind review process, each submission was reviewed by one senior Program Committee member and at least three Program Committee members. Many submissions were extensively and thoroughly discussed by the reviewers. Based on the detailed and critical discussion and reviews, the senior Program Committee members made recommendations. Overall, 98 papers from 341 authors were accepted in the technical program, yielding a 27% acceptance rate. Of these, 39 (10.7%) had long presentations (30 minutes) and 59 (16.3%) had short presentations (15 minutes). The technical program consisted of 22 sessions, covering the general fields of data mining and KDD extensively, including pattern mining, classification, graph mining, applications, machine learning, feature selection and dimensionality reduction, multiple information sources mining, social networks, clustering, text mining, text classification, imbalanced data, privacy-preserving data mining, recommendation, multimedia data mining, stream data mining, data preprocessing and representation.

We were lucky to have three world-class keynote speakers this year. Usama Fayyad, a renowned pioneer in big data entrepreneurship, addressed us on the big picture of big data. Huan Liu, a world-wide leader in social media mining, discussed this exciting new frontier of data mining. Qiang Yang, a famous expert on artificial intelligence and machine learning, talked on how machine learning can address the big data challenge. We were also pleased to have Alexandros Batsakis as an invited speaker from industry. He shared with us the latest developments on big data analytics infrastructure and enterprise applications.

The conference also included six workshops, covering a few exciting and fast-growing hot topics. We also had five very timely and educational tutorials, covering the hot topics of social networks, transfer learning, stream mining, outlier detection, and feature discovery.

In addition to the intellectually inspiring keynote speeches, technical program, workshops and tutorials, we had several dynamic social events to facilitate communication and informal interaction, including a welcome reception, a banquet, and an excursion.

Putting together a conference like PAKDD is never easy. It becomes possible only with tremendous contributions from the organizing team and many volunteers. We thank Jiuyong Li, Kay Chen Tan, and Bo Liu for organizing the workshop program. We also thank Tu Bao Ho and Mengjie Zhang for organizing the tutorial program. We are grateful to Chengqi Zhang for his leadership in the award selection. We owe a big thank-you to the 39 senior Program Committee members, 151 Program Committee members, and the external reviewers for their great contributions and collaboration. We thank Guandong Xu for assembling the proceedings. We also thank the General Chairs, Hiroshi Motoda and Longbing Cao, and the local organization team for their great support. Without the dedicated hard work of so many people, PAKDD 2013 would simply have been mission impossible.

February 2013

Jian Pei  
Vincent S. Tseng

# Organization

## Honorary Co-chairs

Jiawei Han	University of Illinois at Urbana-Champaign, USA
Ramamohanarao Kotagiri	University of Melbourne, Australia
Graham Williams	Australia Taxation Office, Australia

## Conference Co-chairs

Hiroshi Motoda	Osaka University, Japan
Longbing Cao	University of Technology, Sydney, Australia

## Program Committee Co-chairs

Jian Pei	Simon Fraser University, Canada
Vincent S. Tseng	National Cheng Kung University, Taiwan

## Local Arrangements Co-chairs

Vladimir Estivill-Castro	Griffith University (Gold Coast), Australia
Xue Li	University of Queensland, Australia
Richi Nayak	Queensland University of Technology, Australia
Xinhua Zhu	University of Technology, Sydney, Australia

## Workshop Co-chairs

Jiuyong Li	University of South Australia, Australia
Kay Chen Tan	National University of Singapore, Singapore
Bo Liu	Guangdong University of Technology, China

## Tutorial Co-chairs

Tu Bao Ho	Japan Advanced Institute of Science and Technology, Japan
Mengjie Zhang	Victoria University of Wellington, New Zealand

## Awards Chair

Chengqi Zhang	University of Technology, Sydney, Australia
---------------	---

## Sponsorship Co-chairs

Yue Xu	Queensland University of Technology, Australia
Eugene Dubossarsky	Analyst First
Suresh Sood	University of Technology, Sydney, Australia

## Publicity Co-chairs

P.Krishna Reddy	The International Institute of Information Technology, Hyderabad, India
Yifeng Zeng	Aalborg University, Denmark
Xin Wang	University of Calgary, Canada
Zhihong Deng	Peking University, China

## Proceedings Chair

Guandong Xu	Unverisity of Technology, Sydney, Australia
-------------	---

## Registration Co-chairs

Qiang Wu	University of Technology, Sydney, Australia
Can Wang	University of Technology, Sydney, Australia

## Big Data School Co-chairs

Jinyan Li	University of Technology, Sydney, Australia
Xinhua Zhu	University of Technology, Sydney, Australia

## Steering Committee

Graham Williams	Australian Taxation Office, Australia
Tu Bao Ho	Japan Advanced Institute of Science and Technology, Japan
Hiroshi Motoda	Osaka University, Japan (Since 1997)
Rao Kotagiri	University of Melbourne, Australia (Since 1997)
Ning Zhong	Maebashi Institute of Technology, Japan (Since 1999)
Masaru Kitsuregawa	Tokyo University, Japan (Since 2000)
David Cheung	University of Hong Kong, China (Since 2001)
Graham Williams (Treasurer)	Australian National University, Australia (Since 2001)
Ming-Syan Chen	National Taiwan University, Taiwan, ROC (Since 2002)

Kyu-Young Whang	Korea Advanced Institute of Science & Technology, Korea (Since 2003)
Huan Liu	Arizona State University, USA (Since 1998)
Chengqi Zhang	University of Technology Sydney, Australia (Since 2004)
Tu Bao Ho	Japan Advanced Institute of Science and Technology, Japan (Since 2005)
Ee-Peng Lim	Singapore Management University, Singapore (Since 2006)
Jaideep Srivastava	University of Minnesota, USA (Since 2006)
Zhi-Hua Zhou	Nanjing University, China (Since 2007)
Takashi Washio	Institute of Scientific and Industrial Research, Osaka University, Japan (Since 2008)
Thanaruk Theeramunkong	Thammasat University, Thailand (Since 2009)
P. Krishna Reddy	International Institute of Information Technology, Hyderabad (IIIT-H), India (Since 2010)
Joshua Z. Huang	Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China (Since 2011)

## Senior Program Committee Members

James Bailey	University of Melbourne, Australia
Michael Berthold	University of Konstanz, Germany
Nitesh Chawla	University of Notre Dame, USA
Sanjay Chawla	University of Sydney, Australia
Ming-Syan Chen	National Taiwan University, Taiwan
Arbee L. P. Chen	National Chengchi University, Taiwan
Peter Christen	The Australian National University, Australia
Diane Cook	Washington State University, USA
Ian Davidson	UC Davis, USA
Bart Goethals	University of Antwerp, Belgium
Tu-Bao Ho	Japan Advanced Institute of Science and Technology, Japan
Xiaohua Hu	Drexel University, USA
Ming Hua	Facebook, USA
Joshua Huang	Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
Hisashi Kashima	University of Tokyo, Japan
Jiuyong Li	University of South Australia, Australia
Ee-Peng Lim	Singapore Management University, Singapore
Chih-Jen Lin	National Taiwan University, Taiwan
Huan Liu	Arizona State University, USA
Nikos Mamoulis	University of Hong Kong, Hong Kong



Wee Keong Ng	Nanyang Technological University, Singapore
Wen-Chih Peng	National Chiao Tung University, Taiwan
P. Reddy	International Institute of Information Technology, Hyderabad (IIIT-H), India
Kyuseok Shim	Seoul National University, Korea
Masashi Sugiyama	Tokyo Institute of Technology, Japan
Pang-Ning Tan	Michigan State University, USA
Hanghang Tong	IBM T. J. Watson Research, USA
Shusaku Tsumoto	Shimane University, Japan
Wei Wang	University of North Carolina at Chapel Hill, USA
Haixun Wang	Microsoft Research Asia, China
Jianyong Wang	Tsinghua University, China
Ji-Rong Wen	Microsoft Research Asia, China
Xing Xie	Microsoft Research Asia, China
Hui Xiong	Rutgers University, USA
Xifeng Yan	UC Santa Barbara, USA
Jieping Ye	Arizona State University
Jeffrey Yu	The Chinese University of Hong Kong, Hong Kong
Chengqi Zhang	University of Technology, Australia
Zhi-Hua Zhou	Nanjing University, China

## Program Committee Members

Hideo Bannai	Kyushu University, Japan
Marut Buranarach	National Electronics and Computer Technology Center, Thailand
Rui Camacho	Universidade do Porto, Portugal
Tru Cao	Ho Chi Minh City University of Technology, Vietnam
Keith Chan	The Hong Kong Polytechnic University, Hong Kong
Muhammad Cheema	University of New South Wales, Australia
Enhong Chen	University of Science and Technology of China, China
Jake Chen	Indiana University-Purdue University Indianapolis, USA
Jian Chen	Southern China University of Technology, China
Ling Chen	University of Technology Sydney, Australia
Shu-Ching Chen	Florida International University, USA
Songcan Chen	Nanjing University of Aeronautics and Astronautics, China
Yi-Ping Phoebe Chen	La Trobe University, Australia
Zheng Chen	Microsoft Research Asia, China

Hong Cheng	The Chinese University of Hong Kong, Hong Kong
Yiu-ming Cheung	Hong Kong Baptist University, Hong Kong
Bruno Cremilleux	Université de Caen, France
Bin Cui	Peking University, China
Alfredo Cuzzocrea	ICAR-CNR and University of Calabria, Italy
Dao-Qing Dai	Sun Yat-Sen University, China
Bolin Ding	Microsoft Research, USA
Wei Ding	University of Massachusetts Boston, USA
Guozhu Dong	Wright State University, USA
Wei Fan	IBM T. J. Watson Research Center, USA
Eibe Frank	University of Waikato, New Zealand
Joao Gama	Universidade do Porto, Portugal
Dragan Gamberger	Rudjer Boskovic Institute, Croatia
Cong Gao	Nanyang Technological University, Singapore
Jun Gao	Peking University, China
Junbin Gao	Charles Sturt University, Australia
Yong Guan	Iowa State University, USA
Ravi Gupta	Anna University, India
Sung-Ho Ha	Kyungpook National University, Korea
Yi Han	National Defence Technology University, China
Choochart Haruechaiy	National Electronics and Computer Technology Center, Thailand
Jingrui He	IBM Research, USA
Qi He	IBM Research, USA
Chin-Kuan Ho	Multimedia University, Malaysia
Kuo-Wei Hsu	National Chengchi University, Taiwan
Jun Huan	University of Kansas, USA
Jin Huang	Southern China Normal University, China
Daisuke Ikeda	Kyushu University, Japan
Akihiro Inokuchi	Osaka University, Japan
Sanjay Jain	National University of Singapore, Singapore
Shuiwang Ji	Old Dominion University, USA
Ruoming Jin	Kent State University, USA
Toshihiro Kamishima	National Institute of Advanced Industrial Science and Technology, Japan
Hung-Yu Kao	National Cheng Kung University, Taiwan
Panagiotis Karras	Rutgers University, USA
George Karypis	University of Minnesota, USA
Hiroyuki Kawano	Nanzan University, Japan
Latifur Khan	University of Texas at Dallas, USA
Hiroyuki Kitagawa	University of Tsukuba, Japan
Irena Koprinska	University of Sydney, Australia
Marzena Kryszkiewicz	Warsaw University of Technology, Poland
Wai Lam	The Chinese University of Hong Kong, Hong Kong

Yue-Shi Lee	Ming Chuan University, Taiwan
Carson K. Leung	University of Manitoba, Canada
Chengkai Li	The University of Texas at Arlington, USA
Chun-hung Li	Hong Kong Baptist University, Hong Kong
Feifei Li	Florida State University, USA
Jinyan Li	University of Technology Sydney, Australia
Ming Li	Nanjing University, China
Xiaoli Li	Institute for Infocomm Research, Singapore
Xue Li	The University of Queensland, Australia
Xuelong Li	University of London, UK
Zhenhui Li	Pennsylvania State University, USA
Kawuu W. Lin	National Kaohsiung University of Applied Sciences, Taiwan
Shou-de Lin	National Taiwan University, Taiwan
Fei Liu	Bosch Research, USA
Qingshan Liu	NLPR Institute of Automation Chinese Academy of Science, China
Shixia Liu	Microsoft Research Asia, China
Tie-Yan Liu	Microsoft Research Asia, China
David Lo	Singapore Management University, Singapore
Woong-Kee Loh	Sungkyul University, South Korea
Chang-Tien Lu	Virginia Polytechnic Institute and State University, USA
Hua Lu	Aalborg University, Denmark
Shuai Ma	Beihang University, China
Marco Maggini	Università degli Studi di Siena, Italy
Tao Mei	Microsoft Research Asia, China
Wagner Meira	Universidade Federal de Minas Gerais, Brazil
Toshiro Minami	Kyushu University Library, Japan
Pabitra Mitra	Indian Institute of Technology Kharagpur, India
Yang-Sae Moon	Kangwon National University, Korea
Yasuhiko Morimoto	Hiroshima University, Japan
Tsuyoshi Murata	Tokyo Institute of Technology, Japan
Richi Nayak	Queensland University of Technologies, Australia
See-Kiong Ng	Institute for Infocomm Research, A*STAR, Singapore
Wilfred Ng	Hong Kong University of Science and Technology, Hong Kong
Tadashi Nomoto	National Institute of Japanese Literature, Japan
Masayuki Numao	Osaka University, Japan
Manabu Okumura	Japan Advanced Institute of Science and Technology, Japan

Salvatore Orlando	University of Venice, Italy
Yonghong Peng	University of Bradford, UK
Jean-Marc Petit	Universite de Lyon, France
Vincenzo Piuri	Università degli Studi di Milano, Italy
Weining Qian	East China Normal University, China
Chotirat Ratanamatan	Chulalongkorn University, Thailand
Chandan Reddy	Wayne State University, USA
Patricia Riddle	University of Auckland, New Zealand
C. Sekhar	Indian Insitute of Technology, India
Hong Shen	Adelaide University, Australia
Jialie Shen	Singapore Management University, Singapore
Yi-Dong Shen	Chinese Academy of Sciences, China
Andrzej Skowron	University of Warsaw, Poland
Mingli Song	Zhejiang University, China
Yizhou Sun	Northeastern University, USA
Thepchai Supnithi	National Electronics and Computer Technology Center, Thailand
David Taniar	Monash University, Australia
Tamir Tassa	The Open University, Israel
Ivor Tsang	Nanyang Technological University, Singapore
Jeffrey Ullman	Stanford University, USA
Marian Vajtersic	University of Salzburg, Austria
Hui Wang	University of Ulster, UK
Jason Wang	New Jersey Science and Technology University, USA
Lipo Wang	Nanyang Technological University, Singapore
Xin Wang	University of Calgary, Canada
Raymond Wong	Hong Kong University of Science and Technology, Hong Kong
Jian Wu	Zhejiang University, China
Junjie Wu	Beihang University, China
Xindong Wu	University of Vermont, USA
Xintao Wu	University of North Carolina at Charlotte, USA
Xiaofeng Meng	Renmin University of China, China
Seiji Yamada	National Institute of Informatics, Japan
Min Yao	Zhejiang University, China
Mi-Yen Yeh	Academia Sinica, Taiwan
Dit-Yan Yeung	Hong Kong University of Science and Technology, China
Jian Yin	Sun Yat-Sen University, China
Jin Soung Yoo	IUPU, USA
Kennichi Yoshida	University of Tsukuba, Japan
Tetsuya Yoshida	Hokkaido University, Japan
Hwanjo Yu	Pohang University of Science and Technology, South Korea

XIV Organization

Yifeng Zeng	Teesside University, UK
Bo Zhang	Tsinghua University, China
Daoqiang Zhang	Nanjing University of Aeronautics and Astronautics, China
Du Zhang	California State University, USA
Harry Zhang	University of New Brunswick, Canada
Junping Zhang	Fudan University, China
Mengjie Zhang	Victoria University of Wellington, New Zealand
Shichao Zhang	University of Technology, Australia
Wenjie Zhang	University of New South Wales, Australia
Ying Zhang	University of New South Wales, Australia
Zhongfei Zhang	Binghamton University, USA
Zili Zhang	Deakin University, Australia
Peixiang Zhao	Florida State University, USA
Yu Zheng	Microsoft Research Asia, China
Bin Zhou	University of Maryland Baltimore County, USA
Shuigeng Zhou	Fudan University, China
Wenjun Zhou	University of Tennessee - Knoxville, USA
Feida Zhu	Singapore Management University, Singapore
Xingquan Zhu	University of Technology Sydney, Australia

## Table of Contents – Part II

ProCF: Probabilistic Collaborative Filtering for Reciprocal Recommendation .....	1
<i>Xiongcai Cai, Michael Bain, Alfred Krzywicki, Wayne Wobcke, Yang Sok Kim, Paul Compton, and Ashesh Mahidadia</i>	
Product and User Dependent Social Network Models for Recommender Systems .....	13
<i>Min Li, Zhiwei Jiang, Bin Luo, Jiubin Tang, Qing Gu, and Daoxu Chen</i>	
Learning Representative Nodes in Social Networks .....	25
<i>Ke Sun, Donn Morrison, Eric Bruno, and Stéphane Marchand-Maillet</i>	
Tracing Influential Nodes in a Social Network with Competing Information .....	37
<i>Bolei Zhang, Zhuzhong Qian, Xiaoliang Wang, and Sanglu Lu</i>	
ViStrucLizer: A Structural Visualizer for Multi-dimensional Social Networks .....	49
<i>Bing Tian Dai, Agus Trisnajaya Kwee, and Ee-Peng Lim</i>	
Influential Nodes in a One-Wave Diffusion Model for Location-Based Social Networks .....	61
<i>Hao-Hsiang Wu and Mi-Yen Yeh</i>	
Modeling Social Information Learning among Taxi Drivers .....	73
<i>Siyuan Liu, Ramayya Krishnan, Emma Brunskill, and Lionel M. Ni</i>	
Mining User Interests from Information Sharing Behaviors in Social Media .....	85
<i>Tingting Wang, Hongyan Liu, Jun He, and Xiaoyong Du</i>	
Anonymization for Multiple Released Social Network Graphs .....	99
<i>Chih-Jui Lin Wang, En Tzu Wang, and Arbee L.P. Chen</i>	
A New, Fast and Accurate Algorithm for Hierarchical Clustering on Euclidean Distances .....	111
<i>Elio Masciari, Giuseppe Massimiliano Mazzeo, and Carlo Zaniolo</i>	

Clustering Patient Medical Records via Sparse Subspace Representation . . . . .	123
<i>Budhaditya Saha, Duc-Son Pham, Dinh Phung, and Svetha Venkatesh</i>	
A Unified Metric for Categorical and Numerical Attributes in Data Clustering . . . . .	135
<i>Yiu-ming Cheung and Hong Jia</i>	
An Extension of the Infinite Relational Model Incorporating Interaction between Objects . . . . .	147
<i>Iku Ohama, Hiromi Iida, Takuya Kida, and Hiroki Arimura</i>	
Density-Based Clustering Based on Hierarchical Density Estimates . . . . .	160
<i>Ricardo J.G.B. Campello, Davoud Moulavi, and Joerg Sander</i>	
Stock Trend Prediction by Classifying Aggregative Web Topic-Opinion . . . . .	173
<i>Li Xue, Yun Xiong, Yangyong Zhu, Jianfeng Wu, and Zhiyuan Chen</i>	
The Role of Hubs in Cross-Lingual Supervised Document Retrieval . . . . .	185
<i>Nenad Tomašev, Jan Rupnik, and Dunja Mladenić</i>	
Text Document Topical Recursive Clustering and Automatic Labeling of a Hierarchy of Document Clusters . . . . .	197
<i>Xiaoxiao Li, Jiyang Chen, and Osmar Zaiane</i>	
Query-Document Relevance Topic Models . . . . .	209
<i>Meng-Sung Wu, Chia-Ping Chen, and Hsin-Min Wang</i>	
A Two-Stage Approach for Generating Topic Models . . . . .	221
<i>Yang Gao, Yue Xu, Yuefeng Li, and Bin Liu</i>	
Effective Top-Down Active Learning for Hierarchical Text Classification . . . . .	233
<i>Xiao Li, Charles X. Ling, and Huaimin Wang</i>	
Forgetting Word Segmentation in Chinese Text Classification with L1-Regularized Logistic Regression . . . . .	245
<i>Qiang Fu, Xinyu Dai, Shujian Huang, and Jiajun Chen</i>	
CREST: Cluster-based Representation Enrichment for Short Text Classification . . . . .	256
<i>Zichao Dai, Aixin Sun, and Xu-Ying Liu</i>	
Cross Language Prediction of Vandalism on Wikipedia Using Article Views and Revisions . . . . .	268
<i>Khôi-Nguyen Tran and Peter Christen</i>	

An Optimized Cost-Sensitive SVM for Imbalanced Data Learning . . . . .	280
<i>Peng Cao, Dazhe Zhao, and Osmar Zaiane</i>	
A Positive-biased Nearest Neighbour Algorithm for Imbalanced Classification . . . . .	293
<i>Xiuzhen Zhang and Yuxuan Li</i>	
Class Based Weighted K-Nearest Neighbor over Imbalance Dataset . . . . .	305
<i>Harshit Dubey and Vikram Pudi</i>	
ProWSyn: Proximity Weighted Synthetic Oversampling Technique for Imbalanced Data Set Learning . . . . .	317
<i>Sukarna Barua, Md. Monirul Islam, and Kazuyuki Murase</i>	
Differential Privacy Preserving Spectral Graph Analysis . . . . .	329
<i>Yue Wang, Xintao Wu, and Leting Wu</i>	
Sorted Nearest Neighborhood Clustering for Efficient Private Blocking . . . . .	341
<i>Dinusha Vatsalan and Peter Christen</i>	
On Linear Refinement of Differential Privacy-Preserving Query Answering . . . . .	353
<i>Xiaowei Ying, Xintao Wu, and Yue Wang</i>	
A Coupled Clustering Approach for Items Recommendation . . . . .	365
<i>Yonghong Yu, Can Wang, Yang Gao, Longbing Cao, and Xixi Chen</i>	
Location Recommendation Based on Periodicity of Human Activities and Location Categories . . . . .	377
<i>Seyyed Mohammadreza Rahimi and Xin Wang</i>	
Top-N Recommendations by Learning User Preference Dynamics . . . . .	390
<i>Yongli Ren, Tianqing Zhu, Gang Li, and Wanlei Zhou</i>	
Semantic Title Evaluation and Recommendation Based on Topic Models . . . . .	402
<i>Huidong Jin, Lijiu Zhang, and Lan Du</i>	
Video Quality Prediction over Wireless 4G . . . . .	414
<i>Chun Pong Lau, Xiangliang Zhang, and Basem Shihada</i>	
A Self-immunizing Manifold Ranking for Image Retrieval . . . . .	426
<i>Jun Wu, Yidong Li, Songhe Feng, and Hong Shen</i>	
Low-Rank Matrix Recovery with Discriminant Regularization . . . . .	437
<i>Zhonglong Zheng, Haixin Zhang, Jiong Jia, Jianmin Zhao, Li Guo, Fangmei Fu, and Mudan Yu</i>	



Multi-Manifold Ranking: Using Multiple Features for Better Image Retrieval . . . . .	449
<i>Yang Wang, Muhammad Aamir Cheema, Xuemin Lin, and Qing Zhang</i>	
One Pass Concept Change Detection for Data Streams . . . . .	461
<i>Sripirakas Sakthithasan, Russel Pears, and Yun Sing Koh</i>	
Incremental Mining of Significant URLs in Real-Time and Large-Scale Social Streams . . . . .	473
<i>Cheng-Ying Liu, Chi-Yao Tseng, and Ming-Syan Chen</i>	
A Concept-Drifting Detection Algorithm for Categorical Evolving Data . . . . .	485
<i>Fuyuan Cao and Joshua Zhexue Huang</i>	
Framework for Storing and Processing Relational Entities in Stream Mining . . . . .	497
<i>Pawel Matuszyk and Myra Spiliopoulou</i>	
Discovering Semantics from Multiple Correlated Time Series Stream . . . . .	509
<i>Zhi Qiao, Guangyan Huang, Jing He, Peng Zhang, Li Guo, Jie Cao, and Yanchun Zhang</i>	
Matrix Factorization With Aggregated Observations . . . . .	521
<i>Yoshifumi Aimoto and Hisashi Kashima</i>	
An Approach to Identifying False Traces in Process Event Logs . . . . .	533
<i>Hedong Yang, Lijie Wen, and Jianmin Wang</i>	
Split-Merge Augmented Gibbs Sampling for Hierarchical Dirichlet Processes . . . . .	546
<i>Santu Rana, Dinh Phung, and Svetha Venkatesh</i>	
Adaptive Temporal Entity Resolution on Dynamic Databases . . . . .	558
<i>Peter Christen and Ross W. Gayler</i>	
Fuzzy Multi-Sphere Support Vector Data Description . . . . .	570
<i>Trung Le, Dat Tran, and Wanli Ma</i>	
<b>Erratum</b>	
A Coupled Clustering Approach for Items Recommendation . . . . .	E1
<i>Yonghong Yu, Can Wang, Yang Gao, Longbing Cao, and Xixi Chen</i>	
<b>Author Index</b> . . . . .	583

# Table of Contents – Part I

Discovering Local Subgroups, with an Application to Fraud Detection .....	1
<i>Rob M. Konijn, Wouter Duivesteijn, Wojtek Kowalczyk, and Arno Knobbe</i>	
PUF-Tree: A Compact Tree Structure for Frequent Pattern Mining of Uncertain Data .....	13
<i>Carson Kai-Sang Leung and Syed Khairuzzaman Tanbeer</i>	
Frequent Pattern Mining in Attributed Trees .....	26
<i>Claude Pasquier, Jérémy Sanhes, Frédéric Flouvat, and Nazha Selmaoui-Folcher</i>	
Mining Frequent Patterns from Human Interactions in Meetings Using Directed Acyclic Graphs .....	38
<i>Anna Fariha, Chowdhury Farhan Ahmed, Carson Kai-Sang Leung, S.M. Abdullah, and Longbing Cao</i>	
ClaSP: An Efficient Algorithm for Mining Frequent Closed Sequences...	50
<i>Antonio Gomariz, Manuel Campos, Roque Marin, and Bart Goethals</i>	
Efficient Mining of Contrast Patterns on Large Scale Imbalanced Real-Life Data .....	62
<i>Jinju Li, Can Wang, Wei Wei, Mu Li, and Chunming Liu</i>	
Online Cross-Lingual PLSI for Evolutionary Theme Patterns Analysis .....	74
<i>Xin Xin, Kun Zhuang, Ying Fang, and Heyan Huang</i>	
F-Trail: Finding Patterns in Taxi Trajectories .....	86
<i>Yasuko Matsubara, Lei Li, Evangelos Papalexakis, David Lo, Yasushi Sakurai, and Christos Faloutsos</i>	
Mining Appliance Usage Patterns in Smart Home Environment .....	99
<i>Yi-Cheng Chen, Yu-Lun Ko, Wen-Chih Peng, and Wang-Chien Lee</i>	
Computational Models of Stress in Reading Using Physiological and Physical Sensor Data .....	111
<i>Nandita Sharma and Tom Gedeon</i>	
Latent Patient Profile Modelling and Applications with Mixed-Variate Restricted Boltzmann Machine .....	123
<i>Tu Dinh Nguyen, Truyen Tran, Dinh Phung, and Svetha Venkatesh</i>	

MassBayes: A New Generative Classifier with Multi-dimensional Likelihood Estimation . . . . .	136
<i>Sunil Aryal and Kai Ming Ting</i>	
Fast and Effective Single Pass Bayesian Learning . . . . .	149
<i>Nayyar A. Zaidi and Geoffrey I. Webb</i>	
Sparse Reductions for Fixed-Size Least Squares Support Vector Machines on Large Scale Data . . . . .	161
<i>Raghvendra Mall and Johan A.K. Suykens</i>	
Discovery of Regional Co-location Patterns with $k$ -Nearest Neighbor Graph . . . . .	174
<i>Feng Qian, Kevin Chiew, Qinming He, Hao Huang, and Lianhang Ma</i>	
Spectral Decomposition for Optimal Graph Index Prediction . . . . .	187
<i>Liyan Song, Yun Peng, Byron Choi, Jianliang Xu, and Bingsheng He</i>	
Patterns amongst Competing Task Frequencies: Super-Linearities, and the Almond-DG Model . . . . .	201
<i>Danai Koutra, Vasileios Koutras, B. Aditya Prakash, and Christos Faloutsos</i>	
Node Classification in Social Network via a Factor Graph Model . . . . .	213
<i>Huan Xu, Yujiu Yang, Liangwei Wang, and Wenhuan Liu</i>	
Fast Graph Stream Classification Using Discriminative Clique Hashing . . . . .	225
<i>Lianhua Chi, Bin Li, and Xingquan Zhu</i>	
Mining Interesting Itemsets in Graph Datasets . . . . .	237
<i>Boris Cule, Bart Goethals, and Tayena Hendrickx</i>	
Robust Synchronization-Based Graph Clustering . . . . .	249
<i>Junming Shao, Xiao He, Qinli Yang, Claudia Plant, and Christian Böhm</i>	
Efficient Mining of Combined Subspace and Subgraph Clusters in Graphs with Feature Vectors . . . . .	261
<i>Stephan Günnemann, Brigitte Boden, Ines Färber, and Thomas Seidl</i>	
Exploiting Temporal Information in a Two-Stage Classification Framework for Content-Based Depression Detection . . . . .	276
<i>Yu-Chun Shen, Tsung-Ting Kuo, I-Ning Yeh, Tzu-Ting Chen, and Shou-De Lin</i>	
EEG-Based Person Verification Using Multi-Sphere SVDD and UBM . . . . .	289
<i>Phuoc Nguyen, Dat Tran, Trung Le, Xu Huang, and Wanli Ma</i>	

Measuring Reproducibility of High-Throughput Deep-Sequencing Experiments Based on Self-adaptive Mixture Copula . . . . .	301
<i>Qian Zhang, Junping Zhang, and Chenghai Xue</i>	
Mining Representative Movement Patterns through Compression . . . . .	314
<i>Phan Nhat Hai, Dino Ienco, Pascal Poncelet, and Maquelonne Teisseire</i>	
NARGES: Prediction Model for Informed Routing in a Communications Network . . . . .	327
<i>Hooman Homayounfard, Paul J. Kennedy, and Robin Braun</i>	
Mining Usage Traces of Mobile Apps for Dynamic Preference Prediction . . . . .	339
<i>Zhung-Xun Liao, Wen-Chih Peng, and Philip S. Yu</i>	
Leveraging Hybrid Citation Context for Impact Summarization . . . . .	354
<i>Po Hu, Yujing Guo, Dong-Hong Ji, and Jiacong He</i>	
Optimal Allocation of High Dimensional Assets through Canonical Vines . . . . .	366
<i>Wei Wei, Jinyan Li, Longbing Cao, Jingguang Sun, Chunming Liu, and Mu Li</i>	
Inducing Context Gazetteers from Encyclopedic Databases for Named Entity Recognition . . . . .	378
<i>Han-Cheol Cho, Naoaki Okazaki, and Kentaro Inui</i>	
An Optimization Method for Proportionally Diversifying Search Results . . . . .	390
<i>Lin Wu, Yang Wang, John Shepherd, and Xiang Zhao</i>	
Joint Naïve Bayes and LDA for Unsupervised Sentiment Analysis . . . . .	402
<i>Yong Zhang, Dong-Hong Ji, Ying Su, and Hongmiao Wu</i>	
An Unsupervised Learning Model to Perform Side Channel Attack . . . . .	414
<i>Jung-Wei Chou, Min-Huang Chu, Yi-Lin Tsai, Yun Jin, Chen-Mou Cheng, and Shou-De Lin</i>	
Decisive Supervised Learning . . . . .	426
<i>Eileen A. Ni, Da Kuang, and Charles X. Ling</i>	
Learning Overlap Optimization for Domain Decomposition Methods . . . . .	438
<i>Steven Burrows, Jörg Frochte, Michael Völske, Ana Belén Martínez Torres, and Benno Stein</i>	
CLUEKR : CLUstering Based Efficient <i>k</i> NN Regression . . . . .	450
<i>Harshit Dubey and Vikram Pudi</i>	

AREM: A Novel Associative Regression Model Based on EM Algorithm . . . . .	459
<i>Zhonghua Jiang and George Karypis</i>	
One-Class Transfer Learning with Uncertain Data . . . . .	471
<i>Bo Liu, Philip S. Yu, Yanshan Xiao, and Zhifeng Hao</i>	
Time Series Forecasting Using Distribution Enhanced Linear Regression . . . . .	484
<i>Goce Ristanoski, Wei Liu, and James Bailey</i>	
Twin Bridge Transfer Learning for Sparse Collaborative Filtering . . . . .	496
<i>Jiangfeng Shi, Mingsheng Long, Qiang Liu, Guiguang Ding, and Jianmin Wang</i>	
Dimensionality Reduction with Dimension Selection . . . . .	508
<i>Yi Guo, Junbin Gao, and Feng Li</i>	
Multi-View Visual Classification via a Mixed-Norm Regularizer . . . . .	520
<i>Xiaofeng Zhu, Zi Huang, and Xindong Wu</i>	
Mining Specific Features for Acquiring User Information Needs . . . . .	532
<i>Abdulmohsen Algarni and Yuefeng Li</i>	
Ensemble-Based Wrapper Methods for Feature Selection and Class Imbalance Learning . . . . .	544
<i>Pengyi Yang, Wei Liu, Bing B. Zhou, Sanjay Chawla, and Albert Y. Zomaya</i>	
Exploring Groups from Heterogeneous Data via Sparse Learning . . . . .	556
<i>Huawen Liu, Jiuyong Li, Lin Liu, Jixue Liu, Ivan Lee, and Jianmin Zhao</i>	
Multiplex Topic Models . . . . .	568
<i>Juan Yang, Jia Zeng, and William K. Cheung</i>	
Integrating Clustering and Ranking on Hybrid Heterogeneous Information Network . . . . .	583
<i>Ran Wang, Chuan Shi, Philip S. Yu, and Bin Wu</i>	
Learning from Multiple Observers with Unknown Expertise . . . . .	595
<i>Han Xiao, Huang Xiao, and Claudia Eckert</i>	
<b>Author Index</b> . . . . .	607