

Lecture Notes in Artificial Intelligence 7104

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*

Longbing Cao Joshua Zhexue Huang  
James Bailey Yun Sing Koh Jun Luo (Eds.)

# New Frontiers in Applied Data Mining

PAKDD 2011 International Workshops  
Shenzhen, China, May 24-27, 2011  
Revised Selected Papers

## 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

Longbing Cao  
University of Technology Sydney, NSW, Australia  
E-mail: longbing.cao@uts.edu.au

Joshua Zhexue Huang  
Jun Luo  
Chinese Academy of Sciences, SIAT, Shenzhen, China  
E-mail: {zhexue.huang, jun.luo}@siat.ac.cn

James Bailey  
The University of Melbourne, VIC, Australia  
E-mail: baileyj@unimelb.edu.au

Yun Sing Koh  
The University of Auckland, New Zealand  
E-mail: ykoh@cs.auckland.ac.nz

ISSN 0302-9743  
ISBN 978-3-642-28319-2  
DOI 10.1007/978-3-642-28320-8  
Springer Heidelberg Dordrecht London New York

e-ISSN 1611-3349  
e-ISBN 978-3-642-28320-8

Library of Congress Control Number: 2012930911

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

LNCS Sublibrary: SL 7 – Artificial Intelligence

© Springer-Verlag Berlin Heidelberg 2012

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)

# Preface

The 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining in Shenzhen, China, hosted five workshops that allowed researchers and participants to discuss emerging techniques and their application domains extending to previously unexplored areas.

The workshops were selected competitively following a call for workshop proposals. Three of them were sequel workshops (BI, QIMIE, DMHM) that were held in the last few years, while two were new (BDM, AI-TCM), exhibiting both the maturity and innovation of the workshops. The following workshops were organized:

- International Workshop on Behavior Informatics (BI 2011) chaired by Longbing Cao, Jaideep Srivastava, Graham Williams, and Hiroshi Motoda. The main goal of this workshop is to provide an international forum for researchers and industry practitioners to share their ideas, original research results, as well as potential challenges and prospects encountered in behavior informatics.
- Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models Workshop (QIMIE 2011) chaired by Stéphane Lallich and Philippe Lenca. The main focus of QIMIE 2011 is on the theory, the techniques and the practices that can ensure the discovered knowledge of a data mining process is of quality.
- Workshop on Biologically Inspired Techniques for Data Mining (BDM 2011) chaired by Shafiq Alam Burki and Gillian Dobbie. The workshop highlights a relatively new but fast-growing area of data mining which is based on optimization techniques from biological behavior of animals, insects, cultures, social behaviors and biological evolution.
- Workshop on Advances and Issues in Traditional Chinese Medicine Clinical Data Mining (AI-TCM 2011) chaired by Josiah Poon, Xuezhong Zhou, and Junbin Gao. This workshop focuses on usage of data mining such that it becomes a complementary tool to assist traditional chinese medicine clinical research.
- Second Workshop on Data Mining for Healthcare Management (DMHM 2011) chaired by Prasanna Desikan, Jaideep Srivastava, and Ee-Peng Lim. This workshop provides a common platform for discussion of challenging issues and potential techniques in this emerging field of data mining for health care management.

This volume contains the proceedings of the five workshops. Also included are papers from the First PAKDD Doctoral Symposium on Data Mining (DSDM'11). We hope that the published papers propel further interest in the growing field of knowledge discovery in databases (KDD).

Setting up workshops such as these takes a lot of effort. We would like to thank the Program Committee Chairs and their Program Committee members for their time and effort spent to guarantee the high quality of the program.

October 2011

James Bailey  
University of Melbourne, Australia

Yun Sing Koh  
University of Auckland, New Zealand  
(PAKDD 2011 Workshop Co-chairs)

# International Workshop on Behavior Informatics (BI 2011)

## PC Chairs' Message

Due to the behavior implication in normal transactional data, the requirement of deep and quantitative behavior analysis has outstripped the capability of traditional methods and techniques in behavioral sciences. It is more imperative than ever to develop new behavioral analytic technologies that can derive an in-depth understanding of human behaviors beyond the demographic and historical tracking. This leads to the emergence of inter-disciplinary *behavior representation, modeling, mining, analysis and management* (namely, *behavior informatics*). The aim of the Behavior Informatics Workshop is to provide an international forum for researchers and industry practitioners to share their ideas, original research results, as well as potential challenges and prospects encountered in behavior informatics.

Following the call for papers, BI 2011 attracted 40 submissions from 12 different countries, and accepted 16 of them after a double-blind review by at least 3 reviewers for each paper.

The selected papers include the latest work in social network mining, such as user profile modeling, Web user clustering, emotional analysis etc. On behavior mining, papers are related to community detection, group recommendation, and user behavior patterns, etc. Moreover, pattern discovery and clustering are other areas where the authors have utilized and extended behavior informatics.

The Workshop also featured two invited talks: Jeffery Xu Yu on Graph Behaviors: Discovering Patterns When a Graph Evolves, and Jie Tang on Social Prediction: Can We Predict Users' Action and Emotions?

We are very grateful to the General Chair of BI 2011, Philip S. Yu at the University of Illinois at Chicago, who provided continuous support and guidance for the workshop's success. We are particularly thankful to Gang Li at Deakin University, Australia, the Organizing Chair of BI 2011, for the great efforts he made throughout all the phases of organizing BI 2011. Thanks to the authors who made this workshop possible by submitting their work and responding positively to the changes suggested by our reviewers regarding their work. We are also thankful to our Program Committee who dedicated their time and provided us with valuable suggestions and timely reviews. We wish to express our gratitude to the PAKDD 2011 Workshop Chairs and PAKDD 2011 conference organizers who provided us with fantastic support that made this workshop very successful.

Readers are referred to an edited book on *Behavior Computing: Modeling, Analysis, Mining and Decision*, edited by Longbing Cao and Philip S. Yu, published by Springer. Another resource is the Special Issue on Behavior Computing with *Knowledge and Information Systems: An International Journal*, edited by Longbing Cao, Philip S. Yu and Hiroshi Motoda.

Longbing Cao  
Jaideep Srivastava  
Graham Williams  
Hiroshi Motoda

# International Workshop on Behavior Informatics (BI 2011)

## Organization

### General Co-chair

Philip S. Yu

University of Illinois at Chicago, USA

### Workshop Chairs

Longbing Cao

University of Technology Sydney, Australia

Jaideep Srivastava

University of Minnesota, USA

Graham Williams

Australian Taxation Office, Australia

Hiroshi Motoda

Osaka University and AFOSR/AOARD, Japan

### Organizing Chair

Gang Li

Deakin University, Australia

### Program Committee

Sviatoslav Braynov, USA

Guozhu Dong, USA

Gian Luca Foresti, Italy

Xin Geng, P.R. China

Björn Gottfried, Germany

Ken Kaneiwa, Japan

Dionysios Kehagias, Greece

Kazuhiro Kuwabara, Japan

Gang Li, Australia

Ee-Peng Lim, Singapore

Weicheng Lin, Taiwan

Wenjia Niu, P.R. China

Zbigniew Ras, USA

Yasufumi Takama, Japan

David Taniar, Australia

Karl Tuyls, The Netherlands

Jinlong Wang, P.R. China

Lipo Wang, Singapore

Xiaogang Wang, P.R. China

Xiaofeng Wang, P.R. China

Jeffrey Yu, P.R. China

Daniel Zeng, USA



# Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models Workshop (QIMIE 2011)

## PC Chairs' Message

*The Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models Workshop (QIMIE)* focusses on the theory, the techniques and the practices that can ensure the discovered knowledge of a data mining process is of quality. Following QIMIE 2009/PAKDD 2009, QIMIE 2011 was organized in association with PAKDD 2011 (Shenzen, China).

QIMIE 2011 would not have been possible without the work of many people and organizations. We wish to express our gratitude to: Telecom Bretagne, the University of Lyon, the PAKDD 2011 Workshop Chairs (James Bailey and Yun Sing Koh), the Chairs and Co-chairs of PAKDD 2011 (Philip S. Yu, Jianping Fan, David Cheung), the PAKDD 2011 Program Committee Co-chairs (Joshua Huang, Longbing Cao, Jaideep Srivastava), and the QIMIE 2011 Program Committee members. Last but not least, we would like to thank all authors of the submitted papers, the Session Chairs and Longbing Cao for his keynote talk.

Each submission was reviewed by at least three members of the Program Committee. The papers presented in these proceedings were selected after a rigorous review process. The QIMIE 2011 program included one keynote speaker and nine regular paper presentations.

Actionable knowledge discovery is considered as one of the great challenges of the next-generation of knowledge discovery in database studies. The keynote talk by Longbing Cao discussed a general evaluation framework to measure the actionability of knowledge discovered, which covers both technical and business performance evaluation. Metrics and strategies for supporting the extraction of actionable knowledge were discussed.

The regular papers were divided into three sessions, namely, *Clustering*, *Data Structure*, and *Patterns and Rules*.

*Clustering Session* – Ivanescu et al. propose ClasSi, a new ranking correlation coefficient which deals with class label rankings and employs a class distance function to model the similarities between classes. Estivill-Castro introduces a notion of instance easiness to supervised learning and links the validity of a clustering to how its output constitutes an easy instance for supervised learning. An alternative approach for clustering quality evaluation based on the unsupervised measures of Recall, Precision, and F-measure exploiting the descriptors of the data associated with the obtained clusters is proposed by Lamirel et al.

*Data Structure Session* – Hadzic presents a novel structure-preserving way for representing tree-structured document instances as records in a standard

flat data structure to enable the applicability of a wider range of data-mining techniques. To handle fragmentary duplicate pages on the Internet, Fan and Huang fuse some “state-of-the-art” algorithms to reach a better performance.

*Patterns and Rules Session* – Yang et al. propose an evolutionary method to search for interesting association rules. An efficient model based on the notion of multiple constraints is constructed by Surana et al. The periodic-frequent patterns discovered with this model satisfy a downward closure property. Hence, periodic frequent patterns can be efficiently discovered. Wu et al. propose Ap-epi, an algorithm which discovers minimal occurrences of serial episode and NOE-WinMiner which discovers non-overlapping episodes. To address the problem of assessing the information conveyed by a finite discrete probability distribution, Garriga elaborates a measure of certainty which includes a native value for the uncertainty related to unseen events.

Philippe Lenca  
Stéphane Lallich

# Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models Workshop (QIMIE 2011)

## Organization

### Workshop Chairs

Stéphane Lallich  
Philippe Lenca

Université Lyon 2, France  
Lab-STICC, Telecom Bretagne, France

### Program Committee

Hidenao Abe, Japan  
Komate Amphawan, Thailand  
Jérôme Azé, France  
José L. Balezar, Spain  
Bruno Crémilleux, France  
Sven Crone, UK  
Jean Diatta, La Réunion  
Thanh-Nghi Do, Vietnam  
Joao Gama, Portugal  
Ricard Gavaldá, Spain  
Salvatore Gréco, Italy  
Fabrice Guillet, France  
Michael Hahsler, USA  
Shanmugasundaram Hariharan, India  
Martin Holena, Czech Republic  
William Klement, Canada  
Stéphane Lallich, France  
Ludovic Lebart, France

Philippe Lenca, France  
Yannick Le Bras, France  
Ming Li, China  
Patrick Meyer, France  
Amadeo Napoli, France  
David Olson, USA  
Zbigniew Ras, USA  
Jan Rauch, Czech Republic  
Gilbert Ritschard, Switzerland  
Robert Stahlbock, Germany  
Athasit Surarerks, Thailand  
Izabela Szczech, Poland  
Shusaku Tsumoto, Japan  
Kitsana Waiyamai, Thailand  
Dianhui Wang, Australia  
Gary Weiss, USA  
Takahira Yamaguchi, Japan  
Min-Ling Zhang, China

# Workshop on Biologically Inspired Techniques for Data Mining (BDM 2011)

## PC Chairs' Message

For the last few years, biologically inspired data-mining techniques have been intensively used in different data-mining applications such as data clustering, classification, association rule mining, sequential pattern mining, outlier detection, feature selection and bioinformatics. The techniques include neural networks, evolutionary computation, fuzzy systems, genetic algorithms, ant colony optimization, particle swarm optimization, artificial immune system, culture algorithms, social evolution, and artificial bee colony optimization. A huge increase in the number of papers published in the area has been observed in the last decade. Most of these techniques use optimization to speed up the data-mining process and improve the quality of patterns mined from the data.

The aim of the workshop is to highlight the current research related to biologically inspired techniques in different data-mining domains and their implementation in real-life data-mining problems. The workshop provides a platform to researchers from computational intelligence and evolutionary computation and other biologically inspired techniques to get feedback on their work from other data-mining perspectives such as statistical data mining, AI- and machine learning-based data mining.

Following the call for papers BDM 2011 attracted 16 submissions from 8 different countries, with 8 of them accepted after a double-blind review by at least 3 reviewers. The overall acceptance rate for the workshop was 50%.

The selected papers highlight work in PSO-based data clustering and recommender system, PSO-based association rule mining, and genetic algorithm and neural networks-based training schemes for imbalanced classification problems. Another subject covered is the discovery of microRNA precursors in plant genomes based on an SVM detector. Feature selection is another area where the authors proposed a clustering-based framework formed by several unsupervised feature-selection algorithms.

We are thankful to Ajith Abraham, General Chair of BDM 2011, for his constant motivation and guidance throughout all the phases of organizing BDM 2011. Thanks to the authors who made this workshop possible by submitting their work and responding positively to the changes suggested by our reviewers. We are also thankful to our Program Committee who dedicated their time and provided us with valuable suggestions and timely reviews. We wish to express our gratitude to the Workshop Chairs who were always available to answer our queries and provided us with everything we needed to put this workshop together.

Shafiq Alam  
Gillian Dobbie

# Workshop on Biologically Inspired Techniques for Data Mining (BDM 2011)

## Organization

### General Co-chair

Ajith Abraham

Machine Intelligence Research Labs  
(MIR Labs), Scientific Network for  
Innovation and Research Excellence  
(SNIRE), USA

### Workshop Chairs

Shafiq Alam Burki

University of Auckland, New Zealand

Gillian Dobbie

University of Auckland, New Zealand

### Program Committee

Patricia Riddle, New Zealand

Azzam ul Assar, Pakistan

Mengjie Zhang, New Zealand

Kamran Shafi, Australia

Stephen Chen, USA

Kouroush Neshatian, New Zealand

Will Browne, New Zealand

Ganesh K. Venayagamoorthy, USA

Yanjun Yan, USA

Asifullah Khan, Pakistan

Ke Geng, New Zealand

Ming Li, China

Ismail Khalil, Austria

David Taniar, Australia

Redda Alhaj, Canada

Lean Yu, China

Fatos Xhafa, Spain

Xiao-Zhi Gao, Finland

Emilio Corchado, Spain

Michela Antonelli, Italy

Khalid Saeed, Poland

# Workshop on Advances and Issues in Traditional Chinese Medicine Clinical Data Mining (AI-TCM 2011)

## PC Chairs' Message

Traditional Chinese medicine, also known as TCM, has a long history and is considered as one of the major medical approaches in China. Different from its western counterpart, TCM embraces a holistic rather than a reductionist approach. While medications in western medicine often strive to avoid drug interaction, TCM works oppositely by exploiting the useful interaction among multiple herbs that are used in a formula (prescription). The efficacy of a TCM medication derives from the interaction of these herbs. Furthermore, while western medicine has only recently begun to popularize the prescription of personalized medicine according to a patient's genome, Chinese medical doctors have long been practicing this approach. In TCM practice, every prescription is a clinical trial from the viewpoint of western medical practice. In terms of clinical study, TCM clinical practice is a kind of real-world clinical trial, in which all the practical treatments aim to maximize the effectiveness. Therefore, the clinical data generated in the clinical operations are important to hypothesis generation and clinical guideline development.

Historically in TCM, all this knowledge was accumulated and formulated through experience rather than the support of proper scientific research. But we have witnessed a significant change in recent years. There is currently a strong push in the globalization of TCM and practice. This movement encourages the use of different techniques, such as statistical and data-mining methods, to find the underlying biomedical principles and theories, and to unlock the secrets of this "ancient" conservative domain of TCM.

Massive TCM clinical data are now being collected at an increasing pace. Data mining can help unveil the hidden precious knowledge, such that it becomes a complementary tool to assist TCM clinical research. There are many aspects in TCM, e.g., theory discovery, therapy methodology analysis, symptom detection, prescription analysis, herbal drug design, and intelligent diagnosis. It is anticipated that the sharing of experiences will provide new ideas in research directions and give the stimulus for creative breakthroughs. Furthermore, in recent years, many researchers in computer science and statistics have joined in the TCM data-mining field.

Papers in this workshop explored various issues in the field including the exploration of the regularities in the use of herbs and herb-herb interaction in TCM, which are more TCM-focused. There were also papers on other important topics, namely, the choice of interestingness function and feature representation in TCM, which are more data-mining driven. Hence, the workshop was well-balanced between TCM and data mining topics, and they all led to a good and productive discussion.

Josiah Poon  
Xuezhong Zhou  
Junbin Gao

# Workshop on Advances and Issues in Traditional Chinese Medicine Clinical Data Mining (AI-TCM 2011)

## Organization

### Workshop Chairs

Josiah Poon  
Xuezhong Zhou  
Junbin Gao

University of Sydney, Australia  
Beijing Jiaotong University, China  
Charles Sturt University, Australia

### Program Committee

Zhiwei Cao, China  
Huajun Chen, China  
Jianxin Chen, China  
Tao Chen, China  
Yi Feng, China  
Li-Yun He, China  
Caiyan Jia, China  
William Jia, Canada  
Paul Kwan, Australia  
Guozheng Li, China

Shao Li, China  
Clement Loy, Australia  
Yonghong Peng, UK  
Georg Peters, Germany  
Simon Poon, Australia  
Qun Shao, UK  
Daniel Sze, Hong Kong  
Nevin Zhang, Hong Kong  
Runshun Zhang, China



# Second Workshop on Data Mining for Healthcare Management (DMHM 2011)

## PC Chairs' Message

Data mining for healthcare management (DMHM) has been instrumental in detecting patterns of diagnosis, decisions and treatments in healthcare. Data mining has aided in several aspects of healthcare management including disease diagnosis, decision-making for treatments, medical fraud prevention and detection, fault detection of medical devices, healthcare quality improvement strategies and privacy. DMHM is an emerging field where researchers from both academia and industry have recognized the potential of its impact on improved healthcare by discovering patterns and trends in large amounts of complex data generated by healthcare transactions.

This workshop was the second in the series of workshops on Data Mining for Healthcare Management held at PAKDD. The emerging interest in this area has led to the success of these workshops. The workshop served as a critical forum for integrating various research challenges in this domain and promoted collaboration among researchers from academia and industry to enhance the state of the art and help define a clear path for future research in this emerging area.

In response to the call for papers, DMHM 2011 received 9 contributions. We would like to thank the authors for their efforts, since it is their submissions that laid the foundation of a strong technical program. Each submission was reviewed by at least three Program Committee members. Four submissions were selected for presentation. The main selection criterion was the quality of the idea. We would like to thank the members of the Program Committee for taking time to provide insightful critique, and thus ensuring the high quality of the workshop.

The PAKDD community responded very enthusiastically to the DMHM 2011 Workshop, and about 15 people attended the workshop, which brought together data-mining researchers and industry practitioners.

The workshop had an invited talk by Vipin Gopal, Director of Clinical Analytics at Humana. This talk gave an overview of the efforts at Humana, a Fortune 100 health benefits company, in developing advanced analytic solutions for positively impacting care management. Predictive modeling solutions that help identify patients who are at high risk for readmissions, and subsequent interventions to reduce the overall readmissions rates, were discussed.

In their paper titled "Usage of Mobile Phones for Personalized Healthcare Solutions", M. Saravanan, S. Shanthi and S. Shalini introduce one mobile-based system known as the Personalized Mobile Health Service System for Individual's Healthcare, which caters to the specific needs of the user without the constraint on mobility.

Wei Gu, Baijie Wang and Xin Wang, in their paper titled “An Integrated Approach to Multi-Criteria-Based HealthCare Facility Location Planning,” present an integrated approach to healthcare facility planning. A new health accessibility estimation method is developed in order to capture the current characteristics of preventive healthcare services and the problem is formalized as a multi-criteria facility location model.

In “Medicinal Property Knowledge Extraction from Herbal Documents for Supporting Question Answering Systems,” C. Pechsiri, S. Painual and U. Janviriyasopak, aim to automatically extract the medicinal properties of an object, from technical documents as knowledge sources for healthcare problem solving through the question-answering system, especially What-Question, for disease treatment.

Finally, in their paper titled “Robust Learning of Mixture Models and Its Application on Trial Pruning for EEG Signal Analysis,” B. Wang and F. Wan present a novel method based on deterministic annealing to circumvent the problem of the sensitivity to atypical observations associated with the maximum likelihood (ML) estimator via a conventional EM algorithm for mixture models.

DMHM 2011 turned out to be a very successful workshop by all measures. The quality of papers was excellent, the discussion was lively, and a number of interesting directions of research were identified. This is a strong endorsement of the level of interest in this rapidly emerging field of inquiry. For more information on the workshop. Please visit the website: <http://www-users.cs.umn.edu/desikan/pakdd2011/accepted.htm>

Prasanna Desikan  
Ee-Peng Lim  
Jaideep Srivastava

# Second Workshop on Data Mining for Healthcare Management (DMHM 2011)

## Organization

### Workshop Chairs

Prasanna Desikan  
Jaideep Srivastava  
Ee-Peng Lim

Boston Scientific, USA  
University of Minnesota, USA  
Singapore Management University, USA

### Program Committee

Vipin Kumar, USA  
Joydeep Ghosh, USA  
Wei Jiang, Hong Kong  
V.R.K. Subrahmanya Rao, India  
Hui Xiong, USA  
Michael Steinbach, USA  
Deepa Mahajan, USA

Yonghong Tian, China  
Woong-Kee Loh, Korea  
Tieyun Qian, China  
Kalyan Pamarthy, USA  
Jimmy Liu Jiang, Singapore  
Chandan Reddy, USA  
San-Yih Hwang, Taiwan

# First PAKDD Doctoral Symposium on Data Mining (DSDM 2011) Chairs' Message

The First PAKDD Doctoral Symposium on Data Mining (DSDM 2011) was held at Shenzhen, China, May 24, 2011, in conjunction with the 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2011). This was the first edition of the doctoral symposium series at annual PAKDD conferences. This symposium aims to provide a forum for PhD students as well as junior researchers with newly received PhD degrees to present their recent work and seek for constructive feedback and advice from senior researchers, and bridge possible research collaborations with other participants.

This year we received nine submissions, where five were from mainland China, one from Hong Kong, one from Taiwan, and two from the USA. The submissions went through a rigorous reviewing process. Most submissions received four reviews. All the reviews were done in an encouraging and advising way with the purpose of providing constructive suggestions to the authors. The DSDM 2011 Chairs examined all the reviews to further guarantee the reliability and integrity of the reviewing process. Finally, five papers were accepted, one of which was granted the Best Paper Award.

To further realize the goal of this doctoral symposium, we held two invited talks from academia and industry, respectively. Moreover, we held a panel session and invited four senior researchers from the data mining community as our panelists. All the participants and panelists discussed a number of issues related to the trends in data-mining research, data mining in multidisciplinary research and other general questions.

DSDM 2011 would not have been a success without the support of many people. We wish to take this opportunity to thank the Program Committee members for their efforts and engagements in providing a rich and rigorous scientific program as well as the suggestions to the authors. We wish to express our gratitude to the PAKDD 2011 Program Committee Chairs Joshua Huang, Longbing Cao and Jaideep Srivastava for their invaluable support and concern. We are also grateful to the invited speakers Jie Tang and Ping Luo for their insightful and inspirational talks, and to the panelists Longbing Cao, Xintao Wu, Jie Tang and Ping Luo for their discussions and constructive advice.

Last but not least, we also want to thank all authors and all symposium participants for their contributions and support. We hope all participants took this opportunity to share and exchange ideas with one another, and to benefit from the discussions at DSDM 2011.

Ming Li  
Reynold C.K. Cheng  
Mi-Yen Yeh

# First PAKDD Doctoral Symposium on Data Mining (DSDM 2011) Organization

## Symposium Chairs

Ming Li	Nanjing University, China
Reynold C.K. Cheng	University of Hong Kong, Hong Kong
Mi-Yen Yeh	Academia Sinica, Taiwan

## Program Committee

Deng Cai, China	Man-Kwan Shan, Taiwan, ROC
Chien Chin Chen, Taiwan, ROC	Jie Tang, China
Lei Chen, Hong Kong, China	Lei Tang, USA
Jiefeng Cheng, Hong Kong, China	Ivor Tsang, Singapore
Byron Choi, Hong Kong, China	Chi-Yao Tseng, Taiwan, ROC
Chi-Yin Chow, Hong Kong, China	Raymond Wong, Hong Kong, China
Gao Cong, Singapore	Shan-Hung Wu, Taiwan, ROC
Bi-Ru Dai, Taiwan, ROC	De-Chuan Zhan, China
Wei-Shinn Ku, USA	Min-Ling Zhang, China
Shou-De Lin, Taiwan, ROC	Jun Zhu, USA
Fei Tony Liu, Australia	

# Table of Contents

---

## International Workshop on Behavior Informatics (BI 2011)

---

Evaluating the Regularity of Human Behavior from Mobile Phone Usage Logs . . . . .	3
<i>Hyounghyoun Kim and Ji-Hyung Park</i>	
Explicit and Implicit User Preferences in Online Dating . . . . .	15
<i>Joshua Akehurst, Irena Koprinska, Kalina Yacef, Luiz Pizzato, Judy Kay, and Tomasz Rej</i>	
Blogger-Link-Topic Model for Blog Mining . . . . .	28
<i>Flora S. Tsai</i>	
A Random Indexing Approach for Web User Clustering and Web Prefetching . . . . .	40
<i>Miao Wan, Arne Jönsson, Cong Wang, Lixiang Li, and Yixian Yang</i>	
Emotional Reactions to Real-World Events in Social Networks . . . . .	53
<i>Thin Nguyen, Dinh Phung, Brett Adams, and Svetha Venkatesh</i>	
Constructing Personal Knowledge Base: Automatic Key-Phrase Extraction from Multiple-Domain Web Pages . . . . .	65
<i>Yin-Fu Huang and Cin-Siang Ciou</i>	
Discovering Valuable User Behavior Patterns in Mobile Commerce Environments . . . . .	77
<i>Bai-En Shie, Hui-Fang Hsiao, Philip S. Yu, and Vincent S. Tseng</i>	
A Novel Method for Community Detection in Complex Network Using New Representation for Communities . . . . .	89
<i>Yiwen Wang and Min Yao</i>	
Link Prediction on Evolving Data Using Tensor Factorization . . . . .	100
<i>Stephan Spiegel, Jan Clausen, Sahin Albayrak, and Jérôme Kunegis</i>	
Permutation Anonymization: Improving Anatomy for Privacy Preservation in Data Publication . . . . .	111
<i>Xianmang He, Yanghua Xiao, Yujia Li, Qing Wang, Wei Wang, and Baile Shi</i>	

Efficient Mining Top-k Regular-Frequent Itemset Using Compressed Tidsets .....	124
<i>Komate Amphawan, Philippe Lenca, and Athasit Surarerks</i>	
A Method of Similarity Measure and Visualization for Long Time Series Using Binary Patterns .....	136
<i>Hailin Li, Chonghui Guo, and Libin Yang</i>	
A BIRCH-Based Clustering Method for Large Time Series Databases ...	148
<i>Vo Le Quy Nhon and Duong Tuan Anh</i>	
Visualizing Cluster Structures and Their Changes over Time by Two-Step Application of Self-Organizing Maps.....	160
<i>Masahiro Ishikawa</i>	
Analysis of Cluster Migrations Using Self-Organizing Maps.....	171
<i>Denny, Peter Christen, and Graham J. Williams</i>	

---

## Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models Workshop (QIMIE 2011)

---

ClasSi: Measuring Ranking Quality in the Presence of Object Classes with Similarity Information .....	185
<i>Anca Maria Ivanescu, Marc Wichterich, and Thomas Seidl</i>	
The Instance Easiness of Supervised Learning for Cluster Validity .....	197
<i>Vladimir Estivill-Castro</i>	
A New Efficient and Unbiased Approach for Clustering Quality Evaluation .....	209
<i>Jean-Charles Lamirel, Pascal Cuxac, Raghvendra Mall, and Ghada Safi</i>	
A Structure Preserving Flat Data Format Representation for Tree-Structured Data.....	221
<i>Fedja Hadzic</i>	
A Fusion of Algorithms in Near Duplicate Document Detection .....	234
<i>Jun Fan and Tiejun Huang</i>	
Searching Interesting Association Rules Based on Evolutionary Computation .....	243
<i>Guangfei Yang, Yanzhong Dang, Shingo Mabu, Kaoru Shimada, and Kotaro Hirasawa</i>	

An Efficient Approach to Mine Periodic-Frequent Patterns in Transactional Databases . . . . .	254
<i>Akshat Surana, R. Uday Kiran, and P. Krishna Reddy</i>	
Algorithms to Discover Complete Frequent Episodes in Sequences . . . . .	267
<i>Jianjun Wu, Li Wan, and Zeren Xu</i>	
Certainty upon Empirical Distributions . . . . .	279
<i>Joan Garriga</i>	

---

## **Workshop on Biologically Inspired Techniques for Data Mining (BDM 2011)**

---

A Measure Oriented Training Scheme for Imbalanced Classification Problems . . . . .	293
<i>Bo Yuan and Wenhuan Liu</i>	
An SVM-Based Approach to Discover MicroRNA Precursors in Plant Genomes . . . . .	304
<i>Yi Wang, Cheqing Jin, Minqi Zhou, and Aoying Zhou</i>	
Towards Recommender System Using Particle Swarm Optimization Based Web Usage Clustering . . . . .	316
<i>Shafiq Alam, Gillian Dobbie, and Patricia Riddle</i>	
Weighted Association Rule Mining Using Particle Swarm Optimization . . . . .	327
<i>Russel Pears and Yun Sing Koh</i>	
An Unsupervised Feature Selection Framework Based on Clustering . . . . .	339
<i>Sheng-yi Jiang and Lian-xi Wang</i>	

---

## **Workshop on Advances and Issues in Traditional Chinese Medicine Clinical Data Mining (AI-TCM 2011)**

---

Discovery of Regularities in the Use of Herbs in Traditional Chinese Medicine Prescriptions . . . . .	353
<i>Nevin L. Zhang, Runsun Zhang, and Tao Chen</i>	
COW: A Co-evolving Memetic Wrapper for Herb-Herb Interaction Analysis in TCM Informatics . . . . .	361
<i>Dion Detterer and Paul Kwan</i>	



Selecting an Appropriate Interestingness Measure to Evaluate the  
Correlation between Syndrome Elements and Symptoms . . . . . 372  
*Lei Zhang, Qi-ming Zhang, Yi-guo Wang, and Dong-lin Yu*

The Impact of Feature Representation to the Biclustering of  
Symptoms-Herbs in TCM . . . . . 384  
*Simon Poon, Zhe Luo, and Runshun Zhang*

**Second Workshop on Data Mining for Healthcare  
Management (DMHM 2011)**

Usage of Mobile Phones for Personalized Healthcare Solutions . . . . . 397  
*M. Saravanan, S. Shanthi, and S. Shalini*

Robust Learning of Mixture Models and Its Application on Trial  
Pruning for EEG Signal Analysis . . . . . 408  
*Boyu Wang, Feng Wan, Peng Un Mak, Pui In Mak, and Mang I Vai*

An Integrated Approach to Multi-criteria-Based Health Care Facility  
Location Planning . . . . . 420  
*Wei Gu, Baijie Wang, and Xin Wang*

Medicinal Property Knowledge Extraction from Herbal Documents for  
Supporting Question Answering System . . . . . 431  
*Chaveevan Pechsiri, Sumran Painuall, and Uraivan Janviriyasopak*

**First PAKDD Doctoral Symposium on Data Mining  
(DSDM 2011)**

Age Estimation Using Bayesian Process . . . . . 447  
*Yu Zhang*

Significant Node Identification in Social Networks . . . . . 459  
*Chi-Yao Tseng and Ming-Syan Chen*

Improving Bagging Performance through Multi-algorithm Ensembles . . . 471  
*Kuo-Wei Hsu and Jaideep Srivastava*

Mining Tourist Preferences with Twice-Learning . . . . . 483  
*Chen Zhang and Jie Zhang*

Towards Cost-Sensitive Learning for Real-World Applications . . . . . 494  
*Xu-Ying Liu and Zhi-Hua Zhou*

**Author Index . . . . . 507**