

**2018 IEEE/ACM 5th
International Conference on Big
Data Computing Applications and
Technologies (BDCAT 2018)**

**Zurich, Switzerland
17-20 December 2018**



**IEEE Catalog Number: CFP18B46-POD
ISBN: 978-1-5386-5503-0**

**Copyright © 2018 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:	CFP18B46-POD
ISBN (Print-On-Demand):	978-1-5386-5503-0
ISBN (Online):	978-1-5386-5502-3

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

2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT) **BDCAT 2018**

Table of Contents

Message from the UCC/BDCAT 2018 General Chairs	ix
Message from the BDCAT 2018 Program Chairs	xi
Message from the UCC/BDCAT 2018 Poster Chairs	xii
BDCAT 2018 Organizing Committee	xiii
BDCAT 2018 Technical Program Committee	xiv

Session: Big Data Infrastructures I

Spark-DIY: A Framework for Interoperable Spark Operations with High Performance Block-Based Data Models	1
<i>Silvina Caino-Lores (University Carlos III of Madrid), Jesús Carretero (University Carlos III of Madrid), Bogdan Nicolae (Argonne National Laboratory), Orcun Yildiz (Argonne National Laboratory), and Tom Peterka (Argonne National Laboratory)</i>	
CloneHadoop: Process Cloning to Reduce Hadoop's Long Tail	11
<i>Sarthak Kukreti (North Carolina State University) and Frank Mueller (North Carolina State University)</i>	

Session: Big Data Infrastructures II

Fault-Tolerant Query Execution over Distributed Bitmap Indices	21
<i>Sam Burdick (Amazon.com), Jahrme Risner (University of Puget Sound), David Chiu (University of Puget Sound), and Jason Sawin (University of St. Thomas)</i>	
Skipping Unused Events to Speed Up Rollback-Recovery in Distributed Data-Parallel CEP	31
<i>Guilherme F. Lima (Pontifical Catholic University of Rio), Ahmad Slo (University of Stuttgart), Sukanya Bhowmik (University of Stuttgart), Markus Endler (Pontifical Catholic University of Rio), and Kurt Rothermel (University of Stuttgart)</i>	

ntPack: A Software Package for Big Data in Genomics	41
<i>Inanc Birol (BC Cancer Genome Sciences Centre), Hamid Mohamadi (BC Cancer Genome Sciences Centre), and Justin Chu (BC Cancer Genome Sciences Centre)</i>	

Session: Big Data Applications I

Prediction of Air Pollution through Machine Learning Approaches on the Cloud	51
<i>Richard O. Sinnott (University of Melbourne) and Ziyue Guan (University of Melbourne)</i>	
Non-Linear Machine Learning with Active Sampling for MOX Drift Compensation	61
<i>Tamara Matthews (Dublin Institute of Technology), Muhammad Iqbal (National College of Ireland), and Horacio Gonzalez-Velez (National College of Ireland)</i>	

Session: Big Data Applications II

Detecting System Anomalies in Multivariate Time Series with Information Transfer and Random Walk	71
<i>Jongsun Lee (Seoul National University), Hyun-Soo Choi (Seoul National University), Yongkweon Jeon (Seoul National University), Yongsik Kwon (SAP Labs Korea), Donghun Lee (SAP Labs Korea), and Sungroh Yoon (Seoul National University)</i>	
A Decentralized SNS System Based on XMPP with Connection Control in Large-Scale Disasters	81
<i>Hui Yu (Ochanomizu University), Yasunori Owada (National Institute of Information and Communications Technology), and Masato Oguchi (Ochanomizu University)</i>	

Session: Big Data Applications III

A Mobile Application for Dog Breed Detection and Recognition Based on Deep Learning	87
<i>Richard O. Sinnott (University of Melbourne), Fang Wu (University of Melbourne), and Wenbin Chen (University of Melbourne)</i>	
Prediction of Bus Delay over Intervals on Various Kinds of Routes Using Bus Probe Data	97
<i>Tsubasa Yamaguchi (Kyushu University), Mansur As (Kyushu University), and Tsunenori Mine (Kyushu University)</i>	

Session: Big Data Applications IV

Development of a Radiology Decision Support System for the Classification of MRI Brain Scans	107
<i>Alwin Yaoxian Zhang (National University of Singapore), Sean Shao Wei Lam (Singapore Health Services & Duke-NUS Medical School), Nan Liu (Singapore Health Services & Duke-NUS Medical School), Yan Pang (National University of Singapore), Ling Ling Chan (Duke-NUS Medical School & Singapore General Hospital), and Phua Hwee Tang (Duke-NUS Medical School & KK Women's and Children's Hospital)</i>	

Adaptive General Event Popularity Analysis on Streaming Data	116
<i>Qidong Zhang (Hohai University), Yan Tang (Hohai University), Zequan Guo (Hohai University), Tingting Wu (Hohai University), Li Xu (Southeast University), and Xiaofeng Gao (Shanghai Jiao Tong University)</i>	

Session: Big Data Applications V

An Empirical Performance Evaluation of Semantic-Based Similarity Measures in Microblogging Social Media	126
<i>Noufa Alnajran (Manchester Metropolitan University), Keeley Crockett (Manchester Metropolitan University), David McLean (Manchester Metropolitan University), and Annabel Latham (Manchester Metropolitan University)</i>	
An Incremental Community Detection Method in Social Big Data	136
<i>Zhenyu Wu (Nanjing University of Posts and Telecommunications), Jiaying Chen (Nanjing University of Posts and Telecommunications), and Yinuo Zhang (Nanjing University of Posts and Telecommunications)</i>	
GDup: De-Duplication of Scholarly Communication Big Graphs	142
<i>Claudio Atzori (Consiglio Nazionale delle Ricerche), Paolo Manghi (Consiglio Nazionale delle Ricerche), and Alessia Bardi (Consiglio Nazionale delle Ricerche)</i>	

Session: Big Data Applications VI

A Novel Method of Processing Class Imbalance and Its Application in Transaction Fraud Detection	N/A
<i>Youjun Zhang (Tongji University of Shanghai), Guanjun Liu (Tongji University of Shanghai), Lutao Zheng (Tongji University of Shanghai), Chungang Yan (Tongji University of Shanghai), and Changjun Jiang (Tongji University of Shanghai)</i>	
Recurrent Embedding Kernel for Predicting Stock Daily Direction	160
<i>Linh Le (Kennesaw State University) and Ying Xie (Kennesaw State University)</i>	

Session: Big Data Analytics I

Can Temperature Be Used as a Predictor of Data Traffic? A Real Network Big Data Analysis	167
<i>Muhammad Nauman Rafiq (University of Oklahoma), Hasan Farooq (University of Oklahoma), Ahmed Zoha (University of Surrey), and Ali Imran (University of Oklahoma)</i>	
Presence Analytics: Making Sense of Human Social Presence within a Learning Environment	174
<i>Muawya Habib Sarnoub Eldaw (Birkbeck, University of London), Mark Levene (Birkbeck, University of London), and George Roussos (Birkbeck, University of London)</i>	

Session: Big Data Analytics II

An Improved Multi-Objective Evolutionary Approach for Clustering High-Dimensional Data	184
<i>Chao Liu (Beijing University of Technology), Qi Zhao (Beijing University of Technology), Bai Yan (Beijing University of Technology), Saber Elsayed (University of New South Wales Canberra), and Ruhul Sarker (University of New South Wales Canberra)</i>	
A Hierarchical Multi-Metric Framework for Item Clustering	191
<i>Maria Kotouza (Aristotle University of Thessaloniki), Konstantinos Vavliakis (Aristotle University of Thessaloniki), Fotis Psomopoulos (Centre for Research and Technology Hellas), and Pericles Mitkas (Aristotle University of Thessaloniki)</i>	
A Positive Approximation Set Based Accelerating Approach for Condition Attribute Reduction.....	198
<i>Tao Yan (Xi'an Jiaotong University), Chongzhao Han (Xi'an Jiaotong University), and Chengnan Wang (Xi'an Jiaotong University)</i>	

Posters

Large-Scale Data-Driven Financial Risk Modeling Using Big Data Technology	206
<i>Stockinger Kurt (Zurich University of Applied Sciences), Jonas Heitz (Zurich University of Applied Sciences), Nils Bundi (Zurich University of Applied Sciences), and Wolfgang Breymann (Zurich University of Applied Sciences)</i>	
Data Driven Priority Scheduling on Spark Based Stream Processing	208
<i>Tobi Ajila (Carleton University) and Shikaresh Majumdar (Carleton University)</i>	
Proposal and Evaluation of Event Search Method Based on SNS Data Analysis Focusing on Place and Time	211
<i>Ruriko Kudo (Ochanomizu University), Miki Enoki (IBM Research Tokyo), Akihiro Nakao (University of Tokyo), Shu Yamamoto (University of Tokyo), Saneyasu Yamaguchi (Kogakuin University), and Masato Oguchi (Ochanomizu University)</i>	
Urban Hourly Water Demand Prediction Using Human Mobility Data	213
<i>Kamil Smolak (University of Environmental and Life Sciences), Barbara Kasieczka (University of Environmental and Life Sciences), Katarzyna Siła-Nowicka (University of Glasgow), Katarzyna Kopaczyk (University of Environmental and Life Sciences), Witold Rohm (University of Environmental and Life Sciences), and Wiesław Fiałkiewicz (University of Environmental and Life Sciences)</i>	
Implementation of Distributed XA Transactions in MyCat Based on Table Broadcasting Mechanism	215
<i>Yingying Wu (Hohai University) and Zhenghe Liang (Hohai University)</i>	
Author Index	217