Contents – Part I

Data Mining

More Efficient Algorithm for Mining Frequent Patterns with Multiple Minimum Supports .................................................. 3
    Wensheng Gan, Jerry Chun-Wei Lin, Philippe Fournier-Viger, and Han-Chieh Chao

Efficient Mining of Uncertain Data for High-Utility Itemsets ............. 17
    Jerry Chun-Wei Lin, Wensheng Gan, Philippe Fournier-Viger, Tzung-Pei Hong, and Vincent S. Tseng

An Improved HMM Model for Sensing Data Predicting in WSN .......... 31
    Zeyu Zhang, Bailong Deng, Siding Chen, and Li Li

eXtreme Gradient Boosting for Identifying Individual Users Across Different Digital Devices ............................................. 43
    Rongwei Song, Siding Chen, Bailong Deng, and Li Li

Two-Phase Mining for Frequent Closed Episodes ............................ 55
    Guoqiong Liao, Xiaoting Yang, Sihong Xie, Philip S. Yu, and Changxuan Wan

Effectively Updating High Utility Co-location Patterns in Evolving Spatial Databases .................................................. 67
    Xiaoxuan Wang, Lizhen Wang, Junli Lu, and Lihua Zhou

Mining Top-\(k\) Distinguishing Sequential Patterns with Flexible Gap Constraints ......................................................... 82
    Chao Gao, Lei Duan, Guozhu Dong, Haiqing Zhang, Hao Yang, and Changjie Tang

A Novel Chinese Text Mining Method for E-Commerce Review Spam Detection ......................................................... 95
    Xiu Li and Xinwei Yan

Spatial and Temporal Databases

Retrieving Routes of Interest Over Road Networks .......................... 109
    Wengen Li, Jiannong Cao, Jihong Guan, Man Lung Yu, and Shuigeng Zhou
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic-Aware Trajectory Compression with Urban Road Network</td>
<td>124</td>
</tr>
<tr>
<td>Na Ta, Guoliang Li, Bole Chen, and Jianhua Feng</td>
<td></td>
</tr>
<tr>
<td>Discovering Underground Roads from Trajectories Without Road Network</td>
<td>137</td>
</tr>
<tr>
<td>Qiuge Song, Jiali Mao, and Cheqing Jin</td>
<td></td>
</tr>
<tr>
<td>Ridesharing Recommendation: Whether and Where Should I Wait?</td>
<td>151</td>
</tr>
<tr>
<td>Chengcheng Dai</td>
<td></td>
</tr>
<tr>
<td>Keyword-aware Optimal Location Query in Road Network</td>
<td>164</td>
</tr>
<tr>
<td>Jinling Bao, Xingshan Liu, Rui Zhou, and Bin Wang</td>
<td></td>
</tr>
<tr>
<td>Point-of-Interest Recommendations by Unifying Multiple Correlations</td>
<td>178</td>
</tr>
<tr>
<td>Ce Cheng, Jiajin Huang, and Ning Zhong</td>
<td></td>
</tr>
<tr>
<td>Top-k Team Recommendation in Spatial Crowdsourcing</td>
<td>191</td>
</tr>
<tr>
<td>Dawei Gao, Yongxin Tong, Jieying She, Tianshu Song, Lei Chen,</td>
<td></td>
</tr>
<tr>
<td>and Ke Xu</td>
<td></td>
</tr>
<tr>
<td>Explicable Location Prediction Based on Preference Tensor Model</td>
<td>205</td>
</tr>
<tr>
<td>Duoduo Zhang, Ning Yang, and Yuchi Ma</td>
<td></td>
</tr>
</tbody>
</table>

**Recommender Systems**

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Partition Factorization Machines for Context-Aware Recommendations</td>
<td>219</td>
</tr>
<tr>
<td>Shaoqing Wang, Cuilan Du, Kankan Zhao, Cuiping Li, Yangxi Li,</td>
<td></td>
</tr>
<tr>
<td>Yang Zheng, Zheng Wang, and Hong Chen</td>
<td></td>
</tr>
<tr>
<td>A Novel Framework to Process the Quantity and Quality of User Behavior</td>
<td>231</td>
</tr>
<tr>
<td>Penghua Yu, Lanfen Lin, and Yuangang Yao</td>
<td></td>
</tr>
<tr>
<td>RankMBPR: Rank-Aware Mutual Bayesian Personalized Ranking for Item Recommendation</td>
<td>244</td>
</tr>
<tr>
<td>Lu Yu, Ge Zhou, Chuxu Zhang, Junming Huang, Chuang Liu,</td>
<td></td>
</tr>
<tr>
<td>and Zi-Ke Zhang</td>
<td></td>
</tr>
<tr>
<td>Unsupervised Expert Finding in Social Network for Personalized</td>
<td>257</td>
</tr>
<tr>
<td>Recommendation</td>
<td></td>
</tr>
<tr>
<td>Junmei Ding, Yan Chen, Xin Li, Guiquan Liu, Aili Shen,</td>
<td></td>
</tr>
<tr>
<td>and Xiangfu Meng</td>
<td></td>
</tr>
<tr>
<td>An Approach for Clothing Recommendation Based on Multiple Image</td>
<td>272</td>
</tr>
<tr>
<td>Attributes</td>
<td></td>
</tr>
<tr>
<td>Dandan Sha, Daling Wang, Xiangmin Zhou, Shi Feng, Yifei Zhang,</td>
<td></td>
</tr>
<tr>
<td>and Ge Yu</td>
<td></td>
</tr>
</tbody>
</table>
SocialFM: A Social Recommender System with Factorization Machines . . . . 286
   Juming Zhou, Dong Wang, Yue Ding, and Litian Yin

Identifying Linked Data Datasets for sameAs Interlinking Using
Recommendation Techniques . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 298
   Haichi Liu, Ting Wang, Jintao Tang, Hong Ning, Dengping Wei, Songxian Xie, and Peilei Liu

Query-Biased Multi-document Abstractive Summarization via Submodular
Maximization Using Event Guidance . . . . . . . . . . . . . . . . . . . . . . . . . . . . 310
   Rui Sun, Zhenchao Wang, Yafeng Ren, and Donghong Ji

Graph Data Management

Inferring Diffusion Network on Incomplete Cascade Data . . . . . . . . . 325
   Peng Dou, Sizhen Du, and Guojie Song

Anchor Link Prediction Using Topological Information in Social Networks . . 338
   Shuo Feng, Derong Shen, Tiezheng Nie, Yue Kou, and Ge Yu

Collaborative Partitioning for Multiple Social Networks with Anchor Nodes . . 353
   Fenglan Li, Anming Ji, Songchang Jin, Shuqiang Yang, and Qiang Liu

A General Framework for Graph Matching and Its Application in Ontology
Matching . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 365
   Yuda Zang, Jianyong Wang, and Xuan Zhu

Internet Traffic Analysis in a Large University Town: A Graphical
and Clustering Approach . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 378
   Weitao Weng, Kai Lei, Kuai Xu, Xiaoyou Liu, and Tao Sun

Conceptual Sentence Embeddings . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 390
   Yashen Wang, Heyan Huang, Chong Feng, Qiang Zhou, and Jiahui Gu

Inferring Social Roles of Mobile Users Based on Communication
Behaviors . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 402
   Yipeng Chen, Hongyan Li, Jinbo Zhang, and Gaoshan Miao

Sparse Topical Coding with Sparse Groups . . . . . . . . . . . . . . . . . . . . . . . . . . 415
   Min Peng, Qianqian Xie, Jiajia Huang, Jiahui Zhu, Shuang Ouyang, Jimin Huang, and Gang Tian

Information Retrieval

Differential Evolution-Based Fusion for Results Diversification of Web
Search . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 429
   Chunlin Xu, Chunlan Huang, and Shengli Wu
BMF: An Indexing Structure to Support Multi-element Check  
Chenyang Xu, Qin Liu, and Weixiong Rao  

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

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

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

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

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

Context-Aware Entity Summarization  
Jihong Yan, Yanhua Wang, Mings Gao, and Aoying Zhou  

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

Author Index  

XVI Contents – Part I
## Contents – Part II

### Privacy and Trust

Detecting Anomalous Ratings Using Matrix Factorization for Recommender Systems

*Zhihai Yang, Zhongmin Cai, and Xinyuan Chen*

A Novel Spatial Cloaking Scheme Using Hierarchical Hilbert Curve for Location-Based Services

*Ningning Cui, Xiaochun Yang, and Bin Wang*

Efficient Privacy-Preserving Content-Based Image Retrieval in the Cloud

*Kai Huang, Ming Xu, Shaojing Fu, and Dongsheng Wang*

Preserving the $d$-Reachability When Anonymizing Social Networks

*Xiangyu Liu, Jiajia Li, Dahai Zhou, Yunzhe An, and Xiufeng Xia*

Personalized Location Anonymity - A Kernel Density Estimation Approach

*Dapeng Zhao, Jiashong Ma, Xiaoling Wang, and Xiuxia Tian*

Detecting Data-model-oriented Anomalies in Parallel Business Process

*Ning Yin, Shanshan Wang, Hongyan Li, and Lilue Fan*

Learning User Credibility on Aspects from Review Texts

*Yifan Gao, Yuming Li, Yanhong Pan, Jiali Mao, and Rong Zhang*

Detecting Anomaly in Traffic Flow from Road Similarity Analysis

*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

*Ziyu Lin, Qian Xue, and Yongxuan Lai*

Enhanced Query Classification with Millions of Fine-Grained Topics

*Qi Ye, Feng Wang, Bo Li, and Zhimin Liu*

A Hybrid Machine-Crowdsourcing Approach for Web Table Matching and Cleaning

*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 Shiying 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 Yuejun 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
Web-Age Information Management
17th International Conference, WAIM 2016, Nanchang, China, June 3-5, 2016, Proceedings, Part I
Cui, B.; Zhang, N.; Xu, J.; Lian, X.; Liu, D. (Eds.)
2016, XXI, 534 p. 193 illus., Softcover
ISBN: 978-3-319-39936-2