It is our pleasure to welcome you to the proceedings of the 9th International Conference on Advanced Data Mining and Applications (ADMA 2013).

As the power of generating, transmitting, and collecting huge amounts of data grows continuously, information overload is an imminent problem. It generates new challenges for the data-mining research community to develop sophisticated data-mining algorithms as well as successful data-mining applications. ADMA 2013 was held in Hangzhou, China, with the purpose of promoting original research in advanced data mining and applications and providing a dedicated forum for researchers and participants to share new ideas, original research results, case studies, practical development experiences and applications in all aspects related to data mining and applications, the.

The conference attracted 222 submissions from 26 different countries and areas. All papers were peer reviewed by at least three members of the Program Committee composed of international experts in data-mining fields. The Program Committee, together with our Program Committee Co-chairs, did enormous amount of work to select papers through a rigorous review process and extensive discussion, and finally composed a diverse and exciting program including 32 full papers and 64 short papers for ADMA 2013. The ADMA 2013 program was highlighted by three keynote speeches from outstanding researchers in advanced data mining and application areas: Gary G. Yen (Oklahoma State University, USA), Xindong Wu (University of Vermont, USA), and Joshua Zhexue Huang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences).

We would like to thank the support of several groups, without which the organization of the ADMA 2013 would not be successful. These include sponsorship from Zhejiang University, Taizhou University, and University of Technology Sydney. We also appreciate the General Co-chairs for all their precious advice and the Organizing Committee for their dedicated organizing efforts. Finally, we express our deepest gratitude to all the authors and participants who contributed to the success of ADMA 2013.

November 2013

Hiroshi Motoda
Zhaohui Wu
Longbing Cao
Osmar Zaiane
Min Yao
Wei Wang
Organization

ADMA 2013 was organized by Zhejiang University, China

Honorary Chair
Yunhe Pan  Chinese Academy of Engineering, China

Steering Committee Chair
Xue Li  University of Queensland (UQ), Australia

General Co-chairs
Hiroshi Motoda  US Air Force Office of Scientific Research, USA
Zhaohui Wu  Zhejiang University, China
Longbing Cao  University of Technology Sydney, Australia

Program Committee Co-chairs
Osmar Zaiane  University of Alberta, Canada
Min Yao  Zhejiang University, China
Wei Wang  Fudan University, China

Organization Co-chairs
Xiaoming Zhao  Taizhou University, China
Jian Wu  Zhejiang University, China
Guandong Xu  University of Technology Sydney, Australia

Publicity Chair
Liyong Wan  Zhejiang University, China

Registration Chair
Xiaowei Xue  Zhejiang University, China

Web Master
Bin Zeng  Zhejiang University, China
Organization

Steering Committee

Kyu-Young Whang  Korea Advanced Institute of Science and Technology, Korea
Chengqi Zhang  University of Technology, Sydney, Australia
Osmar Zaiane  University of Alberta, Canada
Qiang Yang  Hong Kong University of Science and Technology, China
Jie Tang  Tsinghua University, China
Jie Cao  Nanjing University of Finance and Economics, China

Program Committee

Aixin Sun  Nanyang Technological University, Singapore
Annalisa Appice  University Aldo Moro of Bari, Italy
Atsuyoshi Nakamura  Hokkaido University, Japan
Bin Shen  Ningbo Institute of Technology, China
Bo Liu  QCIS University of Technology, Australia
Daisuke Ikeda  Kyushu University, Japan
Daisuke Kawahara  Kyoto University, Japan
Eiji Uchino  Yamaguchi University, Japan
Faizah Shaari  Polytechnic Sultan Salahudin Abdul Aziz Shah, Malaysia
Feiping Nie  University of Texas, Arlington, USA
Gang Li  Deakin University, Australia
Gongde Guo  Fujian Normal University, China
Guandong Xu  University of Technology Sydney, Australia
Guohua Liu  Yanshan University, China
Hanghang Tong  IBM T.J. Watson Research Center, USA
Haofeng Zhou  Fudan University, China
Jason Wang  New Jersey Institute of Technology, USA
Jianwen Su  UC Santa Barbara, USA
Jinjiu Li  University of Technology Sydney, Australia
Liang Chen  Zhejiang University, China
Manish Gupta  University of Illinois at Urbana-Champaign, USA
Mengchu Zhou  New Jersey Institute of Technology, USA
Mengjie Zhang  Victoria University of Wellington, New Zealand
Michael R. Lyu  The Chinese University of Hong Kong, Hong Kong, China
Michael Sheng  The University of Adelaide, Australia
Philippe Fournier-Viger  University of Moncton, Canada
Qi Wang  Xi’an Institute of Optics and Precision Mechanics of CAS, China
Qi Yu Rochester Institute of Technology, China
Rong Zhu Jiaxing University, China
Sheng Zhong SUNY Buffalo, USA
Shu-Ching Chen Florida International University, USA
Shuliang Wang Wuhan University, China
Songmao Zhang Chinese Academy of Sciences, China
Stefano Ferilli Università di Bari, Italy
Tetsuya Yoshida Hokkaido University, Japan
Tieyun Qian Wuhan University, China
Tomonari Masada Nagasaki University, Japan
Vincent S. Tseng National Cheng Kung University, Taiwan
Wei Hu Nanjing University, China
Wei Wang Fudan University, China
Wynne Hsu National University of Singapore, Singapore
Xiangjun Dong Shandong Institute of Light Industry, China
Xiangliang Zhang King Abdullah University of Science and Technology, Saudi Arabia
Xiao Zheng Southeast University, China
Xin Jin University of Illinois at Urbana-Champaign, USA
Ya Zhang Shanghai Jiao Tong University, China
Yang Gao Nanjing University, China
Yanglan Gan Donghua University, China
Yasuhiko Morimoto Hiroshima University, Japan
Yi Zhuang Zhejiang GongShang University, China
Yin Song University of Technology Sydney, Australia
Yong Zheng DePaul University, USA
Yubao Liu Sun Yat-Sen University, China
Zhaozhong Deng Alibaba Inc., China
Zhihong Wu Nanjing University of Finance and Economics, China
Zhihong Deng Peking University, China
Zhihui Wang Fudan University, China
Zhipeng Xie Fudan University, China
Zijiang Yang York University, Canada

Sponsoring Institutions

College of Computer Science & Technology, Zhejiang University, China
College of Mathematics and information engineering, Taizhou University, China
Advanced Analytics Institute, University of Technology Sydney, Australia
Table of Contents – Part I

Opinion Mining

Mining E-Commerce Feedback Comments for Dimension Rating Profiles ................................................................. 1
  Lishan Cui, Xiuzhen Zhang, Yan Wang, and Lifang Wu

Generating Domain-Specific Sentiment Lexicons for Opinion Mining . . . . 13
  Zaher Salah, Frans Coenen, and Davide Grossi

Effective Comment Sentence Recognition for Feature-Based Opinion Mining ................................................................. 25
  Hui Song, Botian Yang, and Xiaoqiang Liu

Exploiting Co-occurrence Opinion Words for Semi-supervised Sentiment Classification ................................................................. 36
  Suke Li, Jinmei Hao, Yanbing Jiang, and Qi Jing

Behavior Mining

HN-Sim: A Structural Similarity Measure over Object-Behavior Networks .................................................................................... 48
  Jiazhen Nian, Shanshan Wang, and Yan Zhang

Community Based User Behavior Analysis on Daily Mobile Internet Usage .................................................................................... 60
  Jamal Yousaf, Juanzi Li, and Yuanchao Ma

Stream Mining

Tracking Drift Types in Changing Data Streams ........................................ 72
  David Tse Jung Huang, Yun Sing Koh, Gillian Dobby, and Russel Pears

Continuously Extracting High-Quality Representative Set from Massive
Data Streams ......................................................................................... 84
  Xiaokang Ji, Xiuli Ma, Ting Huang, and Shiwei Tang

Change Itemset Mining in Data Streams ............................................. 97
  Minmin Zhang, Gillian Dobby, and Yun Sing Koh
Sequential Data Mining

TKS: Efficient Mining of Top-K Sequential Patterns 109
Philippe Fournier-Viger, Antonio Gomariz, Ted Gueniche, Espérance Mwamikazi, and Rincy Thomas

When Optimization Is Just an Illusion 121
Muhammad Marwan Muhammad Fuad

Accurate and Fast Dynamic Time Warping 133
Hailin Li and Libin Yang

Online Detecting Spreading Events with the Spatio-temporal Relationship in Water Distribution Networks 145
Ting Huang, Xiuli Ma, Xiaokang Ji, and Shiwei Tang

MLSP: Mining Hierarchically-Closed Multi-Level Sequential Patterns 157
Michal Šebek, Martin Hlosta, Jaroslav Zendulka, and Tomáš Hruška

Mining Maximal Sequential Patterns without Candidate Maintenance 169
Philippe Fournier-Viger, Cheng-Wei Wu, and Vincent S. Tseng

Web Mining

Improved Slope One Collaborative Filtering Predictor Using Fuzzy Clustering 181
Tianyi Liang, Jiancong Fan, Jianli Zhao, Yongquan Liang, and Yujun Li

Towards Building Virtual Vocabularies in the Semantic Web 193
Yunqing Wen, Xiang Zhang, Kai Shen, and Peng Wang

Web Mining Accelerated with In-Memory and Column Store Technology 205
Patrick Hennig, Philipp Berger, and Christoph Meinel

Image Mining

Constructing a Novel Pos-neg Manifold for Global-Based Image Classification 217
Rong Zhu, Jianhua Yang, Yonggang Li, and Jie Xu

3-D MRI Brain Scan Feature Classification Using an Oct-Tree Representation 229
Akadej Udomchaiporn, Frans Coenen, Marta García-Fuñana, and Vanessa Sluming
Biometric Template Protection Based on Biometric Certificate and Fuzzy Fingerprint Vault ................................................. 241
Weihong Wang, Youbing Lu, and Zhaolin Fang

Kwankamon Dittakan, Frans Coenen, Rob Christley, and Maya Wardeh

Mixed-Norm Regression for Visual Classification ............................. 265
Xiaofeng Zhu, Jilian Zhang, and Shichao Zhang

Research on Map Matching Based on Hidden Markov Model ........... 277
Jinhui Nie, Hongqi Su, and Xiaohua Zhou

Text Mining

A Rule-Based Named-Entity Recognition for Malay Articles ............... 288
Rayner Alfred, Leow Ching Leong, Chin Kim On, Patricia Anthony, Tan Soo Fun, Mohd Norhisham Bin Razali, and Mohd Hanafi Ahmad Hijazi

Small Is Powerful! Towards a Refinedly Enriched Ontology by Careful Pruning and Trimming ................................................. 300
Shan Jiang, Jiazhen Nian, Shi Zhao, and Yan Zhang

Refine the Corpora Based on Document Manifold ............................ 313
Chengwei Yao, Yilin Wang, and Gencai Chen

Social Network Mining

Online Friends Recommendation Based on Geographic Trajectories and Social Relations ......................................................... 323
Shi Feng, Dajun Huang, Kaisong Song, and Daling Wang

The Spontaneous Behavior in Extreme Events: A Clustering-Based Quantitative Analysis ....................................................... 336
Ning Shi, Chao Gao, Zili Zhang, Lu Zhong, and Jiajin Huang

Restoring: A Greedy Heuristic Approach Based on Neighborhood for Correlation Clustering .................................................. 348
Ning Wang and Jie Li

A Local Greedy Search Method for Detecting Community Structure in Weighted Social Networks ........................................... 360
Bin Liu and Tieyun Qian
Tree-Based Mining for Discovering Patterns of Reposting Behavior in Microblog

Huilei He, Zhiwen Yu, Bin Guo, Xinjiang Lu, and Jilei Tian

An Improved Parallel Hybrid Seed Expansion (PHSE) Method for Detecting Highly Overlapping Communities in Social Networks

Ting Wang, Xu Qian, and Hui Xu

A Simple Integration of Social Relationship and Text Data for Identifying Potential Customers in Microblogging

Guansong Pang, Shengyi Jiang, and Dongyi Chen

An Energy Model for Network Community Structure Detection

Yin Pang and Kan Li

A Label Propagation-Based Algorithm for Community Discovery in Online Social Networks

Yitong Wang, Yurong Zhao, Zhuoxiang Zhao, and Zhicheng Liao

Mining Twitter Data for Potential Drug Effects

Keyuan Jiang and Yujing Zheng

Social-Correlation Based Mutual Reinforcement for Short Text Classification and User Interest Tagging

Rong Li and Ya Zhang

Classification

Graph Based Feature Augmentation for Short and Sparse Text Classification

Guodong Long and Jing Jiang

Exploring Deep Belief Nets to Detect and Categorize Chinese Entities

Yu Chen, Dequan Zheng, and Tiejun Zhao

Extracting Novel Features for E-Commerce Page Quality Classification

Jing Wang, Lanfen Lin, Feng Wang, Penghua Yu, Jiaolong Liu, and Xiaowei Zhu

Hierarchical Classification for Solving Multi-class Problems: A New Approach Using Naive Bayesian Classification

Esra’a Alshdaifat, Frans Coenen, and Keith Dures

Predicting Features in Complex 3D Surfaces Using a Point Series Representation: A Case Study in Sheet Metal Forming

Subhieh El-Salhi, Frans Coenen, Clare Dixon, and Muhammad Sulaiman Khan
Table of Contents – Part I

Automatic Labeling of Forums Using Bloom’s Taxonomy ................. 517
  Vanessa Echeverría, Juan Carlos Gomez, and Marie-Francine Moens

Classifying Papers from Different Computer Science Conferences ........ 529
  Yaakov HaCohen-Kerner, Avi Rosenfeld, Maor Tzidkani, and Daniel Nisim Cohen

Vertex Unique Labelled Subgraph Mining for Vertex Label Classification ........................................ 542
  Wen Yu, Frans Coenen, Michele Zito, and Subhieh El Salhi

A Similarity-Based Grouping Method for Molecular Docking in Distributed System ........................................ 554
  Ruisheng Zhang, Guangcai Liu, Rongjing Hu, Jiaxuan Wei, and Juan Li

A Bag-of-Tones Model with MFCC Features for Musical Genre Classification .................................................... 564
  Zengchang Qin, Wei Liu, and Tao Wan

The GEPSO-Classification Algorithm ........................................ 576
  Weihong Wang, Dandan Jin, Qu Li, Zhaolin Fang, and Jie Yang

Author Index .................................................. 585
# Table of Contents – Part II

## Clustering

- Semi-supervised Clustering Ensemble Evolved by Genetic Algorithm for Web Video Categorization .................................................. 1  
  *Amjad Mahmood, Tianrui Li, Yan Yang, and Hongjun Wang*

- A Scalable Approach for General Correlation Clustering .............. 13  
  *Yubo Wang, Linli Xu, Yucheng Chen, and Hao Wang*

- A Fast Spectral Clustering Method Based on Growing Vector Quantization for Large Data Sets ................................................... 25  
  *Xiujun Wang, Xiao Zheng, Feng Qin, and Baohua Zhao*

- A Novel Deterministic Sampling Technique to Speedup Clustering Algorithms ......................................................................................... 34  
  *Sanguthevar Rajasekaran and Subrata Saha*

- Software Clustering Using Automated Feature Subset Selection ........ 47  
  *Zubair Shah, Rashid Naseem, Mehmet A. Orgun, Abdun Mahmood, and Sara Shahzad*

- The Use of Transfer Algorithm for Clustering Categorical Data ........ 59  
  *Zhengrong Xiang and Lichuan Ji*

- eDARA: Ensembles DARA ................................................................. 71  
  *Chung Seng Kheau, Rayner Alfred, and HuiKeng Lau*

- Efficient Mining Maximal Variant and Low Usage Rate Biclusters without Candidate Maintenance in Real Function-Resource Matrix: The DeCluster Algorithm ......................................................... 83  
  *Lihua Zhang, Miao Wang, Zhengjun Zhai, and Guoqing Wang*

## Association Rule Mining

- MEIT: Memory Efficient Itemset Tree for Targeted Association Rule Mining ....................................................................................... 95  
  *Philippe Fournier-Viger, Espérance Mwamikazi, Ted Gueniche, and Usef Faghihi*

## Pattern Mining

- Mining Frequent Patterns in Print Logs with Semantically Alternative Labels ....................................................................................... 107  
  *Xin Li, Lei Zhang, Enhong Chen, Yu Zong, and Guandong Xu*
### Part II

#### Minimising K-Dominating Set in Arbitrary Network Graphs

*Guangyuan Wang, Hua Wang, Xiaohui Tao, Ji Zhang, and Jinhua Zhang*

#### Regression

**Logistic Regression Bias Correction for Large Scale Data with Rare Events**

*Zhen Qiu, Hongyan Li, Hanchen Su, Gaoyan Ou, and Tengjiao Wang*

**An Automatical Moderating System for FML Using Hashing Regression**

*Peichao Zhang and Minyi Guo*

**Batch-to-Batch Iterative Learning Control Based on Kernel Independent Component Regression Model**

*Ganping Li, Jun Zhao, Fuyang Zhang, and Zhizhen Ni*

#### Prediction

**Deep Architecture for Traffic Flow Prediction**

*Wenhao Huang, Haikun Hong, Man Li, Weisong Hu, Guojie Song, and Kunqing Xie*

**Compact Prediction Tree: A Lossless Model for Accurate Sequence Prediction**

*Ted Gueniche, Philippe Fournier-Viger, and Vincent S. Tseng*

**Generalization of Malaria Incidence Prediction Models by Correcting Sample Selection Bias**

*Orlando P. Zacarias and Henrik Boström*

**Protein Interaction Hot Spots Prediction Using LS-SVM within the Bayesian Interpretation**

*Juhong Qi, Xiaolong Zhang, and Bo Li*

**Predicting the Survival Status of Cancer Patients with Traditional Chinese Medicine Symptom Variation Using Logistic Regression Model**

*Min Wan, Liying Fang, Mingwei Yu, Wenshuai Cheng, and Pu Wang*

#### Feature Extraction

**Exploiting Multiple Features for Learning to Rank in Expert Finding**

*Hai-Tao Zheng, Qi Li, Yong Jiang, Shu-Tao Xia, and Lianshan Zhang*

**Convolution Neural Network for Relation Extraction**

*ChunYang Liu, WenBo Sun, WenHan Chao, and WanXiang Che*
# Table of Contents – Part II

**Extracting Fuzzy Rules from Hierarchical Heterogeneous Neural Networks for Cardiovascular Diseases Diagnosis**

*YuanLian Cui and MingChui Dong*

- 243

**kDMI: A Novel Method for Missing Values Imputation Using Two Levels of Horizontal Partitioning in a Data set**

*Md. Geaur Rahman and Md Zahidul Islam*

- 250

## Identification

**Traffic Session Identification Based on Statistical Language Model**

*Xinyan Lou, Yang Liu, and Xiaohui Yu*

- 264

**Role Identification Based on the Information Dependency Complexity**

*Weidong Zhao, Haitao Liu, and Xi Liu*

- 276

**Detecting Professional Spam Reviewers**

*Junlong Huang, Tieyun Qian, Guoliang He, Ming Zhong, and Qingxi Peng*

- 288

**Chinese Comparative Sentence Identification Based on the Combination of Rules and Statistics**

*Quanchao Liu, Heyan Huang, Chen Zhang, Zhenzhao Chen, and Jiajun Chen*

- 300

## Privacy Preservation

**Utility Enhancement for Privacy Preserving Health Data Publishing**

*Lengdong Wu, Hua He, and Osmar R. Zaiane*

- 311

**Optimizing Placement of Mix Zones to Preserve Users’ Privacy for Continuous Query Services in Road Networks**

*Kamenyi Domenic M., Yong Wang, Fengli Zhang, Yankson Gustav, Daniel Adu-Gyamfi, and Nkatha Dorothy*

- 323

## Applications

**Comparison of Cutoff Strategies for Geometrical Features in Machine Learning-Based Scoring Functions**

*Shirley W.I. Siu, Thomas K.F. Wong, and Simon Fong*

- 336

**Bichromatic Reverse Ranking Query in Two Dimensions**

*Zhao Zhang, Qiangqiang Kang, Cheqing Jin, and Aoying Zhou*

- 348

**Passive Aggressive Algorithm for Online Portfolio Selection with Piecewise Loss Function**

*Li Gao, Weiguo Zhang, and Qiang Tang*

- 360
<table>
<thead>
<tr>
<th>Title</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mining Item Popularity for Recommender Systems</td>
<td>372</td>
</tr>
<tr>
<td>Jian Ren Zhang, Xiaofeng Zhu, Xianxian Li, and Shichao Zhang</td>
<td></td>
</tr>
<tr>
<td>Exploring an Ichthyoplankton Database from a Freshwater Reservoir in Legal Amazon</td>
<td>384</td>
</tr>
<tr>
<td>Michel de A. Silva, Daniela Queiroz Trevisan, David N. Prata, Elineide E. Marques, Marcelo Lisboa, and Monica Prata</td>
<td></td>
</tr>
<tr>
<td>A Pre-initialization Stage of Population-Based Bio-inspired Metaheuristics for Handling Expensive Optimization Problems</td>
<td>396</td>
</tr>
<tr>
<td>Muhammad Marwan Muhammad Fuad</td>
<td></td>
</tr>
<tr>
<td>A Hybrid-Sorting Semantic Matching Method</td>
<td>404</td>
</tr>
<tr>
<td>Kan Li, Wensi Mu, Yong Luan, and Shaohua An</td>
<td></td>
</tr>
<tr>
<td>Improving Few Occurrence Feature Performance in Distant Supervision for Relation Extraction</td>
<td>414</td>
</tr>
<tr>
<td>Hui Zhang and Yuanhao Zhao</td>
<td></td>
</tr>
<tr>
<td>Cluster Labeling Extraction and Ranking Feature Selection for High Quality XML Pseudo Relevance Feedback Fragments Set</td>
<td>423</td>
</tr>
<tr>
<td>Minjuan Zhong, Changxuan Wan, Dexi Liu, Shumei Liao, and Siwen Luo</td>
<td></td>
</tr>
<tr>
<td>Informed Weighted Random Projection for Dimension Reduction</td>
<td>433</td>
</tr>
<tr>
<td>Jaydeep Sen and Harish Karnick</td>
<td></td>
</tr>
<tr>
<td>Protocol Specification Inference Based on Keywords Identification</td>
<td>443</td>
</tr>
<tr>
<td>Yong Wang, Nan Zhang, Yan-mei Wu, and Bin-bin Su</td>
<td></td>
</tr>
<tr>
<td>An Adaptive Collaborative Filtering Algorithm Based on Multiple Features</td>
<td>455</td>
</tr>
<tr>
<td>Yan-Qiu Zhang, Hai-Tao Zheng, and Lan-Shan Zhang</td>
<td></td>
</tr>
<tr>
<td><strong>Machine Learning</strong></td>
<td></td>
</tr>
<tr>
<td>Ensemble of Unsupervised and Supervised Models with Different Label Spaces</td>
<td>466</td>
</tr>
<tr>
<td>Yueyun Jin, Weilin Zeng, Hankz Hankui Zhuo, and Lei Li</td>
<td></td>
</tr>
<tr>
<td>Cost-Sensitive Extreme Learning Machine</td>
<td>478</td>
</tr>
<tr>
<td>Enhui Zheng, Cong Zhang, Xueyi Liu, Huijuan Lu, and Jian Sun</td>
<td></td>
</tr>
<tr>
<td>Multi-Objective Optimization for Overlapping Community Detection</td>
<td>489</td>
</tr>
<tr>
<td>Jingfei Du, Jianyang Lai, and Chuan Shi</td>
<td></td>
</tr>
<tr>
<td>Endmember Extraction by Exemplar Finder</td>
<td>501</td>
</tr>
<tr>
<td>Yi Guo, Junbin Gao, and Yanfeng Sun</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>EEG-Based User Authentication in Multilevel Security Systems</td>
<td>513</td>
</tr>
<tr>
<td>Tien Pham, Wanli Ma, Dat Tran, Phuoc Nguyen, and Dinh Phung</td>
<td></td>
</tr>
<tr>
<td>A New Fuzzy Extreme Learning Machine for Regression Problems with</td>
<td>524</td>
</tr>
<tr>
<td>Outliers or Noises</td>
<td></td>
</tr>
<tr>
<td>Enhui Zheng, Jinyong Liu, Huijuan Lu, Ling Wang, and Le Chen</td>
<td></td>
</tr>
<tr>
<td>Author Index</td>
<td>535</td>
</tr>
</tbody>
</table>