Contents

Data Analytics for Targeted Healthcare

Conversation Intention Perception Based on Knowledge Base. . . . . . . . 3

Yi-Zheng Chen, Hua-Kang Li, and Yi Liu

An Empirical Algorithm for Bias Correction Based on GC Estimation
for Single Cell Sequencing. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

Bo Xu, Tengpeng Li, Yi Luo, Ruotao Xu, and Hongmin Cai

Joint Tree of Combinatorial Maps . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 22

Tao Wang, Congyan Lang, and Songhe Feng

An Optimization Method of Fusing Multiple Decisions in Object Detection . . . 29

Zhu Teng and Baopeng Zhang

Data Mining and Decision Analytics for Public Health and Wellness

Mining Typical Order Sequences from EHR for Building Clinical Pathways . . . 39

Shoji Hirano and Shusaku Tsumoto

Aspect-Based Sentiment Analysis of Amazon Reviews for Fitness
Tracking Devices . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 50

Alaa Shafaee, Hassan Issa, Stefan Agne, Stephan Baumann, and Andreas Dengel

Biologically Inspired Techniques for Data Mining

An Efficient Drug-Target Interaction Mining Algorithm in Heterogeneous
Biological Networks . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 65

Congcong Li, Jing Sun, Yun Xiong, and Guangyong Zheng

Comparison of Hemagglutinin and Neraminidase of Influenza
A Virus Subtype H1N1, H5N1, H5N2, and H7N9 Using Apriori Algorithm . . . 77

Dae Young Kim, Hye-Jun Kim, Junhyeok Bae, and Taeseon Yoon

Privacy Preserving Association Rule Mining Using Binary Encoded NSGA-II . . 87

Peng Cheng, Jeng-Shyang Pan, and Chun-Wei Lin

Detecting the Data Group Most Prone to a Specific Disguise Value . . . . . . . 100

Wen-Yang Lin and Wen-Yu Feng
rRNA of Alphaproteobacteria Rickettsiales and mtDNA Pattern Analyzing with Apriori & SVM ................................................................. 112

Seung Jae Lim, Shin Hyo Bang, Dae Seop Kim, and Taeseon Yoon

Mobile Data Management, Mining, and Computing on Social Networks

Team Formation with the Communication Load Constraint in Social Networks . . 125
Yui-Chieh Teng, Jun-Zhe Wang, and Jiun-Long Huang

Information Diffusion Pattern Mining over Online Social Media ............. 137
Eric Hsueh-Chan Lu and Hui-Ju Hung

Mining Frequent Progressive Usage Patterns Across Multiple Mobile Broadcasting Channels ................................................................. 149
Bijay Prasad Jaysawal and Jen-Wei Huang

Efficient Gossiping Dissemination in Mobile Peer-to-Peer Environment Based on Neighborhood Detection ................................. 156
Yu-Jen Lin, Bo-Heng Chen, Kun-Ta Chuang, and Chao-Chun Chen

Predicting Locations of Mobile Users Based on Behavior Semantic Mining . . 168
Huei-Yu Lung, Chih-Heng Chung, and Bi-Ru Dai

Query Prediction by Currently-Browsed Web Pages and Its Applications . . 181
Hui-Ju Hung and Pu-Jen Cheng

An Emergency System in Mobile Devices ............................................ 194
Chung-Hua Chu and Wen-Hsiang Chang

Study of Mobile Value-Added Services with Augmented Reality ............. 199
Chung-Hua Chu, Shu-Lin Wang, and Bi-Chi Tseng

Mining Uncertain Sequence Data on Hadoop Platform .......................... 204
Zi-Yun Sun, Ming-Che Tsai, and Hsiao-Ping Tsai

Big Data Science and Engineering on E-Commerce

Two Algorithms Under Stochastic Gradient Descent Framework for Recommender Systems ................................................................. 219
Tian-Hsiang Huang and Vladimir Nikulin

Scalable Textual Similarity Search on Large Document Collections Through Random Indexing and K-means Clustering ............................. 231
Ali Cevahir

Complementary Product Selection in E-Commerce ................................. 239
Yian Chen, Yuh-Ming Chiu, and Song Han
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Short-Term Bookmarking System for Collecting User-Interest Data</td>
<td>247</td>
</tr>
<tr>
<td>Chu-Cheng Hsieh, Yoni Medoff, and Naren Chittar</td>
<td></td>
</tr>
<tr>
<td>Cloud Service Discovery</td>
<td></td>
</tr>
<tr>
<td>Automatic Self-Suspended Task for a MapReduce System on Cloud</td>
<td>257</td>
</tr>
<tr>
<td>Computing</td>
<td></td>
</tr>
<tr>
<td>Tzu-Chi Huang, Ce-Kuen Shieh, Sheng-Wei Huang, Chui-Ming Chiu,</td>
<td></td>
</tr>
<tr>
<td>and Tyng-Yeu Liang</td>
<td></td>
</tr>
<tr>
<td>Tag Completion and Refinement for Web Service via Low-Rank Matrix</td>
<td>269</td>
</tr>
<tr>
<td>Completion</td>
<td></td>
</tr>
<tr>
<td>Lei Chen, Geng Yang, Zhengyu Chen, Fu Xiao, and Xuxia Li</td>
<td></td>
</tr>
<tr>
<td>Mobile Sensing, Mining and Visualization for Human Behavior Inference</td>
<td></td>
</tr>
<tr>
<td>A Differential Approach for Identifying Important Student Learning</td>
<td>281</td>
</tr>
<tr>
<td>Behavior Patterns with Evolving Usage over Time</td>
<td></td>
</tr>
<tr>
<td>John S. Kinnebrew, Daniel L.C. Mack, Gautam Biswas,</td>
<td></td>
</tr>
<tr>
<td>and Chih-Kai Chang</td>
<td></td>
</tr>
<tr>
<td>Mining Top-K Relevant Stay Regions from Historical Trajectories</td>
<td>293</td>
</tr>
<tr>
<td>Yung-Hsiang Lin, Chien-Hsiang Lai, and Po-Ruey Lei</td>
<td></td>
</tr>
<tr>
<td>Associative Classification for Human Activity Inference on Smart</td>
<td>305</td>
</tr>
<tr>
<td>Phones</td>
<td></td>
</tr>
<tr>
<td>Yu-Hsiang Peng, Gunarto Sindoro Njoo, Shou-Chun Li, and Wen-Chih Peng</td>
<td></td>
</tr>
<tr>
<td>Personalized Smartphone Wearing Behavior Analysis</td>
<td>318</td>
</tr>
<tr>
<td>Yen-Hsuan Lin and Yi-Ta Chuang</td>
<td></td>
</tr>
<tr>
<td>Activity Recognition Using Discriminant Sequence Patterns in Smart</td>
<td>329</td>
</tr>
<tr>
<td>Environments</td>
<td></td>
</tr>
<tr>
<td>Been-Chian Chien and Rong-Sing Huang</td>
<td></td>
</tr>
<tr>
<td>Scalable Data Analytics: Theory and Applications</td>
<td></td>
</tr>
<tr>
<td>Using Knowledge Graph to Handle Label Imperfection</td>
<td>345</td>
</tr>
<tr>
<td>Yi Liu, Huakang Li, and Yizheng Chen</td>
<td></td>
</tr>
<tr>
<td>Learning to Display in Sponsored Search</td>
<td>357</td>
</tr>
<tr>
<td>Xin Xin and Heyan Huang</td>
<td></td>
</tr>
<tr>
<td>Models for Distributed, Large Scale Data Cleaning</td>
<td>369</td>
</tr>
<tr>
<td>Vincent J. Maccio, Fei Chiang, and Douglas G. Down</td>
<td></td>
</tr>
</tbody>
</table>
XVIII    Contents

A Study of Rumor Spreading with Epidemic Model Based on Network Topology ................................................................. 381
  Dawei Meng, Lizhi Wan, and Lei Zhang

GS4: Generating Synthetic Samples for Semi-Supervised Nearest Neighbor Classification .......................................................... 393
  Panagiotis Moutafis and Ioannis A. Kakadiaris

A Scalable Data Analytics Algorithm for Mining Frequent Patterns from Uncertain Data ............................................................. 404

Parallel Time Series Modeling - A Case Study of In-Database Big Data Analytics ................................................................. 417
  Hai Qian, Shengwen Yang, Rahul Iyer, Xixuan Feng, Mark Wellons, and Caleb Welton

A Biclustering-Based Classification Framework for Microarray Analysis ................................................................. 429
  Baljeet Malhotra, Daniel Dahlmeier, and Naveen Nandan

Parallel Implementation of a Density-Based Stream Clustering Algorithm Over a GPU Scheduling System ................................................................. 441
  Marwan Hassani, Ayman Tarakji, Lyubomir Georgiev, and Thomas Seidl

Algorithms for Large-Scale Information Processing in Knowledge Discovery

Three-way Indexing ZDDs for Large-Scale Sparse Datasets ................................................................. 457
  Hiroshi Aoki, Takahisa Toda, and Shin-ichi Minato

Local Feature Selection by Formal Concept Analysis for Multi-class Classification ................................................................. 470
  Madori Ikeda and Akihiro Yamamoto

Ensemble Clustering of High Dimensional Data with FastMap Projection ................................................................. 483
  Imran Khan, Joshua Zhexue Huang, Nguyen Thanh Tung, and Graham Williams

A General Framework for Parallel Unary Operations on ZDDs ................................................................. 494
  Shogo Takeuchi, Takahisa Toda, and Shin-ichi Minato

On the Size of the Zero-Suppressed Binary Decision Diagram that Represents All the Subtrees in a Tree ................................................................. 504
  Norihito Yasuda, Masaaki Nishino, and Shin-ichi Minato
Data Mining in Social Networks

A Unified Fusion Framework for Time-Related Rank in Threaded Discussion Communities ........................................... 513

Qiang You, Weiming Hu, Ou Wu, and Haiqiang Zuo

Multi-group-based User Perceptions for Friend Recommendation in Social Networks ........................................... 525

Trung L.T. Nguyen and Tru H. Cao

STC: A Joint Sentiment-Topic Model for Community Identification .......... 535

Baoguo Yang and Suresh Manandhar

Considerations About Multistep Community Detection .................. 549

Antonio A. Gentile, Angelo Corallo, Cristian Bisconti, and Laura Fortunato

An Edge-Centric Approach for Change Point Detection in Dynamic Networks . 562

Yongsheng Cheng and Xiaokang Lin

Toward Automatic Censorship Detection in Microblogs .................... 572

Donn Morrison

Competition Component Identification on Twitter .......................... 584

Cheng-Huang Yang, Ji-De Chen, and Hung-Yu Kao

Social Network Community Detection Using Strongly Connected Components ................................................... 596

Wookey Lee, James J. Lee, and Jinho Kim

Data Mining in Biomedical informatics and Healthcare

Automatic Segmentation and Quantitative Analysis of Gray Matter on MR Images of Patients with Epilepsy Based on Unsupervised Learning Methods .................................................. 607

Rui Wang, Jie Wang, Jing Ding, and Su Zhang

An Author Topic Analysis of Tobacco Regulation Investigators ............ 616

Ding Cheng Li, Janet Okamoto, Scott Leischow, and Hongfang Liu

Mining Severe Drug-Drug Interaction Adverse Events Using Semantic Web Technologies: A Case Study .................... 628

Guoqian Jiang, Hongfang Liu, Harold R. Solbrig, and Christopher G. Chute

Cancer Based Pharmacogenomics Network for Drug Repurposing ........ 639

Liwei Wang, Hongfang Liu, Christopher G. Chute, and Qian Zhu
Linked Vaccine Adverse Event Data Representation from VAERS for Biomedical Informatics Research .......................... 652
  Cui Tao, Puqiang Wu, and Yuji Zhang

**Pattern Mining and Application of Big Data**

QRM: A Probabilistic Model for Search Engine Query Recommendation . . . 665
  JianGuo Wang and Joshua Zhexue Huang

Toward Mining User Traversal Patterns in the Indoor Environment . . . . . 677
  Shan-Yun Teng, Tzu-Yuan Chung, Kun-Ta Chuang, and Wei-Shinn Ku

Myocardial Infarction Classification by Morphological Feature Extraction from Big 12-Lead ECG Data .......................... 689
  Julia Tzu-Ya Weng, Jyun-Jie Lin, Yi-Cheng Chen, and Pei-Chann Chang

Automatic Restaurant Information and Keyword Extraction by Mining Blog Data for Chinese Restaurant Search .......................... 700
  Chien-Li Chou, Min-Ho Tsai, Chien-Ho Chao, Hsiao-Jung Lin, Hua-Tsung Chen, Suh-Yin Lee, and Chien-Peng Ho

The Application of Association Rules in Clinical Disease: The Relationship Between Recovery After Operation of Endovascular Aneurysm Repairing and Chronic .................................................. 712
  Lin Hui, Chun-Che Shih, Huan-Chao Keh, Po-Yuan Yu, Yuan-Cheng Cheng, and Nan-Ching Huang

Human Action Recognition in Video Under Clutter and Moving Background . . . 722
  Der-Jyh Duh, Cheng-Chung Kan, Shu-Yuan Chen, and Chia-Ming Lu

Big Data Generation: Application of Mobile Healthcare .......................... 735
  Huan-Chao Keh, Lin Hui, Kuang-Yi Chou, Yuan-Cheng Cheng, Po-Yuan Yu, and Nan-Ching Huang

A Comparison of Feature-Combination for Example-Based Super Resolution . . . 744
  Shwu-Huey Yen, Jen-Hui Tsao, and Wan-Ting Liao

Comparative Analysis of 3D-Culture System for Murine Neonatal Heart Regeneration: A Systematic Approach for Big Gene Expression Data .......................... 754
  Julia Tzu-Ya Weng, Yi-Cheng Chen, Pei-Chann Chang, Shin-Ping Huang, and Yu-Wei Chiu

Data Collection and Analysis from Social Network – Profile Analyzer System (PAS) .......................... 765
  Wei-Hao Chang, Bing Li, and Xin Fang

Analyzing Personality Traits Based on User Behaviours of Using Microblog . . . 773
  Xin Fang, Bing Li, and Wei-Hao Chang
BigSAM: Mining Interesting Patterns from Probabilistic Databases of Uncertain Big Data

Fan Jiang, Carson Kai-Sang Leung, and Richard Kyle MacKinnon

MBPD: Motif-Based Period Detection

Rasaq Otunba, Jessica Lin, and Pavel Senin

A Model-Based Multivariate Time Series Clustering Algorithm

Pei-Yuan Zhou and Keith C.C. Chan

Mining Time-Aware Transit Patterns for Route Recommendation in Big Check-in Data

Hsun-Ping Hsieh and Cheng-Te Li

Author Index
Trends and Applications in Knowledge Discovery and Data Mining
2014, XXI, 833 p. 348 illus., Softcover
ISBN: 978-3-319-13185-6