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>\textit{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>\textit{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>\textit{Chengcheng Dai}</td>
<td></td>
</tr>
<tr>
<td>Keyword-aware Optimal Location Query in Road Network</td>
<td>164</td>
</tr>
<tr>
<td>\textit{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>\textit{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>\textit{Dawei Gao, Yongxin Tong, Jiying She, Tianshu Song, Lei Chen, and Ke Xu}</td>
<td></td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>SocialFM: A Social Recommender System with Factorization Machines</td>
<td>286</td>
</tr>
<tr>
<td><em>Juming Zhou, Dong Wang, Yue Ding, and Litian Yin</em></td>
<td></td>
</tr>
<tr>
<td>Identifying Linked Data Datasets for sameAs Interlinking Using</td>
<td>298</td>
</tr>
<tr>
<td><em>Haichi Liu, Ting Wang, Jintao Tang, Hong Ning, Dengping Wei, Songxian Xie, and Peilei Liu</em></td>
<td></td>
</tr>
<tr>
<td>Query-Biased Multi-document Abstractive Summarization via Submodular Maximization Using Event Guidance</td>
<td>310</td>
</tr>
<tr>
<td><em>Rui Sun, Zhenchao Wang, Yafeng Ren, and Donghong Ji</em></td>
<td></td>
</tr>
</tbody>
</table>

**Graph Data Management**

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inferring Diffusion Network on Incomplete Cascade Data</td>
<td>325</td>
</tr>
<tr>
<td><em>Peng Dou, Sizhen Du, and Guojie Song</em></td>
<td></td>
</tr>
<tr>
<td>Anchor Link Prediction Using Topological Information in Social Networks</td>
<td>338</td>
</tr>
<tr>
<td><em>Shuo Feng, Derong Shen, Tiezheng Nie, Yue Kou, and Ge Yu</em></td>
<td></td>
</tr>
<tr>
<td>Collaborative Partitioning for Multiple Social Networks with Anchor Nodes</td>
<td>353</td>
</tr>
<tr>
<td><em>Fenglan Li, Anming Ji, Songchang Jin, Shuqiang Yang, and Qiang Liu</em></td>
<td></td>
</tr>
<tr>
<td>A General Framework for Graph Matching and Its Application in Ontology Matching</td>
<td>365</td>
</tr>
<tr>
<td><em>Yuda Zang, Jianyong Wang, and Xuan Zhu</em></td>
<td></td>
</tr>
<tr>
<td>Internet Traffic Analysis in a Large University Town: A Graphical and Clustering Approach</td>
<td>378</td>
</tr>
<tr>
<td><em>Weitao Weng, Kai Lei, Kuai Xu, Xiaoyou Liu, and Tao Sun</em></td>
<td></td>
</tr>
<tr>
<td>Conceptual Sentence Embeddings</td>
<td>390</td>
</tr>
<tr>
<td><em>Yashen Wang, Heyan Huang, Chong Feng, Qiang Zhou, and Jiahui Gu</em></td>
<td></td>
</tr>
<tr>
<td>Inferring Social Roles of Mobile Users Based on Communication Behaviors</td>
<td>402</td>
</tr>
<tr>
<td><em>Yipeng Chen, Hongyan Li, Jinbo Zhang, and Gaoshan Miao</em></td>
<td></td>
</tr>
<tr>
<td>Sparse Topical Coding with Sparse Groups</td>
<td>415</td>
</tr>
<tr>
<td><em>Min Peng, Qianqian Xie, Jiajia Huang, Jiahui Zhu, Shuang Ouyang, Jimin Huang, and Gang Tian</em></td>
<td></td>
</tr>
</tbody>
</table>

**Information Retrieval**

<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Differential Evolution-Based Fusion for Results Diversification of Web Search</td>
<td>429</td>
</tr>
<tr>
<td><em>Chunlin Xu, Chunlan Huang, and Shengli Wu</em></td>
<td></td>
</tr>
</tbody>
</table>
## 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, Yanzhe 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

Xiaohua Li, Ning Wang, Kanggui Peng, Xiaochun Yang, and Ge Yu

Low Overhead Log Replication for Main Memory Database System

Jinwei Guo, Chendong Zhang, Peng Cai, Minqi Zhou, and Aoying Zhou

Diversification of Keyword Query Result Patterns

Cem Aksoy, Ananya Dass, Dimitri Theodoratos, and Xiaoying Wu

Efficient Approximate Substring Matching in Compressed String

Yutong Han, Bin Wang, and Xiaochun Yang

Top-K Similarity Search for Query-By-Humming

Peipei Wang, Bin Wang, and Shiying Luo

Social Media

Restricted Boltzmann Machines for Retweeting Behaviours Prediction

Xiang Li, Lijuan Xie, Yong Tan, and Qiuli Tong

Cross-Collection Emotion Tagging for Online News

Li Yu, Xue Zhao, Chao Wang, Haiwei Zhang, and Ying Zhang

Online News Emotion Prediction with Bidirectional LSTM

Xue Zhao, Chao Wang, Zhifan Yang, Ying Zhang, and Xiaojie Yuan

Learning for Search Results Diversification in Twitter

Ying Wang, Zhunchen Luo, and Yang Yu

Adjustable Time-Window-Based Event Detection on Twitter

Qinyi Wang, Jieying She, Tianshu Song, Yongxin Tong, Lei Chen, and Ke Xu

User-IBTM: An Online Framework for Hashtag Suggestion in Twitter

Jia Li and Hua Xu

Unifying User and Message Clustering Information for Retweeting Behavior Prediction

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

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

**Author Index**
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