

2019 IEEE 4th International Conference on Big Data Analytics (ICBDA 2019)

**Suzhou, China
15 – 18 March 2019**



**IEEE Catalog Number: CFP19TPE-POD
ISBN: 978-1-7281-1283-1**

**Copyright © 2019 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 is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP19TPE-POD
ISBN (Print-On-Demand):	978-1-7281-1283-1
ISBN (Online):	978-1-7281-1282-4

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
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2019 the 4th IEEE International Conference on Big Data Analysis

(ICBDA 2019)

Table of Contents

Preface	ix
Committees	x
 Clustering Algorithm Design and Analysis	
A Clustering Algorithm Based on Minimum Spanning Tree and Density	1
<i>Jiayao Chen, Jing Lu</i>	
Movement Pattern Extraction Based on a Non-parameter Sub-trajectory Clustering Algorithm	5
<i>He Ailin, Liu Zhong, Zhou Dechao</i>	
Vessel Sailing Patterns Analysis from S-AIS Data Dased on K-means Clustering Algorithm	10
<i>Zhong Hanyang, Song Xin, Yan Zhenguo</i>	
Research on Differentially Private Bayesian Classification Algorithm for Data Streams	14
<i>Guozheng Zhang, Shuyu Li</i>	
Bayesian Approach to the Classification of BMI Time Series Data from Babyhood to Junior High School Age of Japanese Children	21
<i>Toshiaki Aida, Chiyori Haga</i>	
A Traffic Speed Imputation Method Based on Self-adaption and Clustering	26
<i>Xiayu Qiu, Yaying Zhang</i>	
 Data Analysis and Calculation	
Censorious Young: Knowledge Discovery from High-throughput Movie Rating Data with LME4	32
<i>Zhiyi Chen, Shengxin Zhuy, Qiang Niuy, Xin Lu</i>	
An Original-Destination Matrices Calculation Method for Public Traffic and Rail Transit Data.....	37
<i>BaoKui Cao, XueMing Li, Wei Deng, ShanLong Jiang, Shuai Zhao</i>	
Diminished-1 Fermat Number Transform for Integer Convolutional Neural Networks.....	47
<i>Zhu Baozhou, Nauman Ahmed, Johan Peltenburg, Koen Bertels and Zaid Al-Ars</i>	
An Approach for Dynamic Scheduling of Data Analysis Algorithms.....	53
<i>Jun Gui, Zeyu Zheng, Yuan Gao, Zhaobo Qin</i>	
Generating Classification Rules for Large DataSets.....	58
<i>Ching-Sheng Tu, Deng-Yi Wu, Shie-Jue Lee, Chih-Hung Wu</i>	

Detecting Pickpocketing Offenders by Analyzing Beijing Metro Subway Data	62
<i>Haishuo Gu, Peng Chen, Yaqi Guo, Minggang Yao, Huigen Yang, Jiaqi Hou</i>	
Maritime Traffic Data Visualization: A Brief Review	67
<i>Kai Wang, Maohan Liang, Yan Li, Jingxian Liu, Ryan Wen Liu</i>	
Cross-Linguistic Twitter Analysis of Discussion Themes before, during and after Ramadan.....	73
<i>Aamna AlShehhi, Justin Thomas, Roy Welsch, Zeyar Aung</i>	
A New Method with POI Data Applied in the Determination of Urban Functional Agglomeration Regions	79
<i>Zhu Lifang, Zhou Li, Liu Yang, Chen Yixin</i>	
Data Mining and Data Management	
Synthetic Comparison and Management for Data with Multi Storage Types	83
<i>Li Yan, Cheng Long, Xie Weiqi</i>	
Value Evaluation of Data Assets: Progress and Enlightenment.....	88
<i>Zifeng Li, Yuan Ni, Xia Gao, Gongshan Cai</i>	
Query Optimization of Massive Social Network Data Based on HBase	94
<i>Congkai Bao, Meiyang Cao</i>	
Privacy-preserving Recommendation for Location-based Services	98
<i>Qiuyi Lyu, Yu Ishimaki, Hayato Yamana</i>	
Spatio-Temporal Vessel Trajectory Smoothing Using Empirical Mode Decomposition and Wavelet Transform.....	106
<i>Xinyi Li, Zikun Feng, Yan Li, Zhao Liu, Ryan Wen Liu</i>	
Discovery and Research of Communication Relation Based on Communication Rules of Ultrashort Wave Radio Station	112
<i>Changkun Liu, Xinrong Wu, Changhua Yao, Lei Zhu, Yindong Zhou, Han Zhang</i>	
Predicting Repeat Offenders with Machine Learning: A Case Study of Beijing Theives and Burglars	118
<i>Peng Chen, Justin Kurland, Shaochong Shi</i>	
Process and Application of Data Mining in the University Library	123
<i>Yanping Cheng, Qingyu Liu</i>	
Real-Time Vessel Trajectory Data-Based Collision Risk Assessment in Crowded Inland Waterways .	128
<i>Zikun Feng, Haojie Yang, Xinyi Li, Yan Li, Zhao Liu, and Ryan Wen Liu</i>	
Research on the Process Data Mining of Online Wisdom Learning.....	135
<i>Jie Jiang, Lei Zeng, Lin Zhi</i>	
Big Data Analysis and Key Technologies	
Big Data and Safety Management Methods: The Reduction Model of Hot Work Number	140
<i>Guozhi Zhang, Yunlong Wang, Bo Mu, Tingchun Wang</i>	
Design and Development of Meteorological Popular Science Games based on Somatosensory Interaction Technology in the Background of Big Data.....	144
<i>Xiaoqiang Hu, Zhiying Zhu, Jianzeng Chen</i>	

Research and Application of Big Data Analysis Platform for Oil Production Engineering in Huabei Oilfield	148
<i>Wang Yuanling, Zhang Zhanmin, Du Lihong, Xie Weiyi, Min Guizhi, Zhang Dengwen, Li Yun, Yang Bo, Wu Jiaheng, Yu Limin</i>	
SparkJNI: A Toolchain for Hardware Accelerated Big Data Apache Spark	152
<i>Tudor Alexandru Voicu and Zaid Al-Ars</i>	
Design and Development of IoT Plugin for HPCC Systems	158
<i>K.S.Amogh Vardhan, Manjunath Jakaraddi, Shobha G, Jyoti Shetty, Arjuna Chala, Dan Camper</i>	
Research on the Multi-objective Optimization of China's Financial Industrial Structure Based on the Evolutionary Algorithm	163
<i>Yang Gao, Yunhai Li, Weiyi Pan, Yaojun Wang</i>	
The Hindrance of Administrative Barriers to Integrated Investment Markets: Take All Listed Companies as an Example between Sichuan Province and Chongqing Municipality	168
<i>Cheng Qian-chang</i>	
Big Data Science and Applications	
Modeling for Optimal Investment Strategy of Educational Donation	172
<i>Jinxiao Wang, Zhiquan Shao</i>	
Analysis on the Development Path of Smart City in the Era of Big Data	177
<i>Sen Xiong, Bei Ye</i>	
Detection on Abnormal Usage of Spectrum by Electromagnetic Data Mining	182
<i>Xu'nan Liu, Rong Shi, Binbin He, Min Chen</i>	
Study on Pretreatment Methods of Stellar Spectral Data	188
<i>Bin Jiang, Bingrui Chen, Ziliang Zhao</i>	
An Automated Big Data Accuracy Assessment Tool	193
<i>Goutam Mylavarapu, Johnson P Thomas, K Ashwin Viswanathan</i>	
Deep Product Quantization for Large-Scale Image Retrieval	198
<i>Qi Zhai, Mingyan Jiang</i>	
Big Data Analytics for Higher Education in the Cloud Era	203
<i>Ali Al Hadwer, Dan Gillis, Davar Rezaia</i>	
Research on Art Education Digital Platform Based on Big Data	208
<i>Zhen Xiong, Lin Zhi, Jie Jiang</i>	
Study on the Influence of Organizational Innovation on Organizational Performance under the Circumstances of Big Data—Intermediary Role of Dynamic Capability	212
<i>Li Xiaoran, Shi Chunsheng, Wang Chenggang</i>	
System Modeling and Optimization	
A Path Planning Model based on Spatio-Temporal State Vector from Vehicles Trajectories	216
<i>Rui Yao, Zhiming Ding, Yang Cao, Fujie Ren</i>	
The Principal Component Linear Regression Prediction Model of the Impact of Enterprise Patent Activities on the Income of Main Business: A Case Study of Tianjin, China	221
<i>Tian Jun, Wang Juan</i>	

Recommendation of Item Category Based on Random Walk Model	226
<i>Yilong Shi, Li Jin, Jianfeng Chen, Ke Zeng</i>	
Using Feature Visualisation for Explaining Deep Learning Models in Visual Speech Recognition	231
<i>Timothy Israel Santos, Andrew Abel</i>	
Activity Modelling Using Journey Pairing of Taxi Trajectory Data	236
<i>Shuhui Gong, John Cartlidge, Ruibin Bai, Yang Yue, Qingquan Li, Guoping Qiu</i>	
Propagation Model of Derivative Rumor Considering Propagation Error and Malicious Tampering....	241
<i>Zhenyu Zhao, Xi Chen</i>	
Modeling and Detecting Drift in User Interest Based on Hierarchical Classification	246
<i>Hui Chen</i>	
Information Network and Data Security	
Use of Machine Learning in Detecting Network Security of Edge Computing System	252
<i>Size Hou, Xin Huang</i>	
Identification of Network-wide Super Nodes in High Speed Networks	257
<i>Aiping Zhou, Qi Zhao, Jin Qian, Chengang Zhu</i>	
From the Goldbach Conjecture Cipher to the Han Ge CODE	263
<i>Li Junchi</i>	
A Hybrid Data Security System of Internet of Things.....	269
<i>Haowei Song, Hai-Ning Liang, Lu Liu, Jieming Ma, Xin Huang</i>	
A Personalized Differential Privacy Protection Method for Repeated Queries	274
<i>Senyou Li, Xinsheng Ji, Wei You</i>	
A Transfer Learning Approach for Network Intrusion Detection	281
<i>Peilun Wu, Hui Guo, Richard Buckland</i>	
An Efficient ID-based Fair Off-line Electronic Cash Scheme without Bilinear Pairings.....	286
<i>Wei Cui, Xuepeng Song, Qian Jia</i>	
Research on Pyramid Schemes Communication Network Based on Exponential Random Graph Models	291
<i>Pihu Feng, Duoyong Sun, Zhan Feng Li</i>	
Image Processing and Application	
Extraction of Chromatographic Characteristics Based on Wavelet Ridge and Morphology	297
<i>Xiaoming Lin, Tianzhi Liang, Liang Cao</i>	
Cascaded CNN for Real-time Tongue Segmentation Based on Key Points Localization.....	303
<i>Wei Yuan, Changsong Liu</i>	
Noise-Contrastive Estimation Based on Relative Neighbour Sampling for Unsupervised Image Embedding Learning.....	308
<i>Haoyu Lv, Jiangzhong Cao, Qingyun Dai, Guanshu Liang, Qi Zhao</i>	

E-commerce Application and Financial Market Analysis

Identifying Malicious Dealers in Goods and Services Tax.....	312
<i>Priya Mehta, Jithin Mathews, S.V. Kasi Visweswara Rao, K. Sandeep Kumar, K.Suryamukhi, Ch. Sobhan Babu</i>	
A Selection of Advanced Technologies for Demand Forecasting in the Retail Industry	317
<i>Jiaxing Wang, G. Q. Liu, Lu Liu</i>	
Why Popular? A Quantitative Analysis of Catering Consumption within Mega City	321
<i>Yu Wang, Salan Jiao, Liancheng Zhang, Yan Cao</i>	
Candlestick Analysis in Forecasting U.S. Stock Market: Are They Informative and Effective	325
<i>Haoxuan Qiu, Fanzhuoqun Liu</i>	
Reliability Evaluation of Power-Gas Coupled System with Direct Load Control	329
<i>Peng Liu, Hui Hou, Xixiu Wu, Jinrui Tang, Guorong Zhu, Xianqiang Li, Aihong Tang</i>	
Predicting Financial Markets Using the Wisdom of Crowds	334
<i>Anasse Bari, Pantea Peidaee, Aniruddh Khera, Jianghao Zhu, Hongting Chen</i>	
A Reliable E-commerce Business Model Using Blockchain Based Product Grading System.....	341
<i>Ching-Nung Yang, Yi-Cheng Chen, Shih-Yu Chen, Song-Yu Wu</i>	
Review and Prospect of Data Asset Research: Based on Social Network Analysis Method	345
<i>Zifeng Li, Yuan Ni, Lu Yang, Yudong Gao</i>	

Computer Science and Engineering Applications

Efficient Learning Strategy of Chinese Characters Based on Usage Frequency and Structural Relationship	351
<i>Kun Zhao, Deshun zhao, Jun Yang, Wei Ma, Bo Yu</i>	
The Impact of Sentiment Analysis on Social Media to Assess Customer Satisfaction: Case of Rwanda	356
<i>Marius Ngaboyamahina, Sun Yi</i>	
Finding Composition Skyline Based on Standard Deviation.....	360
<i>Xiaowei Cui, Leigang Dong</i>	
Named Entity Recognition of Chinese Electronic Medical Records Based on Cascaded Conditional Random Field.....	364
<i>Xiaoyu Chen, Shenghui Shi, Siyan Zhan, Daguang Jiang, Xiaoyong Lin</i>	
Privacy Module for Distributed Electronic Health Records(EHRs) Using the Blockchain	369
<i>Richard Nuetey Nortey, Li Yue, Promise Ricardo Agdedanu, Michael Adjeisah</i>	
Crowdsourcing and Its Application for Traffic Survey Work.....	375
<i>Takahiro Koita, Shota Suzuki</i>	
Bayesian Inference Using Expected A Posterior Estimation for Prediction of Thermal Index and Power Consumption of Air Conditioner Based on Mean-field Theory.....	379
<i>Yohei Saika, Masahiro Nakagawa</i>	
Research on Multi-Dimensional Analysis Method of Power Equipment Condition Monitoring Based on OLAP	384
<i>Li Xi, Wang Hongkai, Li Jinhua, Pei Xubin, Yu Zhanpeng</i>	

A New Method Applied in the Regional Evaluation of Urban Comprehensive Carrying Capacity.....	389
<i>Zhu Lifang, Liu Yang, Zhou Li1, Wang Zhenqin</i>	
Implementation of Elementary Chinese Language Learning Application in WeChat Mini Programs...	394
<i>Xijie Wang, Minyong Shi, Chunfang Li</i>	
Research on Assisted Generation Method of Emergency Management Structured Case for Natural Text	399
<i>Kai Wang, Xi Chen</i>	
Research on Interaction Tracking between Community Discovery and theme Evolution Based on DBLP Scientific Research Cooperation Network.....	403
<i>Jianyun Ye, Guanghui Yan, Zongren Li, Yuxuan Liu</i>	
Neural Network and Recommendation System	
Convolution Recurrent Neural Networks Based Dynamic Transboundary Air Pollution Predictiona....	410
<i>Peijiang Zhao, Koji Zettsu</i>	
GKB: A Predictive Analytics Framework to Generate Online Product Recommendations	414
<i>Nicholas Greenquist, Doruk Kilitcioglu, Anasse Bari</i>	
Appearance Frequency-Based Ranking Method for Improving Recommendation Diversity	420
<i>Seiki Miyamoto, Takumi Zamami, Hayato Yamana</i>	
Neural Network-Based Automatic Reconstruction of Missing Vessel Trajectory Data	426
<i>Maohan Liang, Ryan Wen Liuy, Qianru Zhong, Jingxian Liu, Jinfeng Zhang</i>	
An Improved Collaborative Filtering Recommendation Algorithm	431
<i>Wang Hong-xia</i>	
Author Index.....	436