Contents – Part II

Privacy and Trust

Detecting Anomalous Ratings Using Matrix Factorization for Recommender Systems ........................................ 3
    Zhihai Yang, Zhongmin Cai, and Xinyuan Chen

A Novel Spatial Cloaking Scheme Using Hierarchical Hilbert Curve for Location-Based Services ........................................ 15
    Ningning Cui, Xiaochun Yang, and Bin Wang

Efficient Privacy-Preserving Content-Based Image Retrieval in the Cloud ........................................ 28
    Kai Huang, Ming Xu, Shaojing Fu, and Dongsheng Wang

Preserving the \( d \)-Reachability When Anonymizing Social Networks ........................................ 40
    Xiangyu Liu, Jiajia Li, Dahai Zhou, Yunzhe An, and Xiufeng Xia

Personalized Location Anonymity - A Kernel Density Estimation Approach ........................................ 52
    Dapeng Zhao, Jiansong Ma, Xiaoling Wang, and Xiuxia Tian

Detecting Data-model-oriented Anomalies in Parallel Business Process ........................................ 65
    Ning Yin, Shanshan Wang, Hongyan Li, and Lilue Fan

Learning User Credibility on Aspects from Review Texts ........................................ 78
    Yifan Gao, Yuming Li, Yanhong Pan, Jiali Mao, and Rong Zhang

Detecting Anomaly in Traffic Flow from Road Similarity Analysis ........................................ 92
    Xinran Liu, Xingwu Liu, Yuanhong Wang, Juhua Pu, and Xiangliang Zhang

Query Processing and Optimization

PACOKS: Progressive Ant-Colony-Optimization-Based Keyword Search over Relational Databases ........................................ 107
    Ziyu Lin, Qian Xue, and Yongxuan Lai

Enhanced Query Classification with Millions of Fine-Grained Topics ........................................ 120
    Qi Ye, Feng Wang, Bo Li, and Zhimin Liu

A Hybrid Machine-Crowdsourcing Approach for Web Table Matching and Cleaning ........................................ 132
    Chunhua Li, Pengpeng Zhao, Victor S. Sheng, Zhixu Li, Guanfeng Liu, Jian Wu, and Zhiming Cui
An Update Method for Shortest Path Caching with Burst Paths Based on Sliding Windows .......................................................... 145
   Xiaohua Li, Ning Wang, Kanggui Peng, Xiaochun Yang, and Ge Yu

Low Overhead Log Replication for Main Memory Database System ....... 159
   Jinwei Guo, Chendong Zhang, Peng Cai, Minqi Zhou, and Aoying Zhou

Diversification of Keyword Query Result Patterns .......................... 171
   Cem Aksoy, Ananya Dass, Dimitri Theodoratos, and Xiaoying Wu

Efficient Approximate Substring Matching in Compressed String ........ 184
   Yutong Han, Bin Wang, and Xiaochun Yang

Top-K Similarity Search for Query-By-Humming ............................ 198
   Peipei Wang, Bin Wang, and Shiyiing Luo

Social Media

Restricted Boltzmann Machines for Retweeting Behaviours Prediction ..... 213
   Xiang Li, Lijuan Xie, Yong Tan, and Qiuli Tong

Cross-Collection Emotion Tagging for Online News ....................... 225
   Li Yu, Xue Zhao, Chao Wang, Haiwei Zhang, and Ying Zhang

Online News Emotion Prediction with Bidirectional LSTM ................ 238
   Xue Zhao, Chao Wang, Zhifan Yang, Ying Zhang, and Xiaojie Yuan

Learning for Search Results Diversification in Twitter ....................... 251
   Ying Wang, Zhunchen Luo, and Yang Yu

Adjustable Time-Window-Based Event Detection on Twitter ................ 265
   Qinyi Wang, Jieying She, Tianshu Song, Yongxin Tong, Lei Chen, and Ke Xu

User-IBTM: An Online Framework for Hashtag Suggestion in Twitter .... 279
   Jia Li and Hua Xu

Unifying User and Message Clustering Information for Retweeting Behavior Prediction .......................................................... 291
   Bo Jiang, Jiguang Liang, Ying Sha, Lihong Wang, Zhixin Kuang, Rui Li,
   and Peng Li

KPCA-WT: An Efficient Framework for High Quality Microblog
   Extraction in Time-Frequency Domain ........................................ 304
   Min Peng, Xinyuan Dai, Kai Zhang, Guanyin Zeng, Jiahui Zhu,
   Shuang Ouyang, Qianqian Xie, and Gang Tian
Big Data Analytics

Active Learning Method for Constraint-Based Clustering Algorithms. .... 319
  Lijun Cai, Tinghao Yu, Tingqin He, Lei Chen, and Meiqi Lin

An Effective Cluster Assignment Strategy for Large Time Series Data. .... 330
  Damir Mirzanurov, Waqas Nawaz, JooYoung Lee, and Qiang Qu

AdaWIRL: A Novel Bayesian Ranking Approach for Personal Big-Hit
Paper Prediction ......................................................... 342
  Chuxu Zhang, Lu Yu, Jie Lu, Tao Zhou, and Zi-Ke Zhang

Detecting Live Events by Mining Textual and Spatial-Temporal Features
from Microblogs ......................................................... 356
  Zhejun Zheng, Beihong Jin, Yanling Cui, and Qiang Ji

A Label Correlation Based Weighting Feature Selection Approach
for Multi-label Data. .................................................... 369
  Lu Liu, Jing Zhang, Peipei Li, Yuhong Zhang, and Xuegang Hu

Valuable Group Trajectory Pattern Mining Directed by Adaptable Value
Measuring Model ....................................................... 380
  Xinyu Huang, Tengjiao Wang, Shun Li, and Wei Chen

DualPOS: A Semi-supervised Attribute Selection Approach for Symbolic
Data Based on Rough Set Theory ........................................ 392
  Jianhua Dai, Huifeng Han, Hu Hu, Qinghua Hu, Jinghong Zhang,
  and Wentao Wang

Semi-supervised Clustering Based on Artificial Bee Colony Algorithm
with Kernel Strategy ....................................................... 403
  Jianhua Dai, Huifeng Han, Hu Hu, Qinghua Hu, Bingjie Wei,
  and Yuejue Yan

Distributed and Cloud Computing

HMNRS: A Hierarchical Multi-source Name Resolution Service
for the Industrial Internet ................................................ 417
  Yang Liu, Guoqiang Fu, and Xinchi Li

Optimizing Replica Exchange Strategy for Load Balancing
in Multienant Databases ................................................ 430
  Teng Liu, Qingzhong Li, Lanju Kong, Lei Liu, and Lizhen Cui

ERPC: An Edge-Resources Based Framework to Reduce Bandwidth Cost
in the Personal Cloud .................................................... 444
  Shaoduo Gan, Jie Yu, Xiaoling Li, Jun Ma, Lei Luo, Qingbo Wu,
  and Shasha Li
Multidimensional Similarity Join Using MapReduce ............................. 457  
Ye Li, Jian Wang, and Leong Hou U

Real-Time Logo Recognition from Live Video Streams Using an Elastic Cloud Platform ................................................................. 469  
Jianbing Ding, Hongyang Chao, and Mansheng Yang

Profit Based Two-Step Job Scheduling in Clouds ................................. 481  
Shuo Zhang, Li Pan, Shijun Liu, Lei Wu, and Xiangxu Meng

A Join Optimization Method for CPU/MIC Heterogeneous Systems .......... 493  
Kailai Zhou, Hong Chen, Hui Sun, Cuiping Li, and Tianzhen Wu

GFSF: A Novel Similarity Join Method Based on Frequency Vector ........... 506  
Ziyu Lin, Daowen Luo, and Yongxuan Lai

Erratum to: Web-Age Information Management (Part I and II) ............... E1  
Bin Cui, Nan Zhang, Jianliang Xu, Xiang Lian, and Dexi Liu

Demo Papers

SHMS: A Smart Phone Self-health Management System Using Data Mining. . 521  
Chuanhua Xu, Jia Zhu, Zhixu Li, Jing Xiao, Changqin Huang, and Yong Tang

MVUC: An Interactive System for Mining and Visualizing Urban Co-locations ................................................................. 524  
Xiao Wang, Hongmei Chen, and Qing Xiao

SNEExtractor: A Prototype for Extracting Semantic Networks from Web Documents ................................................................................ 527  
Chi Zhang, Yanhua Wang, Chengyu Wang, Wenliang Cheng, and Xiaofeng He

Crowd-PANDA: Using Crowdsourcing Method for Academic Knowledge Acquisition ........................................................................ 531  
Zhaoan Dong, Jiaheng Lu, and Tok Wang Ling

LPSMon: A Stream-Based Live Public Sentiment Monitoring System ....... 534  
Kun Ma, Zijie Tang, Jialin Zhong, and Bo Yang

DPBT: A System for Detecting Pacemakers in Burst Topics .................... 537  
Guozhong Dong, Wu Yang, Feida Zhu, and Wei Wang

CEQA - An Open Source Chinese Question Answer System Based on Linked Knowledge ....................................................... 540  
Zeyu Du, Yan Yang, Qinming Hu, and Liang He
OSSRec: An Open Source Software Recommendation System Based on Wisdom of Crowds. ......................................................... 544

Mengwen Chen, Gang Yin, Chenxi Song, Tao Wang, Cheng Yang, and Huaimin Wang

Author Index ................................................................. 549
## Data Mining

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>More Efficient Algorithm for Mining Frequent Patterns with Multiple</td>
<td>3</td>
</tr>
<tr>
<td>Minimum Supports</td>
<td></td>
</tr>
<tr>
<td>Wensheng Gan, Jerry Chun-Wei Lin, Philippe Fournier-Viger,</td>
<td></td>
</tr>
<tr>
<td>and Han-Chieh Chao</td>
<td></td>
</tr>
<tr>
<td>Efficient Mining of Uncertain Data for High-Utility Itemsets</td>
<td>17</td>
</tr>
<tr>
<td>Jerry Chun-Wei Lin, Wensheng Gan, Philippe Fournier-Viger,</td>
<td></td>
</tr>
<tr>
<td>Tzung-Pei Hong, and Vincent S. Tseng</td>
<td></td>
</tr>
<tr>
<td>An Improved HMM Model for Sensing Data Predicting in WSN</td>
<td>31</td>
</tr>
<tr>
<td>Zeyu Zhang, Bailong Deng, Siding Chen, and Li Li</td>
<td></td>
</tr>
<tr>
<td>eXtreme Gradient Boosting for Identifying Individual Users Across</td>
<td>43</td>
</tr>
<tr>
<td>Different Digital Devices</td>
<td></td>
</tr>
<tr>
<td>Rongwei Song, Siding Chen, Bailong Deng, and Li Li</td>
<td></td>
</tr>
<tr>
<td>Two-Phase Mining for Frequent Closed Episodes</td>
<td>55</td>
</tr>
<tr>
<td>Guoqiong Liao, Xiaoting Yang, Sihong Xie, Philip S. Yu,</td>
<td></td>
</tr>
<tr>
<td>and Changxuan Wan</td>
<td></td>
</tr>
<tr>
<td>Effectively Updating High Utility Co-location Patterns in Evolving</td>
<td>67</td>
</tr>
<tr>
<td>Spatial Databases</td>
<td></td>
</tr>
<tr>
<td>Xiaoxuan Wang, Lizhen Wang, Junli Lu, and Lihua Zhou</td>
<td></td>
</tr>
<tr>
<td>Mining Top-k Distinguishing Sequential Patterns with Flexible Gap</td>
<td>82</td>
</tr>
<tr>
<td>Constraints</td>
<td></td>
</tr>
<tr>
<td>Chao Gao, Lei Duan, Guozhu Dong, Haiqing Zhang, Hao Yang,</td>
<td></td>
</tr>
<tr>
<td>and Changjie Tang</td>
<td></td>
</tr>
<tr>
<td>A Novel Chinese Text Mining Method for E-Commerce Review Spam</td>
<td>95</td>
</tr>
<tr>
<td>Detection</td>
<td></td>
</tr>
<tr>
<td>Xiu Li and Xinwei Yan</td>
<td></td>
</tr>
</tbody>
</table>

## Spatial and Temporal Databases

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retrieving Routes of Interest Over Road Networks</td>
<td>109</td>
</tr>
<tr>
<td>Wengen Li, Jiannong Cao, Jihong Guan, Man Lung Yu, and Shuigeng Zhou</td>
<td></td>
</tr>
</tbody>
</table>
Semantic-Aware Trajectory Compression with Urban Road Network  
*Na Ta, Guoliang Li, Bole Chen, and Jianhua Feng*

Discovering Underground Roads from Trajectories Without Road Network  
*Qiuge Song, Jiali Mao, and Cheqing Jin*

Ridesharing Recommendation: Whether and Where Should I Wait?  
*Chengcheng Dai*

Keyword-aware Optimal Location Query in Road Network  
*Jinling Bao, Xingshan Liu, Rui Zhou, and Bin Wang*

Point-of-Interest Recommendations by Unifying Multiple Correlations  
*Ce Cheng, Jiajin Huang, and Ning Zhong*

Top-k Team Recommendation in Spatial Crowdsourcing  
*Dawei Gao, Yongxin Tong, Jiying She, Tianshu Song, Lei Chen, and Ke Xu*

Explicable Location Prediction Based on Preference Tensor Model  
*Duoduo Zhang, Ning Yang, and Yuchi Ma*

**Recommender Systems**

Random Partition Factorization Machines for Context-Aware Recommendations  
*Shaoqing Wang, Cuilan Du, Kankan Zhao, Cuiping Li, Yangxi Li, Yang Zheng, Zheng Wang, and Hong Chen*

A Novel Framework to Process the Quantity and Quality of User Behavior Data in Recommender Systems  
*Penghua Yu, Lanfen Lin, and Yuangang Yao*

RankMBPR: Rank-Aware Mutual Bayesian Personalized Ranking for Item Recommendation  
*Lu Yu, Ge Zhou, Chuxu Zhang, Junming Huang, Chuang Liu, and Zi-Ke Zhang*

Unsupervised Expert Finding in Social Network for Personalized Recommendation  
*Junmei Ding, Yan Chen, Xin Li, Guiquan Liu, Aili Shen, and Xiangfu Meng*

An Approach for Clothing Recommendation Based on Multiple Image Attributes  
*Dandan Sha, Daling Wang, Xiangmin Zhou, Shi Feng, Yifei Zhang, and Ge Yu*
XXII Contents – Part I

BMF: An Indexing Structure to Support Multi-element Check ................. 441
  Chenyang Xu, Qin Liu, and Weixiong Rao

Efficient Unique Column Combinations Discovery Based on Data Distribution ................................................. 454
  Chao Wang, Shupeng Han, Xiangrui Cai, Haiwei Zhang, and Yanlong Wen

Event Related Document Retrieval Based on Bipartite Graph ............... 467
  Wenjing Yang, Rui Li, Peng Li, Meilin Zhou, and Bin Wang

SPedia: A Semantics Based Repository of Scientific Publications Data .... 479
  Muhammad Ahtisham Aslam and Naif Radi Aljohani

A Set-Based Training Query Classification Approach for Twitter Search ... 491
  Qingli Ma, Ben He, Jungang Xu, and Bin Wang

NBLucene: Flexible and Efficient Open Source Search Engine ............... 504
  Zhaohua Zhang, Benjun Ye, Jiayi Huang, Rebecca Stones, Gang Wang, and Xiaoguang Liu

Context-Aware Entity Summarization ........................................... 517
  Jihong Yan, Yanhua Wang, Ming Gao, and Aoying Zhou

Erratum to: BMF: An Indexing Structure to Support Multi-element Check ... E1
  Chenyang Xu, Qin Liu, and Weixiong Rao

Author Index .............................................................................. 531
Web-Age Information Management
17th International Conference, WAIM 2016, Nanchang, China, June 3-5, 2016, Proceedings, Part II
Cui, B.; Zhang, N.; Xu, J.; Lian, X.; Liu, D. (Eds.)
2016, XXII, 552 p. 203 illus., Softcover
ISBN: 978-3-319-39957-7