

# Table of Contents

## Research Track

### Information Retrieval

Faster MaxScore Document Retrieval with Aggressive Processing . . . . .	1
<i>Kun Jiang, Xingshen Song, and Yuexiang Yang</i>	
Finding Novel Patents Based on Patent Association . . . . .	5
<i>Ling Feng, Zhiyong Peng, Bin Liu, and Dunren Che</i>	
Optimizing Top-k Retrieval: Submodularity Analysis and Search Strategies . . . . .	18
<i>Chaofeng Sha, Keqiang Wang, Dell Zhang, Xiaoling Wang, and Aoying Zhou</i>	

### Recommender System

Improving Recommendations with Collaborative Factors . . . . .	30
<i>Penghua Yu, Lanfen Lin, Feng Wang, Jing Wang, and Meng Wang</i>	
Evolutionary Personalized Hashtag Recommendation . . . . .	34
<i>Jianjun Yu and Yi Shen</i>	
PathSimExt: Revisiting PathSim in Heterogeneous Information Networks . . . . .	38
<i>Leong Hou U., Kun Yao, and Hoi Fong Mak</i>	
AdaMF: Adaptive Boosting Matrix Factorization for Recommender System . . . . .	43
<i>Yanghao Wang, Hailong Sun, and Richong Zhang</i>	

### Query Processing and Optimization

A Generic Approach for Bulk Loading Trie-Based Index Structures on External Storage . . . . .	55
<i>Dongzhe Ma and Jianhua Feng</i>	
Density-Based Local Outlier Detection on Uncertain Data . . . . .	67
<i>Keyan Cao, Lingru Shi, Guoren Wang, Donghong Han, and Mei Bai</i>	
Indexing Uncertain Data for Supporting Range Queries . . . . .	72
<i>Rui Zhu, Bin Wang, and Guoren Wang</i>	

A Distributed Quorum System for Ensuring Bounded Staleness of Key-Value Stores ..... 84  
*Hiroshi Horii, Miki Enoki, and Tamiya Onodera*

**Data Mining I**

Multimodal Data Fusion in Text-Image Heterogeneous Graph for Social Media Recommendation ..... 96  
*Yu Xiong, Daling Wang, Yifei Zhang, Shi Feng, and Guoren Wang*

Accuracy Estimation of Link-Based Similarity Measures and Its Application ..... 100  
*Yinglong Zhang, Cuiqing Li, Chengwang Xie, and Hong Chen*

An Efficient Influence Maximization Algorithm to Discover Influential Users in Micro-blog ..... 113  
*Qian Ma and Jun Ma*

**Data and Information Quality**

Detecting Errors in Numeric Attributes ..... 125  
*Grace Fan, Wenfei Fan, and Floris Geerts*

Distributed Entity Resolution Based on Similarity Join for Large-Scale Data Clustering ..... 138  
*Tiezhen Nie, Wang-chien Lee, Derong Shen, Ge Yu, and Yue Kou*

**Data Mining II**

Exploring the Intervention Problem with the Networked Poisson Process in a Real Heterogeneous Social Network ..... 150  
*Wang Yue*

An Infinite Latent Generalized Linear Model ..... 155  
*Jianbo Luo and Jiangtao Ren*

ITCI: An Information Theory Based Classification Algorithm for Incomplete Data ..... 167  
*Yicheng Chen, Jianzhong Li, and Jizhou Luo*

**Information Extraction**

Discovering Informative Contents of Web Pages ..... 180  
*Qifeng Fan, Chunwei Yan, Lifu Huang, and Lian'en Huang*

Chinese Evaluation Phrase Extraction Based on Cascaded Model ..... 192  
*Yashen Wang, Chong Feng, Quanchao Liu, and Heyan Huang*

Tracking Topics on Revision Graphs of Wikipedia Edit History . . . . .	204
<i>Bonan Li, Jianmin Wu, and Mizuho Iwaihara</i>	

## XML, Semi-structured and Uncertain Data

Holistic Subgraph Search over Large Graphs . . . . .	208
<i>Peng Peng, Lei Zou, Dong Wang, and Dongyan Zhao</i>	
Clustering Query Results to Support Keyword Search on Tree Data . . . .	213
<i>Cem Aksoy, Ananya Dass, Dimitri Theodoratos, and Xiaoying Wu</i>	
An Efficient Conditioning Method for Probabilistic Relational Databases . . . . .	225
<i>Hong Zhu, Caicai Zhang, Zhongsheng Cao, Ruiming Tang, and Mengyuan Yang</i>	

## Session 9: Mobile and Pervasive Computing

A Proactive Complex Event Processing Method Based on Parallel Markov Decision Processes . . . . .	237
<i>Yongheng Wang and Kening Cao</i>	
Social-Aware KNN Search in Location-Based Social Networks . . . . .	242
<i>Huiqi Hu, Jianhua Feng, Sitong Liu, and Xuan Zhu</i>	
A Spatial-temporal Topic Segmentation Model for Human Mobile Behavior . . . . .	255
<i>Xingxing Xing, Man Li, Weisong Hu, Wenhao Huang, Guojie Song, and Kunqing Xie</i>	

## Data Mining III

Real-Time Predicting Bursting Hashtags on Twitter . . . . .	268
<i>Shoubin Kong, Qiaozhu Mei, Ling Feng, and Zhe Zhao</i>	
Fast Approximation of Shortest Path on Dynamic Information Networks . . . . .	272
<i>Junting Jin, Xiaowei Shi, Cuiping Li, and Hong Chen</i>	
Detecting of PIU Behaviors Based on Discovered Generators and Emerging Patterns from Computer-Mediated Interaction Events . . . . .	277
<i>Yaxin Yu, Ke Yan, Xinhua Zhu, and Guoren Wang</i>	

## Stream, Time-Series

Characterizing Tweeting Behaviors of Sina Weibo Users via Public Data Streaming . . . . .	294
<i>Kai Zhang, Qian Yu, Kai Lei, and Kuai Xu</i>	

Time Series Classification Using Multi-Channels Deep Convolutional  
Neural Networks ..... 298  
*Yi Zheng, Qi Liu, Enhong Chen, Yong Ge, and J. Leon Zhao*

A Query Approach of Supporting Variable Physical Window in  
Large-Scale Smart Grid ..... 311  
*Yan Wang, Qingxu Deng, Wei Liu, and Baoyan Song*

**Security and Privacy I**

P<sup>3</sup>RN:Personalized Privacy Protection Using Query Semantics over  
Road Networks ..... 323  
*Xiao Pan, Lei Wu, Chunhui Piao, and Xiaoshuo Xu*

A Novel Privacy-Preserving Group Matching Scheme in Social  
Networks ..... 336  
*Jialin Chi, Zhiquan Lv, Min Zhang, Hao Li, Cheng Hong, and  
Dengguo Feng*

Highly Efficient Indexing for Privacy-Preserving Multi-keyword Query  
over Encrypted Cloud Data ..... 348  
*Fanguan Cheng, Qian Wang, Qianwen Zhang, and Zhiyong Peng*

**Semantic Web**

Parallel Mining of OWL 2 EL Ontology from Linked Data ..... 360  
*Huiying Li and Qiang Sima*

TraPath: Fast Regular Path Query Evaluation on Large-Scale RDF  
Graphs ..... 372  
*Xin Wang, Guozheng Rao, Longxiang Jiang, Xuedong Lyu,  
Yajun Yang, and Zhiyong Feng*

**Cloud Computing**

BF-Matrix: A Secondary Index for the Cloud Storage..... 384  
*Xu Cheng, Hongyan Li, Yue Wang, Tengjiao Wang, and  
Dongqing Yang*

A Cluster Based Schema Design for Multi-tenant Database..... 397  
*Jiacai Ni and Jianhua Feng*

Tenant-Oriented Composite Authentication Tree for Data Integrity  
Protection in SaaS ..... 402  
*Li Lin, Li Qingzhong, Kong Lanju, and Shi Yuliang*

Efficient Graph Similarity Join with Scalable Prefix-Filtering Using MapReduce .....	415
<i>Jun Pang, Yu Gu, Jia Xu, Yubin Bao, and Ge Yu</i>	

## New Hardware

SSD-Aware Temporary Data Management Policy for Improving Query Performance .....	419
<i>Zhiliang Guo, Jiangtao Wang, Chunling Wang, and Xiaofeng Meng</i>	
A Novel Index Structure for Multi-key Search .....	431
<i>Dongyu Wei, Xin Pan, Chuan Shi, and Yueguo Chen</i>	
Load-Balanced Breadth-First Search on GPUs .....	435
<i>Zhe Zhu, Jianjun Li, and Guohui Li</i>	

## Crowdsourcing

Leveraging Attributes and Crowdsourcing for Join .....	448
<i>Jianhong Feng, Jianhua Feng, and Huiqi Hu</i>	
Truth Discovery Based on Crowdsourcing .....	453
<i>Chen Ye, Hongzhi Wang, Hong Gao, Jianzhong Li, and Hui Xie</i>	

## Data Mining IV

Sarcasm Detection in Social Media Based on Imbalanced Classification .....	459
<i>Peng Liu, Wei Chen, Gaoyan Ou, Tengjiao Wang, Dongqing Yang, and Kai Lei</i>	
Centroid-Based Classification of Categorical Data .....	472
<i>Lifei Chen and Gongde Guo</i>	
ChronoSAGE: Diversifying Topic Modeling Chronologically .....	476
<i>Tomonari Masada and Atsuhiko Takasu</i>	

## Social Computing

Comments-Oriented Summarization in Blogosphere Using a Two-Stage Sentence Similarity Measure .....	480
<i>Hongjie Li, Lifu Huang, Qifeng Fan, and Lian'en Huang</i>	
Popularity Prediction of Burst Event in Microblogging .....	484
<i>Xiaoming Zhang, Zhoujun Li, Wenhan Chao, and Jiali Xia</i>	
Identifying Gender of Microblog Users Based on Message Mining .....	488
<i>Faliang Huang, Chaoxiong Li, and Li Lin</i>	

**Information Retrieval II**

Continuous Temporal Top-*k* Query over Versioned Documents ..... 494  
*Chao Lan, Yong Zhang, Chunxiao Xing, and Chao Li*

Theme-Aware Social Strength Inference from Spatiotemporal Data ..... 498  
*Ningnan Zhou, Xiao Zhang, and Shan Wang*

Organizing Sightseeing Tweets Based on Content Relatedness and Sharability ..... 510  
*Qiang Ma and Keisuke Hasegawa*

**Information Retrieval III**

Efficient Diverse Rank of Hot-Topics-Discussion on Social Network ..... 522  
*Tao Zhu, Yumin Lin, Ji Cheng, and Xiaoling Wang*

Encoding Document Semantic into Binary Codes Space ..... 535  
*Zheng Yu, Xiang Zhao, and Liping Wang*

LSG: A Unified Multi-dimensional Latent Semantic Graph for Personal Information Retrieval ..... 540  
*Yang Huangfu, Kuien Liu, Wen Zhang, Peng Zhou, Yanjun Wu, Qing Wang, and Jia Zhu*

**Data Mining V**

HRank: A Path Based Ranking Method in Heterogeneous Information Network ..... 553  
*Yitong Li, Chuan Shi, Philip S. Yu, and Qing Chen*

Action-Scene Model for Recognizing Human Actions from Background in Realistic Videos ..... 566  
*Wen Qu, Yifei Zhang, Shi Feng, Daling Wang, and Ge Yu*

Automatically Learning and Specifying Association Relations between Words ..... 578  
*Jun Zhang, Qing Li, Xiangfeng Luo, and Xiao Wei*

**Information Retrieval IV**

A Chinese Question Answering System for Specific Domain ..... 590  
*Tanche Li, Yu Hao, Xiaoyan Zhu, and Xian Zhang*

Top-*k* Spatio-textual Similarity Search ..... 602  
*Sitong Liu, Yaping Chu, Huiqi Hu, Jianhua Feng, and Xuan Zhu*

A Query Suggestion Interface with Features of Queries and Search Results .....	615
<i>Shuhei Shogen, Takehiro Yamamoto, and Katsumi Tanaka</i>	
Finding Photo Sets of Events by Minimizing Misrecognition from Neighbor Events .....	620
<i>Bei Liu, Makoto P. Kato, and Katsumi Tanaka</i>	
<b>Data Mining VI</b>	
Online Community Transition Detection .....	633
<i>Biying Tan, Feida Zhu, Qiang Qu, and Siyuan Liu</i>	
PPML: Penalized Partial Least Squares Discriminant Analysis for Multi-Label Learning .....	645
<i>Zongjie Ma, Huawei Liu, Kaile Su, and Zhonglong Zheng</i>	
Authorship Attribution with Very Few Labeled Data: A Co-training Approach .....	657
<i>Mengdi Fan, Tiejun Qian, Li Chen, Bin Liu, Ming Zhong, and Guoliang He</i>	
<b>Privacy and Security II</b>	
Edges Protection in Multiple Releases of Social Network Data .....	669
<i>Liangwen Yu, Yonggang Wang, Zhengang Wu, Jiawei Zhu, Jianbin Hu, and Zhong Chen</i>	
Trust Prediction with Temporal Dynamics .....	681
<i>Guoyong Cai, Jiliang Tang, and Yiming Wen</i>	
Trust Inference Path Search Combining Community Detection and Ant Colony Optimization .....	687
<i>Yao Ma, Hongwei Lu, Zaobin Gan, and Yizhu Zhao</i>	
<b>Information Retrieval V</b>	
ExNa: An Efficient Search Pattern for Search Engines .....	699
<i>Xiao Wei, Xiangfeng Luo, Qing Li, and Jun Zhang</i>	
A Correlation-Based Semantic Model for Text Search .....	703
<i>Jing Sun, Bin Wang, and Xiaochun Yang</i>	
Finding Relatedness between Research Papers Using Similarity and Dissimilarity Scores .....	707
<i>Qamar Mahmood, Muhammad Abdul Qadir, and Muhammad Tanvir Afzal</i>	

Context-Dependent Sentiment Classification Using Antonym Pairs and Double Expansion . . . . .	711
<i>Zhifei Zhang, Duoqian Miao, and Bo Yuan</i>	

## Data Mining VII

A Novel Knowledge Network Framework for Financial News Navigation . . . . .	723
<i>Lili Zhou, Hanchao Wang, Lei Zhang, Enhong Chen, and Jun Chen</i>	
How Do People Communicate through Different Social Connections? . . .	728
<i>Keke Cai, Yu Zhao, Jie Tang, Li Zhang, and Zhong Su</i>	
From Trajectories to Path Network: An Endpoints-Based GPS Trajectory Partition and Clustering Framework . . . . .	740
<i>Hua Yuan, Yu Qian, Baojun Ma, and Qiang Wei</i>	

## Industrial Track

An Intelligent Search Platform for Business News . . . . .	744
<i>Hanchao Wang, Lili Zhou, Yu Zong, Lei Zhang, Enhong Chen, Xin Li, and Jun Chen</i>	
Structured Sparse Linear Model for Social Trust Prediction . . . . .	756
<i>Deng Yi, Yin Zhang, Yuqi Wang, and Baogang Wei</i>	
Finding Time-Dependent Hot Path from GPS Trajectories . . . . .	768
<i>Yijiao Chen, Kaixi Yang, Hao Hu, Zhangqing Shan, Renchu Song, and Weiwei Sun</i>	
A Two-Phase Model for Retweet Number Prediction . . . . .	781
<i>Gang Liu, Chuan Shi, Qing Chen, Bin Wu, and Jiayin Qi</i>	
Finding Vacant Taxis Using Large Scale GPS Traces . . . . .	793
<i>Zhen Qiu, Hongyan Li, Shenda Hong, Yiyong Lin, Nana Fan, Gaoyan Ou, Tengjiao Wang, and Lilue Fan</i>	
Logo Detection and Recognition Based on Classification . . . . .	805
<i>Yifei Zhang, MingMing Zhu, Daling Wang, and Shi Feng</i>	

## Demonstration Track

Business Intelligence by Connecting Real-Time Indoor Location to Sales Records . . . . .	817
<i>Jessie Danqing Cai</i>	
Search 360 - Question Answering in Business Intelligence . . . . .	824
<i>Falk Brauer, Bob Massarczyk, and Nicolas Kuchmann-Beauger</i>	



HybridPG: A SSD-Friendly High Performance Hybrid Storage Design for PostgreSQL .....	828
<i>Chunling Wang, Jiangtao Wang, Zhiliang Guo, and Xiaofeng Meng</i>	
SmartCare:Location Based e-Care Assistant on Azure .....	832
<i>Ji Cheng, Mengdi Wang, Bo Li, and Xiaoling Wang</i>	
Leveraging Focused Locations for Web Search .....	835
<i>Xiaoxiang Zhang, Peiquan Jin, Sheng Lin, Shouhong Wan, and Lihua Yue</i>	
<b>Author Index</b> .....	<b>839</b>