

# **2016 IEEE International Conference on Big Data Analysis (ICBDA 2016)**

**Hangzhou, China  
12 – 14 March 2016**



**IEEE Catalog Number: CFP16TPE-POD  
ISBN: 978-1-4673-9592-2**

**Copyright © 2016 by the Institute of Electrical and Electronics Engineers, Inc  
All Rights Reserved**

*Copyright and Reprint Permissions:* Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

***\*\*\*This publication is a representation of what appears in the IEEE Digital Libraries. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP16TPE-POD
ISBN (Print-On-Demand):	978-1-4673-9592-2
ISBN (Online):	978-1-4673-9591-5

**Additional Copies of This Publication Are Available From:**

Curran Associates, Inc  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: (845) 758-0400  
Fax: (845) 758-2633  
E-mail: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# Table of Contents

## Data Mining

01-A115	A Behavior Mining Based Hybrid Recommender System <i>Zhiyuan Fang, Lingqi Zhang, Kun Chen</i>	1
02-A222E2	New Combination Algorithms in Commercial Area Data Mining and Clustering <i>Dongqi Fu, Suoju He</i>	6
03-A232E	Research on Audit Log Association Rule Mining Based on Improved Apriori Algorithm <i>Maocai Cheng, Kaiyong Xu, Xuerong Gong</i>	11
04-A236	Research on Counter-terrorism based on Big Data <i>Sanjun Nie, Duoyong Sun</i>	18
05-A255E2	Research of Association Rule Algorithm based on Data Mining <i>Changxin Song</i>	23
06-A262	Sentiment Analysis in A Cross-Media Analysis Framework <i>Yonas Woldemariam</i>	27
07-A280	Evaluation Algorithm for Clustering Quality Based on Information Entropy <i>Liang Xingxing, Xiu Baoxin, Fan Changjun, Chen Chao</i>	32
08-A245	Unsupervised Feature Selection for Text Classification via Word Embedding <i>Weikang Rui, Jinwen Liu, Yawei Jia</i>	37
09-A113	What is your Mother Tongue?: Improving Chinese Native Language Identification by Cleaning Noisy Data and Adopting BM25 <i>Lan Wang, Masahiro Tanaka, Hayato Yamana</i>	42
10-A295	Using Social Media Mining Technology to Assist in Price Prediction of Stock Market <i>Yaojun Wang, Yaoqing Wang</i>	48
11-A296E	Apriori-based Diagnostical Analysis of Passings in the Football Game <i>Liu Tianbiao, Hohmann Andreas</i>	52
12-A261	RHadoop-based Fuzzy Data Mining: Architecture, Design and System Implementation <i>Ping Sun, Lei Xu, Hongfei Fan</i>	56

## Database and Data Modeling

13-A253	Keyword Query Approach over RDF Data Based on Tree Template <i>Qiang Sima, Huiying Li</i>	61
14-A260E	Agricultural Data Modeling Method based on Semantics <i>Lin Bi, Xiao-qiang Di, Ying Zhang</i>	67
15-A254	Finding Database Contention Hotspots under Large-Scale Workloads– A Big Data Approach <i>An Lei, Ping Sun, Weixiong Rao, Hong Min, Gong Su, Zhenhua Dong</i>	72

## Big Data Analysis and Processing

16-A110E	Big Data Analysis Based on POT Method for Design Flood Prediction <i>Dan Li, Shenglian Guo, Jiabo Yin</i>	77
17-A222E3	Original Sustainability Measurement Based on Data Analysis and Processing <i>Dongqi Fu, Wenbo Zhang</i>	82
18-A256	Research and Implementation of Big Data Preprocessing System Based on Hadoop <i>Huadong Dai, Shu Zhang, Li Wang, Yan Ding</i>	90
19-A260	Classification of Customer Requirements on Map Reduce-based Naive Bayes <i>Xinggong Wang, Liang Xue, Buyun Sheng, Zheng Xiao</i>	95
20-A271	Enhanced KStore With the Use of Dictionary and Trie for Retail Business Data <i>Jiandong Li, Runhe Huang</i>	99
21-A287	A Scalable and Productive Workflow-based Cloud Platform for Big Data Analytics <i>Chao Chen, Yuzhong Yan, Lei Huang, Xishuang Dong</i>	104
22-A297	Fuzzy Risk Mode and Effect Analysis Based on Raw Driving Data for Pay-How-You-Drive Vehicle Insurance <i>Wei Nai, Fangqi Zhang, Decun Dong, Yijun Chen, Yi Yu, Wenyi Zheng</i>	109
23-A298	Effective Presenting Method for Different Driving Styles Based on Hexagonal Eye Diagram Applied in Pay-How-You-Drive Vehicle Insurance <i>Wei Nai, Fangqi Zhang, Decun Dong, Yijun Chen, Yi Yu, Wenyi Zheng</i>	114
24-A285	Mean-field Theory of Power Demand Prediction in Small-sized System Via Information on Timetable and Environmental Data <i>Yohei Saika</i>	120
25-A264	Cardiovascular Risk Prediction Method Based on Test Analysis and Data Mining Ensemble System <i>Shan Xu, Haoyue Shi, Xiaohui Duan, Tiangang Zhu, Peihua Wu, Dongyue Liu</i>	126
26-A293	Study of Data Analysis Model Based on Big Data Technology <i>Jinhua Chen, Yuxin Wang, Qin Jiang, Jing Tang</i>	131

## Data Security and Management

27-A114	A Multilevel Deep Learning Method for Big Data Analysis and Emergency Management of Power System <i>X. Z. Wang, J. Zhou, Z. L. Huang, X. L. Bi, Z. Q. Ge, L. Li</i>	137
28-A299	Research of Food Safety Risk Assessment Methods based on Big Data <i>Yongjun Ma, Yangyang Hou, Yushan Liu, Yonghao Xue</i>	142
29-A247	Robust Feature Selection via Simultaneous Capped $\ell_2$ -Norm and $\ell_{2,1}$ -Norm Minimization <i>Gongmin Lan, Chenping Hou, Dongyun Yi</i>	147

30-A250	Forecasting Accuracy Analysis based on Two New Heuristic Methods and Holt-Winters-Method	152
	<i>Dianjun Fang, Yin Zhang, Klaus Spicher</i>	
31-A278	Framework of Integrated Big Data: A Review	158
	<i>Zhikui Chen, Fangming Zhong, Xu Yuan, Yueming Hu</i>	
32-A270	A Preliminary Data Transfer Model for Private Remote Sensing Cloud	163
	<i>Zhenju Li, Xuejun Li, Tao Liu, Sheng Yang, Jianwei Xie</i>	
33-A295E	Research and Implementation of Distributed Storage System Based on Big Data	168
	<i>Ke Ma</i>	

### **Data Visualization and Data Application**

34-A248	Social Relation Extraction with Improved Distant Supervised and Word Embedding Features	172
	<i>Jinwen Liu, Weikang Rui, Liping Zhang, Yawei Jia</i>	
35-A110E2	Research on Analysis System of City Price Based on Big Data	177
	<i>Bi Puyun, LI Miao</i>	
36-A224	Design and Implementation of Data Integrity Check of Subway Log Collection System	181
	<i>Qi Tao, Li Yuan, Lei Youxun</i>	
37-A232	Design and Implementation of Integrating Geographic Location into Web Map Based on Scripting Layer for Android	185
	<i>Yiliang Xing, Huajun Lei</i>	
38-A263	A Tourist Flows Analysis System Based on Phone Big Data	189
	<i>Dandan Lu, Yongde Zhong</i>	
39-A117	2D Approach Measuring Multidimensional Data Pattern in Big Data Visualization	194
	<i>Jinson Zhang, Mao Lin Huang</i>	
40-A252	PCNET: Exploratory Visual Analysis of Large-Scale Network	200
	<i>Jian Huang, Yingmei Wei, Xiaolei Du</i>	
41-A255	OBAVUM: An Ontology-based Approach to Visualizing Urban Mobility Data	206
	<i>Thiago Sobral, Vera Costa, Jos éBorges, Tânia Fontes, Teresa Galvão</i>	

### **Social Networks and Network Engineering**

42-A224E	Key Technology Based on Network Pharmacology of Complex Networks	212
	<i>Xian Zhou, Liping Huang, Wangping Xiong</i>	
43-A240	A Reconstructed Event-based Framework for Analyzing Community Evolution	217
	<i>Jinxu Zhu, Jun Liu, Xuewu Zhang, Yafang Zhao</i>	
44-A234	Theme Evolution and Structure in Twitter: A Case Study of South African Student Protests of 2015	221

*Richard Millham*

- 45-A294 Efficient Spam Detection across Online Social Networks 225  
*Hailu Xu, Weiqing Sun, Ahmad Javaid*

### **Computer Science and Engineering**

- 46-A242 Neural Incremental Attribute Learning Based on Principal Component Analysis 231  
*Ting Wang, Wei Zhou, Xiaoyan Zhu, Fangzhou Liu, Sheng-Uei Guan*
- 47-A268E A Secure Private Cloud Storage System Based on Virtual Isolation Mechanism 236  
*Jiajia Miao, Aihua Bao, Yanjie Niu, Zhizhong Liu, Handong Mao*
- 48-A270E Large-scale Mobile Wireless Sensor Network Data Fusion Algorithm 241  
*Yinggao Yue, Jianqing Li, Hehong Fan, Qin Qin*
- 49-A277 Trust Assurance Privacy Preserving Framework for Moving Objects in Participatory Sensing 246  
*Dorothy Kalui, Xi Guo, Dezheng Zhang, Yonghong Xie, Xiaotong Zhang*
- 50-A244 Providing Caches for Reduce Tasks in a Map Reduce Cloud 252  
*Tzu-Chi Huang, Kuo-Chih Chu, Jhe-Ru Chen, Xue-Yan Zeng, Ce-Kuen Shieh*
- 51-A116 Improving Mobility Prediction Performance with State Based Prediction Method When the User Departs from Routine 256  
*Airui Li, Qiujian Lv, Yuanyuan Qiao, Jie Yang*
- 52-A119 Sample Capacity for Autonomous Underwater Vehicle Range Accreditation 263  
*Xiaojun Guo, Shaojing Su, Zhiping Huang, Yimeng Zhang, Chunwu Liu*
- 53-A237 Visualizing Big Graphs with Labels through Edge Bundling 268  
*Wu-Quan Wang, Zhuang Cai, Kang Zhang*
- 54-A246 Learning Partial Differential Equations for Saliency Detection 273  
*Zhenyu Zhao, Chenping Hou, Yi Wu, Yuanyuan Jiao*
- 55-A257 Path Restoring Algorithm based on Vehicle Trajectory Mining 278  
*Yuanhui Sun, Yaying Zhang*
- 56-A275 Active Learning for Cost-Sensitive Classification Using Logistic Regression Model 284  
*Siyuan Zhou, Ya Zhang*
- 57-A103 A Universal Framework For Flexible Cloud Computing 288  
*Xinyi Zhou, Nianqing Tang, Yin Kuang*
- 58-A221E Recommendation Method for Mobile Network Based on User Characteristics and User Trust Relationship 292  
*Peng Yu*
- 59-A247E Online Human Action Recognition Based on Improved Dynamic Time Warping 298  
*Hongbo Pan, Ji Li*

60-A280E	A Novel Gene Regulatory Network Construction Method based on Singular Value Decomposition <i>An Baiyi, Shen Wei</i>	303
61-A2101	The Design of Three-tier Criteria based Intelligent Essay Rating System <i>Wei Qun, Tu Xiaofan</i>	307
62-A243	Towards Building a Distributed File System for Remote Sensing Image Process <i>Zhenju Li, Xuejun Li, Tao Liu, Jianwei Xie</i>	311
63-A233	Second-trimester Triple Serum Tests in Women Screening for Down's Syndrome: A Meta Analysis <i>Kai Li, Lei Ma, Yan Xiang, Xiaodong Han, Long Li, Baosheng Zhu</i>	315
64-A286	An Efficient End-to-End Mutual Authentication Scheme for 2G-GSM System <i>Mohammed Ramadan, Fagen Li, Chun Xiang Xu, Ahmed Abdalla, Hisham Abdalla</i>	319

### **Economics and Management Education**

65-A118	Research and Application of One-Key Publishing Technologies for Meteorological Service Products <i>Chao Li, YinTian Liu, Rui Li, Haiqing Zhang</i>	325
66-A104E	Stock Portfolio Analysis based on Margin of Safety and Competitive Edge Evaluation in the American Wine and Liquor Industry <i>Qinli Fang, Yizeng Chen</i>	330
67-A2102	Research on Tobacco Supply Chain System <i>Qun Wei, Xiaofan Tu</i>	337

<b>Author Index</b>		341
---------------------	--	-----