

# Contents

## Pattern Mining and Application of Big Data

ProbitUCB: A Novel Method for Review Ranking . . . . .	3
<i>Wanying Ding, Yue Shang, Dae Hoon Park, Lifan Guo, and Xiaohua Hu</i>	
From Cluster-Based Outlier Detection to Time Series Discord Discovery . . . .	16
<i>Nguyen Huy Kha and Duong Tuan Anh</i>	
Web Site Audience Segmentation Using Hybrid Alignment Techniques . . . .	29
<i>Vinh-Trung Luu, Germain Forestier, Frédéric Fondement, and Pierre-Alain Muller</i>	
Mining Massive-Scale Spatiotemporal Trajectories in Parallel: A Survey . . . .	41
<i>Pengtao Huang and Bo Yuan</i>	
Manifold Regularized Symmetric Joint Link Model for Overlapping Community Detection . . . . .	53
<i>Hao Chen, Xianchao Zhang, Wenxin Liang, and Feng Ding</i>	
High Dimensional Explicit Feature Biased Matrix Factorization Recommendation . . . . .	66
<i>Weibin Sun, Xianchao Zhang, Wenxin Liang, and Zengyou He</i>	
A Simhash-Based Generalized Framework for Citation Matching in MapReduce . . . . .	78
<i>Pengsen Wang, Bin Wu, Xiaoming Li, Lin Wang, and Bai Wang</i>	
A Cloud Based Type-2 Diabetes Mellitus Lifestyle Self-Management System . . . . .	91
<i>Shih-Hao Chang and Chih-Ning Li</i>	
Construction of a Prediction Model for Nephropathy Among Obese Patients Using Genetic and Clinical Features . . . . .	104
<i>Guan-Mau Huang, Yi-Cheng Chen, and Julia Tzu-Ya Weng</i>	
A Research of Applying Association Rules and Decision Tree to Endometriosis . . . . .	113
<i>Lin Hui, Huan-Chao Keh, Nan-Ching Huang, Chiung-Tzu Chang, and Yi-Fan Yang</i>	

A Dynamic Feature Selection Based LDA Approach to Baseball Pitch Prediction . . . . .	125
<i>Phuong Hoang, Michael Hamilton, Joseph Murray, Corey Stafford, and Hien Tran</i>	

**Quality Issues, Measures of Interestingness and Evaluation of Data Mining Models**

A Study of Interestingness Measures for Associative Classification on Imbalanced Data. . . . .	141
<i>Guangfei Yang and Xuejiao Cui</i>	
Evaluation of Community Mining Algorithms in the Presence of Attributes . . .	152
<i>Reihaneh Rabbany and Osmar R. Zaïane</i>	
Leveraging the Common Cause of Errors for Constraint-Based Data Cleansing . . . . .	164
<i>Ayako Hoshino, Hiroki Nakayama, Chihiro Ito, Kyota Kanno, and Kenshi Nishimura</i>	
Analyzing User Behaviors Based on Temporal Patterns of Sequential Pattern Evaluation Indices on Twitter . . . . .	177
<i>Hidenao Abe</i>	
Model Selection of Symbolic Regression to Improve the Accuracy of PM <sub>2.5</sub> Concentration Prediction. . . . .	189
<i>Guangfei Yang and Jian Huang</i>	
Internal Clustering Evaluation of Data Streams . . . . .	198
<i>Marwan Hassani and Thomas Seidl</i>	
Feature Maximization Based Clustering Quality Evaluation: A Promising Approach . . . . .	210
<i>Jean-Charles Lamirel and Shadi Al Shehabi</i>	

**Data Analytics for Evidence-Based Healthcare**

Integrating Content Centric Networking and Web Content Mining: A Future Efficient Internet Architecture for Healthcare . . . . .	225
<i>Rabia Bashir and Sajjad Akbar</i>	
Citation Enrichment Improves Deduplication of Primary Evidence . . . . .	237
<i>Miew Keen Choong, Sarah Thorning, and Guy Tsafnat</i>	
Learning Entry Profiles of Children with Autism from Multivariate Treatment Information Using Restricted Boltzmann Machines. . . . .	245
<i>Pratibha Vellanki, Dinh Phung, Thi Duong, and Svetha Venkatesh</i>	

## Vietnamese Language and Speech Processing

Fast Dependency Parsing Using Distributed Word Representations . . . . .	261
<i>Phuong Le-Hong, Thi-Minh-Huyen Nguyen, Thi-Luong Nguyen, and My-Linh Ha</i>	
Modeling Vietnamese Speech Prosody: A Step-by-Step Approach Towards an Expressive Speech Synthesis System . . . . .	273
<i>Dang-Khoa Mac and Do-Dat Tran</i>	
<b>Author Index . . . . .</b>	<b>289</b>

Trends and Applications in Knowledge Discovery and  
Data Mining

PAKDD 2015 Workshops: BigPMA, VLSP, QIMIE, DAEBH,  
Ho Chi Minh City, Vietnam, May 19-21, 2015. Revised  
Selected Papers

Li, X.; Cao, T.; Lim, E.-P.; Zhou, Z.-H.; Ho, T.-B.; Cheung,  
D. (Eds.)

2015, XV, 289 p. 91 illus., Softcover

ISBN: 978-3-319-25659-7