Table of Contents

2017 IEEE Sixth International Congress on Big Data

BigData Congress 2017

Research Track Papers
Applications Track Papers
Short Paper Track Papers
Visionary Track Papers
Message from the General Chairs
Message from the Program Chairs
Organizing Committee
Technical Program Committee
IEEE Computer Society Technical Committee on Services Computing
Services Society

2017 IEEE Sixth International Congress on Big Data

(BigData Congress 2017)

BigData Congress Research Track

Research Session 1: Machine Learning for Big Data

Comprehensive Predictions of Tourists’ Next Visit Location Based on Call Detail Records and LTSM Machine Learning(#11725) (not uploaded)

Bayesian Networks with Structural Restrictions: Parallelization, Performance, and Efficient Cross-Validation(#11704)
Hao Peng (University of Georgia US) Zhe Jin (University of Georgia) John Miller (University of Georgia)
Building An SVM Classifier for Automated Selection of Big Data(#11707)
Junhua Ding (East Carolina University US) Jiabin Wang (East Carolina University) Xiaojun Kang (China University of Geosciences) Xin-Hua Hu (East Carolina University)

Research Session 2: Big Data Analytics Framework

Adaptive Concurrent Query Execution Framework for an Analytical In-Memory Database System(#12020)
Harshad Deshmukh (University of Wisconsin - Madison US) Hakan Memisoglu (University of Wisconsin - Madison) Jignesh Patel (University of Wisconsin - Madison)

Toward Scalable Indexing and Search on Distributed and Unstructured Data(#12059)
Alexandru Iulian Orhean (Illinois Institute of Technology US) Kyle Chard (University of Chicago) Itua Ijagbone (Illinois Institute of Technology) Dongfang Zhao (University of Nevada) Ioan Raicu (Illinois Institute of Technology)

Scalable, Efficient Anonymization with INCOGNITO - Framework & Algorithm (#11691)
Antorweep Chakravorty (University of Stavanger NO) Chunming Rong, K. R. Jayaram, Shu Tao

Research Session 3: Big Data Privacy and Protection

Differentially Private Trajectory Analysis for Points-of-Interest Recommendation (#12074)
Chao Li (University of Pittsburgh US) Balaji Palanisamy (University of Pittsburgh US) James Joshi (University of Pittsburgh)

Privacy-Preserving Inductive Learning with Decision Trees(#12269)
Stacey Truex (Georgia Institute of Technology) Ling Liu (Georgia Institute of Technology US) Mehmet Emre Gursoy (Georgia Institute of Technology) Lei Yu (Georgia Institute of Technology)

SafeNVM: A Non-Volatile Memory Store with Thread-Level Page Protection(#12233)
Pradeep Kumar (The George Washington University) H. Howie Huang (The George Washington University)

Research Session 4: Social Big Data

Maximizing the Effectiveness of Advertising Campaigns on Twitter(#12058)
Yan Mei (Macquarie University AU) Weiliang Zhao (Macquarie University) Jian Yang (Macquarie University)

Big Social Data Analytics in Football: Predicting Spectators and TV Ratings from Facebook Data(#12062)
Nicolai H. Egebjerg (Copenhagen Business School) Niklas Hedegaard (Copenhagen Business School)
Research Session 5: Big Data Systems

An Integrated Software System for Supporting Real-Time Near-Infrared Spectral Big Data Analysis and Management(#12157)
Liping Zhao (University of Manchester and Key Lab of Beam Bioengineering, CAS), Shupeng Hu (University of Manchester), Yuejin Wu (Key Lab of Beam Bioengineering, CAS), Yanqing Lin (Key Lab of Beam Bioengineering, CAS), Jing Liu (Key Lab of Beam Bioengineering, CAS), Shuang Fan (Key Lab of Beam Bioengineering, CAS), Qi Wang (Key Lab of Beam Bioengineering, CAS), Zhuopin Xu (Key Lab of Beam Bioengineering, CAS), Xiaojun Zeng (University of Manchester), Yu Wang (Anhui University)

A Model-Driven Methodology for Big Data Analytics-as-a-Service(#12149)
Claudio Agostino Ardagna (Università degli Studi di Milano IT) Valerio Bellandi (Università degli Studi di Milano IT) Michele Bezzi (SAP France FR) Paolo Ceravolo (Università degli Studi di Milano) Ernesto Damiani (Khalifa University - EBTIC) Cedric Hebert (SAP France)

Bleach: A Distributed Stream Data Cleaning System(#11708)
Yongchao Tian (Eurecom) Pietro Michiardi (Eurecom FR) Marko Vukolic (IBM Research)

Research Session 6: Big Data Performance

Zero-day Attack Identification in Streaming data using Semantics and Spark(#12156)
Sai Pallaprolu (University of Maryland, Baltimore County) Muthu Thevar (University of Maryland, Baltimore County) Rishi Sankineni (University of Maryland, Baltimore County) George Karabatis (University of Maryland, Baltimore County) Jianwu Wang (University of Maryland, Baltimore County US)

Big Data Processing: Scalability with Extreme Single-Node Performance(#12153) (Not Received) (not uploaded)

Parallel Compression and Indexing of Large-Scale Geospatial Raster Data with GPGPUs(#11730)
Nathalie Kaligirwa (University of Oklahoma) Eleazar Leal (University of Oklahoma US) Le Gruenwald (University of Oklahoma US) Jianting Zhang (City College of New York US) Simin You (CUNY Graduate Center)
Research Session 7: Resource Management in Big Data

Scheduling Recurring Distributed Dataflow Jobs Based on Resource Utilization and Interference(#11754)

Resource Allocation in cloud-based distributed cameras(#12057)
Bikash Agrawal (DNVGL) Jayachander Surbiryala (University of Stavanger NO) Chunming Rong (University of Stavanger NO)

Resource Management for Parallel Processing Frameworks with Load Awareness at Worker Side(#11729)
Son-Hai Ha (Orange - EURECOM) Patrick Brown (Orange) Pietro Michiardi (EURECOM FR)

Research Session 8: Big Data Processing

CEPaaS: Complex Event Processing as a Service(#12050)
Wilson A. Higashino (Western University CA) Miriam A. M. Capretz (Western University CA) Luiz F. Bittencourt (University of Campinas BR)

Towards Elucidating the Structural Principles of Host-Pathogen Protein-Protein Interaction Networks: A bioinformatics survey(#12187)
Huaming Chen (University of Wollongong AU) Jiangning Song (Monash University) Geng Sun (University of Wollongong AU) Jun Shen (University of Wollongong AU) Lei Wang (University of Wollongong)

OB-Tree: Accelerating Data Cleaning in Out-of-Core Column-Store Databases(#11700)
Feng Yu (Youngstown State University), Brandon J. Latronics (Youngstown State University), Tyler Matacic (Youngstown State University), Eric S. Jones (Youngstown State University)

Research Session 9: Big Data Learning

A deep-learning based precipitation forecasting approach using multiple environmental factors(#11475)
Pengcheng Zhang (Hohai University, Nanjing, P.R.China CN) Lei Zhang (Hohai University, Nanjing, P.R.China) Hareton Leung (Hong Kong Polytechnic University) Jimin Wang (Hohai University, Nanjing, P.R.China)

What are my Users Looking for when preparing a Big Data Campaign(#12041)
Ernesto Damiani (Khalifa University, United Arab Emirates) Claudio Ardagna (Università degli Studi di Milano, Italy IT) Guido Lena Cota (Università degli Studi di Milano, Italy) Paolo Ceravolo (Università degli Studi di Milano, Italy) Muhammad Muneeb Kiani (Università degli Studi di Milano, Italy IT)

Content Embedding Regularized Matrix Factorization for Recommender Systems (#12247)
Hao Wu (Yunnan University) Zhengxin Zhang (Yunnan University) Kun Yue (Yunnan University) Binbin Zhang (Yunnan University) Ruichao Zhu (Yunnan University)
**Research Session 10: Big Data Case Study**

Spatio-Temporal Mining Approach for enhancing Satellite Data Availability: A Case Study on Blue Green Algae (#11747)
VINAY BODDULA (University of Georgia US) Lakshmish Ramaswamy (University of Georgia) Deepak Mishra (University of Georgia)

Real-Time Anomaly Detection of Short Time-Scale GWAC Survey Light Curves (#12042)
Tianzhi Feng (Beijing University, China CN) Zhihui Du (Tsinghua University, China) Yankui Sun Tsinghua University, China Jianyan Wei (National Astronomical Observatories of China, China) Jing Bi (Beijing University, China US) Jason Liu (Florida International University, USA)

**BigData Congress Visionary Track**

Parallel Processing over Spatial-Temporal Datasets from Geo, Bio, Climate and Social Science Communities: A Research Roadmap (#12342)
Sushil K. Prasad, Daniel Aghajanian, Michael McDermott, Dhara Shah (Georgia State University); Mohamed Mokbel (University of Minnesota); Satish Puri (Marquette University); Sergio J. Rey (University of California Riverside); Shashi Shekhar and Yiqun Xe (University of Minnesota); Ranga Raju Vatsavai (North Carolina State University); Fusheng Wang, Yanhui Liang and Hoang Vo (Stony Brook University); Shaowen Wang (University of Illinois at Urbana-Champaign)

Data Collection and Management in IoT Environments: Challenges and Future Directions (#12336)

Adding Support for Theory in Open Science Big Data (12071)
John A. Miller (University of Georgia), Hao Peng (University of Georgia) and Michael E. Cotterell (University of Georgia)

Towards A Privacy-Aware Quantified Self Data Management Framework
Bhavani Thuraisingham, and Murat Kantarcioğlu (The University of Texas at Dallas), Elisa Bertino (Purdue University), and Maribel Fernande (University of London, UK)

**BigData Congress Application Track**

Session 1: Big Data QoS

De-centralized Reputation-based Trust Model to Discriminate between Cloud Providers Capable of Processing Big Data (#12049)
A Multivariate Clustering Approach for Infrastructure Failure Predictions(#12036)
Simon Luo (DATA61 and University of Tasmania) Victor W. Chu (Nanyang Technological University) Jianlong Zhou (DATA61) Fang Chen (DATA61) Raymond K. Wong (UNSW) Weidong Huang (University of Tasmania) Microblog Sentiment Classification using Parallel SVM in Apache Spark(#11684)
Bo Yan (Beijing Institute of Technology) Zijiang Yang (York University CA) Yitian Ren (Beijing Institute of Technology) Xing Tan (York University) Eric Liu (Unionville Montessori Private School)

Session 2: Machine Learning in Big Data

Drowsy Driving Warning System based on GS1 Standards with Machine Learning(#12046)
Seongjoo Moon (Mando, Pankyo, Korea) Moran Min (Graduate School of Software, Korea US) Jangwoo Nam (Solutionlink, Seocho, Korea KR) Jaeyou Park (Graduate School of Software, Korea) Doheon Lee (Mando, Korea) Daeyoung Kim (Graduate School of Software, Seoul, Korea)

An Effective Distributed GHSOM Algorithm for Unsupervised Clustering on Big Data(#11732)
Chui-Hui Chiu (National Chengchi University Taipei, Taiwan) Jin-Jie Chen (National Chengchi University Taipei, Taiwan) Fang Yu (National Chengchi University Taipei, Taiwan TW)

Integrating Extractive and Abstractive Models for Long Text Summarization(#12190)
Shuai Wang (National University of Defense Technology) Xiang Zhao (National University of Defense Technology, China) Bo Li (National University of Defense Technology) Bin Ge, Daquan Tang (National University of Defense Technology)

Session 3: Big Data Processing

SCMAT: A Mechanism Presuming SCMs to Efficiently Enable both OLAP and OLTP (#12035)
Takamitsu Shioi (School of Computing, Tokyo institute of Technology JP) Kenji Hatano (Doshisha University) Haruo Yokota (School of Computing, Tokyo institute of Technology JP)

A Real-time Scheduling Strategy Based on Processing Framework of Hadoop(#11694)
Fangbing Chen (, Shenzhen Graduate School, Peking University CN) Ji Liu (Shenzhen Graduate School, Peking University) Yuesheng Zhu (Shenzhen Graduate School, Peking University)

Developing a Real-time Data Analytics Framework For Twitter Streaming Data(#11738)
Babak Yadranjaghdam (East Carolina University US) Seyedfaraz Yasrobi (East Carolina University US) Nasseh Tabrizi (East Carolina University)
Session 4: Big Data Mining

Locating Sources in Online Social Networks via Random Walk(#12259)
Zhijian Zhang (Yunnan University) Kun Yue (Yunnan University) Zhengbao Sun (Yunnan University) Ling Liu (Yunnan Police Officer Academy) Weiyi Liu (Yunnan University)

Big Data and Analysis of Data Transfers for International Research Networks Using NetSage (#11753)
Alberto Gonzalez (University of Hawaii Mānoa US) Jason Leigh (University of Hawaii Mānoa) Sean Peisert (Lawrence Berkley National Lab) Brian Tierney (Lawrence Berkley National Lab) Andrew Lee (Indiana University) Jennifer M. Schopf (Indiana University)

PAS4SCF - A Publication Analytics System for Services Conference Federation (SCF 2004-2016)(#12216)
Liang-Jie Zhang (Kingdee Research) Huan Chen (Kingdee Research CN)

Session 5: Big Data Algorithms

I-sampling: A New Block-Based Sampling Method for Large-Scale Dataset(#11718)
Yulin HE (Shenzhen University CN) Joshua Zhexue Huang, Hao LONG, Qiang WANG, Chenghao WEI

Leveraging Stacked Denoising Autoencoder in Prediction of Pathogen-Host Protein-Protein Interactions(#11695)
Huaming Chen (University of Wollongong AU) Jun Shen (University of Wollongong AU) Lei Wang (University of Wollongong) Jiangning Song (Monash University)

Experiences with the Twitter Health Surveillance (THS) System(#11734)
Manuel Rodriguez (UPR, Mayaguez US)

Session 6: Big Data Case Studies

Data Governance Framework for Big Data Implementation with a Case of Korea(#11585)
Hee-Young Kim (Hankuk University of Foreign Studies) June-Suh Cho (Hankuk University of Foreign Studies KR)

Educational and Scientific Applications of Climate Model Diagnostic Analyzer(#12037)
Seungwon Lee (JPL) Chengxing Zhai (JPL) Terence Kubar (JPL/UCLA) Benyang Tang (JPL) Lei Pan (JPL) Juijin Li (JPL) Qihao Bao (CMU)

Implementation of Disaster Prevention-Oriented Information Service Platform of XRAIN on DIAS 11693
Session 7: Big Data Models

New E-Commerce User Interest Patterns(#11709) (not uploaded)
Matthias Volk (Magdeburg Research and Competence Cluster DE) Abed Elrahman Shareef (Magdeburg Research and Competence Cluster) Naoum Jamous (Magdeburg Research and Competence Cluster)
Klaus Turowski (Magdeburg Research and Competence Cluster)

Big Data Analysis and Visualization for the Smart Grid(#11697)
Wilson Rivera (University of Puerto Rico US) Alejandro Sanchez (University of Puerto Rico)

Session 8: Big Data Analysis

Compliance analysis of configurable business process model based on extend CTL(#11690)
Huang Yiwang (School of Data Science, Tongren University)

Any-time methods for time-series prediction with missing observations(#12217)
Yanjie Wei (Harbin Institute of Technology US) Hongzhi Wang (Harbin Institute of Technology) Shengfei Shi (Harbin Institute of Technology) Hong Gao (Harbin Institute of Technology) Jianzhong Li (Harbin Institute of Technology)

BigData Congress Theme Track

Session 1: Big Data Management

TeenRead: An Adolescents Reading Recommendation System towards Online Bibliotherapy(#12218) (not uploaded)
Yunxing Xin (Tsinghua University CN) Yongqiang Chen (Tsinghua University) Li Jin (Tsinghua University)
Yici Cai (Tsinghua University) Ling Feng (Tsinghua University)

A Resilient Index Graph for Querying Large Biological Scientific Data(#12205)
Liang Li (CNIC CN) Zhihong Shen (CNIC CN) Jianhui Li (CNIC CN) Dongjiang Liu (CNIC CN) Huajin Wang, Lipeng Wang (Institute of Microbiology) Qinglan Sun (Institute of Microbiology)

SPARQL Query Parallel Processing: A Survey(#12222)
Jiaying Feng (Tianjin University) Chenhong Meng (Tianjin University) Jiaming Song (Tianjin University)
Session 2: Big Data Management

Burst-based Event Classification on Weakly Labeled Time Series Data of Sensors(#12220)
Hanbo Zhang (Fudan University CN) Yawen Wang (Fudan University) Peng Wang (Fudan University)
Wei Wang (Fudan University)

Bootstrapping a Blockchain Based Ecosystem for Big Data Exchange(#12202)
Jinshuan Chen (Renmin University of China CN) Yunzhi Xue (Institute of Software, Chinese Academy of Sciences)

The Performance of SQL-on-Hadoop Systems An Experimental Study (#12188)
xiongpai qin (Renmin University of China CN) yueguo chen (Renmin University of China CN) jun chen (Renmin University of China) shuai li (Renmin University of China) Jiesi liu (Renmin University of China)
Huijie zhang (Renmin University of China)

Session 3: Real-Time Big Data Analytics

Real Time Social Attitude Expression Prediction(#12213)
Jiawei Zhang (UIC) Philip S. Yu (UIC US) Charu C. Aggarwal (IBM) Limeng Cui (UCAS)

A hybrid learning framework for imbalanced stream classification(#12240)
Wenbin Zhang (University of Maryland, Baltimore County) Jianwu Wang (University of Maryland, Baltimore County US)

Personalized recommendation algorithm based on LFM with QoS constraint(#12215)
Ying JIN (Department of Management of Hefei University) Yiwen ZHANG (The Key Laboratory of Intelligent Computing & Signal Processing, Ministry of Education, Anhui Univer)

BigData Congress Short Paper Track

Session 1: Big Data Processing

Clustering Data Streams with Adaptive Forgetting(#12075)
Gopi Chand Nutakki (Student US) Olfa Nasraoui (Professor US)
Big Data Pre-Processing: Closing the Data Quality Enforcement Loop(#12069)
Ikbal Taleb (Concordia University) Mohamed Adel Serhani (UAE University AE)

Cloud-based and Big data-enabled Brokerage System for Smart Grids(#12052)
Howard Martinez (University of Puerto Rico PR) Manuel Rodriguez (University of Puerto Rico US)

Graph Computation and Its Applications in Smart Grid (#12028)
Guangyi Liu (GEIRI,North America) Kewen Liu (GEIRI) Di Shi (GEIRI,North America) Xi Chen (GEIRI,North America) wendong Zhu (GEIRI, North America) Zhiwei Wang

Session 2: QoS of Big Data

Open Medical Big Data and Open Consent and their Impact on Privacy(#12109)
Jingquan Li (Texas A&M University-San Antonio US)

Enhancing Energy, Health and Security Policy by Extracting, Enriching and Interfacing Next Generation Data in the Transport domain (A study on the use of Big Data in cross-sectoral policy development) (#12072)
Susan Grant-Muller (University of Leeds GB) Frances Hodgson (University of Leeds) Nicholas Malleson (University of Leeds) Robert Snowball (Transport Development, Newcastle City Council)

Managing Emerging Risks and Liabilities in Data-Enabled Solutions(#12053)
Olaitan Olaleye (Philips R&D) Talmai Oliveira (Philips R&D) Alexandru Darie (Philips R&D)

Ecosystem Design of Big Data through Previous Study Analysis in the World: Policy Design for Big Data as Public Goods#11719)

Challenges of software testing for astronomical big data (#12110) (not uploaded)
Lixiao Zhou (National Astronomical Observatories, Chinese Academy of Sciences) Maohai Huang (National Astronomical Observatories, Chinese Academy of Sciences)

Session 3: Big Data Applications

A distributed system for fining high profit areas over big taxi trip data with MongoDB and Spark(#12070) (not uploaded)
Fadhilah Putri (Pusan National University) Joonho Kwon (Pusan National University)

Comparison of Gambling Industry through Study Analysis in the World(#11716)

Big data Analysis on Gambling related News in South Korea(#11715)
Semantic Analysis of Transportation related Researches(#11717)
SiYeun Lee (The Korea Transport Institute KR) HyeJung Moon (Institute of Lottery Policy KR)

Session 4: Satellite Session

Business Graphing for Internet-enabled Enterprises(#12331)
Bo Hu (Kingdee ) Jian Wang (Wuhan University CN) Yutao Ma (Wuhan University CN) Liang-Jie Zhang (Kingdee )

An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends(#12317)
Zibin Zheng (Sun Yat-sen University, China ) Shaoan Xie (Sun Yat-sen University China ) Hongning Dai (Macau University of Science and Technology, Macau) Xianping Chen (National Engineering Research Center of Digital Life , Sun Yat-sen ) Huaimin Wang (National Laboratory for Parallel & Distributed Processing National University of Defense Technology, )

A New Sleep Mechanism Based on SDN Framework(#12314)
Yu Peizhao (Wuhan University of Tech ) Xu Lei (Wuhan University of Science and Technology US) Gu Jinguang (Wuhan University of Science and Technology CN) He Heng (Wuhan University of Science and Technology )

CloudSec : A Novel Approach to Verifying Security Conformance at the Bottom of the Cloud(#12257)
Liu Xiaochen (Beihang University )

Cutting parameters optimization based on ITLBO algorithm with Bigdata driven(#12243)
(not uploaded)

wang chen (Shanghai university,China) dong yao (Shanghai university,China)