

Proceedings of the 3rd ASE International Conference on Big Data Science and Computing

2014, Beijing, China

BIGDATASCIENCE '14

Conference Chairs: Dr. Alvin Chin, Prof. Justin Zhan, Prof. Wei Ding, Prof. Junjie Wu

Program Chairs: Prof. Wei Xu, Dr. Fei Wang

Sponsors: ASE, Tsinghua University, DICE, ILAB NCAT, HPC Advisory Council, Microsoft, Baidu



**The Association for Computing Machinery
2 Penn Plaza, Suite 701
New York New York 10121-0701**

ACM COPYRIGHT NOTICE. Copyright © 2014 by the Association for Computing Machinery, Inc. Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Publications Dept., ACM, Inc., fax +1 (212) 869-0481, or permissions@acm.org.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, +1-978-750-8400, +1-978-750-4470 (fax).

Notice to Past Authors of ACM-Published Articles

ACM intends to create a complete electronic archive of all articles and/or other material previously published by ACM. If you have written a work that was previously published by ACM in any journal or conference proceedings prior to 1978, or any SIG Newsletter at any time, and you do NOT want this work to appear in the ACM Digital Library, please inform permissions@acm.org, stating the title of the work, the author(s), and where and when published.

ACM ISBN: 978-1-4503-2891-3

Proceedings of BigDataScience 2014 Beijing: Third International Conference on Big Data Science and Computing, Beijing, China

Table of Contents

Keynotes

1. Big Data Research from Management and Business Perspective: Thoughts for Debating
Wayne Huang, Ohio State University
2. Utility of Context Awareness for Behavioural Health
Guanling Chen, University of Massachusetts at Lowell
3. Unlocking Value from Open Data
Zhen Liu, Microsoft
4. Computing Issues for Big Data – Theory, Systems and Applications
Keynote Talk
Chunming Hu, Beihang University
5. Connecting People Through Mobile Social Networks
Alvin Chin (Microsoft)
6. Scaling Distributed Machine Learning with the Parameter Server
Keynote Talk
Mu Li (Baidu)

Session 1: Big Data Applications – I

7. Real Time Event Detection in Sina Weibo
Paper
Xinjiang Lu (Northwestern Polytechnic University), Zhiwen Yu (Northwestern Polytechnic University)
Bin Guo (Northwestern Polytechnic University), Jiafan Zhang (Northwestern Polytechnic University)
Alvin Chin (Microsoft), Jilei Tian (Microsoft), Yang Cao (Microsoft)
8. A Feature Set for Streams and a Demonstration on High-Frequency Financial Tick Data
Paper
Terry Lyons (University of Oxford), Hao Ni (University of Oxford), Harald Oberhauser (University of Oxford)
9. The Value of Using Open-Source Scalable Technologies for Social Science and Social Network Analysis
Paper
Eugene Ch'ng (University of Nottingham Ningbo)
10. Using Hadoop to Implement a Semantic Method of Assessing The Quality of Research Medical Datasets

Paper

Stephen Bonner (University of Huddersfield), Grigoris Antoniou (University of Huddersfield), Laura Moss (University of Glasgow), Ibad Kureshi (Durham University), David Corsar (University of Aberdeen)

Session 2: Big Data Science, Mining and Analytics – I

11. Community Detection for Clustered Attributed Graphs via a Variational EM

Paper

Xiangyong Cao (Xi'an Jiaotong University), Xiangyu Chang (Xi'an Jiaotong University), Zongben Xu (Xi'an Jiaotong University)

Session 3: Big Data Mining, Science and Analytics – II

12. Detecting Weak Public Transport Connections from Cellphone and Public Transport Data

Paper

Thomas Holleczeck (SingTel), Liang Yu (A-STAR), Joseph K. Lee (University of British Columbia), Oliver Senn (Singapore-MIT Alliance for Research and Technology), Kristian Kloeckl (MIT), Carlo Ratti (MIT), Patrick Jaillet (MIT)

13. Galactica: A GPU Parallelized Database Accelerator

Paper

Keh Kok Yong (Mimos Berhad), Ettikan K. Karuppiah (Mimos Berhad), Simon See (NVIDIA)

14. Scalable Hierarchical Recommendations Using Spatial Autocorrelation

Paper

Ayushi Dalmia (IIIT Hyderabad), Joydeep Das (The Heritage Academy), Prosenjit Gupta (Heritage Institute of Technology), Subhashis Majumder (Heritage Institute of Technology), Debarshi Dutta (University of Southampton)

15. Recommending Friends in Local Social Networks: An Envelope of Algorithms

Paper

Yuewen Liu (Xi'an Jiaotong University), Xiangyu Chang (Xi'an Jiaotong University), Wei Huang (Ohio University)

Session 4: Big Data Applications – II

16. Creating a Chemistry of Sciences with Big Data

Paper

Yike Guo (Imperial College London), David Johnson (Imperial College London)

17. Hidden Aspect Rating Discovery from Text Reviews of E-Commerce Web Sites

Paper

Yinqing Xu (The Chinese University of Hong Kong), Wai Lam (The Chinese University of Hong Kong)

18. Data Readiness Level For Unstructured Data with a Focus on Unindexed Text Data

Paper

Yang Lu (North Carolina A&T State University), Xing Fang (North Carolina A&T State University)
Justin Zhan (North Carolina A&T State University)

19. Clustering Experiments on Big Transactional Data for Market Segmentation

Paper

Ashishkumar Singh (Monash University), Grace Rumantir (Monash University), Annie South (Australia
New Zealand Bank), Blair Bethwaite (Monash University)

Symposium: 1st International Symposium On Recent Advances In Big Data

20. SQLPhi: A SQL-Based Database Engine for Intel Xeon Phi Coprocessors

Stuart Olsen (Texas State University), Brian Romoser (Texas State University), Ziliang Zong (Texas
State University)

21. Bigdata Analysis with Interactive Visualization Using R Packages

Wonhee Cho (Kookmin University), Yoojin Lim (Kookmin University), Hwangro Lee (Kookmin
University), Mohan Krishna Varma Nandimandalam (Kookmin University), Moonsoo Lee (Kookmin
University), Eunmi Choi (Kookmin University)

22. CAP Plus for Big Data

Justin Shi (Temple University)

23. Identity Tracking in Big Data - Preliminary Research Using In-Memory Data Graph Models for
Record Linkage and Probabilistic Signature Hashing for Approximate String Matching in Big Health
and Human Services Databases

Joseph Jupin (Temple University), Justin Shi (Temple University)

24. On The Reproducibility and Scalability of Extreme Scale Applications

Justin Shi (Temple University)

25. Paging Method Switching for QEMU-KVM Guest Machine

Yu Zhang (Technische Universitat Chemnitz) Rene Oertel (Technische Universitat Chemnitz),
Wolfgang Rehm (Technische Universitat Chemnitz)

26. FAST: Fast Anonymization of Big Data Streams

Esmaeil Mohammadian (DNSL), Morteza Noferesti (DNSL), Rasool Jalili (DNSL)

Posters

27. Fast Mode Regression in Big Data Analysis

Keming Yu (Brunel University), Joanne Ford (Brunel University), Frauke Becker (Brunel University)

28. IFLINE - The Student's Social Network
Arthur Bortone (Federal Institute of Rio de Janeiro), Carlos Oliveira (Federal Institute of Rio de Janeiro), Henric Johnson (Blekinge Institute of Technology)
29. Big Data and Smart Grid
Xin Miao (China Electric Power Research Institute)
30. Scholarly Collaboration in the HASTAC Social Network
Marco T. Bastos (Duke University)
31. Community Detection for Clustered Attributed Graphs via a Variational EM Algorithm
Xiangyong Cao (Xi'an Jiaotong University), Xiangyu Chang (Xi'an Jiaotong University), Zongben Xu (Xi'an Jiaotong University)
32. Galactica: A GPU Parallelized Database Accelerator
Keh Kok Yong (MIMOS Berhad), Ettikan K. Karupiah (MIMOS Berhad), Simon See (NVIDIA Corporation)
33. Scalable Hierarchical Recommendations Using Spatial Autocorrelation
Ayushi Dalmia (IIIT Hyderabad), Joydeep Das (The Heritage Academy), Prosenjit Gupta (Heritage Institute of Technology), Subhashis Majumder (Heritage Institute of Technology), Debarshi Dutta (University of Southampton)
34. VANET-based Secure Value-Added Services
Shi-Jinn Horng (National Taiwan University of Science and Technology, and Southwest Jiaotong University), Shiang-Feng Tzeng (National Taiwan University of Science and Technology, and Southwest Jiaotong University)
35. FAST: Fast Anonymization of Big Data Streams
Esmaeil Mohammadian (Sharif University of Technology), Morteza Noforesti (Sharif University of Technology), Rasool Jalili (Sharif University of Technology)
36. An Optimized Method to Allocate and Re-allocate Aggregated Data in a Data Storage
Vivian Lee (Fujitsu Laboratories of Europe), Bo Hu (Fujitsu Laboratories of Europe)
37. The Value of Using Big Data Technologies for Social Network Analysis in Computational Social Science
Eugene Ch'ng (University of Nottingham Ningbo)
38. Detecting Weak Public Transport Connections from Cellphone and Public Transport Data
Thomas Holleczeck (SingTel), Liang Yu (A*STAR), Joseph K. Lee (University of British Columbia), Oliver Senn (Singapore-MIT Alliance for Research and Technology), Carlo Ratti (MIT), Patrick Jaillet (MIT)

39. Recommending Friends in Local Social Networks: An Envelope of Algorithms

Yuewen Liu (Xi'an Jiaotong University), Xiangyu Chang (Xi'an Jiaotong University), Wayne Wei Huang (Ohio State University)

40. Data Readiness Level For Unstructured Data with a Focus on Unindexed Text Data

Yang Lu (North Carolina A&T State University), Xing Fang (North Carolina A&T State University), Justin Zhan (North Carolina A&T State University)