Contents – Part II

Machine Learning

Combining Node Identifier Features and Community Priors for Within-Network Classification

Qi Ye, Changlei Zhu, Gang Li, and Feng Wang

An Active Learning Approach to Recognizing Domain-Specific Queries From Query Log

Weijian Ni, Tong Liu, Haohao Sun, and Zhensheng Wei

Event2vec: Learning Representations of Events on Temporal Sequences

Shenda Hong, Meng Wu, Hongyan Li, and Zhengwu Wu

Joint Emoji Classification and Embedding Learning

Xiang Li, Rui Yan, and Ming Zhang

Target-Specific Convolutional Bi-directional LSTM Neural Network for Political Ideology Analysis

Xitian Li, Wei Chen, Tengjiao Wang, and Weijing Huang

Boost Clickbait Detection Based on User Behavior Analysis

Hai-Tao Zheng, Xin Yao, Yong Jiang, Shu-Tao Xia, and Xi Xiao

Recommendation Systems

A Novel Hybrid Friends Recommendation Framework for Twitter

Yan Zhao, Jia Zhu, Mengdi Jia, Wenyuan Yang, and Kai Zheng

A Time and Sentiment Unification Model for Personalized Recommendation

Qinyaung Wang, Hongzhi Yin, and Hao Wang

Personalized POI Groups Recommendation in Location-Based Social Networks

Fei Yu, Zhijun Li, Shouxu Jiang, and Xiaofei Yang

Learning Intermediary Category Labels for Personal Recommendation

Wenli Yu, Li Li, Jingyuan Wang, Dengbao Wang, Yong Wang, Zhanbo Yang, and Min Huang

Skyline-Based Recommendation Considering User Preferences

Shuhei Kishida, Seiji Ueda, Atsushi Keyaki, and Jun Miyazaki
Improving Topic Diversity in Recommendation Lists: Marginally
or Proportionally? .............................................................. 142
   Xiaolu Xing, Chaofeng Sha, and Junyu Niu

**Distributed Data Processing and Applications**

Integrating Feedback-Based Semantic Evidence to Enhance Retrieval
Effectiveness for Clinical Decision Support .......................... 153
   Chenhao Yang, Ben He, and Jungang Xu

Reordering Transaction Execution to Boost High Frequency Trading
Applications ................................................................. 169
   Ningnan Zhou, Xuan Zhou, Xiao Zhang, Xiaoyong Du, and Shan Wang

Bus-OLAP: A Bus Journey Data Management Model for Non-on-time
Events Query .............................................................. 185
   Tinghai Pang, Lei Duan, Jyrki Nummenmaa, Jie Zuo, and Peng Zhang

Distributed Data Mining for Root Causes of KPI Faults
in Wireless Networks ...................................................... 201
   Shiliang Fan, Yubin Yang, Wenyang Lu, and Ping Song

Precise Data Access on Distributed Log-Structured Merge-Tree .......... 210
   Tao Zhu, Huiqi Hu, Weining Qian, Aoying Zhou, Mengzhan Liu,
   and Qiong Zhao

Cuttle: Enabling Cross-Column Compression in Distributed Column Stores . . 219
   Hao Liu, Jiang Xiao, Xianjun Guo, Haoyu Tan, Qiong Luo,
   and Lionel M. Ni

**Machine Learning and Optimization**

Optimizing Window Aggregate Functions via Random Sampling .......... 229
   Guangxuan Song, Wenwen Qu, Yilin Wang, and Xiaoling Wang

Fast Log Replication in Highly Available Data Store .................. 245
   Donghui Wang, Peng Cai, Weining Qian, Aoying Zhou, Tianze Pang,
   and Jing Jiang

New Word Detection in Ancient Chinese Literature ................... 260
   Tao Xie, Bin Wu, and Bai Wang

Identifying Evolutionary Topic Temporal Patterns Based on Bursty
Phrase Clustering ......................................................... 276
   Yixuan Liu, Zihao Gao, and Mizuho Iwaihara
Personalized Citation Recommendation via Convolutional Neural Networks .................................................. 285
Jun Yin and Xiaoming Li

A Streaming Data Prediction Method Based on Evolving Bayesian Network .................................................. 294
Yongheng Wang, Guidan Chen, and Zengwang Wang

A Learning Approach to Hierarchical Search Result Diversification .................................................. 303
Hai-Tao Zheng, Zhuren Wang, and Xi Xiao

Demo Papers

TeslaML: Steering Machine Learning Automatically in Tencent .................................................. 313
Jiawei Jiang, Ming Huang, Jie Jiang, and Bin Cui

DPHSim: A Flexible Simulator for DRAM/PCM-Based Hybrid Memory .................................................. 319
Dezhi Zhang, Peiquan Jin, Xiaoliang Wang, Chengcheng Yang, and Lihua Yue

CrowdIQ: A Declarative Crowdsourcing Platform for Improving the Quality of Web Tables .................................................. 324
Yihai Xi, Ning Wang, Xiaoyu Wu, Yuqing Bao, and Wutong Zhou

OICPM: An Interactive System to Find Interesting Co-location Patterns Using Ontologies .................................................. 329
Xuguang Bao, Lizhen Wang, and Qing Xiao

BioPW: An Interactive Tool for Biological Pathway Visualization on Linked Data .................................................. 333
Yuan Liu, Xin Wang, and Qiang Xu

ChargeMap: An Electric Vehicle Charging Station Planning System .................................................. 337
Longlong Xu, Wutao Lin, Xiaorong Wang, Zhenhui Xu, Wei Chen, and Tengjiao Wang

Topic Browsing System for Research Papers Based on Hierarchical Latent Tree Analysis .................................................. 341
Leonard K.M. Poon, Chun Fai Leung, Peixian Chen, and Nevin L. Zhang

A Tool of Benchmarking Realtime Analysis for Massive Behavior Data .................................................. 345
Mingyan Teng, Qiao Sun, Buqiao Deng, Lei Sun, and Xiongpai Qin

Interactive Entity Centric Analysis of Log Data .................................................. 349
Qiao Sun, Xiongpai Qin, Buqiao Deng, and Wei Cui
A Tool for 3D Visualizing Moving Objects .......................... 353  
Weiwei Wang and Jianqiu Xu

Author Index ............................................................... 359
Contents – Part I

Tutorials

Meta Paths and Meta Structures: Analysing Large Heterogeneous Information Networks .......................... 3
Reynold Cheng, Zhipeng Huang, Yudian Zheng, Jing Yan, Ka Yu Wong, and Eddie Ng

Spatial Data Processing and Data Quality

TrajSpark: A Scalable and Efficient In-Memory Management System for Big Trajectory Data .................. 11
Zhigang Zhang, Cheqing Jin, Jiali Mao, Xiaolin Yang, and Aoying Zhou

A Local-Global LDA Model for Discovering Geographical Topics from Social Media .......................... 27
Siwei Qiang, Yongkun Wang, and Yaohui Jin

Team-Oriented Task Planning in Spatial Crowdsourcing ......................................................... 41
Dawei Gao, Yongxin Tong, Yudian Ji, and Ke Xu

Negative Survey with Manual Selection: A Case Study in Chinese Universities ............................... 57
Jianguo Wu, Jianwen Xiang, Dongdong Zhao, Huanhuan Li, Qing Xie, and Xiaoyi Hu

Element-Oriented Method of Assessing Landscape of Sightseeing Spots by Using Social Images .......... 66
Yizhu Shen, Chenyi Zhuang, and Qiang Ma

Sifting Truths from Multiple Low-Quality Data Sources .......................................................... 74
Zizhe Xie, Qizhi Liu, and Zhifeng Bao

Graph Data Processing

A Community-Aware Approach to Minimizing Dissemination in Graphs ........................................ 85
Chuxu Zhang, Lu Yu, Chuang Liu, Zi-Ke Zhang, and Tao Zhou

Time-Constrained Graph Pattern Matching in a Large Temporal Graph ........................................ 100
Yanxia Xu, Jinjing Huang, An Liu, Zhixu Li, Hongzhi Yin, and Lei Zhao

Efficient Compression on Real World Directed Graphs ............................................................. 116
Guohua Li, Weixiong Rao, and Zhongxiao Jin
Keyphrase Extraction Using Knowledge Graphs .......................... 132
   Wei Shi, Weiguo Zheng, Jeffrey Xu Yu, Hong Cheng, and Lei Zou

Semantic-Aware Partitioning on RDF Graphs ............................. 149
   Qiang Xu, Xin Wang, Junhu Wang, Yajun Yang, and Zhiyong Feng

An Incremental Algorithm for Estimating Average Clustering Coefficient
Based on Random Walk .................................................... 158
   Qun Liao, Lei Sun, He Du, and Yulu Yang

Data Mining, Privacy and Semantic Analysis

Deep Multi-label Hashing for Large-Scale Visual Search Based
on Semantic Graph ......................................................... 169
   Chunlin Zhong, Yi Yu, Suhua Tang, Shin’ichi Satoh, and Kai Xing

An Ontology-Based Latent Semantic Indexing Approach
Using Long Short-Term Memory Networks .............................. 185
   Ningning Ma, Hai-Tao Zheng, and Xi Xiao

Privacy-Preserving Collaborative Web Services QoS Prediction
via Differential Privacy .................................................... 200
   Shuju Liu, An Liu, Zhixu Li, Guanfeng Liu, Jiajie Xu, Lei Zhao,
   and Kai Zheng

High-Utility Sequential Pattern Mining with Multiple Minimum
Utility Thresholds .......................................................... 215
   Jerry Chun-Wei Lin, Jiexiong Zhang, and Philippe Fournier-Viger

Extracting Various Types of Informative Web Content via Fuzzy
Sequential Pattern Mining ................................................. 230
   Ting Huang, Ruizhang Huang, Bowei Liu, and Yingying Yan

Exploiting High Utility Occupancy Patterns .............................. 239
   Wensheng Gan, Jerry Chun-Wei Lin, Philippe Fournier-Viger,
   and Han-Chieh Chao

Text and Log Data Management

Translation Language Model Enhancement for Community Question
Retrieval Using User Adoption Answer .................................. 251
   Ming Chen, Lin Li, and Qing Xie

Holographic Lexical Chain and Its Application in Chinese
Text Summarization ....................................................... 266
   Shengluan Hou, Yu Huang, Chaoqun Fei, Shuhan Zhang,
   and Ruqian Lu
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authorship Identification of Source Codes</td>
<td>282</td>
</tr>
<tr>
<td>Chunxia Zhang, Sen Wang, Jiayu Wu, and Zhendong Niu</td>
<td></td>
</tr>
<tr>
<td>DFDS: A Domain-Independent Framework for Document-Level Sentiment Analysis Based on RST</td>
<td>297</td>
</tr>
<tr>
<td>Zhenyu Zhao, Guozheng Rao, and Zhiyong Feng</td>
<td></td>
</tr>
<tr>
<td>Fast Follower Recovery for State Machine Replication</td>
<td>311</td>
</tr>
<tr>
<td>Jinwei Guo, Jiahaow Wang, Peng Cai, Weining Qian, Aoying Zhou, and Xiaohang Zhu</td>
<td></td>
</tr>
<tr>
<td>Laser: Load-Adaptive Group Commit in Lock-Free Transaction Logging</td>
<td>320</td>
</tr>
<tr>
<td>Huan Zhou, Huiqi Hu, Tao Zhu, Weining Qian, Aoying Zhou, and Yukun He</td>
<td></td>
</tr>
<tr>
<td><strong>Social Networks</strong></td>
<td></td>
</tr>
<tr>
<td>Detecting User Occupations on Microblogging Platforms: An Experimental Study</td>
<td>331</td>
</tr>
<tr>
<td>Xia Lv, Peiquan Jin, Lin Mu, Shouhong Wan, and Lihua Yue</td>
<td></td>
</tr>
<tr>
<td>Counting Edges and Triangles in Online Social Networks via Random Walk</td>
<td>346</td>
</tr>
<tr>
<td>Yang Wu, Cheng Long, Ada Wai-Chee Fu, and Zitong Chen</td>
<td></td>
</tr>
<tr>
<td>Fair Reviewer Assignment Considering Academic Social Network</td>
<td>362</td>
</tr>
<tr>
<td>Kaixia Li, Zhao Cao, and Dacheng Qu</td>
<td></td>
</tr>
<tr>
<td>Viral Marketing for Digital Goods in Social Networks</td>
<td>377</td>
</tr>
<tr>
<td>Yu Qiao, Jun Wu, Lei Zhang, and Chongjun Wang</td>
<td></td>
</tr>
<tr>
<td>Change Detection from Media Sharing Community</td>
<td>391</td>
</tr>
<tr>
<td>Naoki Kito, Xiangmin Zhou, Dong Qin, Yongli Ren, Xiuzhen Zhang, and James Thom</td>
<td></td>
</tr>
<tr>
<td>Measuring the Similarity of Nodes in Signed Social Networks with Positive and Negative Links</td>
<td>399</td>
</tr>
<tr>
<td>Tianchen Zhu, Zhaohui Peng, Xinghua Wang, and Xiaoguang Hong</td>
<td></td>
</tr>
<tr>
<td><strong>Data Mining and Data Streams</strong></td>
<td></td>
</tr>
<tr>
<td>Elastic Resource Provisioning for Batched Stream Processing System in Container Cloud</td>
<td>411</td>
</tr>
<tr>
<td>Song Wu, Xingjun Wang, Hai Jin, and Haibao Chen</td>
<td></td>
</tr>
<tr>
<td>An Adaptive Framework for RDF Stream Processing</td>
<td>427</td>
</tr>
<tr>
<td>Qiong Li, Xiaowang Zhang, and Zhiyong Feng</td>
<td></td>
</tr>
</tbody>
</table>
Investigating Microstructure Patterns of Enterprise Network in Perspective of Ego Network ................................................................. 444
    Xiutao Shi, Liqiang Wang, Shijun Liu, Yafang Wang, Li Pan, and Lei Wu

Neural Architecture for Negative Opinion Expressions Extraction ........... 460
    Hui Wen, Minglan Li, and Zhili Ye

Identifying the Academic Rising Stars via Pairwise Citation Increment Ranking ................................................................. 475
    Chuxu Zhang, Chuang Liu, Lu Yu, Zi-Ke Zhang, and Tao Zhou

Fuzzy Rough Incremental Attribute Reduction Applying Dependency Measures ................................................................. 484
    Yangming Liu, Suyun Zhao, Hong Chen, Cuiping Li, and Yanmin Lu

Query Processing

SET: Secure and Efficient Top-k Query in Two-Tiered Wireless Sensor Networks ................................................................. 495
    Xiaoying Zhang, Hui Peng, Lei Dong, Hong Chen, and Hui Sun

Top-k Pattern Matching Using an Information-Theoretic Criterion over Probabilistic Data Streams ................................................................. 511
    Kento Sugiura and Yoshiharu Ishikawa

Sliding Window Top-K Monitoring over Distributed Data Streams ........... 527
    Zhijin Lv, Ben Chen, and Xiaohui Yu

Diversified Top-k Keyword Query Interpretation on Knowledge Graphs ........... 541
    Ying Wang, Ming Zhong, Yuanyuan Zhu, Xuhui Li, and Tiejyun Qian

Group Preference Queries for Location-Based Social Networks ........... 556
    Yuan Tian, Peiquan Jin, Shouhong Wan, and Lihua Yue

A Formal Product Search Model with Ensembled Proximity ........... 565
    Zepeng Fang, Chen Lin, and Yun Liang

Topic Modeling

Incorporating User Preferences Across Multiple Topics into Collaborative Filtering for Personalized Merchant Recommendation ................................................................. 575
    Yunfeng Chen, Lei Zhang, Xin Li, Yu Zong, Guiquan Liu, and Enhong Chen

Joint Factorizational Topic Models for Cross-City Recommendation ........... 591
    Lin Xiao, Zhang Min, and Zhang Yongfeng
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aligning Gaussian-Topic with Embedding Network for Summarization Ranking</td>
<td>610</td>
</tr>
<tr>
<td>Linjing Wei, Heyan Huang, Yang Gao, Xiaochi Wei, and Chong Feng</td>
<td></td>
</tr>
<tr>
<td>Improving Document Clustering for Short Texts by Long Documents via a Dirichlet Multinomial Allocation Model</td>
<td>626</td>
</tr>
<tr>
<td>Yingying Yan, Ruizhang Huang, Can Ma, Liyang Xu, Zhiyuan Ding, Rui Wang, Ting Huang, and Bowei Liu</td>
<td></td>
</tr>
<tr>
<td>Intensity of Relationship Between Words: Using Word Triangles in Topic Discovery for Short Texts</td>
<td>642</td>
</tr>
<tr>
<td>Ming Xu, Yang Cai, Hesheng Wu, Chongjun Wang, and Ning Li</td>
<td></td>
</tr>
<tr>
<td>Context-Aware Topic Modeling for Content Tracking in Social Media</td>
<td>650</td>
</tr>
<tr>
<td>Jinjing Zhang, Jing Wang, and Li Li</td>
<td></td>
</tr>
<tr>
<td>Author Index</td>
<td>659</td>
</tr>
</tbody>
</table>
Web and Big Data
First International Joint Conference, APWeb-WAIM 2017, Beijing, China, July 7-9, 2017, Proceedings, Part II
Chen, L.; Jensen, C.S.; Shahabi, C.; Bin, W.; Lian, X. (Eds.)
2017, XXI, 362 p. 136 illus., Softcover
ISBN: 978-3-319-63563-7