2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA 2018)

Shanghai, China
9-12 March 2018
Table of Contents

Preface ........................................................................................................................................................................ viii
Committees ..................................................................................................................................................................... ix

**Big Data Analysis and Methods**

The Effect Analysis of Critical Subgroup Based on K-core on Public Outbreak as Big Data .......................... 1
Yu Gong, Mingchun Zheng

College Data Analysis Based on Multi-learning Method ......................................................................................... 5
Jia Shuai, Shaocong Mo, Yanyao Yang, Xu Han, Zhongyuan Chen, Zefeng Xie, Chao Dai

Leveraging Big Data Analytics to Improve Decision Making in South African public Universities ............... 9
Mmatshuene Anna Segooa, Billy Mathias Kalema

Big Data Analytics Using Association Rules in eLearning .................................................................................. 14
Korn Poonsirivong, Chanintorn Jittawiriaynukoon

The Research on Industrial Big Data Information Security Risks ................................................................. 19
Feng Xiaorong, Jia Shizhun, Mai Songtao

Study on the Application of Big Data in Accurate Marketing of Cross-border Ecommerce in China ............ 24
Huiqun Liu, Xiao Wang

An Effective Selecting Approach for Social Media Big Data Analysis –Taking Commercial Hotspot Exploration with Weibo Check-in Data as An Example ......................................................... 28
Qingwu Hu, Yuan Zhang

The Logistics Performance Evaluation Index System in the Transportation Industry Based on Big Data ...... 33
Ran Ding, Peng Ren

Centrality Analysis of Online Social Network Big Data .................................................................................. 38
Hui Li

Real-Time Big Data Warehousing and Analysis Framework .............................................................................. 43
Abbas Raza Ali

**Data Mining**

Mining Spatio-temporal Patterns in Multivariate Spatial Time Series .......................................................... 50
Gene P. K. Wu, Keith C. C. Chan

Association Rules Mining Analysis of App Usage Based on Mobile Traffic Flow Data ........................... 55
Yang Liu, Ke Yu, Xiaofei Wu, Yazhou Shi, Yaowen Tan

Analysis and Prediction of Influenza in the UAE based on Arabic Tweets .................................................. 61
Balsam Alkouz, Zaher Al Aghbari

An Algorithmic Approach to Handle Circular Trading in Commercial Taxation System ....................... 67
Jithin Mathews, Priya Mehta, Ch. Sobhan Babu, S.V. Kasi Visweswara Rao
Research on Kano Model Based on Online Comment Data Mining.................................................................76
Huaming Song, Chao Chen, Qin Yu

A Brief Analysis of the Key Technologies and Applications of Educational Data Mining on Online Learning Platform ......................................................................................................................83
Wei Zhang, Shiming Qin

Data Theory and Engineering

Research and Implementation of Data Fusion Method Based on RDF ........................................................87
Bin Yu, Chen Zhang, Yitong Li, Jiangyan Sun

Research hotspots of "Course of Action" Based on Bibliometrics.................................................................92
Chao Chen, Changjun Fan, Zili Li, Li Zeng

The Simulation of Open One-side Uncertain Probability for Fusion Model of Data Uncertainty and Data Relation Uncertainty................................................................................................................97
Ruqi Ye, Mingxue Liao, Tianyu Cui, Pin Lv

Generalizing Application Agnostic Remaining Useful Life Estimation Using Data-Driven Open Source Algorithms ..........................................................................................................................102
Bernhard Schlegel, Artur Mrowca, Peter Wolf, Bernhard Sick, Sebastian Steinhorst

Spatio-temporal Prediction of Shopping Behaviours Using Taxi Trajectory Data .........................................112
John Cartlidge, Shuhui Gong, Ruibin Bai, Yang Yue, Qingquan Li, and Guoping Qiu

Automated Prediction of Shopping Behaviours Using Taxi Trajectory Data and Social Media Reviews........117
Shuhui Gong, John Cartlidge, Ruibin Bai, Yang Yue, Qingquan Li and Guoping Qiu

Reliability Enhanced Data Processing System Capable of Runtime Fault Recovery ..................................122
Beichen Shen, Libiao Peng, Yongle Xie

Evaluate Academic Influence of Research Teams in Scholarly Data .............................................................128
Kaiyu Liu, Hong Qiao, Mingchun Zheng

Learning Subconcepts in User-Generated Text from Social Media Sites: A Case of Exploring Food Concept...133
Jie Wang

Effective NBA Player Signing Strategies Based on Salary Cap and Statistics Analysis ...............................138
Ramya Nagarajan, Yihang Zhao, Lin Li

Dimensional Modeling of Medical Data Warehouse Based on Ontology ..................................................144
Shuxia Ren, Teng Wang, Xu Lu

Application of Ontology in Medical Heterogeneous Data Integration .........................................................150
Shuxia Ren, Xu Lu, Teng Wang

MES: A Memory-Efficient Key-Value Storage With User-Level Network Stack ........................................156
Ninggou Yang, Junpeng Liu, Peng Zhang, Chao Zheng, Qingyun Liu

Big Data Processing and Application

A Parallel Algorithm for Matrix Fast Exponentiation Based on MPI ..........................................................162
Ji Zhang, Xu-chuan Zhou

Building the Federation of Cloud Service for Big Data ..............................................................................166
Jian Shu, Changyong Liang, Binyou Wang, Jian Xu

Big Data and Foreign Language E-learning on Production-oriented Approach .............................................170
Wu Yanhui
A Noise Classification Algorithm Based on SAMME and BP Neural Network .......................................................... 274
  Chen Guoqiang, Zhang Yuying, Tang Deyou

A Message Authentication Scheme for VANETs Based on Trapdoor Hash Function ................................................. 279
  Yang Xiaodong, An Faying, Yang Ping, Xiao Likun, Li Yutong, Ma Tingchun, Wang Caifen

Research and Simulation Implementation of Mobile Phone Communication Based on SL4A ........................................ 283
  Huaqun Lei

Cloud Computing and Recommendation System
Research and Implementation of POI Recommendation System Integrating Temporal Feature .................. 287
  Jinxin Liu, Xu Jiao, Youzhi Jin, Xinyu Liu, Li Liu

Research on Artist Intelligent Recommendation Method Based on Mapping Knowledge Domains-Data Mining
Based on Artist Association Analysis .......................................................................................................................... 293
  Ge Su, Ke Ma, Fulian Yin

Design and Implementation of Scientific Research Information Recommendation System ............................. 297
  Ziejia Pan, Ying Jiang, Tong Wang, Yudan Lin, Hongjun Li, Jiting Xu, Xin Zhao

Access Control of Cloud Users Credible Behavior Based on IPv6 ........................................................................ 303
  Sisi Cao, Wenbo Zheng, Shaocong Mo, Xin Jinf, Chao Daik, Zeyu Fan, Yili Qu, Jiangwei Zhou, Fei Deng

Performance Evaluation of Memory Usage Costs for Commercial Cloud Services ................................................... 307
  Takahiro Koita, Yu Manabe

Attribute-based Proxy Re-encryption Method for Revocation in Cloud Storage: Reduction of Communication
Cost at Re-encryption .................................................................................................................................................. 312
  Yoshiko Yasumura, Hiroki Imabayashi, Hayato Yamana

Verification Method of Data Quality in Science and Technology Cloud in Shaanxi Province ........................... 319
  Bin Yu, Chen Zhang, Zhou Hua Tang, JiangYan Sun

Computer and Information Engineering
App Usage Prediction Based on an Extended Feature Selection Method RSPEC_MIC ........................................... 324
  Yazhou Shi, Fang Wei, Ke Yu, Xiaofei Wu

The Evolution of Network Topology Structure of Chinese Stock Market ............................................................... 329
  Xuewei Cao, Yongbin Shi, Penghao Wang, Liujun Chen, Yougui Wang

Improved Approaches for Financial Market Forecasting Based on Stationary Time Series Analysis .......... 334
  Hongbo Sun, Jing Xu

A Topic-based Cross-language Retrieval Model with PLSA and TF-IDF ......................................................... 340
  Zhuo-Liang Huo, Jun-Feng Wu, Yao Lu, Chang-Zhou Li

Mixed Feature Expansion-based Short Text Classification ....................................................................................... 345
  Zhenrong Deng, Baojun Zhang, Jingjing Zhang

A Fast Registration from 3D CT Images to 2D X-ray Images ............................................................................... 351
  Hanqian Pan, Chunlin Zhou, Qi Zhu, and Delei Zheng

Improved Deep Belief Network Model and Its Application in Named Entity Recognition of Chinese Electronic
Medical Records ......................................................................................................................................................... 356
  Wusuo Li, Shenghui Shi, Ziqiao Gao, Wei Wei, Qunxiang Zhu, Xiaoyong Lin, Daguang Jiang, Shang Gao

The Effect of the Family Planning Policy on Population Loss in China’s Provinces .............................................. 361
  Wang Xiang, Hu Die, Wang Yuandi, Hu Ruifeng

Enhancing Privacy Using Location Semantics in Location Based Services ......................................................... 368
  Mingjie Ma
A Multi-tenant Play Script Management System for SAAS .................................................................374
  XueKe Liu, MinYong Shi, ChunFang Li

Detection of Dependence Between Two Groups of Variables in the Presence of Another Group ..........379
  Ah Hin Pooi

Analyzing Collusive and Opportunistic Voting Strategies in an Academic Department P & T Decision Using Low-Dimensional Spectral Embedding ....................................................................................383
  Sanjoy Das

Image Classification of Fine-grained Fashion Image Based on Style Using Pre-Trained Convolutional Neural Network ................................................................................................................................................ 387
  Yian Seo, Kyung-shik Shin

Power Electronics Technology and Reliability Prediction

PEMFC Power Prediction Based on Deep Auto-encoder and LS-SVMR ..................................................391
  Sheng Hong, Longlong Sun, JiaWei Yin, Ting Yu, Yifan Wang, Wei Zhu

An Improved Parallel Maximum Flow Approach for Vulnerability Analysis of Power System ........397
  Yanlin Shi, Dafang Zhang, Zulong Diao

Early Anomaly Detection in Wind Turbine Bolts Breaking Problem – Methodology and Application ....402
  Chung-Wei Wu, Mei Chen

Short Term Load Forecasting Using Wavelet Augmented Non-linear Autoregressive Neural Networks-A Single Customer Level Perspective ...................................................................................... 407
  Solomon Netsanet, Jianhua Zhang, Dehua Zheng

Remaining Useful Life Prediction of Electric Vehicle Lithium-Ion Battery Based on Particle Filter Method...412
  Zachary Bosire Omariba, Lijun Zhang, Dongbai Sun

Data Clustering and Algorithm

Study on Initial Sensitivity of Fuzzy Clustering Algorithm for Big Data Processing ................................417
  Enwen Li, Linong Wang, Bin Song, Ruyi Cai

K-Means Clustering Algorithm Based on Improved Cuckoo Search Algorithm and Its Application ....422
  Shuce Ye, Xiaoli Huang, Yinyin Teng, Yuxia Li

Clustering of Public Transport Operation using K-Means .....................................................................427
  Rabiah Abdul Kadir, Yasuki Shima, Riza Sulaiman, Fathelalem Ali

SparkSNN: A Density-based Clustering Algorithm on Spark ...................................................................433
  Aman Mani Aryal, SuJing Wang

State Grid Office System User Clustering Analysis Based on K-Means Algorithm .................................438
  Ting Li, Peng Ye, Shikai Zheng

Protocol Reverse Based on Hierarchical Clustering and Probability Alignment from Network Traces ....443
  FanZhi Meng, Chunrui Zhang, Guo Wu

PCA-Based Hierarchical Clustering of AIS Trajectories with Automatic Extraction of Clusters ........448
  Jing Cao, Maohan Liang, Yan Li, Jinwei Chen, Huanhuan Li, Ryan Wen Liu, and Jingxian Liu

Clustering Driving Trip Trajectory Data Based on Pattern Discovery Techniques .................................453
  Gene F. K. Wu, Keith C. C. Chan

Overlapping community detection algorithm based on Markov chain clustering ....................................458
  GuoLiang Deng, ZhiXin Ma, XianBo Li

Author Index ..............................................................................................................................................463