# Table of Contents

Message from Big Data Service 2017 General Chairs ................................................................. x
Message from Big Data Service 2017 Program Chairs ................................................................. xi
Big Data Service 2017 Organizing Committee ............................................................................. xii
Big Data Service 2017 Program Committee ................................................................................. xiii
Smart City Big Data Workshop Chairs and Committee ................................................................. xv
Big Data Security and Service Workshop Chairs and Committee .................................................... xvi
Big Data Quality Assurance Workshop Chairs and Committee ...................................................... xvii
Keynotes ........................................................................................................................................ xviii

---

**Session 1: Intelligent Data Mining**

Enriched Over_Sampling Techniques for Improving Classification of Imbalanced Big Data ................................................................. 1  
  *Sachin Subhash Patil and Shefali Pratap Sonavane*

Improved Sentiment Classification by Multi-Modal Fusion ................................................................. 11  
  *Lige Gan, Rachid Benlamri, and Richard Khoury*

Towards Automatic Linkage of Knowledge Worker’s Claims with Associated Evidence from Screenshots ......................................................................................... 17  
  *Paul Jones, Dakota Medd, Sreekanth Ramakrishnan, Rajat Shah, Joann Keyton, and Nagiza Samatova*

**Session 2: Frameworks for Big Data Processing**

CaPaR: A Career Path Recommendation Framework ........................................................................... 23  
  *Bharat Patel, Varun Kakuste, and Magdalini Eirinaki*

A Goal-Oriented Big Data Analytics Framework for Aligning with Business ........................................... 31  
  *Grace Park, Lawrence Chung, Liping Zhao, and Sam Supakkul*
When Rule Engine Meets Big Data: Design and Implementation of a Distributed Rule Engine Using Spark .................................................................41
  Jindou Zhang, Jinxing Yang, and Jing Li

Session 3: Big Data Processing and Mining
Balanced Parallel Frequent Pattern Mining over Massive Data Stream .................................................................50
  Xi Fu, Lei Shi, and Jing Li
Data Allocation of Large-Scale Key-Value Store System Using Kinetic Drives ..........................................................60
  Xiang Cao, Manas Minglani, and David Hung-Chang Du
Scaling Collaborative Filtering to Large-Scale Bipartite Rating Graphs Using Lenskit and Spark .................................................................70
  Christos Sardianos, Iraklis Varlamis, and Magdalini Eirinaki
Small Boxes Big Data: A Deep Learning Approach to Optimize Variable Sized Bin Packing .................................................................80
  Feng Mao, Edgar Blanco, Mingang Fu, Rohit Jain, Anurag Gupta, Sebastien Mancel, Rong Yuan, Stephen Guo, Sai Kumar, and Yayang Tian

Session 4: Big Data Analysis
Impact of Core Charging Station's Cease Operation in the Entire Charging Station System: A Case Study in Shenzhen .................................................................90
  Zhiyong Tian, Lai Tu, Yi Wang, Fan Zhang, and Chen Tian
MD-Miner: Behavior-Based Tracking of Network Traffic for Malware-Control Domain Detection .................................................................96
  Jia-Hao Sun, Tzung-Han Jeng, Chien-Chih Chen, Hsiu-Chuan Huang, and Kuo-Sen Chou

Smart City Big Data Workshop

Session 1: Smart City Big Data
Viral Marketing for Smart Cities: Influencers in Social Network Communities .................................................................106
  Madhura Kaple, Ketki Kulkarni, and Katerina Potika
A Deep Learning Framework for Smart Street Cleaning .........................................................................................112
  Chandni Balchandani, Rakshit Koravadi Hatwar, Parteek Makkar, Yanki Shah, Pooja Yelure, and Magdalini Eirinaki
Smart Dairies — Enablement of Smart City at Gross Root Level .............................................................................118
  Jaya Shankar Vuppalapati, Santosh Kedari, Ananth Ilapakurthy, Anitha Ilapakurti, and Chandrasekar Vuppalapati
iDispenser — Big Data Enabled Intelligent Dispenser .........................................................................................124
  Anitha Ilapakurti, Jaya Shankar Vuppalapati, Santosh Kedari, Sharat Kedari, Chitanshu Chauhan, and Chandrasekar Vuppalapati
### Session 5: Data Analytics and Visualization

- Engineering Scalable Distributed Services for Real-Time Big Data Analytics ................................................................. 131
  *Sahar Jambi and Kenneth M. Anderson*
- Answering How-to-Reach Query in Big Attributed Graph Databases ...................................................................................... 141
  *Duncan Yung and Shi-Kuo Chang*
- On Building a Big Data Analysis System for California Drought .......................................................................................... 149
  *Pengcheng Zhang, Jerry Gao, A. G. Thomas, K. P. Alagupackiam, K. Mannava, P. I. Bosco, and Sen Chiao*
- Rapidly Generate and Visualize the Digest of Massive Time Series Data .................................................................................. 157
  *Zhiwei Gao, Tiangang Zou, Lulu Zhang, Juan Chen, Jing Guo, and Weixing Ji*
- Visualization Techniques for Electrical Grid Smart Metering Data: A Survey ............................................................................ 165
  *Maria Stefan, Jose G. Lopez, Morten H. Andreasen, and Rasmus L. Olsen*

### Session 6: Big Data Application

- Data Reduction as a Service in Smart City Architecture ........................................................................................................ 172
  *Aseel Alkhelaiwi and Dan Grigoras*
- Wind Farm Operation and Maintenance Optimization Using Big Data .................................................................................... 179
  *Jan Helsen, Cédric Peeters, Peter Doro, Eveline Ververs, and Pieter Jan Jordaens*
- Automated Medical Diagnosis from Clinical Data ...................................................................................................................... 185
  *Vishnu S. Pendyala and Silvia Figueira*

### Session 7: Big Data Framework, Technology, and Solutions

- A Distributed Multi-exemplar Affinity Propagation Clustering Algorithm Based on MapReduce .................................................. 191
  *Yu-Bo Yang, Chang-Dong Wang, and Jian-Huang Lai*
- CENSE: A Cognitive Navigation System for People with Special Needs .................................................................................. 198
  *Akhila Kishore, Anhad Bhasin, Arun Balaji, Chandrasekar Vuppalapati, Divyesh Jadav, Preethi Anantharaman, and Shriutee Gangras*
- Transparent Link: A Framework of Anonymizing MSA-Dataset Based on Probabilistic Graphical Model ...................................... 204
  *Li-E Wang, Yameng Guo, and Xianxian Li*

### Session 8: Invited Papers

- An Advanced Inventory Data Mining System for Business Intelligence .................................................................................... 210
  *Qifeng Zhou, Bin Xia, Wei Xue, Chunqiu Zeng, Ruyuan Han, and Tao Li*
- A Review of Performance Evaluation on 2D Face Databases ..................................................................................................... 218
  *Gabriel Castaneda and Taghi M. Khoshgoftaar*
- Data-Driven Water Quality Analysis and Prediction: A Survey .................................................................................................. 224
  *Gaganjot Kang, Jerry Zeyu Gao, and Gang Xie*
PROV-TE: A Provenance-Driven Diagnostic Framework for Task Eviction in Data Centers

Abdulaziz Albati, David McKee, Paul Townend, Lydia Lau, and Jie Xu

**Big Data Security and Service Workshop**

**Session 1: Security Services for Smart City**

Smart City Service Acceleration on FPGAs

Youngsoo Kim, Marcus Garcia, Alan Chen, Nathan Wong, Shrikant Jadhav, and Clay S. Gloster

Analyzing Movie Scripts as Unstructured Text

Seong-Ho Lee, Hye-Yeon Yu, and Yun-Gyung Cheong

Smart Illegal Dumping Detection

Akshay Dabholkar, Bhushan Muthiyani, Shilpa Srinivasan, Swetha Ravi, Hyeran Jeon, and Jerry Gao

Intelligent Learning Systems Design for Self-Defense Education

Hyeran Jeon, Kaikai Liu, Younghee Park, Jerry Gao, Gong Chen, and Jim Kao

**Session 2: Big Data for Security**

Big Data Cloud Deduplication Based on Verifiable Hash Convergent Group Signcryption

Ei Mon Cho and Takeshi Koshiba

DDoS Attack Mitigation in Internet of Things Using Software Defined Networking

M. Ejaz Ahmed and Hyoungshick Kim

Graph-Theory Based Simplification Techniques for Efficient Biological Network Analysis

Euiseong Ko, Mingon Kang, Hyung Jae Chang, and Donghyun Kim

**Big Data Quality Assurance Workshop**

**Session 1: Big Data Quality Assurance and Validation**

Big Data Validation Case Study

Chunli Xie, Jerry Gao, and Chuanqi Tao

A Practical Approach to Validation of Buildings’ Sensor Data: A Commissioning Experience Report

Claudio Giovanni Mattera, Sanja Lazarova-Molnar, Hamid Reza Shaker, and Bo Narregaard Jørgensen

Evolution Slicing-Based Change Impact Analysis

Wanzhi Wen, Jianping Chen, Jiaqi Yuan, and Xiaoyong Chen
Session 2: Quality Assurance and Validation for Big Data-Based Applications

A Quality Evaluation Approach to Search Engines of Shopping Platforms .................................................................299
   Chen Hao, Tao Chuanqi, and Jerry Gao

A Viewpoint on Location-Based Mobile Apps Testing ..................................................................................................305
   Wajeh Addin Mohsen, Chuanqi Tao, Qian-mu Li, and Jerry Gao

A Survey on Quality Assurance Techniques for Big Data Applications ........................................................................313
   Pengcheng Zhang, Xuewu Zhou, Wenrui Li, and Jerry Gao

Author Index ..................................................................................................................................................................320