2018 IEEE International Conference on

BIG DATA

DECEMBER 10–13 | SEATTLE, WA

Sponsored by

2018 IEEE International Conference on Big Data

Organization Committee

Conference Co-Chairs

Dr. Donald Kossmann: Microsoft Research, USA
Prof. Bing Liu: University of Illinois at Chicago, USA

Program Co-Chairs

Dr. Naoki Abe: IBM T.J. Watson Research Center, USA
Prof. Huan Liu: Arizona State University, USA
Prof. Calton Pu: Georgia Institute of Technology, USA

Program Vice Co-Chairs

Prof. Kisung Lee: Louisiana State University, USA
Prof. Jiliang Tang: Michigan State University, USA

Industry and Government Program Committee Co-Chairs

Dr. Nesreen Ahmed: Intel Labs, USA
Dr. Mu Qiao: IBM Almaden Research Center, USA
Dr. Yang Song: Google AI, USA

Workshop Co-Chairs

Prof. Jingrui He: Arizona State University, USA
Prof. Jeffrey Saltz: Syracuse University, USA

Tutorial Co-Chairs

Prof. Honggang Wang: University of Massachusetts, Dartmouth, USA
Dr. Lingfei (Teddy) Wu: IBM Research AI, USA

Big Data Cup Chairs

Prof. Mohammad Al Hasan (Hasan): IUPUI, USA
Prof. Yicheng Tu: South Florida University, USA

Poster Chairs

Prof. Michael Gubanov: University of Texas at San Antonio, USA
Prof. Xia (Ben) Hu: Texas A&M University, USA

Sponsorship Chair

Prof. Xiaohua Tony Hu: Drexel University, USA

Local Arrangement Chairs

Dr. Mengwen Liu: Amazon, USA
Dr. Zunyan Xiong: T-Mobile, USA

Registration Chair

Prof. Yuan An: Drexel University, USA

Publicity Chairs

Prof. Hyoil Han: Illinois State University, USA
Prof. Alfredo Cuzzocrea: University of Trieste, Italy
Prof. Dominik Slezak: University of Warsaw, Poland

Student Travel Award Chair

Prof. Feng Chen: SUNY at Albany, USA

BigData Steering Committee

Dr. Amr Awadallah: Cloudera, USA
Dr. Xueqi Cheng: Chinese Academy of Science, China
Prof. Yi-ke Guo: Imperial College, UK
Prof. Jimmy Lin: University of Waterloo, Canada
Prof. Xiaohua Tony Hu (Chair) (xh29@drexel.edu): Drexel University, USA
Dr. Raghunath Nambiar: Cisco Systems, USA
Prof. Jian Pei: Simon Fraser University, Canada
Prof. Vijay Raghavan: University of Louisiana at Lafayette, USA
Prof. Amit Sheth: Wright State University, USA
Prof. Matthew Smith: Leibniz Universität Hannove, Germany
Dr. Shusaku Tsumoto: Shimane University, Japan
Prof. Athanasios Vasilakos: Lulea University of Technology(LTU), Sweden
Prof. Qiang Yang: Hong Kong University of Science and Technology, China
Pro. Wei Wang: University of California at Los Angle, USA

Program Committee

Main Conference Senior PC Committee

Name	Organization	Country
Karl Aberer	EPFL	Switzerland
Danilo Ardagna	Politecnico di Milano, Italy	Italy
Peter Baumann	Jacobs University	Germany
Albert Bifet	LTCI, T él écom ParisTech	France
Nozha Boujemaa	INRIA	France
Ricardo J. G. B. Campello	University of Newcastle	Australia
Barbara Carminati	University of Insubria, Italy	Italy
Philip Chan	Florida Institute of Technology	USA
Edward Chang	Google China	China
Nitesh Chawla	University of Notre Dame	USA
Sanjay Chawla	Qatar Computing Research Institute, HBKU	Qatar
Keke Chen	Wright State University	USA
Lei Chen	Hong Kong University of Science and Technology	Hong Kong
Shu-Ching Chen	Florida International University	USA
Wei Chen	Microsoft Research Asia	China
Reynold Cheng	University of Hong Kong	China
Xueqi Cheng	Chinese Academy of Science	CHINA
Vijil Chenthamarakshan	IBM Research	USA
Gao Cong	Nanyang Technological University, Singapore	Singapore
Bin Cui	Peking University	China
Alfredo Cuzzocrea	ICAR-CNR and University of Calabria, Italy	Italy
Ernesto Damiani	Universita degli Studi di Milano, Italy	Italy
Gautam Das	University of Texas at Arlington	USA
Xiaoyong DU	Renmin University of China	China
Joao Eduardo Ferreira	University of Sao Paulo	Brazil
Avigdor Gal	Technion - Israel Institute of Technology	Israel
Dimitrios Geogakopoulos	Swinburne U	Australia
Dimitrios Gunopulos	National and Kapodistrian University of Athens	Greece
Jiawei Han	University of Illinois at Urbana-Champaign	USA
Wook-Shin Han	Postech	Korea
Bingsheng He	National University of Singapore	Singapore
Teruo Higashino	Osaka University	Japan

Jimmy Xiangji Huang	York University	Canada
Hai Jin	Huazhong University of Science & Technology, China	China
Panos Kalnis	KAUST	Saudi Arabia
Vana Kalogeraki	Athens University of Economics and Business	Greece
George Karypis	University of Minnesota	USA
Latifur Khan	The University of Texas at Dallas	USA
Taghi Khoshgoftaar	Florida Atlantic University	USA
Irwin King	The Chinese University of Hong Kong	Hong Kong
Dongwon Lee	Penn State	USA
Guoliang Li	Tsinghua University	China
Ee-Peng LIM	Singapore Management University	Singapore
Eric Lo	Chinese University of Hong Kong	Hong Kong
Jay Lofstead	Sandia National Laboratories	USA
Bradley Malin	Vanderbilt University	USA
Hiroshi Mamitsuka	Kyoto University / Aalto University	Japan / Finland
Wagner Meira	Universidade Federal de Minas Gerais	Brazil
John Miller	University of Georgia	USA
Shinichi Morishita	University of Tokyo	Japan
Dimitrios Nikolopoulos	Queen's University Belfast	UK
Beng Chin Ooi	National University of Singapore	Singapore
M. Tamer Ozsu	University of Waterloo	Canada
Balaji Palanisamy	University of Pittsburgh	US
Barbara Pernici	Politecnico Milano	Italy
Evaggelia Pitoura	University of Ioannina	Greece
Lakshmish Ramaswamy	The University of Georgia	USA
Huzefa Rangwala	GEORGE MASON UNIVERSITY	USA
Berthold Reinwald	IBM Research - Almaden	USA
Rizos Sakellariou	University of Manchester	UK
Pierangela Samarati	Universita` degli Studi di Milano, Italy	Italy
Amit Sheth	Knoesis Center	USA
Kyuseok Shim	Seoul National University	Korea, Republic of
Alkis Simitsis	HPE Labs	USA
Domenico Talia	University of Calabria	Italy
Jie Tang	Tsinghua University	China
Hanghang Tong	Arizona State University	USA
Masashi Toyoda	Institute of Industrial Science, the University of Tokyo	Japan

Hong-Linh Truong	TU Wien	Austria
Vincent S. Tseng	National Chiao Tung University	Taiwan
Vassilis J. Tsotras	University of California, Riverside	USA
Fei Wang	Cornell Medical School	USA
Ke Wang	Simon Fraser University	Canada
Liqiang Wang	University of Central Florida	USA
Wei Wang	University of California, Los Angeles, USA	USA
Ji-Rong Wen	Renmin University of China	China
Feng Xia	Dalian University of Technology	China
Xing Xie	Microsoft Research Asia	China
Guansong Xu	UTS	Australia
Jianliang Xu	Hong Kong Baptist University	China
Jian Yang	Macquarie University	Australia
Jianwei Yin	Zhejiang University	China
Philip S. Yu	University of Illinois at Chicago	USA
Xin Yuan	Florida State University	USA
Jiayu Zhou	Michigan state university	USA
Hill Zhu	Florida Atlantic University	USA
Yanmin Zhu	Shanghai Jiao Tong University, China	China

Main Conference PC members

Name	Organization	Country
James Abello	DIMACS/Rutgers, USA	USA
Ankit Agrawal	Northwestern University	USA
Gail-Joon Ahn	Arizona State University	USA
Walid Aref	Purdue University	USA
Chris Argenta	Applied Research Associates, Inc.	USA
Antonio Badia	University of Louisville	USA
Nathalie Baracaldo	IBM Almaden Research, USA	USA
Gustavo Batista	Universidade de Sao Paulo	Brazil
Roberto J. Bayardo	Google, USA	USA
David Belanger	Stevens Institute of Technology	USA
Martin Berzins	University of Utah	USA
Larbi Boubchir	University of Paris 8	France
Mario Bravetti	University of Bologna	Italy
Hoang Bui	Western Illinois University	USA

Ali R. Butt	Virginia Tech	USA
Suren Byna	Lawrence Berkeley National Lab	USA
Jiannong Cao	The Hong Kong Polytechnic University	Hong Kong
Lei Cao	MIT	USA
Cinzia Cappiello	Politecnico di Milano	Italy
Cornelia Caragea	Kansas State University	USA
David Carrera	Technical University of Catalonia, USA	USA
Paolo Ceravolo	Universita', degli Studi di Milano	Italy
Chengliang Chai	Tsinghua University	China
Abhishek Chandra	University of Minnesota	USA
Lijun Chang	The University of Sydney	Australia
Lin-Ching Chang	Dept. of Electrical Engineering and Computer Science, the Catholic University of America	USA
Rong Chang	IBM T.J. Watson Research Center, USA	USA
Dong Chen	Florida International University	USA
Feng Chen	Computer Science Department, University at Albany – SUNY	USA
Haopeng Chen	Shanghai Jiao Tong University	China
Muhao Chen	University of California, Los Angeles	USA
Shiping Chen	CSIRO, Australia	Australia
Yong Chen	Texas Tech University	USA
Yueguo Chen	Renmin University of China, Key Laboratory of Data Engineering and Knowledge Engineering	China
Zehua Chen	Taiyuan University of Technology	China
Zhengzhang Chen	NEC Laboratories America	USA
Zhiyuan Chen	University of Maryland Baltimore County	USA
Malolan Chetlur	IBM India,	India
Kenneth Chiu	Binghamton University	USA
Wonik Choi	Inha University	Korea
Michele Ciavotta	Universit à Milano Bicocca	Italia
Andrea Clematis	IMATI - CNR	Italy
Pietro Colombo	University of Insubria	Italy
Ayse Coskun	Boston University	USA
Alexandru Costan	INRIA	France
Daniel Crichton	Jet Propulsion Laboratory, Caltech	USA
Edward Curry	Insight Centre for Data Analytics at National University of Ireland, Galway	Ireland
Brian D. Davison	Lehigh University	USA
Miyuru Dayarathna	WSO2 Inc.	USA

Noel De Palma	University Joseph Fourier	France
Eduard Deagut	Temple University	USA
Boris Delibasic	University of Belgrade	Serbia
Alex Delis	Univ. of Athens and NYU Abu Dhabi	Greece
Tyler Derr	MSU	USA
Marios Dikaiakos	University of Cyprus	Cyprus
Junhua Ding	University of North Texas	USA
Wanying Ding	VIPSHOP(US) Inc.	USA
Xiaoning Ding	New Jersey Inst. of Technology	USA
Nemanja Djuric	Uber ATG	USA
Matthieu Dorier	Argonne National Laboratory	USA
Zhicheng Dou	Renmin University of China	China
Boxin Du	ASU	USA
Nick Duffield	Texas A&M University	USA
Dmitry Duplyakin	University of Utah	USA
Roee Ebenstein	Google	USA
Magdalini Eirinaki	San Jose State University	USA
Toshio Endo	Tokyo Institute of Technology	Japan
Miki Enoki	IBM Research - Tokyo	Japan
Ju Fan	Renmin University of China	China
Yi Fang	Santa Clara University	USA
Jose Fortes	University of Florida	USA
Dmitriy Fradkin	Siemens	USA
Vanessa Frias-Martinez	University of Maryland	USA
Ada Wai-Chee Fu	The Hong Kong University of Science and Technology	Hong Kong
Yun Fu	Northeastern University	China
Paul Suganthan G C	Google	USA
Jacob Gao	Pinterest	USA
Hancheng Ge	Amazon	USA
Felix Gessert	University of Hamburg	Germany
Mohamed Ghalwash	IBM T.J. Watson Research Center	USA
Gabriele Gianini	Universita degli Studi di Milano, Italy and EBTIC at Khalifa University, Abu Dhabi	UAE
Harald Gjermundrod	University of Nicosia	Cyprus
Anastasios Gounaris	Aristotle University of Thessaloniki	Greece
Jane Greenberg	Drexel University	USA
Paul Grefen	Eindhoven University of Technology Eindhoven University of Technology	Netherlands

Clemens Grelck	University of Amsterdam	Netherlands
Le Gruenwald	University of Oklahoma	USA
Michael Gubanov	Florida State University	USA
Jayant Gupchup	Microsoft	USA
Amarnath Gupta	San Diego Supercomputing Center	USA
Vijay K. Gurbani	Illinois Institute of Technology	USA
Jialong Han	Tencent AI Lab	China
Masatoshi Hanai	Nanyang Technological University	Singapore
Mohammad Hasan	Indiana University Purdue University Indianapolis	USA
Daqing He	University of Pittsburgh	USA
Keijo Heljanko	Aalto University	Finland
Nguyen Ho	Department of Computer Science, Aalborg University	Denmark
Hiroshi Horii	IBM Research - Tokyo	Japan
Vivian Hu	Ryserson University/University of Calgary	Canada
Xia Hu	Texas A&M University	USA
Chao Huang	ND	USA
Fabrice Huet	INRIA-I3S-CNRS	France
Marty Humphrey	University of Virginia	USA
Hiroshi Inoue	IBM Research - Tokyo	Japan
Kazuaki Ishizaki	IBM Research	Japan
Arash Jalal Zadeh Fard	Microfocus - Vertica	USA
Saltz Jeffrey	Syracuse University	USA
Ruoming Jin	Kent State University	USA
Milos Jovanovic	University of Belgrade	Serbia
Alexander Jung	Aalto University	Finland
David Kaeli	Northeastern University	USA
Zoi Kaoudi	Qatar Computing Research Institute	Qatar
Hamid Karimi	Michigan State University	USA
Kiyokuni Kawachiya	IBM Research - Tokyo	Japan
Hideyuki Kawashima	Keio University	Japan
Yiping Ke	Nanyang Technological University	Singapore
Vlado Keselj	Dalhausie University	Canada
Jinha Kim	Oracle Labs	USA
Xiangjie Kong	Dalian University of Technology	China
Amey Kulkarni	FPGA Engineer, DSP and Computer Vision Focus, Velodyne LiDAR, Inc.	USA
Eren Kursun	Columbia University	USA

Alberto Laender	UFMG & Kunumi	Brazil
Zhiling Lan	Illinois Institute of Technology	USA
Jack Lange	University of Pittsburgh	USA
Alexey Lastovetsky	University College Dublin	Ireland
Alexander Lazovik	University of Groningen	Netherlands
Jinsoo Lee	Oracle Labs	USA
Wang-Chien Lee	The Pennsylvania State University	USA
Chengkai Li	University of Texas at Arlington	USA
Jundong Li	Arizona State University	USA
Min Li	JD.COM	USA
Pan Li	Case Western Reserve University	US
Sheng Li	Adobe Research	USA
Xue Li	School of Information Technology and Electrical Engineering, The University of Queensland	Australia
Yandong Li	university of central florida	USA
Zhanhuai Li	Polytechnical University, Xian	China
Zhixu Li	Soochow University	China
Zhoujun Li	BAUU	China
Defu Lian	University of Electronic Science and Technology of China	China
Hongwei Liang	Simon Fraser University	Canada
Lipyeow Lim	University of Hawaii at Manoa	USA
Kaixiang Lin	Michigan State University	USA
An Liu	Soochow University	China
Chengfei Liu	Swinburne University of Technology	Australia
Haishan Liu	Tencent	China
Hongfu Liu	Northeastern University	China
Yang Liu	Wilfrid Laurier University	Canada
Zhining Liu	University of Electronic Science and Technology of China	China
Zitao Liu	TAL AI Lab	USA
Cheng Long	Queen's University Belfast	UK
Shiyong Lu	Wayne State University	USA
Wei Lu	Renmin University	China
feng luo	Clemson University	USA
Qiong Luo	Hong Kong University of Science and Technology	Hong Kong
Kwan-Liu Ma	University of California, Davis	USA
Yao Ma	Michigan State University	USA
Sanjay Madria	Department of Computer Science, Missouri University Science and Technology	USA

Tiziana Margaria	University of Limerick and Lero - The Irish Software Research Centre	Ireland
Fabrizio Marozzo	Universit à della Calabria	Italy
Mohammad Masud	United Arab Emirates University	USA
George Mathew	MIT Lincoln Laboratory	USA
Satoshi Matsuoka	Riken Center for Computational Science / Tokyo Tech.	Japan
Edgar Meij	University of Amsterdam	Netherlands
Christoph Meinel	Hasso-Plattner-Institute, Germany	Germany
Wagner Meira, Jr.	UFMG	Brazil
Ningfang Mi	Northeastern Universty	USA
Taneli Mielikainen	Oath	USA
Alessandro Moschitti	University of Trento	Italy
Sebastien Mosser	Universit? Nice-Sophia Antipolis	France
Abdullah Mueen	Microsoft Research	USA
Aibek Musaev	University of Alabama	USA
Azad Naik	Microsoft	USA
Hidemoto Nakada	National Institute of Advanced Industrial Science and Technology (AIST)	Japan
Surya Nepal	CSIRO	Australia
Jingchao Ni	Penn State University	USA
Jian-Yun Nie	Univeristy of Montreal	Canada
Nikola S. Nikolov	University of Limerick	Ireland
Alexandros Ntoulas	LinkedIn	USA
Dana Petcu	West University of Timisoara, Romania	Romania
Sumit Purohit	Pacific Northwest National Laboratory	USA
Baojun Qiu	Chaoda Foodmall Group	China
Judy Qiu	School of Informatics and Computing, Indiana University	USA
Shannon Quinn	University of Georgia	USA
Christoph Quix	Fraunhofer FIT	Germany
Vladan Radosavljevic	Uber Advanced Technology Group	USA
Jan Ramon	INRIA Lille	France
Stephan Reiff-Marganiec	University of Leicester, UK	UK
Abdelmounaam Rezgui	New Mexico Tech	USA
Philip Rhodes	University of Mississippi	USA
Uwe Roehm	The University of Sydney	Australia
Aki-Hiro Sato	Kyoto University	Japan
Bruno Schulze	National Lab. for Scientific Computing, Brazil	Brazil
Matthias Schunter	Intel	USA

Yingxia Shao	Peking University	China
Saeedeh Shekarpour	knoesis center	USA
Haiying Shen	Clemson University	USA
Xiang Sheng	Syracuse University	USA
Conglei Shi	Airbnb	USA
Weidong Shi	University of Houston	USA
Lidan Shou	Zhejiang University, China	China
Fengguang Song	Indiana University-Purdue University Indianapolis	USA
Guojie Song	Peking University	China
Shaoxu Song	Tsinghua University	China
Torsten Suel	Polytechnic Institute of New York University, USA	USA
Aixin Sun	Nanyang Technological University, Singapore	Singapore
Guangzhong Sun	University of Science and Technology of China	China
Hailong Sun	Beihang University	China
Ichigaku Takigawa	Hokkaido University	Japan
Douglas Talbert	Tennessee Technological University	USA
Pang-Ning Tan	Michigan State University, USA	USA
Gabriel Tanase	Graphen Inc.	USA
Shanjiang Tang	Tianjing University	China
Yusuke TANIMURA	National Institute of Advanced Industrial Science and Technology and University of Tsukuba	Japan
Dimitrios Tsoumakos	Deparment of Informatics, Ionian University	Greece
Mauricio Tsugawa	Microsoft	USA
Guan-hua Tu	Michigan State University	USA
Takanori Ueda	IBM Research - Tokyo	Japan
Takeaki Uno	National Institute of Informatics	Japan
Ana Lucia Varbanescu	University of Amsterdam	Netherlands
Maksims Volkovs	Layer6 AI	Canada
Slobodan Vucetic	Temple University	USA
Milan Vukicevic	University of Belgrade	Serbia
Thomas Walsh	Kronos Inc.	USA
Jianguo Wang	University of California, San Diego	USA
Jianwu Wang	University of Maryland, Baltimore County	USA
Senzhang Wang	Nanjing University of Aeronautics and Astronautics	China
Suhang Wang	PSU	USA
Ting Wang	http://x-machine.github.io/	USA
Xiting Wang	Microsoft Research Asia	China

Zhiwei Wang	Michigan State University	USA
Xiaokai Wei	Facebook	USA
Zhi Wei	NJIT	USA
Ran Wolff	Yahoo Research	Israel
Ka-Chun Wong	City University of Hong Kong	Hong Kong
Raymond Wong	University of New South Wales	Australia
Stefan Wrobel	University of Bonn, Germany	Germany
Le Wu	Hefei University of Technology	China
Liang Wu	Arizona State University	USA
Lingfei Wu	IBM T.J. Watson Research Center, USA	USA
Sai Wu	Zhejiang University	China
Houping Xiao	GSU	USA
Hongbo Xu	Chinese Academy of Sciences, China	China
Jian Xu	Alibaba Group	USA
Jianpeng Xu	Michigan state university	USA
Jun Xu	ICT, Chinese Academy of Science	China
Weijia Xu	TACC	USA
Wenjian Xu	Hong Kong Polytechnic University	Hong Kong
Yuhong Yan	Concordia University, Canada	Canada
Haiqin Yang	Hang Seng Management College	Hong Kong
Xiaoyan Yang	Yitu Singapore	Singapore
Yu Yang	Simon Fraser University	Canada
Zhi Yang	Peking University	China
Junjie Yao	East China Normal University	China
Yuan Yao	Nanjing University	China
Peifeng Yin	IBM	USA
Yiming Ying	The University at Albany, State University of New York	USA
Qi Yu	Rochester Institute of Technology	USA
Weikuan Yu	Florida State University	USA
Carlo Zaniolo	UCLA	USA
Chunqiu Zeng	Google Inc.	USA
Chengxiang Zhai	University of Illinois at Urbana-Champaign	USA
Allan Nengsheng Zhang	Singapore Institute of Manufacturing Technology	Singapore
Chao Zhang	UIUC	USA
Da Zhang	University of Miami	USA
Fuzheng Zhang	Microsoft Research Asia	China
Kunpeng Zhang	University of Maryland, College Park	USA

Meihui Zhang	Beijing Institute of Technology	China
Min Zhang	Tsinghua University, China	China
Rui Zhang	IBM Research - Almaden	USA
Si Zhang	ASU	USA
Wenjie Zhang	University of New South Wales	Wales
Xiangliang Zhang	King Abdullah University of Science and Technology	Saudi Arabia
Ya Zhang	Shanghai Jiao Tong University	China
Jessie Zhao	York University	Canada
Juan Zhao	Vanderbilt University	USA
Ming Zhao	Arizona State University	USA
Weiliang Zhao	Macquarie University	Australia
Xiangyu Zhao	Michigan State University	USA
Zhi-Dan Zhao	University of Electronic Science and Technology of China, Chengdu 610051, China	China
Lecheng Zheng	Arizona State University	USA
Yudian Zheng	Twitter Inc.	US
Amelie Chi Zhou	Shenzhen University	China
Dawei Zhou	ASU	USA
Fang Zhou	Temple University	USA
Guangyou Zhou	Central China Normal University	China
Nianjun (Joe) Zhou	IBM T.J. Watson Research Center, USA	USA
Yang Zhou	Auburn University	USA
Yao Zhou	ASU	USA
Yongluan Zhou	University of Copenhagen	Denmark
Yanmin Zhu	Shanghai Jiao Tong University, China	China

Industry and Governement Program PC members

Name	Organization	Country
Khalifeh Aljadda	Home Depot	USA
Mansurul Bhuiyan	Walmart Labs	USA
Bin Bi	Microsoft Research	USA
Cheng Bo	JD.COM American Technologies	USA
Mihai CapotÄ*f*	Intel Labs, Intel Corporation	USA
Mehdi Dadfarnia	NIST	USA
Vachik Dave	Indiana University Purdue University Indianapolis	USA
Sameh Elnikety	Microsoft Research	USA

GermÃ n Flores	IBM Research - Almaden	USA
Vijay Gadepally	Massachusetts Institute of Technology	USA
Assefaw Gebremedhin	Washington State University	USA
Simon-Pierre Genot	BayWa AG	Germany
Raghav Gupta	Google Research	USA
Mahantesh Halappanavar	Pacific Northwest National Laboratory	USA
Bo Hu	LinkedIn	USA
Ziang Hu	Huawei Technologies	USA
Lei Huang	IBM Research - Almaden	USA
Divyesh Jadav	IBM Research - Almaden	USA
Nilesh Jain	Intel Corp	USA
Madian Khabsa	Apple	USA
Christine Klymko	Lawrence Livermore National Laboratory	USA
Eunyee Koh	Adobe Research	USA
Mohammed Korayem	CareerBuilder	USA
Tim La Fond	Lawrence Livermore National Laboratory	USA
John Boaz Lee	Worcester Polytechnic Institute	USA
Min Li	JD.COM	USA
Yuan Ling	Amazon Alexa	USA
Haibin Liu	IBM Watson	USA
Mengwen Liu	Drexel University	USA
Guixiang Ma	uic	USA
Anup Rao	Adobe Research	USA
Abhinav Rastogi	Google AI	USA
Ryan Rossi	Adobe Research	USA
Sherif Sakr	University of New South Wales	Australia
Pararth Shah	Google Research	USA
Xiaolin Shi	Snap Inc	USA
Koichi Shirahata	Fujitsu	Japan
Shaden Smith	Intel Labs	USA
Srivathsan Srinivasagopalan	Visa	USA
Antonino Tumeo	Pacific Northwest National Laboratory	USA
Yida Wang	Amazon	USA
Lijie Wen	Tsinghua University	China
Theodore Willke	Intel Labs	USA
Michael Wolf	Sandia National Labs	USA

Zhaohui Wu	Microsoft	USA
Yinglong Xia	Huawei Research America	USA
Jianbo Ye	Amazon Lab126	USA
Mo Yu	Google	USA
Bo Zhang	IBM Watson & Cloud Platform	USA
Nairan Zhang	Facebook	USA
Haozhen Zhao	Ankura	USA
Rong Zhou	Google	USA
Shucheng Zhu	Google	USA
Xia Zhu	Parallel Computing Lab, Intel Corporation	USA

IEEE Big Data 2018 Program Schedule

Seattle, WA, USA
December 10 - December 13, 2018

Keynote Lecture: **60 minutes** (about 45 minutes for talk and 15 minutes for Q and A)
Main conference regular paper: **25 minutes** (about 20 minutes for talk and 5 minutes for Q and A)
Main conference short paper: **15 minutes** (about 11 minutes for talk and 4 minutes for Q and A)

All conference activities take place at the Westin Seattle located at 1900 5th Avenue. Seattle, WA.

Sunday, December 9, 2018	
3:00pm – 8:00 pm Location:	**Registration** GRAND REGISTRATION *Floor 4*

Day 1: Monday, December 10, 2018	
7:20am-6:00 pm Location:	**Registration** GRAND REGISTRATION *Floor 4*
10:00-10:20 am and 3:30 – 3:50 pm Location:	**Coffee Break** GRAND FOYER *Floor 4*
2:00 – 6:00 pm Location:	**Poster Session (Set up only)** FIFTH AVENUE ROOM *Floor 4*

Time	Workshops	Session Chair	Location
Whole Day	The 5th Workshop on Big Data Analytic Technology for Bioinformatics and Health Informatics (KDDBHI)	Donghui Wu and Xin Deng	CASCADE II *Floor 2*
Whole Day	The 2nd IEEE Workshop on Human-in-the-loop Methods and Human-Machine Collaboration in Big Data (HMData 2018)	, Senjuti Basu Roy, Lei Chen, Atsuyuki Morishima	PIKE *Lower Floor*
Whole Day	The 2nd Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD)	Zhiyuan Chen, Jianwu Wang, Feng Chen, and Yiming Ying	CASCADE I-B/C *Floor 2*
Whole Day	The Third Workshop on Application of Big Data for computational social science	Akira Ishii, Fujio Toriumi, Hiroki Takikawa, and Mitsuo Yoshida	OLYMPIC *Floor 2*
Whole Day	5th National Symposium for NSF REU Research in Data Science, Systems, and Security	Mohamamd Al Hasan,	PUGET SOUND ROOM *Floor 1*
Whole Day	BDMM workshop/hackathon	Wo Chang	ELLIOTT BAY *Floor 1*
Whole Day	3rd Workshop on Real-time and Stream Analytics in Big Data & Stream Data Management	Sabri SKHIRI	CASCADE I-A *Floor 2*

Whole Day	Symposium on Benchmarking, Measuring and Optimizing (Bench' 18)	Jianfeng Zhan	DENNY / MERCER *Floow Lower*
Whole Day	Big Social Media Data Management and Analysis	Xin Huang	ORCAS *Floor 3*
AM	The 2nd International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID 2018)	Huaglory Tianfield and Stephen McGough	VASHON I *Floor 3*
AM	The 2nd International Workshop on Big Data Analytic for Cybercrime Investigation and Prevention	Andrii Shalaginov, Katrin Franke, and Jan William Johnsen	VASHON II *Floor 3*
AM	Fifth International Workshop on High Performance Big Graph Data Management, Analysis, and Mining (BigGraphs 2018)	Mohammad Al Hasan	BLAKELY *Floor 3*
AM	The Second workshop on Big Data for Economic and Business Forecasting	Wei Shang, Xiangbin Yan,	ST. HELENS *Floor 2*
AM	The 2nd International Workshop on Big Data for Financial News and Data	Quanzhi Li	BAKER *Floor 2*
AM	Workshop on Energy-Efficient Big Data Analytics	Mohammed Alawad	ADAMS *Floor2*
AM	3rd International Workshop on Big Data Transfer Learning (BDTL)	Ming Shao, Tongliang Liu, and Zhengming Ding	WHIDBEY *Floor 3*
AM	BigData Cup Challenges: Road Damage detection and Classfication Challenges	Hiroya Maeda, Yoshihide Sekimoto, Toshikazu Seto, Takehiro Kashiyama, Hiroshi Omata	PINE *Lower Floor*
8:00-10:00	Tutorial 7: Creating Reproducible Bioinformatics Workflows Using BioDepot-workflow- Builder (BwB)	Ling-Hong Hung, Ka Yee Yeung, Wes Lloyd, Eyhab Al-Masri	GRAND I, II, III Floor 4
10:20-12:20	Tutorial 9: Big Data for everyone: Modeling of Data Processing Pipelines for the Industrial Internet of Things	Dominik Riemer, Dominik Riemer, Ljiljana Stojanovic, Philipp Zehnder	GRAND I, II, III *Floor 4*
12:00 - 1:30 pm	**Lunch (On Own)**		
PM	The 1st International Workshop on Big Video Dataset Construction, Management and Applications	Cheng Jin, Haimiao Hu, Shengcai Liao, Mingli Song, Rui Wang, and Mingyu You	GRAND CRESCENT *Floor 4*
PM	International Workshop on Conversational Agents and Chatbots with Machine Learning (ChatbotML 2018)	Huaglory Tianfield	VASHON I *Floor 3*
PM	International Workshop on Big Data Analytics for Cyber Threat Hunting (CyberHunt 2018)	Vasileios Mavroeidis	VASHON II *Floor 3*
PM	7th Workshop on Scalable Cloud Data Management	Felix Gessert and Norbert Ritter	BLAKELY *Floor 3*
PM	6th International Workshop on Distributed Storage and Blockchain Technologies for Big Data	Hui Li, Kenneth Shum, and Bing Zhu	STUART *Floor 2*
PM	4th IEEE Workshop on Big Data Analytics in Supply Chains and Transportation	Allan Nengsheng Zhang and Satish Ukkusuri	ST. HELENS *Floor 2*

PM	The 2nd workshop on Graph Techniques for Adversarial Activity Analytics (GTA3 2.0)	Jiejun Xu and Hanghang Tong	BAKER *Floor 2*
PM	Open Science in Big Data (OSBD) Workshop	Shannon Quinn	ADAMS *Floor 2*
PM	Big Data Engineering and Analytics in Cyber-Physical Systems (BigEACPS)	Akbar Siami Namin	WHIDBEY *Floor 3*
PM	Applications of Big Data in the Transport Industry	John Easton	PINE *Lower Floor*
PM	The 1st International Workshop on Big Data for Marketing Intelligence and Operation Management	Wutao Wei	GRAND I, II, III *Floor 4*

Day 2: Tuesday, December 11, 2018			
7:20-6:00 pm Location:	**Registration** GRAND REGISTRATION *Floor 4*		
8:30 am - 8:45 am Location:	**Opening and Welcome** Conference Chairs, PC Chairs, Industry and Government Program Chairs GRAND I, II, III *Floor 4*		
8:45 am - 9:45 am Location:	**Keynote Decentralized Machine Learning** **by Dr. Blaise Agüera y Arcas** Chair: Yang Song GRAND I, II, III *Floor 4*		
9:45 am - 10:05 am Location:	**Coffee Break** GRAND FOYER *Floor 4* **Poster Session (Set up)** FIFTH AVENUE ROOM *Floor 4*		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
10:05 am - 12:10 pm	L1 Scientific Data Management	Yong Chen	PIKE *Lower Floor*
	L2 Social Web Search and Mining	Weijia Xu	CASCADE II *Floor 2*
	L3 Semantic-based Data Mining	Da Zhang	PINE *Lower Floor*
	L4 Big Data Applications: Machine Learning	Alfredo Cuzzocrea	CASCADE I-B/C *Floor 2*
	I&G Regular 1: Big Data and Machine Learning (1)	IIya Safro	PUGET SOUND ROOM *Lower Floor*
	Tutorial 1: Large-Scale Multi-view Data Analysis	Ming Shao and Zhengming Ding	GRAND CRESCENT *Floor 4*
AM	Analysis of Large-scale Disparate Data	Michael Barton	CASCADE I-A *Floor 2*
Whole Day	BDMM workshop/hackathon	Wo Chang	ELLIOTT BAY *Floor 1*
Whole Day	Special Session: Intelligent Data Mining	Uraz Yavanoglu	WHIDBEY *Floor 3*
Whole Day	Symposium on Benchmarking, Measuring and Optimizing (Bench' 18)	Jianfeng Zhan	DENNY / MERCER *Lower Floor*
Whole Day	First International Workshop on the Internet of Things Data Analytics (IOTDA)	Eyhab Al-Masri, and Yan Bai	ST. HELENS *Floor 2*
12:10 pm - 1:30 pm Location:	**Lunch (provided by Conference)**		

	Lunch Talk: Dan Friedman, Expedia Data Science, (12:30-1:00) GRAND I, II, III *Floor 4*		
1:30 pm - 2:30 pm Location:	**Keynote: Three principles of data science: predictability, computability, and stability (PCS)** **by Prof. Bin Yu** Chair: Naiko Abe GRAND I, II & III *Floor 4*		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
2:30 pm - 4:10 pm	L5 Novel Theoretical Models for Big Data	Noseong Park	PIKE *Lower Floor*
	L6 Big Data Analytics Frameworks	Dirk Van den Poel	CASCADE II *Floor 2*
	L7 Software Systems to Support Big Data Computing	Patrick Koch	PINE *Lower Floor*
	L8 Threat Detection using Big Data Analytics	Hamid Karimi	CASCADE I-B/C *Floor 2*
	I&G Regular 2: Big Data Applications	Ryan Rossi	PUGET SOUND ROOM *Floor 1*
Whole Day	BDMM workshop/hackathon	Wo Chang	ELLIOTT BAY *Floor 1*
Whole Day	Special Session: Intelligent Data Mining	Uraz Yavanoglu	WHIDBEY *Floor 3*
Whole Day	Symposium on Benchmarking, Measuring and Optimizing (Bench' 18)	Jianfeng Zhan	DENNY / MERCER *Lower Floor*
Whole Day	First International Workshop on the Internet of Things Data Analytics (IOTDA)	Eyhab Al-Masri, and Yan Bai	ST. HELENS *Floor 2*
PM	4th International Workshop on Methodologies to Improve Big Data projects	Jeffrey Saltz	GRAND CRESCENT *Floor 4*
PM	Big Data for Digital Twins	Arne Berre, Ljiljana Stojanovic, and Nenad Stojanovic	BLAKELY *Floor 3*
PM	Special session: Information Granulation in Data Science and Scalable Computing	Shusaku Tsumoto, Dominik Slezak, Tzung-Pei Hong, S. L. Wang	ADAMS *Floor 2*
PM	Special Session: HealthCare Data	Ozgun Pinarer, Sultan Turhan	OLYMPIC *Floor 2*
PM	The Second Annual Workshop on Big Data Analytics in the Legal Industry	Jianping Zhang, Nathaniel Huber-Fliflet, Haozhen Zhao	CASCADE I-A *Floor 2*
PM	Advances in High Dimensional (AdHD) Big Data	Sotiris Tasoulis	STUART *Floor 2*
4:10 pm - 4:30 pm Location:	**Coffee Break** GRAND FOYER *Floor 4*		

	Poster Session Sets Up and Displays FIFTH AVENUE ROOM *Floor 4*		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
4:30 pm - 6:30 pm	S1 Big Data Infrastructure (1)	Hsing Bung Chen	PIKE *Lower Floor*
	S2 Big Data Applications (1)	Xiangliang Zhang	CASCADE II *Floor 2*
	S3 Spatio-temporal and Stream Data Management	Jiliang Tang	PINE *Floor lowe*
	S4 Big Data Security and Privacy	Huan Liu	CASCADE I-B/C *Floor 2*
	I&G Short 1: Big Data Algorithms & Systems (1)	Shaikh Arifuzzaman	PUGET SOUND ROOM *Floor 1*
	Tutorial 6: Anomaly Detection in Cyber Physical Systems	Maggie Cheng	VASHON I & II *Floor 3*
Whole Day	BDMM workshop/hackathon	Wo Chang	ELLIOTT BAY *Floor 1*
Whole Day	Special Session: Intelligent Data Mining	Uraz Yavanoglu	WHIDBEY & ORCAS *Floor 3*
Whole Day	Symposium on Benchmarking, Measuring and Optimizing (Bench' 18)	Jianfeng Zhan	DENNY / MERCER *Floow Lower*
PM	4th International Workshop on Methodologies to Improve Big Data projects	Jeffrey Saltz	GRAND CRESCENT *Floor 4*
PM	First International Workshop on the Internet of Things Data Analytics (IOTDA)	Eyhab Al-Masri, and Yan Bai	ST. HELENS *Floor 2*
PM	Special session: Information Granulation in Data Science and Scalable Computing	Shusaku Tsumoto, Dominik Slezak, Tzung-Pei Hong, S. L. Wang	ADAMS *Floor 2*
PM	Special Session: HealthCare Data	Ozgun Pinarer, Sultan Turhan	OLYMPIC *Floor 2*
PM	The Second Annual Workshop on Big Data Analytics in the Legal Industry	Jianping Zhang, Nathaniel Huber-Fliflet, Haozhen Zhao	CASCADE I-A *Floor 2*
PM	Advances in High Dimensional (AdHD) Big Data	Sotiris Tasoulis	STUART *Floor 2*

Day 3: Wednesday, December 12, 2018			
7:30am-6:00 pm Location:	**Registration** GRAND REGISTRATION *Floor 4*		
8:40 am - 8:45 am Location:	**Opening** GRAND I, II, III *Floor 4*		
8:45 am - 9:45 am Location:	**Keynote: Transformational Role of Big Data in Society 5.0** **by Prof. Masaru Kitsuregawa** Chair: Calton Pu GRAND I, II, III *Floor 4*		
9:45 am - 10:05 am Location:	**Coffee Break** GRAND FOYER *Floor 4* **Poster Session Displays** FIFTH AVENUE ROOM *Floor 4*		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
10:05 am - 12:10 pm	L9 Recommendation Systems and Stream Data Management	Wanying Ding	PINE *Lower Floor*
	L10 Link and Graph Mining	Alfredo Cuzzocrea	CASCADE II *Floor 2*
	L11 Big Data Applications: Industry and Business	Hancheng Ge	PIKE *Lower Floor*
	L12 Big Data Applications: Health and Science Discovery	João Eduardo Ferreira	CASCADE I-B/C *Floor 2*
	I&G Regular 3: Big Data Platforms & Frameworks	Abdeltawab Hendawi	PUGET SOUND ROOM *Floor 1*
	Tutorial 2: Analysis of Complex Rare Categories	Dawei Zhou, Jingrui He	GRAND CRESCENT *Floor 4*
Whole Day	Workshop: Computational Archival Science	Mark Hedges, Richard Marciano, and Victoria Lemieux	ELLIOTT BAY *Floor 1*
Whole Day	Symposium on Benchmarking, Measuring and Optimizing (Bench' 18)	Jianfeng Zhan	DENNY / MERCER *Floow Lower*
Whole Day	Bigdata Cup Challenges: FEMH Voice Data Challenge	Yu Tsao	CASCADE I-A *Floor 2*
AM	Big Data and AI for Air Quality Estimation, Forecasting, and Health Advice	Victor OK LiVictor OK Li, Jacqueline CK Lam, Mihaela van der Schaar, and Ingmar Cox	ST. HELENS *Floor 2*
12:10 pm - 1:30 pm Location:	**Lunch (provided by Conference)** **Lunch Talk: AutoDL: Automated Deep Learning for Open and Inclusive AI, Dr. Jun Huan, Baidu Big Data Lab (12:30-1pm)**		

	GRAND I, II, III *Floor 4*		
1:30 pm - 2:30 pm Location:	**Keynote: On Metric Learning for Complex Data Analysis** **by Prof. Aidong Zhang** Chair: Bing Liu GRAND I, II, III *Floor 4*		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
2:30 pm - 4:10 pm	L13 New Computational Models for Big Data	Da Zhang	PINE *Lower Floor*
	L14 HPC Platforms for Big Data	Sumit Purohit	CASCADE II *Floor 2*
	L15 Big Data Search (1)	Ben Harris	PIKE *Lower Floor*
	L16 Web Data and Crowdsourcing	Yang Zhou	CASCADE I-B/C *Floor 2*
	I&G Regular 4: Big Data and Machine Learning (2)	Shreyansh Gandi	PUGET SOUND ROOM *Floor 1*
Whole day	Workshop: Computational Archival Science	Mark Hedges, Richard Marciano, and Victoria Lemieux	ELLIOTT BAY *Floor 1*
Whole Day	Symposium on Benchmarking, Measuring and Optimizing (Bench' 18)	Jianfeng Zhan	DENNY / MERCER *Floow Lower*
Whole Day	Bigdata Cup Challenges: FEMH Voice Data Challenge	Yu Tsao	CASCADE I-A *Floor 2*
PM	5th Workshop on Advances in Software and Hardware for Big Data Science (ASH)	Hui Zhang, Weijia Xu, and Hongfeng Yu	ST. HELENS *Floor 2*
PM	The 5th International Workshop on Privacy and Security of Big Data - PSBD	Alfredo Cuzzocrea	OLYMPIC *Floor 2*
4:10 pm - 4:30 pm Location:	**Coffee Break** GRAND FOYER *Floor 4* **Poster Session Displays** *FIFTH AVENUE ROOM* *Floor 4*		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
4:30 pm - 6:30 pm	S5 Big Data Infrastructure (2)	Xintao Wu	PINE *Lower Floor*
	S6 Big Data Applications (2)	Noseong Park	CASCADE II *Floor 2*
	S7 Big Data Management	Qing Wang	PIKE *Lower Floor*
	S8 Link and Graph Mining	Da Zhang	CASCADE I-B/C

			Floor 2
	I&G Short 2: Massive Processing and Experience	Miroslav Hodak	PUGET SOUND ROOM *Floor 1*
	Tutorial 5: Big Data Analytics for Societal Event Forecasting	Liang Zhao, Liang Zhao	GRAND CRESCENT *Floor 4*
Whole day	Workshop: Computational Archival Science	Mark Hedges, Richard Marciano, and Victoria Lemieux	ELLIOTT BAY *Floor 1*
Whole Day	Symposium on Benchmarking, Measuring and Optimizing (Bench' 18)	Jianfeng Zhan	DENNY / MERCER *Floow Lower*
Whole Day	Bigdata Cup Challenges: FEMH Voice Data Challenge	Yu Tsao	CASCADE I-A *Floor 2*
PM	5th Workshop on Advances in Software and Hardware for Big Data Science (ASH)	Hui Zhang, Weijia Xu, and Hongfeng Yu	ST. HELENS *Floor 2*
PM	The 5th International Workshop on Privacy and Security of Big Data - PSBD	Alfredo Cuzzocrea	OLYMPIC *Floor 2*
7:00 – 9:00 pm Location	**Banquet (Ticket required)** *Chair:* Conference Chairs, PC Co-chairs, I&G PC Co-chairs, 1. *Best Paper Award, Best Application Paper Award,* PC Co-chairs 2. *Best Industry and Government Application Paper,* I&G PC Co-chairs 3. *Hackathon Awards:* Wo Chan 4. *Big Data Cup Challenges Awards:* Hiroya Maeda, Yu Tsao GRAND I, II, III *Floor 4*		

Day 4: Thursday, December 13, 2018

07:30-6:00pm Location:	**Registration** GRAND REGISTRATION *Floor 4*		
8:30 am - 8:45 am Location:	**Opening** GRAND I, II, III *Floor 4*		
8:45 am - 9:45 am Location:	**Keynote: Big Data for Speech and Language Processing** **by Dr. Xuedong Huang** Chair: Huan Liu GRAND I, II, III *Floor 4*		
9:45 am - 10:05 am Location:	**Coffee Break** GRAND FOYER *Floor 4* **Poster Session Displays** GRAND FOYER *Floor 4*		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
10:05 am - 12:10 pm	L17 Big Data Applications: Deep Learning	Liqiang Wang	PINE *Lower Floor*
	L18 Mobile and IoT Data	Rajeev Agrawal	CASCADE II *Floor 2*
	L19 Big Data Applications: Society	Dhruv Sharma	PIKE *Lower Floor*
	L20 Big Data Search (2)	Lingfei Wu	CASCADE I-B/C *Floor 2*
	I&G Short 3: Big Data Algorithms & Systems (2)	Haozhen Zhao	PUGET SOUND ROOM *Floor 1*
	Tutorial 4: Recent Progress in Zeroth Order Optimization and Its Applications to Adversarial Robustness in Deep Learning	Pin-Yu Chen, Sijia Liu	GRAND CRESCENT *Floor 4*
Whole day	The 3rd IEEE International Workshop on Big Spatial Data (BSD 2018)	Chengyang Zhang	ELLIOTT BAY *Floor 1*
Whole day	The Second International Workshop on Automation in Machine Learning and Big Data	Tao Wang, Patrick Koch	CASCADE I-A *Floor 2*
AM	Symposium on Benchmarking, Measuring and Optimizing (Bench' 18)	Jianfeng Zhan	DENNY / MERCER *Lower Floor*
AM	Workshop on Identifying and Combating Disinformation in Big Data	Tyler Smith and Brian Isle	OLYMPIC *Floor 2*
12:10 pm - 1:20 pm Location:	**Lunch (provided by the conference)** **Lunch Talk: Squirrel AI, the machine that regularly outperforms human teachers and redefines education, Richard Tong, Yixue Inc. (12:30-1pm)** GRAND I, II, III		

	Floor 4		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
1:20 pm - 3:05 pm	L21 Big Data Search (3)	Dinesh Singh	PINE *Floor loer*
	L22 Privacy and Security	Yanfang (Fanny) Ye	CASCADE II *Floor 2*
	S9 Big Data Foundation (1)	Tony Hu	PIKE *Lower Floor*
	S10 Experiences and Case Studies with Big Data	Jiliang Tang	CASCADE I-B/C *Floor 2*
	Tutorial 3: High-Performance SVD for big data	Andreas Stathopoulos, Eloy Romero	GRAND CRESCENT *Floor 4*
Whole day	The 3rd IEEE International Workshop on Big Spatial Data (BSD 2018)	Chengyang Zhang	ELLIOTT BAY *Floor 1*
Whole day	The Second International Workshop on Automation in Machine Learning and Big Data	Tao Wang, Patrick Koc	CASCADE I-A *Floor 2*
PM	Workshop on Big Data for CyberSecurity (BigCyber-2018)	Karuna Joshi	OLYMPIC *Floor 2*
3:05 pm - 3:25 pm Location:	**Coffee Break** GRAND FOYER		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
3:25 pm - 5:25 pm	S11 Big Data Analytics	Alfredo Cuzzocrea	*PINE* *Lower Floor*
	S12 Recommendation Systems and Stream Data Mining	Esin Saka	CASCADE II *Floor 2*
	S13 Big Data Search and Mining	Ming Shao	PIKE *Lower Floor*
	S14 Big Data Foundation (2)	Dinesh Singh	CASCADE I-B/C *Floor 2*
	Tutorial 8: Managing Big Structured Data for Unsupervised Feature Representation Learning	Lingfei Wu, Ian Yen	GRAND CRESCENT *Floor 4*
Whole day	The 3rd IEEE International Workshop on Big Spatial Data (BSD 2018)	Chengyang Zhang	ELLIOTT BAY *Floor 1*
Whoel day	The Second International Workshop on Automation in Machine Learning and Big Data	Tao Wang, Patrick Koc	CASCADE I-A *Floor 2*
PM	Workshop on Big Data for CyberSecurity (BigCyber-2018)	Karuna Joshi	OLYMPIC *Floor 2*

Keynote Lectures

Keynote: Decentralized Machine Learning

Speaker:
Blaise Agüera y Arcas, Distinguished Scientist, Google AI, USA

Abstract:
In the past decade we have seen very rapid growth in two fields: cloud services, and neural networks. These two are connected, in that logs from services are the fuel that has powered data-hungry deep learning algorithms. However, there are several forces on the other side of the coin, pushing neural capabilities onto the device and out of the cloud. These include: the development of power-efficient on-device neural processors; scaling laws relating energy density, size, and bandwidth; and an increasing demand for data privacy. This talk will address these trends, technologies designed to address them (including Federated Learning, quantization, and device-friendly architectures like MobileNet), and the product landscape emerging from these new developments.

Short Bio:
Blaise Agüera y Arcas leads a team at Google focusing on Machine Intelligence for mobile devices—including both basic research and new products. His group works extensively with deep neural nets for machine perception, distributed learning, and agents, as well as collaborating with academic institutions on connectomics research. Until 2014 he was a Distinguished Engineer at Microsoft, where he worked in a variety of roles, from inventor to strategist, and led teams with strengths in interaction design, prototyping, computer vision and machine vision, augmented reality, wearable computing and graphics. Blaise has given TED talks on Seadragon and Photosynth (2007, 2012) and Bing Maps (2010). In 2008, he was awarded MIT's prestigious TR35 ("35 under 35").

Keynote: Big Data for Speech and Language Processing

Speaker:
Xuedong Huang, Microsoft Technical Fellow of Microsoft Cloud and AI, Microsoft, USA

Abstract:
Amongst all creatures the human species stands unique in Darwin's natural selection process. It is no exaggeration that speech and language helped to differentiate human intelligence from animal intelligence in the evolution process. The impact of big data and cloud to speech and language evolution is foundational to realize the society's AI vision. This talk will review how Microsoft achieved human parity on both conversational speech recognition and news machine translation research tasks and highlight significant challenges remaining to make speech and language production services mainstream in our AI journey.

Short Bio:
Dr. Xuedong Huang is a Microsoft Technical Fellow in AI and Research. He leads Microsoft's Speech and Language Group. In 1993, Huang joined Microsoft to found the company's speech technology group. As the general manager of Microsoft's spoken language efforts, he helped to bring speech to the mass market by introducing SAPI to Windows in 1995 and Speech Server to the enterprise call center in 2004. He served as General Manager for MSR Incubation and Chief Architect for Bing and Ads. In 2015, he returned to AI and Research to lead the advanced technology group. In 2016, he led the team achieving a historical conversational speech recognition human parity milestone on the Switchboard task. He helped to advance AI across Microsoft's whole AI Stack: Solutions: AI for customer support (Project Toronto), Voice Assistant Cortana, Microsoft Translator; APIs: Cognitive Services on Azure; Engines: Speech, Machine translation, Gesture, and NLP; Deep Learning infrastructure: Cognitive Toolkit (CNTK) and GPU Cluster (Project Philly)

He was on the faculty of School of Computer Sciences at Carnegie Mellon University before joining Microsoft. He received Alan Newell research excellence leadership medal in 1992 and IEEE Best Paper Award in 1993. He was named the Asian American Engineer of the Year (2011), and Wired Magazine's 25 Geniuses (2016). He holds over 100 patents and published over 100 papers & 2 books. He received his PhD, MS, and BS in Computer Science from the University of Edinburgh, Tsinghua University, and Hunan University respectively. He has been elected fellow of IEEE and ACM.

Keynote: Transformational Role of Big Data in Society 5.0

Speaker:

Masaru Kitsuregawa, Professor and Director, University of Tokyo and National Institute of Informatics, Japan

Abstract:
Japan is launching 'Society 5.0', the vision for a future smarter society. One of the fundamental pillars of Society 5.0 is to help the society become smarter in a data-driven way. Through the advance of Internet of Things (IoT), the rapidly growing big data is substantially transforming our society, for example, through smarter commercial products and services. In this talk, we will focus on the role of big data in providing smarter services for societal benefits, with special emphases on disaster management and socialized healthcare. In accordance to Sustainable Development Goals (a United Nations initiative), our solution frameworks are being deployed both in Japan and partner developing countries.

Short Bio:
Dr. Masaru Kitsuregawa is the Director General of National Institute of Informatics (NII), and a professor at the University of Tokyo. He received his PhD degree in information engineering from the University of Tokyo in 1983. He has been working in the area of high performance database system and systems for big data. He has served as President of Information Processing Society of Japan (IPSJ) and a science advisor for Ministry of Education, Culture, Sports, Science and Technology, Japan. He is a fellow of the ACM, IEEE, IPSJ and IEICE.

Keynote: Three principles of data science: predictability, computability, and stability (PCS)

Speaker:
Bin Yu, Departments of Statistics and of Electrical Engineering & Computer Sciences, University of California at Berkeley, USA

Abstract:
In this talk, I'd like to discuss the intertwining importance and connections of three principles of data science in the title and the PCS workflow that is built on the three principles. The principles will be demonstrated in the context of two collaborative projects in neuroscience and genomics for interpretable data results and testable hypothesis generation.

Short Bio:
Dr. Bin Yu is Chancellor's Professor in the Departments of Statistics and of Electrical Engineering & Computer Sciences at the University of California at Berkeley. Her current research interests focus on statistics and machine learning theory, methodologies and algorithms for solving high-dimensional data problems. Her group is engaged in interdisciplinary research with scientists from genomics, neuroscience, and precision medicine.

She obtained her B.S. degree in Mathematics from Peking University in 1984, her M.A. and Ph.D. degrees in Statistics from the University of California at Berkeley in 1987 and 1990, respectively. She is Member of the U.S. National Academy of Sciences and Fellow of the American Academy of Arts and Sciences. She was a Guggenheim Fellow in 2006, and the Tukey Memorial Lecturer of the Bernoulli Society in 2012. She was President of IMS (Institute of Mathematical Statistics) in 2013-2014 and the Rietz Lecturer of IMS in 2016. She received the E. L. Scott Award from COPSS (Committee of Presidents of Statistical Societies) in 2018.

Keynote: On Metric Learning for Complex Data Analysis

Speaker:
Aidong Zhang, SUNY Distinguished Professor and Program Director, State University of New York at Buffalo and National Science Foundation, USA

Abstract:
Comparing and measuring similarities or distances between pairs of instances is a basic but important step toward successes of many data mining and machine learning approaches. In this talk, I will discuss how both linear and nonlinear metric learning can be approached to capture various important relationships for complex data sets and how the learned metrics can be used for complex data analysis.

Short Bio:
Dr. Aidong Zhang is a SUNY Distinguished Professor of Computer Science and Engineering at the State University of New York (SUNY) at Buffalo where she served as the Department Chair from 2009 to 2015 and has also held adjunct professor positions in both Biomedical Engineering and Biomedical Informatics Departments. She is currently on leave and serving as a Program Director in the Information & Intelligent Systems Division of the Directorate for Computer & Information Science & Engineering, at the National Science Foundation. Her research interests include data mining/data science, machine learning, bioinformatics, and health informatics. She has authored over 300 research publications in these areas. Dr. Zhang currently serves as the Editor-in-Chief of the IEEE Transactions on Computational Biology and Bioinformatics (TCBB). She served as the founding Chair of ACM Special Interest Group on Bioinformatics, Computational Biology and Biomedical Informatics during 2011-2015 and is currently the Chair of its advisory board. She is also the founding and steering chair of ACM international conference on Bioinformatics, Computational Biology and Health Informatics. She has served as editors for several other journal editorial boards and has also chaired or served on numerous program committees of international conferences and workshops. Dr. Zhang is an ACM Fellow and an IEEE Fellow.

Conference Paper Presentations

Regular Paper Sessions

	L1 Scientific Data Management	
BigD358	Accelerating a Distributed CPD Algorithm for Large Dense, Skewed Tensors	Kareem Aggour, Alex Gittens, and Bülent Yener
BigD396	Optimizing Lossy Compression with Adjacent Snapshots for N-body Simulation Data	Sihuan Li, Sheng Di, Xin Liang, Zizhong Chen, and Franck Cappello
BigD475	Error-Controlled Lossy Compression Optimized for High Compression Ratios of Scientific Datasets	Xin Liang, Sheng Di, Dingwen Tao, Sihuan Li, Shaomeng Li, Hanqi Guo, Zizhong Chen, and Franck Cappello
BigD529	Cloud based Real-Time and Low Latency Scientific Event Analysis	Chen Yang, Zhihui Du, and Xiaofeng Meng
BigD611	Alleviating I/O Inefficiencies to Enable Effective Model Training Over Voluminous, High-Dimensional Datasets	Daniel Rammer, Walid Budgaga, Thilina Buddhika, Shrideep Pallickara, and Sangmi Lee Pallickara
	L2 Social Web Search and Mining	
BigD425	On Learning Psycholinguistics Tools for English-based Creole Languages using Social Media Data	Pei-Chi LO and Ee-Peng LIM
BigD440	Semi-supervised Multi-instance Learning for Flu Shot Adverse Event Detection	Junxiang Wang, Liang Zhao, and Yanfang Ye
BigD477	One-Shot Learning on Attributed Sequences	Zhongfang Zhuang, Xiangnan Kong, Elke Rundensteiner, Aditya Arora, and Jihane Zouaoui
BigD508	FauxBuster: A Content-free Fauxtography Detector Using Social Media Comments	Daniel Zhang, Lanyu Shang, Biao Geng, Shuyue Lai, Ke Li, Hongmin Zhu, Md Tanvir Amin, and Dong Wang
BigD658	Influence Maximization in Social Networks With Non-Target Constraints	Madhavan Padmanabhan, Naresh Somisetty, Samik Basu, and A Pavan
	L3 Semantic-based Data Mining	
BigD305	Differentially Private Semi-Supervised Learning With Known Class Priors	Anh Pham and Jing Xi
BigD437	Automated Extraction of Personal Knowledge from Smartphone Push Notifications	Yuanchun Li, Ziyue Yang, Yao Guo, Xiangqun Chen, Yuvraj Agarwal, and Jason Hong
BigD483	A Data-Centric Approach for Image Scene Localization	Abdullah Alfarrarjeh, Seon Ho Kim, Shivnesh Rajan, Akshay Deshmukh, and Cyrus Shahabi
BigD509	Lifelong Memory Networks with Knowledge Learning from Big Data for Aspect Sentiment Classification	Shuai Wang, Guangyi Lv, Sahisnu Mazumder, Geli Fei, and Bing Liu
BigD753	Improved Dynamic Memory Network for Dialogue Act Classification with Adversarial Training	Yao Wan, Wenqiang Yan, Jianwei Gao, Zhou Zhao, Jian Wu, and Philip S. Yu
	L4 Big Data Applications: Machine Learning	
BigD721	Time-Aware Subgroup Matrix Decomposition: Imputing Missing Data Using Forecasting Events	Xi Yang and Min Chi
BigD495	Exploiting Knowledge Graph to Improve Text-based Prediction	Shan Jiang, Chengxiang Zhai, and Qiaozhu Mei
BigD526	A Minimax Approach for Classification with Big-data	Krishnan Raghavan, Jagannathan Sarangapani, and VA Samaranayake
BigD445	Transfer learning for time series classification	Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller

BigD593	Knowledge-guided Bayesian Support Vector Machine for High-Dimensional Data with Application to Analysis of Genomics Data	Wenli Sun, Changgee Chang, Yize Zhao, and Qi Long

	L5 Novel Theoretical Models for Big Data	
BigD357	Linear Models with Many Cores and CPUs: A Stochastic Atomic Update Scheme	Edward Raff and Jared Sylvester
BigD409	Best-Choice Edge Grafting for Efficient Structure Learning of Markov Random Fields	Walid Chaabene and Bert Huang
BigD564	Detecting Latent Structure Uncertainty with Structural Entropy	So Hirai and Kenji Yamanishi
BigD580	Time Series Classification Using a Neural Network Ensemble	Soukaina Filali Boubrahimi and Rafal Angryk
	L6 Big Data Analytics Frameworks	
BigD336	Online Density Estimation over Streaming Data: A Local Adaptive Solution	Zhong Chen, Zhide Fang, Jiabin zhao, Wei Fan, Andrea Edwards, and Kun Zhang
BigD397	Learning-based Automatic Parameter Tuning for Big Data Analytics Frameworks	Liang Bao, Xin Liu, and Weizhao Chen
BigD738	A Reinforcement Learning Based Resource Management Approach for Time-critical Workloads in Distributed Computing Environment	Zixia Liu, Hong Zhang, Bingbing Rao, and Liqiang Wang
BigD671	A Method-Level Test Generation Framework for Debugging Big Data Applications	Huadong Feng, Jaganmohan Chandrasekaran, Yu Lei, Raghu Kacker, and D. Richard Kuhn
	L7 Software Systems to Support Big Data Computing	
BigD239	ARCHIE: Data Analysis Acceleration with Array Caching in Hierarchical Storage	Bin Dong, Teng Wang, Houjun Tang, Quincey Koziol, Kesheng Wu, and Suren Byna
BigD350	Practical Cross Program Memoization with KeyChain	Craig Mustard and Alexandra Fedorova
BigD403	Dynamic and Transparent Memory Sharing for Accelerating Big Data Analytics Workloads in Virtualized Cloud	Wenqi Cao and Ling Liu
BigD236	Benchmarking API Costs of Network Sampling Strategies	Michele Coscia and Luca Rossi
	L8: Threat Detection using Big Data Analytics	
BigD314	A Unified Unsupervised Gaussian Mixture Variational Autoencoder for High Dimensional Outlier Detection	Weixian Liao, Yifan Guo, Xuhui Chen, and Pan Li
BigD534	Actionable Objective Optimization for Suspicious Behavior Detection on Large Bipartite Graphs	Tong Zhao, Matthew Malir, and Meng Jiang
BigD538	Phishing URL Detection with Oversampling based on Text Generative Adversarial Networks	Ankesh Anand, Kshitij Gorde, Joel Moniz, Noseong Park, Tanmoy Chakraborty, and Bei-Tseng Chu
BigD677	GCI: A Transfer Learning Approach for Detecting Cheats of Computer Game	Bo Dong, Md Shihabul Islam, Swarup Chandra, Latifur Khan, and Bhavani Thuraisingham

	L9 Recommendation Systems and Stream Data Management	
BigD566	PER: A Probabilistic Attentional Model for Personalized Text Recommendations	Lei Zheng, Yixue Wang, Lifang He, Sihong Xie, Fengjiao Wang, and Philip S. Yu
BigD626	StreamGuard: A Bayesian Network Approach to Copyright Infringement Detection Problem in Large-scale Live Video Sharing Systems	Daniel Zhang, Lixing Song, Qi Li, Yang Zhang, and Dong Wang
BigD600	BigSR: real-time expressive RDF stream reasoning on modern Big Data platforms	Xiangnan Ren, Olivier Curé, Hubert Naacke, and Guohui Xiao
BigD758	A Sketch-Based Naive Bayes Algorithms for Evolving Data Streams	Maroua Bahri, Silviu Maniu, and Albert Bifet

BigD309	Concept-Driven Load Shedding: Reducing Size and Error of Voluminous and Variable Data Streams	Nikos R. Katsipoulakis, Alexandros Labrinidis, and Panos K. Chrysanthis
	L10 Link and Graph Mining	
BigD334	AURORA: Auditing PageRank on Large Graphs	Jian Kang, Meijia Wang, Nan Cao, Yinglong Xia, Wei Fan, and Hanghang Tong
BigD414	ImVerde: Vertex-Diminished Random Walk for Learning Imbalanced Network Representation	Jun Wu, Jingrui He, and Yongming Liu
BigD569	Fast and Accurate Mining of Node Importance in Trajectory Networks	Tilemachos Pechlivanoglou and Manos Papagelis
BigD624	Constructing Influence Trees from Temporal Sequence of Retweets: An Analytical Approach	Ayan Kumar Bhowmick, G. Sai Bharath Chandra, Yogesh Singh, and Bivas Mitra
BigD665	Mining top-k Popular Datasets via a Deep Generative Model	Uchenna Akujuobi, Ke Sun, and Xiangliang Zhang
	L11 Big Data Applications: Industry and Business	
BigD528	Mining Illegal Insider Trading of Stocks: A Proactive Approach	Sheikh Rabiul Islam, Sheikh Khaled Ghafoor, and William Eberle
BigD547	Profiling Driver Behavior for Personalized Insurance Pricing and Maximal Profit	Bing He, Dian Zhang, Siyuan Liu, Hao Liu, Dawei Han, and Lionel M. Ni
BigD685	An Unsupervised Learning Based Approach for Mining Attribute Based Access Control Policies	Leila Karimi and James Joshi
BigD285	Realtime Robustification of Interdependent Networks under Cascading Attacks	Zhen Chen, Hanghang Tong, and Lei Ying
BigD643	Situation-Based Interaction Learning for Personality Prediction on Facebook	Lei Zhang, Liang Zhao, Xuchao Zhang, Wenmo Kong, Zitong Sheng, and Chang-Tien Lu
	L12 Big Data Applications: Health & Science Discovery	
BigD653	Technology Enablers for Big Data, Multi-Stage Analysis in Medical Image Processing	Shunxing Bao, Prasanna Parvathaneni, Yuankai Huo, Yogesh Barve, Andrew Plassard, Yuang Yao, Hongyang Sun, Ilwoo Lyu, David Zald, Bennett Landman, and Aniruddha Gokhale
BigD456	A Structured Learning Approach with Neural Conditional Random Fields for Sleep Staging	Karan Aggarwal, Swaraj Khadanga, Shafiq Joty, Louis Kazaglis, and Jaideep Srivastava
BigD402	Dynamic Prediction of ICU Mortality Risk Using Domain Adaptation	Tiago Alves, Alberto Laender, Adriano Veloso, and Nivio Ziviani
BigD353	Large-Scale Validation of Hypothesis Generation Systems via Candidate Ranking	Justin Sybrandt, Micheal Shtutman, and Ilya Safro
BigD354	Are Abstracts Enough for Hypothesis Generation?	Justin Sybrandt, Angelo Carrabba, Alexander Herzog, and Ilya Safro

	L13 New Computational Models for Big Data	
BigD504	Semi-supervised Deep Representation Learning for Multi-View Problems	Vahid Noroozi, Lei Zheng, Sara Bahaadini, Sihong Xie, Weixiang Shao, and Philip S. Yu
BigD545	Projection-SVM: Distributed Kernel Support Vector Machine for Big Data using Subspace Partitioning	Dinesh Singh and Krishna Mohan C
BigD602	Hybridization of Active Learning and Data Programming for Labeling Large Industrial Datasets	Mona Nashaat, Aindrila Ghosh, Shaikh Quader, Chad Marston, Jean-Francois Puget, and James Miller
BigD717	DANN: Incorporating Prior Domain Knowledge into Model Training	Nikhil Muralidhar, Mohammad Raihanul Islam, Manish Marwah, Anuj Karpatne, and Naren Ramakrishnan

	L14 HPC Platforms for Big Data	
BigD234	An Empirical Analysis on Expressibility of Vertex Centric Graph Processing Paradigm	Siyuan Liu and Arijit Khan
BigD294	Column Cache: Buffer Cache for Columnar Storage on HDFS	Takeshi Yoshimura, Tatsuhiro Chiba, and Hiroshi Horii
BigD431	Scalable Manifold Learning for Big Data with Apache Spark	Frank Schoeneman and Jaroslaw Zola
BigD598	Mira: Sharing Resources for Distributed Analytics at Small Timescales	Michael Kaufmann, Kornilios Kourtis, Adrian Schuepbach, and Martina Zitterbart
	L15 Big Data Search (1)	
BigD571	Fast Bag-Of-Words Candidate Selection in Content-Based Instance Retrieval Systems	Michal Siedlaczek, Qi Wang, Yen-Yu Chen, and Torsten Suel
BigD423	Efficient Discovery of Weighted Frequent Itemsets in Very Large Transactional Databases: A Re-visit	RAGE UDAY KIRAN
BigD302	Optimal k-Nearest-Neighbor Query Processing via Multiple Lower Bound Approximations	Christian Beecks and Max Berrendorf
BigD328	Scalable Construction of Text Indexes with Thrill	Timo Bingmann, Simon Gog, and Florian Kurpicz
	L16 Web Data and Crowdsourcing	
BigD379	Explaining Aggregates for Exploratory Analytics	Fotis Savva, Christos Anagnostopoulos, and Peter Triantafillou
BigD318	HYPE: Massive Hypergraph Partitioning with Neighborhood Expansion	Christian Mayer, Ruben Mayer, Sukanya Bhowmik, Lukas Epple, and Kurt Rothermel
BigD573	Influence Maximization in Evolving Multi-Campaign Environments	Iouliana Litou and Vana Kalogeraki
BigD481	Truth Inference on Sparse Crowdsourcing Data with Local Differential Privacy	Haipei Sun, Boxiang Dong, Wendy Hui Wang, Ting Yu, and Zhan Qin

	L17 Big Data Applications: Deep Learning	
BigD401	IL-Net: Using Expert Knowledge to Guide the Design of a Furcated Neural Networks	Khushmeen Sakloth, Wesley Beckner, Jim Pfaendtner, and Garrett Goh
BigD760	Learning Informative and Private Representations via Generative Adversarial Networks	Tsung-Yen Yang, Christopher Brinton, Prateek Mittal, Mung Chiang, and Andrew Lan
BigD406	Two Birds with One Network: Unifying Event Prediction and Time-to-failure Modeling	Karan Aggarwal, Onur Atan, Ahmed Farahat, Chi Zhang, Kosta Ristovski, and Chetan Gupta
BigD292	Market Abnormality Period Detection via Co-movement Attention Model	Yue Wang, Chenwei Zhang, Shen Wang, Philip S. Yu, Lu Bai, and Lixin Cui
BigD298	Optimizing Taxi Carpool Policies via Reinforcement Learning and Spatio-Temporal Mining	Ishan Jindal, Zhiwei (Tony) Qin, Xuewen Chen, Matthew Nokleby, and Jieping Ye
	L18 Mobile and IoT Data	
BigD371	Enabling of Predictive Maintenance in the Brownfield through Low-Cost Sensors, an IIoT-Architecture and Machine Learning	Patrick Strauß, René Wöstmann, Markus Schmitz, and Jochen Deuse
BigD240	Using Smart Card Data to Model Commuters' Response Upon Unexpected Train Delays	Xiancai Tian and Baihua Zheng
BigD542	Hot Spot Analysis for Big Trajectory Data	Panagiotis Nikitopoulos, Aris-Iakovos Paraskevopoulos, Christos Doulkeridis, Nikos Pelekis, and Yannis Theodoridis
BigD722	Fusion of Terrain Information and Mobile Phone Location Data for Flood Area Detection in Rural Areas	Takahiro Yabe, Kota Tsubouchi, and Yoshihide Sekimoto

BigD296	Benchmarking Anomaly Detection Algorithms in an Industrial Context: Dealing with Scarce Labels and Multiple Positive Types	David Renaudie, Maria A. Zuluaga, and Rodrigo Acuna-Agost
	L19 Big Data Applications: Society	
BigD387	Integrating the University of São Paulo Security Mobile App to the Electronic Monitoring System	João Eduardo Fereira, José Antônio Visintin, Jun Okamoto Jr., Mauro Cesar Bernardes, Adriano Paterlini, Alexander Csóka Roque, and Moisés Ramalho Miguel
BigD476	RiskSens: A Multi-view Learning Approach to Identifying Risky Traffic Locations in Intelligent Transportation Systems Using Social and Remote Sensing	Yang Zhang, Yiwen Lu, Daniel Zhang, Lanyu Shang, and Dong Wang
BigD278	A Bayesian Approach to Residential Property Valuation Based on Built Environment and House Characteristics	Zhicheng Liu, Shuai Yan, Jun Cao, Tanhua Jin, Jiabo Tang, Junyan Yang, and Qiao Wang
BigD590	Inferring Housing Demand based on Express Delivery Data	Qingyang Li, Zhiwen Yu, Bin Guo, and Xinjiang Lu
BigD655	The unbanked and poverty: Predicting area-level socio-economic status from M-Money transactions	Gregor Engelmann, James Goulding, and Gavin Smith
	L20 Big Data Search (2)	
BigD364	An Efficient System for Subgraph Discovery	Aparna Joshi, Yu Zhang, Petko Bogdanov, and Jeong-Hyon Hwang
BigD457	Candidate List Maintenance in High Utility Sequential Pattern Mining	Scott Buffett
BigD469	ParIS: The Next Destination for Fast Data Series Indexing and Query Answering	Botao Peng, Themis Palpanas, and Panagiota Fatourou
BigD498	Learning Multiclassifiers with Predictive Features Varied along with Data Distribution	Xuan-Hong Dang, Omid Askarisichani, and Ambuj K. Singh

	L21 Big Data Search (3)	
BigD307	Revisiting Exact kNN Query Processing with Probabilistic Data Space Transformations	Atoshum Samuel Cahsai, Christos Anagnostopoulos, Nikos Ntarmos, and Peter Triantafillou
BigD356	Adaptive Data Pruning for Support Vector Machines	Yasuhiro Fujiwara, Junya Arai, Sekitoshi Kanai, Yasutoshi Ida, and Naonori Ueda
BigD662	A Multi-Criteria Experimental Ranking of Distributed SPARQL Evaluators	Damien Graux, Louis Jachiet, Pierre Genevès, and Nabil Layaïda
BigD740	Fast Clustering with Flexible Balance Constraints	Hongfu Liu, Ziming Huang, Yun Fu, Qi Chen, Mingqin Li, and Lintao Zhang
	L22 Privacy and Security	
BigD531	There goes Wally: Anonymously sharing your location gives you away	Apostolos Pyrgelis, Nicolas Kourtellis, Ilias Leontiadis, Joan Serra, and Claudio Soriente
BigD247	Distributed Machine Learning Meets Blockchain: A Decentralized, Secure, and Privacy-preserving Realization	Xuhui Chen, Jinlong Ji, Changqing Luo, Weixian Liao, and Pan Li
BigD388	PrivacyZone: A novel approach to protecting location privacy of mobile users	Emre Yigitoglu, Mehmet Emre Gursoy, Ling Liu, Margaret Loper, Bhuvan Bamba, and Kisung Lee
BigD238	Do Bitcoin Users Really Care About Anonymity? A Graph Analysis on Bitcoin Transaction Graphs	Anil Gaihre, Yan Luo, and Hang Liu

Short Paper Sessions

	S1 Big Data Infrastructure (1)	

BigD291	Analyzing Alibaba's Co-located Datacenter Workloads	Yue Cheng, Ali Anwar, and Xuejing Duan
BigD394	OverSketch: Approximate Matrix Multiplication for the Cloud	Vipul Gupta, Shusen Wang, Thomas Courtade, and Kannan Ramchandran
BigD459	POSUM: A Portfolio Scheduler for MapReduce Workloads	Maria Voinea, Alexandru Uta, and Alexandru Iosup
BigD563	Scalable Distributed Top-k Join Queries in Topic-Based Pub/Sub Systems	Nikos Zacheilas, Dimitris Dedousis, and Vana Kalogeraki
BigD735	Spark-uDAPL: Cost-Saving Big Data Analytics on Microsoft Azure Cloud with RDMA Networks	Xiaoyi Lu, Dipti Shankar, Haiyang Shi, and Dhabaleswar K. (DK) Panda
BigD682	Parallel DBSCAN Algorithm Using a Data Partitioning Strategy with Spark Implementation	Dianwei Han
BigD689	Sync-on-the-fly: A Parallel Framework for Gradient Descent Algorithms on Transient Resources	Guoyi Zhao, Lixin Gao, and David Irwin
	S2 Big Data Applications (1)	
BigD248	A Longitudinal Social Network Clustering Method Based on Tie Strength	Zhiyong Zhang, Mao Ye, Yijie Huang, and Nan Sun
BigD263	Personalized heart failure severity estimates using passive smartphone data	Ayse Cakmak, Erik Reinertsen, Herman Taylor, Amit Shah, and Gari Clifford
BigD280	Data-driven Blockbuster Planning on Online Movie Knowledge Library	Ye Liu, Jiawei Zhang, Chenwei Zhang, and Philip S. Yu
BigD310	Deep Convolutional Neural Networks for Log Event Classification on Distributed Cluster Systems	Rui Ren, Jiechao Cheng, Yan Yin, Jianfeng Zhan, and Lei Wang
BigD321	Social-Media aided Hyperlocal Help-Network Matching & Routing during Emergencies	Dheeraj Kumar, Takahiro Yabe, and Satish Ukkusuri
BigD362	Visual Reasoning of Feature Attribution with Deep Recurrent Neural Networks	Chuan Wang, Takeshi Onishi, Keiichi Nemoto, and Kwan-Liu Ma
BigD380	Entropy-Isomap: Manifold Learning for High-dimensional Dynamic Processes	Frank Schoeneman, Varun Chandola, Nils Napp, Olga Wodo, and Jaroslaw Zola
BigD389	A Hierarchical Framework for Timely Freeway Accident Detection and Localization	Yasitha Warahena Liyanage, Charalampos Chelmis, and Daphney-Stavroula Zois
	S3 Spatiotemporal and Stream Data Management	
BigD317	Communication Model for Parallel Iterative Stream Processing	Sachini Jayasekara, Xunyun Liu, Shanika Karunasekera, and Aaron Harwood
BigD290	Distributed Execution of Spatial SQL Queries	Konstantinos Giannousis, Konstantina Bereta, Nikolaos Karalis, and Manolis Koubarakis
BigD325	Efficient Processing of Probabilistic Single and Batch Reachability Queries in Large and Evolving Spatiotemporal Contact Networks	Zohreh Raghebi and Farnoush Banaei-Kashani
BigD381	Integrated Real-Time Data Stream Analysis and Sketch-Based Video Retrieval in Team Sports	Lukas Probst, Fabian Rauschenbach, Heiko Schuldt, Philipp Seidenschwarz, and Martin Rumo
BigD345	The content correlation of multiple streaming edges	Michel de Rougemont and Guillaume Vimont
BigD627	A Survey on Trajectory Data Management for Hybrid Transactional and Analytical Workloads	Keven Richly
BigD276	Efficient Principal Subspace Projection of Streaming Data Through Fast Similarity Matching	Andrea Giovannucci, Victor Minden, Cengiz Pehlevan, and Dmitri Chklovskii
	S4 Big Data Security & Privacy	
BigD217	Novel anomaly detection and classification schemes for Machine-to-Machine uplink	Akshay Kumar, Ahmed Abdelhadi, and Charles Clancy
BigD630	dynamicMF: A Matrix Factorization Approach to Monitor Resource Usage in High Performance Computing	Niyazi Sorkunlu, Duc Thanh Anh Luong, and Varun Chandola

BigD589	CVExplorer: Multidimensional Visualization for Common Vulnerabilities andExposures	Vung Pham and Tommy Dang
BigD446	Automated Generation and Selection of Interpretable Features for Enterprise Security	Jiayi Duan, Ziheng Zeng, Alina Oprea, and Shobha Vasudevan
BigD523	Graph Mining-based Trust Evaluation Mechanism with Multidimensional Features for Large-scale Heterogeneous Threat Intelligence	Yali Gao
BigD346	Toward End-to-End Deception Detection in Videos	Hamid Karimi, Jiliang Tang, and Yanen Li
BigD443	Learning Light-Weight Edge-Deployable Privacy Models	Yeon-sup Lim, Mudhakar Srivatsa, Supriyo Chakraborty, and Ian Taylor

	S5 Big Data Infrastructure (2)	
BigD711	GreenDataFlow: Minimizing the Energy Footprint of Global Data Movement	Zulkar Nine, Luigi Di Tacchio, Asif Imran, Tevfik Kosar, Fatih Bulut, and Jinho Hwang
BigD252	Serverless Big Data Processing using Matrix Multiplication as Example	Sebastian Werner, Jörn Kuhlenkamp, Markus Klems, Johannes Müller, and Stefan Tai
BigD604	XOS: An Application-Defined Operating System for Data Center Servers	Chen Zheng, Lei Wang, Sally A. McKee, Jianfeng Zhan, and Lixin Zhang
BigD470	Experimental Characterizations and Analysis of Deep Learning Frameworks	Yanzhao Wu, Wenqi Cao, Semih Sahin, and Ling Liu
BigD617	Culster-based Data Reduction for Persistent Homology	Anindya Moitra, Nick Malott, and Philip Wilsey
BigD645	GeoMatch: Efficient Large-Scale Map Matching on Apache Spark	Ayman Zeidan, Eemil Lagerspetz, Kai Zhao, Petteri Nurmi, Sasu Tarkoma, and Huy Vo
	S6 Big Data Applications (2)	
BigD465	Predicting Individual-Level Call Arrival from Online Account Customer Activity	Somayeh Moazeni
BigD494	A Subspace Pre-learning Approach to Fast High-Accuracy Machine Learning of Large XOR PUFs with Component-Differential Challenges	Ahmad O. Aseeri, Yu Zhuang, and Mohammed Saeed Alkatheiri
BigD506	Scalable Classification of Univariate and Multivariate Time Series	Saeed Karimi-Bidhendi, Faramarz Munshi, and Ashfaq Munshi
BigD535	Short-term local weather forecast using dense weather station by deep neural network	Kazuo Yonekura, Hitoshi Hattori, and Taiji Suzuki
BigD654	Distributed Learning of Deep Sparse Neural Networks for High-dimensional Classification	Shweta Garg, Krishnan Raghavan, Jagannathan Sarangapani, and Samaranayake V.A.
BigD659	Twitter Health Surveillance (THS) System	Manuel Rodriguez-Martinez and Cristian Garzon-Alfonso
BigD746	Land Cover Classification at the Wildland Urban Interface using High-Resolution Satellite Imagery and Deep Learning	Mai H. Nguyen, Jessica Block, Daniel Crawl, Vincent Siu, Akshit Bhatnagar, Federico Rodriguez, Alison Kwan, Namrita Baru, and Ilkay Altintas
BigD755	Distributed Reverse DNS Geolocation	Ovidiu Dan, Vaibhav Parikh, and Brian D. Davison
	S7 Big Data Management	
BigD351	FairGAN: Fairness-aware Generative Adversarial Networks	Depeng Xu, Shuhan Yuan, Lu Zhang, and Xintao Wu
BigD365	Aggregation of Linked Data: a case study in the cultural heritage domain	Nuno Freire, Enno Meijers, René Voorburg, Roland Cornelissen, Antoine Isaac, and Sjors de Valk
BigD287	Steering Top-k Influencers in Dynamic Graphs via Local Updates	Vijaya Krishna Yalavarthi and Arijit Khan

BigD384	A Universal Namespace Approach to Support Metadata Management and Efficient Data Convergence of HPC and Cloud Scientific Workflows	Hsing-Bung Chen
BigD621	Optimized Storing of Workflow Outputs through Mining Association Rules	Debasish Chakroborti, Manishankar Mondal, Banani Roy, Chanchal K. Roy, and Kevin A. Schneider
BigD703	Exploring Size-Speed Trade-Offs In Static Index Pruning	Juan Rodriguez and Torsten Suel
BigD666	PACAS: Privacy-Aware, Data Cleaning-as-a-Service	Yu Huang, Mostafa Milani, and Fei Chiang
BigD644	An Application of Storage-Optimal MatDot Codes for Coded Matrix Multiplication: Fast k-Nearest Neighbors Estimation	Utsav Sheth, Sanghamitra Dutta, Malhar Chaudhari, Haewon Jeong, Yaoqing Yang, Jukka Kohonen, Teemu Roos, and Pulkit Grover
	S8 Link and Graph Mining	
BigD301	Dynamic Network Embeddings: From Random Walks to Temporal Random Walks	Giang Nguyen, John Boaz Lee, Ryan Rossi, Nesreen Ahmed, Eunyee Koh, and Sungchul Kim
BigD436	Detecting Highly Overlapping Community Structure by Model-based Maximal Clique Expansion	Said Jabbour, Nizar Mhadhbi, Badran Raddaoui, and Lakhdar Sais
BigD480	Local Partition in Rich Graphs	Scott Freitas, Nan Cao, Yinglong Xia, Duen Horng Chau, and Hanghang Tong
BigD533	Motif-Preserving Dynamic Local Graph Cut	Dawei Zhou, Jingrui He, Hasan Davulcu, and Ross Maciejewski
BigD697	Deep Learning for Predicting Dynamic Uncertain Opinions in Network Data	Xujiang Zhao, Feng Chen, and Jin-Hee Cho
BigD698	Density-aware Local Siamese Autoencoder Network Embedding with Autoencoder Graph Clustering	Yang Zhou, Amnay Amimeur, Chao Jiang, Dejing Dou, Ruoming Jin, and Pengwei Wang
BigD731	Enumerating Top-k Quasi-Cliques	Seyed-Vahid Sanei-Mehri, Apurba Das, and Srikanta Tirthapura

	S9 Big Data Foundations (1)	
BigD366	Efficient Dimensionality Reduction for Sparse Binary Data	Rameshwar Pratap, Raghav Kulkarni, and Ishan Sohony
BigD369	Effective Outlier Detection based on Bayesian Network and Proximity	Sha Lu, Lin Liu, Jiuyong Li, and Thuc Duy Le
BigD405	Hash-Grams On Many-Cores and Skewed Distributions	Edward Raff and Mark McLean
BigD451	AdaDIF: Adaptive Diffusions for Efficient Semi-supervised Learning over Graphs	Dimitris Berberidis, Athanasios Nikolakopoulos, and Georgios B. Giannakis
BigD482	Source Free Domain Adaptation Using an Off-the-Shelf Classifier	Arun Reddy Nelakurthi, Ross Maciejwski, and Jingrui He
BigD501	Modeling Road Traffic Dynamics Using Big Data	Fan Yang, Alina Vereshchaka, and Wen Dong
BigD701	Queryable Compression on Time-Evolving Social Networks with Streaming	Michael Nelson, Sridhar Radhakrishnan, and Chandra Sekharan
	S10 Experiences and Case Studies with Big Data	
BigD245	Predicting Perceived Cycling Safety Levels Using Open and Crowdsourced Data	Jiahui Wu, Lingzi Hong, and Vanessa Frias-Martinez
BigD594	Defining an Alert Mechanism for Detecting likely threats to National Security	Pedro Cardenas Canto, Georgios Theodoropoulos, and Boguslaw Obara
BigD568	NetClips: A Framework for Video Analytics in Sports Broadcast	Masoumeh Izadi, Shangjing Wu, Aiden Chia, and Bernard Cheng
BigD337	Session Expert: a Lightweight Conference Session Recommender System	Jinfeng Yi, Qi Lei, Junchi Yan, and Wei Sun

BigD719	ThousandSunny: A Large-Scale Neural Network Training System For Online Advertising	Quanchang Qi, Guangming Lu, Jun Zhang, Lichun Yang, and Haishan Liu
BigD339	Speed Accuracy Trade-off for Pedestrian and Vehicle Detection using Localized Big Data	Yeongro Yun, Youngseok Park, Chanhee Woo, and Sejoon Lim
BigD667	An Integrated Knowledge Graph to Automate GDPR and PCI DSS Compliance	Lavanya Elluri, Ankur Nagar, and Karuna Pande Joshi

	S11 Big Data Analytics	
BigD330	StageMap: Extracting and Summarizing Progression Stages in Event Sequences	Yuanzhe Chen, Abishek Puri, Linping Yuan, and Huamin Qu
BigD390	Monitoring the shape of weather, soundscapes, and dynamical systems: a new statistic for dimension-driven data analysis on large data sets	Henry Kvinge, Elin Farnell, Michael Kirby, and Chris Peterson
BigD669	Scalable K-Core Decomposition for Static Graphs Using a Dynamic Graph Data Structure	Alok Tripathy, Fred Hohman, Duen Horng Chau, and Oded Green
BigD614	SynthNotes: A Generator Framework for High-volume, High-fidelity Synthetic Mental Health Notes	Edmon Begoli, Kris Brown, Sudarshan Srinivas, and Suzanne Tamang
BigD461	CAM: A Combined Attention Model for Natural Language Inference	Amit Gajbhiye, Sardar Jaf, Noura Al Moubayed, Steven Bradley, and A. Stephen McGough
BigD484	All-in-One Urban Mobility Mapping Application with Optional Routing Capabilities	Rebekah Thompson, Jose Stovall, Daniel Velasquez, Viswa Sri Rupa Anne, Alex Samoylov, and Mina Sartipi
BigD633	Spatio-Temporal Attention based recurrent neural network for next poi prediction	Basmah Altaf, Lu Yu, and Xiangliang Zhang
BigD499	Context-Aware Deep Sequence Learning with Multi-View Factor Pooling for Time Series Classification	Sreyasee Das Bhattacharjee, William J. Tolone, Ashish Mahabal, Mohammed Elshambakey, Isaac Cho, and S. G. Djorgovski
	S12 Recommendation Systems & Stream Data Mining	
BigD385	Top-N-Rank: A Truncated List-wise Ranking Approach for Large-scale Top-N Recommendation	Junjie Liang, Jinlong Hu, Shoubin Dong, and Vasant Honavar
BigD395	Predicting computational reproducibility of data analysis pipelines in large population studies using collaborative filtering	Soudabeh Barghi, Lalet Scaria, Ali Salari, and Tristan Glatard
BigD751	Context Aware Recommender System for Large Scaled Flash Sale Sites	Wanying Ding, Ran Xu, Ying Ding, Yue Zhang, and Chuanjiang Luo
BigD741	Dynamic Online Performance Optimization in Streaming Data Compression	Kade Gibson, Dongeun Lee, Jaesik Choi, and Alexander Sim
BigD383	Learning Fast and Slow - A Unified Batch/Stream Framework	Jacob Montiel, Albert Bifet, Viktor Losing, Jesse Read, and Talel Abdessalem
BigD393	Clustering-Driven and Dynamically Diversified Ensemble for Drifting Data Streams	Lukasz Korycki and Bartosz Krawczyk
BigD537	AnySC: Anytime Set-wise Classification of Variable Speed Data Streams	Jagat Sesh Challa, Poonam Goyal, Vijay M Giri, Dhananjay Mantri, and Navneet Goyal
BigD583	Correlated Anomaly Detection from Large Streaming Data	Zheng Chen, Xinli Yu, Yuan Lin, Xiaohua Hu, and Erjia Yan
	S13 Big Data Search and Mining	
BigD293	Identifying Pros and Cons of Product Aspects Based on Customer Reviews	Ebad Ahmadzadeh and Philip Chan
BigD473	FastTopK: A Fast Top-K Trajectory Similarity Query Processing Algorithm for GPUs	Hamza Mustafa, Eleazar Leal, and Le Gruenwald

BigD497	Multi-Attribute Topic Feature Construction for Social Media-based Prediction	Alex Morales, Nupoor Gandhi, Man-pui Sally Chan, Sophie Lohmann, Travis Sanchez, Lyle Ungar, Dolores Albaracin, and Chengxiang Zhai
BigD386	Dynamic Partition Forest: An Efficient and Distributed Indexing Scheme for Similarity Search based on Hashing	Yangdi Lu, Yang Bo, Wenbo He, and Amir Nabatchian
BigD444	Improving Query Execution Performance in Big Data using Cuckoo Filter	Sharafat Ibn Mollah Mosharraf and Muhammad Abdullah Adnan
BigD530	DLA: a Distributed, Location-based and Apriori-based Algorithm for Biological Sequence Pattern Mining	Eirini Stamoulakatou, Andrea Gulino, and Pietro Pinoli
	S14 Big Data Foundations (2)	
BigD536	Scalable Bottom-up Subspace Clustering using FP-Trees for High Dimensional Data	Tuan Doan, Jianzhong Qi, Sutharshan Rajasegarar, and Christopher Leckie
BigD565	Biomedical Data Classification using Random Projection Ensembles	Sotiris Tasoulis, Aristidis Vrahatis, Spiros Georgakopoulos, and Vassilis Plagianakos
BigD581	Representation Learning for Question Classification via Topic Sparse Autoencoder and Entity Embedding	Dingcheng Li, Jingyuan Zhang, and Ping Li
BigD699	Scaling up Inference in MLNs with Spark	Maminur Islam, Khan Mohammad Al Farabi, Somdeb Sarkhel, and Deepak Venugopal
BigD708	Topological approaches to skin disease image analysis	Yu-Min Chung, Chuan-Shen Hu, Austin Lawson, and Clifford Smyth
BigD730	DeepFP: A Deep Learning Framework For User Fingerprinting via Mobile Motion Sensors	Sara Amini, Vahid Noroozi, Sara Bahaadini, Philip S. Yu, and Chris Kanich
BigD442	Securing Behavior-based Opinion Spam Detection	Shuaijun Ge, Guixiang Ma, Sihong Xie, and Philip S. Yu

Industry and Government Paper Presentations

Regular Paper Sessions

	I&G Regular1: Big Data and Machine Learning (1)	
N210	CUImage: A Neverending Learning Platform on a Convolutional Knowledge Graph of Billion Web Images	Ping Luo, Zhanglin Peng, Lingyun Wu, and Jiamin Ren
N205	Relational Similarity Machines (RSM): A Similarity-based Learning Framework for Graphs	Ryan Rossi, Nesreen Ahmed, Rong Zhou, and Hoda Eldardiry
N213	Learning Effective Embeddings for Machine Generated Emails with Applications to Email Category Prediction	Yu Sun, Lluis Garcia-Pueyo, James Wendt, Marc Najork, and Andrei Broder
N217	Scheduling Large-scale Distributed Training via Reinforcement Learning	Zhanglin Peng, Jiamin Ren, Ruimao Zhang, Lingyun Wu, Xinjiang Wang, and Ping Luo
N221	STIPA: A Memory Efficient Technique for Interval Pattern Discovery	Amit Kumar and Dhaval Patel

	I&G Regular2: Big Data Applications	
N209	Bridging the Gap between Big Data System Software Stack and Applications: A Case Study of Semiconductor Wafer Fabrication Foundries	Chia-Ping Tsai, Hung-Chang Hsiao, Yu-Chang Chao, Michael Hsu, and Andy RK Chang
N226	AISTAR: An Intelligent System for Online IT Ticket Automation Recommendation	Qing Wang, Chunqiu Zeng, S. S. Iyengar, Larisa Shwartz, Genady Ya Grabarnik, and Tao Li
N256	Reacting to Variations in Product Demand: An Application for Conversion Rate (CR) Prediction in Sponsored Search	Marcello Tallis and Pranjul Yadav
N257	A Smart System for Selection of Optimal Product Images in E-Commerce	Abon Chaudhuri, Paolo Messina, Samrat Kokkula, Aditya Subramanian, Abhinandan Krishnan, Shreyansh Gandhi, Alessandro Magnani, and Venkatesh Kandaswamy
N265	Data models for service failure prediction in supply-chain networks	Monika Sharma, Tristan Glatard, Eric Gelinas, Mariam Tagmouti, and Brigitte Jaumard

	I&G Regular3: Big Data Platforms & Frameworks	
N211	Learning to Simplify Distributed Systems Management	Christopher Streiffer, Ramya Raghavendra, Theophilus Benson, and Mudhakar Srivatsa
N220	Parallel Polyglot Query Processing on Heterogeneous Cloud Data Stores with LeanXcale	Boyan Kolev, Oleksandra Levchenko, Esther Pacitti, Patrick Valduriez, Ricardo Vilaca, Rui Goncalves, Ricardo Jimenez-Peris, and Pavlos Kranas
N229	Build and Execution Environment (BEE): an Encapsulated Environment Enabling HPC Applications Running Everywhere	Jieyang Chen, Qiang Guan, Xin Liang, Paul Bryant, Patricia Grubel, Allen McPherson, Li-Ta Lo, Timothy Randles, Zizhong Chen, and James Ahrens
N241	High-Throughput Adaptive Data Virtualization via Context-Aware Query Routing	Amirhossein Aleyasen, Mohamed Soliman, Lyublena Antova, Florian Mike Waas, and Marianne Winslett
N259	Finding Data Should be Easier than Finding Oil	Evgeny Kharlamov, Martin Skjaeveland, Theofilos Mailis, Ernesto Jimenez-Ruiz, Guohui Xiao, Ahmet Soylu, Hallstein Lie, and Arild Waaler

	I&G Regular4: Big Data and Machine Learning (2)	

N237	Predicting Age & Gender of Mobile Users at Scale - A Distributed Machine Learning Approach	Kajanan Sangaralingam, Nisha Verma, Aravind Ravi, Anindya Datta, and Varun Chugh
N228	Character Recognition by Deep Learning: An Enterprise Solution	Khaled Bouaziz, Thiagarajan Ramakrishnan, Srinivasan Raghavan, Kyle Grove, Awny Al-Omari, and Choudur Lakshminarayan
N243	Efficient Super Resolution for Large-Scale Images using Attentional GAN	Brooke Cowan, Xinxin Li, Shervin Minaee, and Harsh Nilesh Pathak
N244	ANNOTATE: orgANizing uNstructured cOntenTs viA Topic labEls	Deepak Ajwani, Bilyana Taneva, Sourav Dutta, Pat Nicholson, Ghasem Nobari, and Alessandra Sala

Short Paper Sessions

	I&G Short1: Big Data Algorithms & Systems (1)	
N249	A Batched Multi-Armed Bandit Approach to Dynamic News Headline Testing	Yizhi Mao, Miao Chen, Abhinav Wagle, Junwei Pan, Michael Natkovich, and Don Matheson
N233	In situ TensorView: In situ Visualization of Convolutional Neural Networks	Xinyu Chen, Qiang Guan, Li-Ta Lo, Simon Su, Zhengyong Ren, James Ahrens, and Trilce Estrada
N248	Acquire, adapt, and anticipate: continuous learning to block malicious domains	Ignacio Arnaldo and Kalyan Veeramachaneni
N236	Predicting Individual Level Consumer Brand Preferences Using Persistent Mobility Patterns	Aravind Ravi, Kajanan Sangaralingam, and Anindya Datta
N264	I4TSPS: a Visual-Interactive Web System for Industrial Time Series Pre-processing	Kevin Villalobos, Jon Vadillo, Borja Diez, Borja Calvo, and Arantza Illarramendi
N245	E-commerce Product Query Classification Using Implicit User's Feedback from Clicks	Yiu-Chang Lin and Ankur Datta
N208	Focusing on the Big Picture: Insights into an End-to-End Systems Approach to Deep Learning for Satellite Imagery	Ritwik Gupta, Carson Sestili, Javier Vazquez-Trejo, and Matthew Gaston
N261	Towards Semantic Simplification of Analytical Workflows at Siemens (Extended Abstract)	Evgeny Kharlamov, Gulnar Mehdi, Ognjen Savkovic, Guohui Xiao, Steffen Lamparter, Arild Waaler, and Ian Horrocks

	I&G Short2: Massive Processing & Experience	
N246	Explainable Text Classification in Legal Document Review: A Case Study of Explainable Predictive Coding	Rishi Chhatwal, Peter Gronvall, Nathaniel Huber-Fliflet, Robert Keeling, Jianping Zhang, and Haozhen Zhao
N234	Performance Prediction using Neural Network and Confidence Intervals: a Gas Turbine application	Silvia Cisotto and Randa Herzallah
N250	Identifying Distracted and Drowsy Drivers	Sujay Yadawadkar, Brian Mayer, Sanket Lokegaonkar, Mohammad Raihanul Islam, Miao Song, Mike Mollenhauer, and Naren Ramakrishnan
N227	A Complete Data Science Work-flow For Insurance Field	Mohammed Ghesmoune, Hanane Azzag, Mustapha Lebbah, Salima Benbernou, Mourad Ouziri, and Tarn Duong
N255	NetDP: An Industrial-Scale Distributed Network Representation Framework for Default Prediction in Ant Credit Pay	Jianbin Lin, Zhiqiang Zhang, Jun Zhou, Xiaolong Li, Jingli Fang, Yanming Fang, Quan Yu, and Yuan Qi
N253	Performance Implications of Big Data in Scalable Deep Learning: On the Importance of Bandwidth and Caching	Miro Hodak, David Ellison, Peter Seidel, and Ajay Dholakia
N235	Spatio-temproal prediction of crimes using network analytic approach	Saroj Dash, Ilya Safro, and Ravisutha Srinivasamurthy

N212	A Generic and Scalable Pipeline for Large-Scale Analytics of Continuous Operational Aircraft Engine Data	Florent Forest, Jérôme Lacaille, Mustapha Lebbah, and Hanene Azzag

	I&G Short3: Big Data Algorithms & Systems (2)	
N222	Using Real-World Store Data for Foot Traffic Forecasting	Soheila Abrishami and Piyush Kumar
N218	Distributed NoSQL Data Stores: Performance Analysis and a Case Study	Abdeltawab Hendawi, Jayant Gupta, Liu Jiayi, Ankur Teredesai, Ramakrishnan Naveen, Shah Mohak, and Mohamed Ali
N247	Augmenting Software Project Managers with Predictions from Machine Learning	Kalyan Veeramachaneni and Benjamin Schreck
N254	ChieF : A Change Pattern based Interpretable Failure Analyzer	Dhaval Patel, Lam Nguyen, Akshay Rangamani, Shrey Shrivastava, and Jayant kalagnanam
N240	Big Data Streaming Analytics for QoE Monitoring in Mobile Networks: A Practical Approach	Diego F. Rueda, Dahyr Vergara, and David Reniz
N214	Large Scale Open Source Video Recommender Tool Using Metadata Surrogates	George Mathew, Steven Smith, and John Passarelli
N242	A Deterministic Self-Organizing Map Approach and its Application on Satellite Data based Cloud Type Classification	Wenbin Zhang, Jianwu Wang, Daeho Jin, Lazaros Oreopoulos, and Zhibo Zhang
N225	Root Cause Detection using Dynamic Dependency Graphs from Time Series Data	Syed Yousaf Shah, Xuan-Hong Dang, and Petros Zerfos

Tutorials

TUTORIAL 1: Large-Scale Multi-view Data Analysis

Presenters:

Zhengming Ding, Ph.D. Candidate
427 Richards Hall, 360 Huntington Ave., Boston, MA 02115
Email: allanding@ece.neu.edu
Ming Shao, Assistant Professor
285 Old Westport Road, Dartmouth, MA 02747-2300, Dion 303A
Email: mshao@umassd.edu
Yun Fu, Associate Professor
403 Dana Research Center, 360 Huntington Ave., Boston, MA 02115
Email: yunfu@ece.neu.edu

Abstract:

Multi-view data are extensively accessible nowadays, since various types of features, view-points and different sensors. For example, the most popular commercial depth sensor Kinect uses both visible light and near infrared sensors for depth estimation; automatic driving uses both visual and radar sensors to produce real-time 3D information on the road; and face analysis algorithms prefer face images from different views for high-fidelity reconstruction and recognition. All of them tend to facilitate better data representation in different application scenarios. Essentially, multiple features attempt to uncover various knowledge within each view to alleviate the final tasks, since each view would preserve both shared and private information. This becomes increasingly common in the era of "Big Data" where the data are in large-scale, subject to corruptions, generated from multiple sources, and have complex structures. While these problems attracted substantial research attention recently, a systematic overview of multi-view learning for Big Data analysis has never been given. In face of big data and challenging real-world applications, we summarize and go through the most recent multi-view learning techniques appropriate to different data driven problems. Specifically, our tutorial covers most multi-view data representation approaches, centered around two major applications along with Big Data, i.e., multi-view clustering, multi-view classification. In addition, it discusses the current and upcoming challenges. This would benefit the community in both industry and academia from literature review to future directions.

TUTORIAL 2: Analysis of Complex Rare Categories

Presenters:

Dawei Zhou, PhD student
School of Computing, Informatics, Decision Systems Engineering Arizona State University
Email: Dawei.Zhou@asu.edu
Jingrui He, Associate Professor
School of Computing, Informatics, Decision Systems Engineering Arizona State University
Email: jingrui.heg@asu.edu

Abstract:

In the era of big data, it is often the rare categories that are of great importance in many high-impact applications, ranging from financial fraud detection in online transaction networks to emerging trend detection in social networks, from spam image detection in social media to rare disease diagnosis in medical decision support system. As a result, the detection, characterization, tracking, and representation of rare categories become fundamental learning tasks, that may protect us from malicious behaviors, discover the novelty for scientific studies, and even save lives. The unique challenges of rare category analysis include (1) the highly-skewed class-membership distribution; (2) the nonseparability nature of the rare categories from the majority classes; (3) the data heterogeneity, e.g., the multi-modal representation of examples, and the analysis of similar rare categories across multiple related tasks. In this tutorial, we will provide a concise review of rare category analysis, where the majority classes have a smooth distribution, while the minority classes exhibit a compactness property. In particular, we will start with early developments on rare category analysis, that focus on discovering or characterizing rare examples from static homogeneous data. Then, we will introduce the more recent developments of rare category analysis in complex scenarios, such as rare category detection on time series data sets, rare category tracking in time-evolving graphs, rare category characterization with data heterogeneity. Finally, we will conclude the existing works and share our thoughts regarding the future directions.

TUTORIAL 3: High-Performance SVD for big data

Presenters:

Andreas Stathopoulos, Professor
Computer Science Department, College of William & Mary
Email: andreas@cs.wm.edu
Eloy Romero, Postdoc
Computer Science Department, College of William & Mary
Email: eloy@cs.wm.edu

Abstract:

The singular value decomposition (SVD) is one of the core computations of today's scientific applications and data analysis tools. The main goal is to compute a compact representation of a high dimensional operator, a matrix, or a set of data that best resembles the original in its most important features. Thus, SVD is widely used in scientific computing and machine learning, including low rank factorizations, graph learning, unsupervised learning, compression and analysis of images and text. The popularity of SVD has resulted in an increasing diversity of methods and implementations that exploit specific features of the input data (e.g., dense/sparse matrix, data distributed among the computing devices, data from queries or batch access, spectral decay) and certain constraints on the computed solutions (e.g., few/many number of singular values and singular vectors computed, targeted part of the spectrum, accuracy). The use of the proper method and the customization of the settings can significantly reduce the cost. In this tutorial we present a classification of the most relevant methods in terms of computing cost and accuracy (direct methods, iterative methods, online methods), including the most recent advances in randomized and online SVD solvers. We present what parameters have the biggest impact on the computational cost and the quality of the solution, and some intuition for their tuning. Finally, we discuss the current state of the software on widely used platforms (MATLAB, Python's numpy/scipy and R) as well as high-performance solvers with support for multicore, GPU, and distributed memory.

TUTORIAL 4: Recent Progress in Zeroth Order Optimization and Its Applications to Adversarial Robustness in Deep Learning

Presenters:

Pin-Yu Chen, Research Staff
IBM Research AI
Email: pin-yu.chen@ibm.com
Sijia Liu, Research Staff
IBM Research AI
Email: sijia.liu@ibm.com

Abstract:

Zeroth-order (ZO) optimization is increasingly embraced for solving big data and machine learning problems when explicit expressions of the gradients are difficult or infeasible to obtain. It achieves gradient-free optimization by approximating the full gradient via efficient gradient estimators. Some recent important applications include: a) generation of prediction-evasive, black-box adversarial attacks on deep neural networks, b) online network management with limited computation capacity, c) parameter inference of black-box/complex systems, and d) bandit optimization in which a player receives partial feedback in terms of loss function values revealed by her adversary. This tutorial aims to provide a comprehensive introduction to recent advances in ZO optimization methods in both theory and applications. On the theory side, we will cover convergence rate and iteration complexity analysis of ZO algorithms and make comparisons to their first-order counterparts. On the application side, we will highlight one appealing application of ZO optimization to studying the robustness of deep neural networks - practical and efficient adversarial attacks that generate adversarial examples from a black-box machine learning model. We will also summarize potential research directions regarding ZO optimization, big data challenges and some open-ended machine learning problems.

TUTORIAL 5: Big Data Analytics for Societal Event Forecasting

Presenters:

Liang Zhao, Assistant Professor
Geroge Mason University
Email: lzhao9@gmu.edu
Feng Chen, Assistant Professor
Univerity at Albany - SUNY
Email: chen5@albany.edu

Abstract:

Spatio-temporal societal event forecasting, which has traditionally been prohibitively challenging, is now becoming possible and experiencing rapid growth thanks to the big data from Open Source Indicators (OSI) such as social media, news sources, blogs, economic

indicators, and other meta- data source. Spatio-temporal societal event forecasting benefits the soci- ety in various aspects, such as political crises, humanitarian crises, mass violence, riots, mass migrations, disease outbreaks, economic instability, resource shortages, responses to natural disasters, and others. Differ- ent from traditional event detection that identifies ongoing events, event forecasting focuses on predicting the future events yet to happen. Also different from traditional spatio-temporal prediction on numerical indices, spatio-temporal event forecasting needs to leverage the heterogeneous in- formation from OSI to discover the predictive indicators and mappings to future societal events. The resulting problems typically require the pre- dictive modeling techniques that can jointly handle semantic, temporal, and spatial information, and require a design of efficient algorithms that scale to high-dimensional large real-world datasets. In this tutorial, we will present a comprehensive review of the state-of- the-art methods for spatio-temporal societal event forecasting. First, we will categorize the inputs OSI and the predicted societal events commonly researched in the literature. Then we will review methods for temporal and spatio-temporal societal event forecasting. We will illustrate the basic theoretical and algorithmic ideas and discuss specific applications in all the above settings.

TUTORIAL 6: Anomaly Detection in Cyber Physical Systems

Presenters:

Maggie Cheng, Associate Professor
Illinois Institute of Technology
Email: maggie.cheng@iit.edu

Abstract:

In a large distributed network, devices generate data with three V's: high volume, high velocity, and high variety. The data are generally unstructured and correlated. To quickly and accurately detect anomalies from the massive amount of data is paramount as detection of anomalies can help identify system faults and enables immediate countermeasures to mitigate faults and to stop fault propagation in the network, and yet it is very challenging as it requires effective detection algorithms as well as adequate understanding of the underlying physical process that generated the data. In this tutorial, we will cover elements of anomaly detection in a networked system, basic detection techniques and their applications in Internet of Things (IoT) and Cyber Physical Systems (CPS). First, will will introduce the concept and categories of anomalies; then we focus on the models and algorithms for anomaly detection, and group the existing detection techniques based on the underlying models and approaches used. The statistical property and algorithmic aspects of the detection methods will be discussed. Subsequently, using communication networks and power grids as examples, we will discuss the application of these detection techniques in application domains. Finally, we will discuss the outlook of this research topic and its relation to other areas of study. We will focus on two broadly defined anomaly detection problems: 1) outlier detection from a large dataset, and 2) change point detection from a dynamic process. Both offline and online algorithms will be discussed.

TUTORIAL 7: Creating Reproducible Bioinformatics Workflows Using BioDepot-workflow- Builder (BwB)

Presenters:

Ling-Hong Hung, Research Scientist
Institute of Technology, University of Washington, Tacoma, WA, USA
Email: lhhung@uw.edu
Ka Yee Yeung, Associate Professor
Institute of Technology, University of Washington, Tacoma, WA, USA
Email: kayee@uw.edu
Wes Lloyd, Assistant Professor
Institute of Technology, University of Washington, Tacoma, WA, USA
Email: wlloyd@uw.edu
Eyhab Al-Masri, Assistant Professor
Institute of Technology, University of Washington, Tacoma, WA, USA
Email: ealmasri@uw.edu

Abstract:

Reproducibility is essential for the verification and advancement of scientific research. It is often necessary, not just to recreate the code, but also the software and hardware environment to reproduce results of computational analyses. Software containers like Docker that distribute the entire computing environment are rapidly gaining popularity in bioinformatics. Docker not only allows for the reproducible deployment of bioinformatics workflows, but also facilitates mix-and-match of components from different workflows that have complex and possibly conflicting software requirements. However, configuration and deployment of Docker, a command-line tool, can be exceedingly challenging for biomedical researchers with limited training in programming and technical skills. We developed a drag and drop GUI called the Biodepot-Workflow-Builder (Bwb) to allow users to assemble, replicate, modify and execute Docker workflows. Bwb represents individual software modules as widgets which are dragged onto a canvas and connected together to form a graphical representation of an analytical pipeline. These widgets allow the user interface to interact with software containers such that software tools written in other languages are compatible

and can be used to build modular bioinformatics workflows. We will present a case study using the BwB to create and execute a RNA sequencing data workflow.

TUTORIAL 8: Managing Big Structured Data for Unsupervised Feature Representation Learning

Presenters:

Lingfei Wu, Research Staff
IBM Research AI, Yorktown Heights, NY 10598
Email: lwu@email.wm.edu, wuli@us.ibm.com
Ian E.H. Yen, PhD
Carnegie Mellon University, Pittsburgh, PA 15213
Email: eyan@cs.cmu.edu

Abstract:

In recent years, there have been a surge of interests in learning expressive representation from large-scale structured inputs, ranging from time-series data, string data, text data, and graph data. Transforming such massive variety of structured inputs into a contextpreserving representation in an unsupervised setting, is a grand challenge to the research community. In many problem domains, it is easier to specify a reasonable dissimilarity (or similarity) function between instances, than to construct a feature representation. Even for complex structured inputs, there are many well-developed dissimilarity measures, such as the Edit Distance (Levenshtein distance) between sequences, Dynamic Time Warping measure between time series, Hausdor. distance between sets, and Wasserstein distance between distributions. However, most standard machine models are designed for inputs with a vector feature representation. Through the proposed tutorial, we aim to present a simple yet principled learning framework for generating vector representations of structured inputs such as time-series, strings, text, and graphs from a well-defined distance function. The resulting representation can then be used for any downstream machine learning tasks such as classification and regression problems. We show a comprehensive catalog of the best practices of generating such vector representations, and demonstrating state-of-the-art performance compared to existing best methods through various analytic applications. We will also share our experiences of various challenges in construction of these representations in di.erent applications in time-series, strings, text, and graph domains.

TUTORIAL 9: Big Data for everyone: Modeling of Data Processing Pipelines for the Industrial Internet of Things

Presenters:

Dr. Dominik Riemer, Department Manager, Senior Researcher
FZI Research Center for Information Technology, Karlsruhe, Germany
Email: riemer@fzi.de
Nenad Stojanovic
Nissatech Innovation Center, Nis, Serbia
Email: nenad.stojanovic@nissatech.com
Ljiljana Stojanovic
Fraunhofer Institute of Optronics, System Technologies, and Image Exploitation IOSB, Karlsruhe, Germany
Email: ljiljana.stojanovic@iosb.fraunhofer.de
Philipp Zehnder
FZI Research Center for Information Technology, Karlsruhe, Germany
Email: zehnder@fzi.de

Abstract:

In many application domains such as manufacturing, the integration and continuous processing of real-time sensor data from the Internet of Things (IoT) provides users with the opportunity to continuously monitor and detect upcoming situations. One example is the optimization of maintenance processes based on the current condition of machines (conditionbased maintenance). While continuous processing of events in scalable architectures is already well supported by the existing Big Data tool landscape (e.g., Apache Kafka, Apache Spark or Apache Flink), building such applications requires a still enormous effort which, besides programming skills, demands for a rather deep technical background on distributed, scalable infrastructures. On the other hand, especially small and medium enterprises from the manufacturing domain often do not have the expertise required to build such programs. Therefore, there is a need for more intuitive solutions supporting the development of real-time applications. In this tutorial, we present methods and tools enabling flexible modeling of real-time and batch processing pipelines by domain experts. We will present ongoing standardization efforts and novel, lightweight, semantics-based models that allow to enhance data streams and stream processing algorithms with background knowledge. Furthermore, we look deeper into graphical modeling of processing pipelines, i.e., stream processing programs that can be defined using graphical tool support and are automatically deployed in distributed stream processors. The tutorial is accompanied by a hands-on session and includes many real-world examples and motivating scenarios we gathered from a number of research and industry projects within the last years.

Workshops

Computational Archival Science *Workshop Chairs: Mark Hedges, Victoria Lemieux, Richard Marciano*		
Time	**Title**	**Presenter/Author**
9:00 – 9:10	Welcome	Mark Hedges (King's College London), Victoria Lemieux (U. British Columbia), Richard Marciano (U. Maryland)
9:10 – 9:45	**Keynote:** Reclaiming our Story: Using Digital Archives to Preserve the History of WWII Japanese-American incarceration	Geoff Froh (Densho.org)
9:45 – 10:05	**Coffee Break**	
10:05 – 11:40	**SESSION 1: Computational Thinking & Computational Archival Science**	
10:05 – 10:30	Introducing Computational Thinking into Archival Science Education	William Underwood (U. Maryland)
10:30 – 10:50	Automating the Detection of Personally Identifiable Information (PII) in Japanese-American WWII Incarceration Camp Records	Richard Marciano (U. Maryland)
10:50 – 11:15	Computational Archival Practice: Towards a Theory for Archival Engineering	Kenneth Thibodeau (NARA)
11:15 – 11:40	Stirring the Cauldron: Redefining Computational Archival Science for the Big Data Domain	Nathaniel Payne (U. British Columbia)
11:40 – 12:10	Discussion and Feedback	Michael Kurtz, Bill Underwood (U. Maryland)
12:10 – 1:30	**Lunch**	
1:30 – 2:05	**Keynote:** Computational Archives & Politics: The Consequences of Failing to Implement Best Practices	Jason R. Baron (Drinker, Biddle & Reath LLP)
2:05 – 2:55	**SESSION 2: Machine Learning in Support of Archival Functions**	
2:05 – 2:30	Protecting Privacy in the Archives: Supervised Machine Learning and Born-Digital Records	Tim Hutchinson (U. Saskatchewan)
2:30 – 2:55	Computer-Assisted Appraisal and Selection of Archival Materials	Christopher Lee (U. North Carolina)
2:55 – 4:10	**SESSION 3: Metadata and Enterprise Architecture**	
2:55-3:20	Measuring Completeness as Metadata Quality Metric in Europeana	Péter Király (U. Göttingen)
3:20-3:45	In-place Synchronisation of Hierarchical Archival Descriptions	Mike Bryant (King's College London)
3:45-4:10	The Utility Enterprise Architecture for Records Professionals	Shadrack Katuu (U. South Africa)
4:10 – 4:30	**Coffee Break**	
4:30 – 5:20	**SESSION 4: Data Management**	
4:30 – 4:55	Framing the scope of the common data model for machine-actionable Data Management Plans	João Cardoso (Instituto Superior Técnico Lisboa)
4:55 – 5:20	The Blockchain Litmus Test	Tyler Smith (Adventium Labs)
5:20 – 6:10	**SESSION 5: Social and Cultural Institution Archives**	
5:20-5:45	A Case Study in Creating Transparency in Using Cultural Big Data: The Legacy of Slavery Project	Ryan Cox (Maryland State Archives), Michael Kurtz (U. Maryland)

5:45-6:10	Jupyter Notebooks for Generous Archive Interfaces	Mari Wigham (Netherlands Institute for Sound and Vision)
6:10-6:30	**Closing Remarks**	

4th International Workshop on Methodologies to Improve Big Data projects *WorkshopChairs: Jeff Saltz*		
Time	**Title**	**Presenter/Author**
2:30 pm	A Variability-Aware Design Approach to the Data Analysis Modeling Process	Cristina Tavares / Paulo Alencar / Donald Cowan
2:55 pm	Improving Data Science Projects by Enriching Analytical Models with Domain Knowledge	Heng Zhang / Jeff Saltz
3:20 pm	What is Good Feedback in Big Data Projects for Cyberinfrastructure Diffusion in e-Science?	Kerk Kee / Jamie McCain
3:45 pm	Will Deep Learning Change How Teams Execute Big Data Projects?	Ivan Shamshurin
4:10 pm	**Coffee Break**	
4:30pm	Big Data Augmentation with Data Warehouse: A Survey	Umar Aftab / Ghazanfar Farooq
4:55pm	Comparative Analysis of Large-scale Network Visualization Tools	Abdul Motaleb Faysal / Shaikh Arifuzzaman
5:20pm	An Approach for Validating Quality of Datasets for Machine Learning	Junhua Ding / XinChuan Li
5:45pm	**Closing Remarks**	

The 5th workshop on Big Data Analytic Technology for Bioinformatics and Health Informatics (KDDBHI 2018) *Workshop Chairs: Donghui Wu and Xin Deng*		
Time	**Title**	**Presenter/Author**
8:00 – 8:10	Chairs' Welcome Remarks	Donghui Wu
8:10 – 8:35	Improving Health Big Data Integration via Word2Vec and Long Short-Term Memory Networks	Zhaohui Liang, Jimmy Xiangji Huang, and Honglai Zhang
8:35 – 9:00	Retrofitting Word Embeddings with the UMLS Metathesaurus for Clinical Information Extraction	Mohammed Alawad, S M Shamimul Hasan, Blair Christian, and Georgia Tourassi
9:00 – 9:25	Clinical Text Classification with Word Embedding Features vs. Bag-of-Words Features	Yijun Shao, Stephanie Taylor, Nell Marshall, Craig Morioka, and Qing Zeng-Treitler
9:25 – 10:00	**Panel: NLP and EMR Text Extraction and Understanding**	Panel Chair: Donghui Wu
10:00 – 10:20	**Coffee Break**	
10:20 – 10:50	Mining Discriminative Patterns from fMRI-based Complete Functional Connectivity Networks	Shah Muhammad Hamdi, Yubao Wu, Berkay Aydin, Soukaina Filali Boubrahimi, Rafal Angryk, Lisa Crystal Krishnamurthy, and Robin Morris

10:50 – 11:20	Development of a Radiology Decision Support System for the Classification of MRI Brain Scans	Alwin Yaoxian Zhang, Sean Shao Wei Lam, Nan Liu, James Yan Pang, Ling Ling Chan, and Phua Hwee Tang
11:20 – 11:50	Towards Biological Sequence Data Service with Insights	Huaming Chen, Jun Shen, Lei Wang, and Chi-Hung Chi
11:50 – 12:00	Morning Session Remarks/Afternoon Session Promotion	Donghui Wu
12:00 – 1:30	**Lunch on Your Own**	
1:30 – 2:00	A Novel Deep Learning Pipeline to Analyze Temporal Clinical Data	Terri Workman, Michael Hirezi, Eduardo Trujillo-Rivera, Anita Patel, Julia Heneghan, James Bost, Qing Zeng-Treitler, and Murray Pollack,
2:00 – 2:30	Research Hypothesis Generation Using Link Prediction in a Bipartite Graph	Jung-Hun Kim and Aviv Segev
2:30 – 3:00	Class Imbalance in Cancer Risk Modeling: A Cloud Computing Approach	Aaron Richter and Taghi Khoshgoftaar
3:00 – 3:30	Spatio-Temporal Convolutional Neural Network For Elderly Fall Detection In Depth Video Cameras	Maryam Rahnemoonfar and Hend Alkittawi
3:30 – 3:50	**Coffee Break**	
3:50 – 4:20	SimREC: Analyzing FDA Adverse Event Reporting System Data for Post-approval Research of Biosimilar Drugs	Rithika Lakshminarayanan, Anurag Joshi, and Rahul Majethia
4:20 – 4:50	Optimization Framework for Flavour and Nutrition Balanced Recipe: A Data Driven Approach	Isura Nirmal, Amith Caldera, and Roshan Dela Bandara
4:50 – 5:20	A Hybrid Approach to Identifying Key Factors in Environmental Health Studies	Shi Dong, Zlatan Feric, Xiangyu Li, Sheikh Mokhlesur Rahman, Guanyu Li, Chieh Wu, April Z. Gu, Jennifer Dy, David Kaeli, John Meeker, Ingrid Y. Padilla, Jose Cordero, Carmen Velez Vega, Zaira Rosario, and Akram Alshawabkeh
5:20 – 6:00	**Panel: Transforming Health Care Delivery via Deep Learning and Machine Learning**	Panel Chair: Donghui
	Closing Remarks	

2nd International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID 2018) *Workshop Chairs:* Stephen McGough and Huaglory Tianfield		
Time	**Title**	**Presenter/Author**
08:30-08:40	Arrival and Welcome	Stephen McGough and Huaglory Tianfield
08:40-09:00	An encoding technique for CNN-based network anomaly detection	Taejoon Kim, Sang Suh, Hyunjoo Kim, Jonghyun Kim and Jinoh Kim
09:00-09:20	Generating interpretable network asset clusters for security analytics	Derek Lin and Anying Li

09:20-09:40	Analyzing evolving trends of vulnerabilities in national vulnerability database	Mark Williams, Sumi Dey, Roberto Barranco, Sheikh Naim, Mahmud Hossain and Monika Akbar
09:40-10:00	Detecting unmanaged and unauthorized devices on the network with long short-term memory network	Derek Lin and Baoming Tang
10:00-10:20	**Coffee Break**	
10:20-10:40	Inline detection of domain generation algorithms with context-sensitive word embeddings	**Joewie Koh and Barton Rhodes**
10:40-11:00	Cyberattack prediction through public text analysis and mini-theories	Ian Perera, Jena Hwang, Kevin Bayas, Bonnie Dorr and Yorick Wilks
11:00-11:20	A natural language processing-based trend analysis of advanced persistent threat techniques	Amirreza Niakanlahiji, Jinpeng Wei and Bei-Tseng Chu
11:20-11:40	TMIXT: A process flow for transcribing mixed handwritten and machine-printed text	Fady Medhat, Mahnaz Mohammadi, **Sardar Jaf**, Chris Willcocks, Toby Breckon, Peter Matthews, Stephen McGough, Georgios Theodoropoulos and Boguslaw Obara
11:40-12:00	Longitudinal analysis of linguistic rigidity of value-motivated groups	Mohammad Al Boni, Seth Green, Megan Stiles, Katherine Harton and Donald E. Brown
12:00-12:10	**Closing Remarks**	

6th International Workshop on Distributed Storage and Blockchain Technologies for Big Data

WorkshopChairs: Hui Li, Kenneth Shum, and Bing Zhu

Time	Title	Presenter/Author
2:30 – 2:50 pm	MystikoDB - Blockchain Meets Big Data	Eranga Herath, Wee Keong, Kasun De Zoysa, Newton Fernando, Supun Tharaka, Maurakiri Nathan, and Namal Jayasuriya
2:50 – 3:10 pm	Converging Blockchain and Social Business for Socio-Economic Development	Raghava Mukkamala, Ravi Vatrapu, Pradeep Kumar Ray, Gora Sengupta, and Sankar Halder
3:10 – 3:30 pm	Blockchain Based Log System	Jiansen Huang, Hui Li, and Jiyang Zhang
3:30 – 3:50 pm	**Coffee Break**	
3:50 – 4:10 pm	Combating Workflow Failures with Integrity-based Checkpoints and Blockchain	Omkar Bhide, Raquel Hill, Karan Vahi, Mats Rynge, and Von Welch
	Closing Remarks	

The 2nd International Workshop on Big Data Analytic for Cyber Crime Investigation and Prevention

Monday, December 10th, 2018; Vashon II Room - 3rd floor Workshop Chairs: Andrii Shalaginov, Katrin Franke, Jan William Johnsen; Norwegian University of Science and Technology		
Time	**Title**	**Presenter/Author**
08:00-08:20	**Opening Remarks and Welcome**	Andrii Shalaginov
08:20-08:40	*File Toolkit for Selective Analysis and Reconstruction (FileTSAR) for Large-Scale Networks*	Raymond Hansen
08:40-09:00	*Detection and Characterization of Human Trafficking Networks Using Unsupervised Scalable Text Template Matching*	Lin Li
09:00-09:20	*Intelligent analysis of digital evidences in large-scale logs in power systems attributed to the attacks*	Asif Iqbal
09:20-09:40	*Blockchain evolution: from Bitcoin to Forensic in Smart Grids*	Igor Kotsiuba
09:40-10:00	*Security Analysis of Mobile Money Applications on Android*	Mohammad Husain
10:00-10:20	**Coffee Break**	
10:20-10:40	*Analyzing Digital Evidence Using Parallel k-means with Triangle Inequality on Spark*	Ambika Shrestha Chitrakar
10:40-11.00	*Camera Model Identification Using Convolutional Neural Networks*	Ruslan Dautov
11.00-11.20	*Towards modelling insiders behaviour as rare behaviour to detect malicious RDBMS access*	Imran Khan
11.20-11.40	*Identification of Attack-based Digital Forensic Evidences for WAMPAC Systems*	Asif Iqbal
11.40-12.00	*A Decision Support System for Personality Based Phishing Susceptibility Analysis*	Mohammad Husain
12.00-12.10	**Closing Remarks**	Andrii Shalaginov

4th IEEE Workshop on Big Data Analytics in Supply Chains and Transportation Chairs: Dr Allan Zhang and Prof Satish Ukkusuri Location: ST. HELENS, Floor 2		
Time	**Title**	**Presenter/Author**
2:00 pm - 2:05 pm	**Opening remarks**	
2:05 pm - 2:25 pm	*Artificial Intelligence and Deep Learning Applications for Automotive Manufacturing*	*Andre Luckow, et al.*
2:25 pm - 2:45 pm	*Visualizing the Impact of Severe Weather Disruptions to Air Transportation*	*Cynthia Glass, et al.*
2:45 pm – 3:05 pm	*Estimation of the economic impact of large-scale flooding in the Tokyo metropolitan area"*	*Shaofeng Yang, et al.*
3:05 pm – 3:25 pm	*Forecast UPC-Level FMCG Demand, Part IV: Statistical Ensemble*	*Dazhi. Yang, et al.*
3:25pm - 3:45pm	*Performing literature review using text mining, Part III: Summarizing articles using TextRank*	*Dazhi. Yang, et al.*
3:45pm- 4.05pm	*Urban Dynamic Logistics Pattern Mining with 3D Convolutional Neural Network*	*Rong Wen*
4:05 pm - 4:25 pm	***Coffee Break***	

4:25 pm - 4:45 pm	*Identification of Traffic Accident Clusters using Kulldorff's Space-Time Scan Statistics"*	*Junxian Song, et al.*
4:45 pm – 5:05 pm	*A hybrid predictive model for high-frequency and multi-periodic data in call center of online travel agency"*	*Shufang Hou, et al.*
5:05 pm - 5:25 pm	**Poster and Network**	
	Revenue Optimized Capacity Provisioning for Integrators in Air Freight Industry Under Uncertainty	*Chi Xu, et al.*
	An Intelligent Water Drops Algorithm to Supply-Demand Hub in Industrial Cluster Considering Transportation Mode	*Vahid Kayvanfar, et al.*
	Adaptive Spatio-temporal Mining for Route Planning and Travel Time Estimation	*Rong Wen*
	Closing remarks	

3rd Open Science in Big Data (OSBD) Workshop

(ADAMS, Floor 2)

Workshop Chairs: Shannon Quinn, John Miller, Suchi Bhandarkar, Nicole Lazar, Kyle Johnsen

Time	Title	Presenter/Author
1:30 – 1:35pm	Welcome and Introduction	Shannon Quinn (OSBD Chair)
1:35 – 2:10	"***Automatic Segmentation and Quantification of TB Scale Volumetric Murine Brain Vasculature Data"***	Katherine Scott (3Scan)
2:10 – 2:45	"***Interpretable Machine Learning in Precision Medicine"***	Prof. Su-In Lee (University of Washington)
2:45 – 3:20	"***Setting Up Your Public Data for Success"***	Dr. Rachael Tatman (Kaggle)
3:20 – 3:40	**Coffee Break**	
3:40 – 3:55	"Towards an Open (Data) Science Analytics-Hub for Reproducible Multi-Model Climate Analysis at Scale"	Sandro Fiore, Donatello Elia, Cosimo Palazzo, Alessandro D'Anca, Fabrizio Antonio, Dean Williams, Ian Foster, and Giovanni Aloisio
3:55 – 4:10	"The iEnvironment Platform: Developing an Open Science Software Platform for Integrated Environmental Monitoring and Modeling of Surface Water"	Paulo Alencar, Donald Cowan, Doug Mulholland
4:10 – 4:25	"Scientific Visualization and Reproducibility for "Open" Environmental Science"	Judith Cushing, Denise Lach, Chad Zanocco, and Jonathan Halama
4:25 – 4:40	"Automatic Segmentation and Quantification of TB Scale Volumetric Murine Brain Vasculature Data"	Venkata Vemuri, Hunter Jackson, and Katherine Scott
4:40 – 4:55	"Toward Simple & Scalable 3D Cell Tracking"	Mojtaba Sedigh Fazli, Stephen A. vella, Silvia N.J. Moreno, Gary E. Ward, and Shannon Quinn
4:55 – 5:10	"Parallelizing Bayesian Knowledge Tracing Tool For Large-scale Online Learning Analytics"	Yanjun Pu, Wenjun Wu, and Yong Han

5:10 – 5:25	"MORF: A Framework for Predictive Modeling and Replication At Scale With Privacy-Restricted MOOC Data"	Joshua Gardner, Miguel Andres-Bray, Christopher Brooks, and Ryan Baker
5:25 – 5:40	"Detecting Anomalies in the LCLS Workflow"	Tal Shachaf, Alex Sim, Kesheng Wu, and Wilko Kroeger
5:40 – 5:45	**Closing Remarks**	

The Second Annual Workshop on Big Data Analytics in the Legal Industry

Workshop Chairs: Jianping Zhang, Robert Keeling, Peter Gronvall, Nathaniel Huber-Fliflet, Haozhen Zhao

Time	Title	Presenter
2:30 pm – 2:50 pm	**Opening Remarks**	
2:50 pm – 3:15 pm	Empirical Evaluations of Seed Set Selection Technique Impact on Predictive Coding's Effectiveness	Christian Mahoney
3:15 pm – 3:40 pm	Break up the Family: Protocols for Efficient Recall-Oriented Retrieval Under Legally-Necessitated Dual Constraints	Jeremy Pickens
3:40 pm – 4:05 pm	Technology Assisted Review of Images using Machine Vision	Thanasis Schoinas
4:10 pm – 4:30 pm	**Coffee Break**	
4:30 pm – 4:55 pm	An Empirical Study of the Application of Machine Learning and Keyword Terms to Privilege Document Review	Nathaniel Huber-Fliflet
4:55 pm – 5:20 pm	Unsupervised Threshold Autoencoder to Analyze and Understand Sentence Elements	Xuan-Hong Dang
5:20 pm – 5:45 pm	Smart Contracts: Legal Considerations	Jack Gilcrest
5:45 pm – 6:10 pm	Empirical Study of Deep Learning for Text Classification in Legal Document Review	Fusheng Wei
6:10 pm – 6:30 pm	**Closing Remarks**	

6th Workshop on Scalable Cloud Data Management

Workshop Chairs: Norbert Ritter, Felix Gessert

Time	Title	Presenter/Author
1:30- 1:40pm	**Opening Remarks**	Norbert Ritter, Felix Gessert (University of Hamburg, Germany)
1:40 - 2:20	**Keynote** Push vs Pull: The Future of Real-Time Databases in the Cloud	Wolfram Wingerath (University of Hamburg, Germany)
	Session I: Data Management	
2:20 - 3:50	Polypheny-DB: Towards a Distributed and Self-Adaptive Polystore	Marco Vogt (University of Basel, Switzerland)

	Adapting to Access Locality via Live Data Migration in Globally Distributed Datastores	Aleksey Charapko (University at Buffalo, SUNY, United States)
	Adaptive Time, Monetary Cost Aware Query Optimization on Cloud DataBase	Chenxiao Wang (University of Oklahoma, USA)
3:50 - 4:10	**Coffee Break**	
4:10 - 6:10	**Session II: Cloud Data and Systems**	
	Recurrent Movement of Relational Data Within a Hybrid Cloud	Sean Rooney (IBM Research Zurich, Switzerland)
	ACTOR: Active Cloud Storage with Energy-Efficient On-Drive Data Processing	Zhi Qiao (University of North Texas, United States)
	Skew-Aware Collective Communication for MapReduce Shuffling	Harunobu Daikoku (University of Tsukuba, Japan)
	Reliability Characterization of Solid State Drives in a Scalable Production Datacenter	Shuwen Liang (University of North Texas, United States)
6:10 - 6:15	**Closing Remarks**	

		BSD 2018
Start	**End**	**Event**
7:20	**8:15**	**Registration**
8:15	**09:45**	**Paper Presentation Session: Data Management**
8:15	8:45	**Leveraging Spatio-Temporal Soccer Data to Define a Graphical Query Language for Game Recordings** ● Keven Richly
8:45	9:15	**Concept and Analysis of Information Spaces to improve Prediction-Based Compression** ● Ugur Cayoglu, Frank Tristram, Jörg Meyer, Tobias Kerzenmacher, Peter Braesicke, and Achim Streit
9:15	9:45	**Accelerating Cross-Matching Operation of Geospatial Datasets using a CPU/GPU Hybrid Platform** ● Chao Gao, Furqan Baig, Hoang Vo, Yangyang Zhu, and Fusheng Wang
9:45	**10:05**	**Coffee Break**
10:05	**11:05**	**Paper Presentation Session: Pattern Discovery**
10:05	10:35	**Deriving Real-time City Crowd Flows by Heterogeneous Big Urban Data** ● Bo Tang, Chuan Yang, Long Xiang, and Jian Zeng
10:35	11:05	**Impact of Trajectory Segmentation on Discovering Trajectory Sequential Patterns** ● Somayah Karsoum
11:05	**12:05**	**Paper Presentation Session: Mobile**
11:05	11:35	**A data-driven impact evaluation of Hurricane Harvey from mobile phone data** ● Aude Marzuoli and Fengmei Liu
11:35	12:05	**RiSC: Quantifying change after natural disasters to estimate infrastructure damage with mobile phone data** ● Xavier Andrade, Fabricio Layedra, Carmen Vaca, and Eduardo Cruz
12:10	**13:20**	**Lunch**
13:20	**14:05**	**Paper Presentation Session: Learning and Data Mining**

13:20	13:50	**On Network Embedding for Machine Learning on Road Networks: A Case Study on the Danish Road Network** ● Tobias Skovgaard Jepsen, Christian Søndergaard Jensen, Thomas Dyhre Nielsen, and Kristian Torp
13:50	14:05	**Semantic Segmentation of Complex Road Environments from Aerial Images Using Convolutional Neural Networks (Short)** ● David Schweitzer and Rajeev Agrawal
14:05	**15:05**	**Keynote: John Krumm, Compute Useful Insights from Location Data**
15:05	**15:25**	**Coffee Break**
15:25	**17:10**	**Paper Presentation Session: Data Analysis**
15:25	15:55	**Multi-Class Object Detection from Aerial Images Using Mask R-CNN** ● David Schweitzer and Rajeev Agrawal
15:55	16:25	**Spatio-Temporal Multiple Geo-Location Identification on Twitter** ● Kambiz Ghoorchian and Sarunas Girdzijauskas
16:25	16:40	**Trajectory Cluster Lifecycle Analysis: An Evolutionary Perspective (short)** ● Ivens Portugal, Paulo Alencar, and Donald Cowan
16:40	17:10	**A Practical Expert System with (Near) Real-Time Analysis of Large Spatial Sets of Air Traffic Data** ● Vasily Sidorov, Wee Keong Ng, and Mohamed Faisal Bin Mohamed Salleh
17:10	**17:20**	**Adjourn**

The Second IEEE Workshop on Human-in-the-loop Methods and Human-Machine Collaboration in BigData (HMData 2018)

Workshop Chair: Senjuti Basu Roy (New Jersey Institute of Technology), Lei Chen (HKUST), Atsuyuki Morishima (University of Tsukuba)

The detailed program is available at http://humanmachinedata.org

Time	Title	Presenter/Author
8:45	**Opening**	
8:50	Exploring the Potential of Modern Advanced Metering Infrastructure in Low-Voltage Grid Monitoring Systems	Maria Stefan, Jose Gutierrez Lopez, and Rasmus Løvenstein Olsen
9:20	Time-Lapse Image Generation using Image-Based Modeling by Crowdsourcing	Hidehiko Shishido, Emi Kawasaki, Yutaka Ito, Youhei Kawamura, Toshiya Matsui, and Itaru Kitahara
9:30	A Cache-based Approach to Dynamic Switching between Different Dataflows in Crowdsourcing	Yusuke Suzuki, Masaki Matsubara, Keishi Tajima, Toshiyuki Amagasa, and Atsuyuki Morishima
9:40	Cyber-Physical Disaster Drill: Preliminary Results and Social Challenges of the First Attempts to Unify Human, ICT and AI in Disaster Response	Flavia Fulco, Munenari Inoguchi, and Tomoya Mikami
9:50	Worker Classification based on Answer Pattern for Finding Typical Mistake Patterns	Tomoya Mikami, Masaki Matsubara, Takashi Harada, and Atsuyuki Morishima
10:00	Toward Explainable Recommendations: Generating Review Text from Multicriteria Evaluation Data	Takafumi Suzuki, Satoshi Oyama, and Masahito Kurihara
10:10	**Coffee Break**	
10:30	Realization of Effective Team Management Collaborating between Cloud-based System and On-site Human Activities -A Case Study of	Keiko Tamura, Munenari Inoguchi, Kei Horie, Ryota Hamamoto, and Haruo Hayashi

	Building Damage Inspection at 2018 Hokkaido Eastern Iburi Earthquake-	
11:00	A Learning Effect by Presenting Machine Prediction as a Reference Answer in Self-correction	Masaki Matsubara, Masaki Kobayashi, and Atsuyuki Morishima
11:30	Interactive Machine Learning by Visualization: A Small Data Solution	Huang Li, Shiaofen Fang, Snehasis Mukhopadhyay, Andrew Sakin, and Li Shen
11:40	Finding Evidences by Crowdsourcing	Nadeesha Wijerathna, Masaki Matsubara, and Atsuyuki Morishima
11:50	A Method to Collect Multi-view Images of High Importance Using Disaster Map and Crowdsourcing	Koyo Kobayashi, Hidehiko Shishido, Yoshinari Kameda, and Itaru Kitahara
12:00	**Lunch**	
13:30	**Keynote:** Decision Theoretical Crowdsourcing	Dan Weld (University of Washington)
14:30	A Case Study on Start-up of Dataset Construction: In Case of Recipe Named Entity Corpus	Yoko Yamakata, Keishi Tajima, and Shinsuke Mori
15:00	**Coffee Break (with Posters)**	
15:40	Toward Human-in-the-Loop Collaboration of Software Engineers and Machine Learning Algorithms	Nathalia Nascimento, Paulo Alencar, Carlos Lucena, and Donald Cowan
16:10	Identification of Important Images for Understanding Web Pages	Ying Zhong, Masaki Matsubara, and Atsuyuki Morishima
16:40	A Task Assignment Method Considering Inclusiveness and Activity Degree	Hirotaka Hashimoto, Masaki Matsubara, Yuhki Shiraishi, Daisuke Wakatsuki, Jianwei Zhang, and Atsuyuki Morishima
17:10	Implementation of Effective Field Survey for Damaged Buildings under Harmonious Collaboration between Human and ICT - A Case Study of 2018 Hokkaido Eastern Iburi Earthquake -	Munenari Inoguchi, Keiko Tamura, Kei Horie, Ryota Hamamoto, and Haruo Hayashi
17:40	**Closing Remarks**	

Analysis of Large-scale Disparate Data

Workshop Chairs: Dr. Michael Barton, Dr. Wesley Griffin, Prof. Dhabaleswar Panda, Brian Panneton, Dr. Simon Su

Time	Title	Presenter/Author
8:00	Opening Remarks	
8:10	Application of Comprehensive Data Analysis for Interactive, Hierarchical Views of HPC Workloads	Matthew Dwyer
8:35	Proposal of Scalable and Performing Implementation of Algorithms for Anomaly and Community Detection	Prof. Yaya Sylla
9:00	Visually Analyzing A Billion Tweets: An Application for Collaborative Visual Analytics on Large High-Resolution Display	Dr. Simon Su
9:25	Visual computation and simulation of path loss effects on tactical networks in urban canyon	Dr. Venkat Dasari

9:50	**Coffee Break**	
10:10	Using Big Data Analytics to Create a Predictive Model for Joint Strike Fighter	Ryan Norman
10:35	Collaborative Visual Analytics for Streaming Flight Data	Prof. Bo (Beth) Sun
11:00	Using Cartograms to Visualize Population Normalized Big-Data Sets	Prof. Anthony Breitzman
11:25	**Closing Remarks**	

Advances in High Dimensional (AdHD) Big Data

Workshop Chairs: Sotiris Tasoulis, Vassilis Plagianakos

Time	Title	Presenter/Author
14:30 – 14:55	*Subspace Clustering of Very Sparse High-Dimensional Data*	Hankui Peng, Nicos Pavlidis, Idris Eckley, and Ioannis Tsalamanis
14:55 – 15:20	*Analysis of large sparse graphs using regular decomposition of graph distance matrices*	Hannu Reittu, Lasse Leskelä, Tomi Räty, and Marco Fiorucci
15:20 – 15:45	*An Integrative Analysis of Time-varying Regulatory Networks From High-dimensional Data*	Zi Wang, Yun Guo, and Haijun Gong
15:45 – 16:10	*Segmentation of Time Series in Improving Dynamic Time Warping*	Ruizhe Ma, Azim Ahmadzadeh, Soukaina Filali Boubrahimi, and Rafal Angryk
16:10 – 16:30	**Coffee Break**	
16:30 – 16:55	*dynnode2vec: Scalable Dynamic Network Embedding*	Sedigheh Mahdavi, Shima Khoshraftar, and Aijun An
16:55 – 17:20	*Temporal Graph Offset Reconstruction: Towards Temporally Robust Graph Representation Learning*	Stephen Bonner, John Brennan, Ibad Kureshi, Georgios Theodoropoulos, Stephen McGough, and Boguslaw Obara
17:20 – 17:45	*Divide-and-Conquer Kronecker Product Decomposition for Memory-Efficient Graph Approximation*	Venkata Suhas Maringanti and Ming Shao
17:45 – 18:10	*Automating Relevance Banding in e-Commerce Search using Click Model*	Deependra Singh and Vinay Deolalikar
18:10 – 18:35	*Trustworthy data processing for health analytics tasks*	Kostas Moutselos, Dimosthenis Kyriazis, Vasiliki Diamantopoulou, and Ilias Maglogiannis
18:35 – 18:40	**Closing Remarks**	

The Second IEEE International Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD), December 10

Workshop Chairs: Zhiyuan Chen, Jianwu Wang, Feng Chen, Yiming Ying

Time	Title	Presenter/Authors
8:30-8:35	Welcome, opening remark	Zhiyuan Chen & Feng Chen

	Session 1: Workflow, Benchmark, Scientific Data	
8:35-8:55	Hoagie: A Database and Workload Generator using Published Specifications	Shahram Ghandeharizadeh, Haoyu Huang
8:55-9:20	Deep Learning for Enhancing Fault Tolerant Capabilities of Scientific Workflows	Alok Singh, Ilkay Altintas, Malachi Schram, Nathan Tallent
9:20-9:40	Blockchain Based Provenance Sharing of Scientific Workflows	Wanghu Chen, Xiaoyan Liang, Jing Li, Hongwu Qin, Yuxiang Mu, and Jianwu Wang
9:40-10:00	Toward Scalable Analysis of Multidimensional Scientific Data: A Case Study of Electrode Arrays	Ye Niu, Abdullah Al-Mamun, Hui Lin, Tonglin Li, Yi Zhao, and Dongfang Zhao
10:00-10:20	Coffee Break	
	Session 2: Machine Learning & Analytics I	
10:20-10:45	A study of Exact Ridge Regression for Big Data	Wanchih Chiang, Xiang Liu, Tonglin Zhang, and Baijian Yang
10:45-11:10	A Flexible-blocking Based Approach for Performance Tuning of Matrix Multiplication Routines for Large Matrices with Edge Cases	Md Mosharaf Hossain, Thomas M. Hines, Sheikh Rabiul Islam, Sheikh K. Ghafoor, and Ramakrishnan Kannan
11:10-11:35	Performance and Memory Trade-offs of Deep Learning Object Detection in Fast Streaming High-Definition Images	Aishwarya Srivastava, Dung Nguyen, Siddhant Aggarwal, Andre Luckow, Edward Duffy, Ken Kennedy, Marcin Ziolkowski, and Amy Apon
11:35-12pm	City-Wide Influenza Forecasting based on Multi-Source Data	Kun Su, Yu Xiong, Li Qi, Yu Xia, Baisong Li, Lin Yang, Qin Li, Wenge Tang, Xian Li, Xiaowen Ruan, Shaofeng Lu, Xianxian Chen, Chaobo Shen, Boran Hong, Jiaying Xu, Liang Xu, Mei Han, and Jing Xiao
12-1:30	**Lunch**	
1:30-2:15	**Invited Talk:** Benchmarking for Big Data Applications with the DataBench Framework	Dr. Arne Berre
	Session 3: Optimization and Tuning	
2:15-2:40	Enhancing the Scalability and Performance of Iterative Graph Algorithms on Apache Storm	Sachini Jayasekara, Shanika Karunasekera, and Aaron Harwood
2:40-3:05	One Self-Adaptive Memory Scheduling Algorithm for the Shuffle Process in Spark Platform	Jungang Xu, Shanshan Huang, Renfeng Liu, and Pengfei Li
3:05-3:30	Big data storage technologies: a case study for web-based LiDAR visualization	David Deibe, Margarita Amor, and Ramón Doall
3:30-3:50	Coffee Break	
	Session 4: Machine Learning and Analytics II	
3:50-4:15	Predicting the Computational Cost of Deep Learning Models	Daniel Justus, John Brennan, Stephen Bonner, and Stephen McGough
4:15-4:40	Key based Deep Data Locality on Hadoop	Sungchul lee, Juyeon Jo, and Yoohwan Kim,
4:40-5:00	An Efficient Multi-Objective Genetic Algorithm for Cloud Computing: NSGA-G	Trung-Dung Le, Verena Kantere, and Laurent d'Orazio
5:00-5:20	The Challenge of a Strong Speed-Up of a Bio-Medical Big Data Application	Marco Strutz, Bjoern Lindequist, Hermann Hessling, and Achim Streit

5:20-5:40	In-memory Blockchain: Toward Efficient and Trustworthy Data Provenance for HPC Systems	Abdullah Al-Mamun, Tonglin Li, Mohammad Sadoghi, and Dongfang Zhao
5:40-6:00	Achieving Horizontal Scalability in Density-based Clustering for URLs	Azadeh Faroughi, Reza Javidan, Marco Mellia, Andrea Morichetta, Francesca Soro, and Martino Trevisan
6-6:05	Closing remarks	Zhiyuan Chen & Feng Chen

GTA ^{3}2.0: The 2nd workshop on Graph Techniques for Adversarial Activity Analytics

Workshop Chairs: Jiejun Xu, Hanghang Tong, Andrea Bertozzi, Vince Lyzinski, George Chin

Time	Title	Presenter/Author
1:25pm – 1:30pm	Opening Remarks	Workshop organizers
1:30pm – 2:20pm	Keynote 1	Prof. Carey Priebe
2:20pm – 3:10pm	Keynote 2	Prof. Andrea Bertozzi
3:10pm – 3:25pm	An Empirical Assessment of the Complexity and Realism of Synthetic Social Contact Networks	Kiran Karra, Samarth Swarup, and Justus Graham
3:25pm – 3:40pm	A Chronological Edge-Driven Approach to Temporal Subgraph Isomorphism	Patrick Mackey, Katherine Porterfield, Erin Fitzhenry, Sutanay Choudhury, George Chin Jr.
3:40pm – 4:00pm	**Coffee Break**	
4:00pm – 4:50pm	Keynote 3	Dr. Christopher White
4:50pm – 5:05pm	From Gamergate to FIFA: Identifying Polarized Groups in Online Social Media	Dana Warmsley, Jiejun Xu, Tsai-Ching Lu
5:05pm – 5:20pm	Filtering Methods for Subgraph Matching on Multiplex Networks	Jacob D. Moorman, Qinyi Chen, Thomas K. Tu, Zachary M. Boyd, Andrea L. Bertozzi
5:20pm – 5:35pm	Multi-Channel Large Network Simulation Including Adversarial Activity	Joseph A. Cottam, Sumit Purohit, Patrick Mackey, George Chin Jr.
5:35pm – 5:50pm	Using Web-Scale Graph Analytics to Counter Technical Support Scams	Jonathan Larson, Bryan Tower, Duane Hadfield, Darren Edge, Christopher White
5:50pm – 6:05pm	Trimming the Hairball: Edge Cutting Strategies for Making Dense Graphs Usable	Darren Edge, Jonathan Larson, Markus Mobius, Christopher White
6:05pm – 6:10pm	**Closing Remarks**	

2018 BigGraphs Workshop at IEEE BigData'18

WorkshopChairs: Nesreen Ahmed, Mohammad Al Hasan, Shaikh Arifuzzaman, and Kamesh Madduri

Time	Title	Presenter/Author
8:05 am	**Opening Remarks:** Shaikh Arifuzzaman and Nesreen Ahmed	
8:15 am	*Scalability Analysis of Cluster-based Betweenness Computation in Large Weighted Graphs*	Andrea Castiello, Gianmarco Fucci, Angelo Furno, and Eugenio Zimeo

8:35 am	*PatBinQL, a compact, inference-enabled query language dedicated for RDF stream processing*	Jérémy Lhez, Badre Belabbess, and Olivier Curé
8:55 am	*Single-Source Shortest Path Tree For Big Dynamic Graphs*	Sara Riazi, Sriram Srinivasan, Sajal K. Das, Sanjukta Bhowmick, and Boyana Norris
9:15 am	*RECUT: RE-Compressing partially Unordered Trees*	Stefan Böttcher and Rita Hartel
9:35 am	*Mobility Optimization on Cyber Physical System via Multiple Object Tracking and Mathematical Programming*	Nozomi Hata, Takashi Nakayama, Akira Tanaka, Takashi Wakamatsu, Akihiro Yoshida, Nariaki Tateiwa, Yuri Nishikawa, Jun Ozawa, and Katsuki Fujisawa
10:00 am	**Coffee Break**	
10:20 am	**Keynote by Prof. Jennifer Neville**	
11:10 am	BiasedWalk: Biased Sampling for Representation Learning on Graphs	Duong Nguyen, Fragkiskos Malliaros
11:30 am	Social-Based Classification of Multiple Interactions in Dynamic Attributed Networks	Thiago H. P. Silva, Pedro O. S. Vaz de Melo, Alberto H. F. Laender
11:50 pm	*Automatic Hierarchical Clustering of Static Call Graphs for Program Comprehension*	Gharib Gharibi, Rakan Alanazi, and Yugyung Lee
12:10 pm	**Closing Remarks:** Nesreen Ahmed and Shaikh Arifuzzaman	

The 1st International Workshop on Big Data for Marketing Intelligence and Operation Management *WorkshopChairs: Wutao Wei / Yiwen Zhang / Huaiye Zhang*		
Time	**Title**	**Presenter/Author**
1:30 – 1:35	***Opening Remark***	
1:35 – 2:00	*pRNN: A Recurrent Neural Network based Approach for Customer Churn Prediction in Telecommunication Sector*	Jinlong Hu, Yi Zhuang, Jiang Yang, Lei Lei, Minjie Huang, Runchao Zhu, and Shoubin Dong
2:00 - 2:25	*Enhancing Talent Search by Integrating and Querying Big HR Data*	Otman MANAD, Mehdi BENTOUNSI, and Patrice Darmon
2:25 – 2:50	*Implications of the General Data Protection Directive for Virtual Personal Marketing Assistants*	Bogdan Hoanca, Christina McDowell Marinchak, and Edward Forrest
2:50 – 3:15	*Whales, Dolphins, or Minnows? Towards the Player Clustering in Free Online Games Based on Purchasing Behavior via Data Mining Technique*	Wanshan Yang, Gemeng Yang, Ting Huang, Lijun Chen, and Youjian (Eugene) Liu
3:15 – 3:40	*Know your customer: Detection of Customer Experience (CX) in Social Platforms using Text Categorization*	Leonardo Kuffo, Carmen Vaca, Juan Carlos Bustamante, and Edgar Izquierdo
3:40 – 3:55	**Coffee Break**	

3:55 – 4:10	*Restricted Boltzmann Machines on Recommendation Systems with Implicit Feedback*	Fan Yang and Ying Lu
4:10 – 4: 35	*Application of distributed back propagation neural network for dynamic real-time bidding*	Ankit Desai and Sanjay Chaudhary
4:35 – 5:00	*Graph-based Dynamic Route Planning for Collective Customer Matching*	Ajitesh Srivastava, Kristopher Rollert, Rajgopal Kannan, Viktor Prasanna, Yinglong Xia, Masoud Monajatipoor, and Kartik Lakhotia
	Closing Remarks	

The Second Big Data for Economic and Business Forecasting *WorkshopChairs: Wei Shang, Xiangbin Yan*		
Time	**Title**	**Presenter/Author**
8:20-8:40	A Comparative Study of LSTM and DNN for Stock Market Forecasting	**Dev Shah**
8:40-9:00	Massive text mining for abnormal market trend detection	Ying Li, ting jin, Meng Xi, ShengPeng Liu, and Zhiling Luo
9:00-9:20	The Impact of Antidumping Protection Against China: Evidence From Traditional And New Antidumping Users	**Qin Bao**
9:20-9:40	Forecasting crude oil prices based on an internet search driven model	**Rui Tao**
9:40-10:00	Predicting Stock Price Trend Using Candlestick Chart Blending Technique	**Yoshihisa Udagawa**
10:00-10:20 am	**Coffee Break**	
10:20-10:40	Application of LSTM Neural Network for Urban Road Diseases Trend Forecasting	**Ailin Chen**
10:40-11:00	Finding Skill Similarity Matrix Utilizing Expert Recommended Skill Clusters	**Brian Johnston**
11:00-11:20	Prediction of Movie Playback Based on Ordinal Support Vector Machine Classification	**Wei Shang**
11:20-11:40	Research on the Evolution of the Influence of Opinion Leaders in Social Networking Sites Taking Zhihu.com as an Example	Chenxiao Jin, Bo Liu, Wei Qi, and Juan Qin
11:40-12:00	Research on User Consumption Behavior Prediction Based on Improved XGBoost Algorithm	**Yan Xiangbin**
	Closing Remarks	

Big Data for Digital Twins

Arne Berre, Ljiljana Stojanovic		
Time	**Title**	**Presenter/Author**
2:30-2:50	INVITED TALK: Feeding the Digital Twin: Basics, Models and Lessons Learned from Building an IoT Analytics Toolbox	Dominik Riemer
Semantic Models and Ontologies for Digital Twins		
2:50-3:00	Towards Semantically Enhanced Digital Twin	Evgeny Kharlamov
3:00-3:10	Representing Industrial Data Streams in Digital Twins using Semantic Labeling	Philipp Zehnder
3:10-3:20	Linking an Asset and a Domain Specific Ontology for a Simple Asset Time Series Application	Victor Danilchenko
Digital Twins in practice -Big Data with real-time IoT		
3:20-3:30	Data-driven Digital Twin approach for process optimization: an industry use case	Nenad Stojanovic
3:30-3:40	Simulation-ready digital twin for real-time management of logistics systems	Benjamin Korth
3:40-3:50	Continuous real-time anomaly detection in flexible production: D2Lab-based use case	Milan Jovic
3:50-4:00	MyFitnessDigitalTwin Data Analytics driven Continuous Improvement of the Trainee Performances in the Fitness	Milan Djordje
4:00-4:10	**Concluding discussion – common approaches and issues emerging from the workshop**	

The 5th Workshop on Advances in Software and Hardware for Big Data Sciences *WorkshopChairs: Hui Zhang, Weijia Xu, Hongfeng Yu*		
Time	**Title**	**Presenter/Author**
2:30 – 2:55pm	Scalable Record Linkage	Luke Wolcott
2:55pm – 3:20pm	Performance Analysis of Divide-and-Conquer strategies for Large scale Simulations in R	Ranjini Subramanian
3:20pm – 3:45pm	Integrated HPC Scheduler Data Processing Workflow using Apache Zeppelin	Fang Liu
3:45pm – 4:10pm	Untangling Mathematical Knots with Simulated Annealing and Opposition-Based Learning	Juan Lin
4:10pm – 4:30pm	**Coffee Break**	
4:30pm – 4:55pm	3D Reconstruction of Plant Leaves for High-Throughput Phenotyping	Feiyu Zhu
4:55pm – 5:20pm	A Low-Overhead Integrity Verification for Big Data Transfers	Engin Arslan

5:20pm – 5:45pm	Enabling User Driven Big Data Application on Remote Computing Resources	Weijia Xu
5:45pm – 6:10pm	Scaling Collaborative Filtering with PETSc	Alister
6:10pm – 6:30pm	**Closing Remarks**	

Workshop on Identifying and Combating Disinformation in Big Data

Workshop Chairs: Tyler Smith and Brian Isle

Time	Title	Presenter/Author
8:00	Opening Remarks	Workshop Chairs
8:10	Keynote Presentation	Simon Bracey-Lane
8:30	Political Warfare: Competition in the Cyber Era	Antonios Nestoras
8:50	Leveraging Archival Theory to Develop A Taxonomy of Online Disinformation	Dr. Victoria Lemieux and Tyler Smith
9:10	The Alt Right and Global Information Warfare	Alexander Reid Ross and Emmi Bevensee
9:30	Trustworthiness of Citizen Journalists Videos from the Perspective of Archival Science	Hoda Hamouda
9:50	**Coffee Break**	
10:00	The Strategic Need to Understand Online Memes and Modern Information Warfare Theory	Greg Rowett
10:20	Real World Examples Suggest a Path to Automated Mitigation of Disinformation	Brian Isle
10:40	Multiple Time-Series Data Analysis for Rumor Detection on Social Media	Chandra Mouli Madhav Kotteti, Xishuang Dong, and Lijun Qian
11:00	Fake News: A Method to Measure Distance from Fact	Dr. Char Sample, Dr. Connie Justice, and Dr. Emily Darraj
11:20	Countering Inside Threat Actors in Algorithm-Based Media	Tyler Smith
11:40	**Closing Remarks**	

Big Data and AI for Air Quality Estimation, Forecasting, and Health Advice

WorkshopChairs: Victor OK Li, University of Hong Kong (HKU), Jacqueline CK Lam, HKU, Mihaela van der Schaar, University of Oxford, and Ingemar Cox, University College London

Time	Title	Presenter/Author
9:00 – 9:20 am	A Bayesian LSTM Model to Evaluate the Effects of Air Pollution Control Regulations in China	Yang Han, Jacqueline CK Lam, and Victor OK Li
9:20 – 9:40 am	PM2.5 Forecasting Using Pre-Trained Components	Ming-Chuan Yang and Meng Chang Chen
9:40 – 10:00 am	Guiding the Data Learning Process with Physical Model in Air Pollution Inference	Rui Ma, Xiangxiang Xu, Yue Wang, Hae Young Noh, Pei Zhang, and Lin Zhang

10:00 – 10:20 am	Optimization of Urban Heating Network Design Using Genetic Algorithm	Andong Wang, Victor O.K. Li, Jacqueline C.K. Lam
10:20 – 10:40 am	**Coffee Break**	
10:40 – 11:00 am	Do People Move Away From Bad Air? Preliminary Evidence From Twitter	Zhiyi Lu, Jacqueline C.K. Lam, Victor O.K. Li, Yang Han, and Zafar Gilani
11:00 – 11:20 am	Unsupervised Domain Adaptation with Generative Adversarial Networks for Facial Expression Recognition	Yingruo Fan, Jacqueline C.K. Lam, and Victor O.K. Li
11:20 – 11:40 am	Round Table Discussion – How do we develop Ethical and Socially Beneficial AI for Well-being and Society?	Victor OK Li and Jacqueline CK Lam
11:40 – 12 noon		
12:00 – 12:10 pm	**Closing Remarks**	

Applications of Big Data Technology in the Transport Industry *Workshop Chair: Dr John Easton, University of Birmingham, UK*		
Time	**Title**	**Presenter/Author**
13:30 – 13:45	The Process of Building Railway Digital Twins through the Convergence of IT and OT	Diego Galar, Dammika Seneviratne, and Roberto Villarejo
13:50 – 14:05	Deep Reinforcement Learning Approach for Train Rescheduling Utilizing Graph Theory	Mitsuaki Obara, Takehiro Kashiyama, and Yoshihide Sekimoto
14:10 – 14:25	Towards Encrypting Industrial Data on Public Distributed Networks	Joseph Preece and John Easton
14:30 – 14:45	A Software Framework for Cluster Lifecycle Analysis in Transportation	Ivens Portugal, Paulo Alencar, and Donald Cowan
14:50 – 15:05	A Three-step Agglomerated Machine Learning: An Alternative to Weibull Defect Analysis of Rail Infrastructure	Ahmed Lasisi, Emmanuel Martey, Dominique Guillot, and Nii Attoh-Okine
15:10 – 15:25	Big Data Analytics for Linear Asset Management in Transportation	Roberto Villarejo, Diego Galar, and Dammika Seneviratne
	Break	
15:50 – 16:05	Consolidating Billions of Taxi Rides with AWS EMR and Spark in the Cloud	Alex Kaplunovich
16:10 – 16:25	Big Data and Smart City Planning: The Case of Owl Bus in Seoul	Sounman Hong, Youngrok Kim, and Jeongin Park
16:30 – 16:45	Big data in Railway Rolling Stock Maintenance: Passenger Train Subsystem Maintenance	Dammika Seneviratne, Diego Galar, and Roberto Villarejo
16:50 – 17:05	Review: The Potential use of Blockchain Technology in Railway Applications - An Introduction of a Mobility and Speech Recognition Prototype	Feras Naser
17:10 – 17:25	Modeling and Simulation of HVAC System for Failures in High Speed Trains	Dammika Seneviratne, Diego Galar, and Roberto Villarejo
	Closing Remarks	

Big Data Engineering and Analytics in Cyber-Physical Systems (BigEACPS'18) *WorkshopChairs: Akbar Siami Namin*		
Time	**Title**	**Presenter/Author**
Modeling [1:30 – 2:45]		
1:30 – 1:50	A multi-variable Stacked Long-Short Term Memory Network for Wind Speed Forecasting	Sisheng Liang, Long Nguyen, Fang Jin
1:55 – 2:15	Defending SDN-based IoT Networks Against DDoS Attacks Using Markov Decision Process	Jianjun Zheng, Akbar Siami Namin
2:20 – 2:45	Predicting Customer Behaviors on Energy Consumption: Why past usage data are not enough?	Supadchaya Puangpontip, Rattikorn Hewett
Survey and Position Papers [2:45 – 3:30]		
2:45 – 3:05	Detect Hidden Road Hazards Combining Multiple Social Media Data	Fang Jin, Hongchao Liu
3:10 – 3:30	A Survey of Privacy Concerns in Wearable Devices	Prerit Datta, Akbar Siami Namin, Moitrayee Chatterjee
	Coffee Break [3:30 – 3:50]	
Security and Privacy [3:50 – 4:30]		
3:50 – 4:10	Towards Prediction of Security Attacks on Software Defined Networks: A Big Data Analytics Approach	Emre Unal, Sonali Sen Baidya, Rattikorn Hewett
4:10 – 4:30	Evidence Fusion for Malicous Bot Detection in IoT	Moitrayee Chatterjee, Akbar Siami Namin, Prerit Datta
Data Analytics and Visualization [4:30 – 6:00]		
4:30 – 5:00	IoTNegViz: An Interactive Tool for Visualizing Negative Aspects of IoT	Huyen N. Nguyen, Vinh T. Nguyen, Ngan V.T. Nguyen, Vung Pham, Tommy Dang
5:00 – 5:30	Unleashing the Power of Hashtags in Tweet Analytics with Distributed Framework on Apache Storm	Vibhuti Gupta, Rattikorn Hewett
5:30 – 6:00	IoTViz: Visualizing Emerging Topics in the Internet of Things	Vung Pham, Vinh T. Nguyen, Tommy Dang
	Closing Remarks	

The 1st International Workshop on Big Video Dataset Construction, Management and Applications *Workshop Chairs: Rui Wang, Cheng Jin, Mingyu You, Haimiao Hu, Shengcai Liao, and Mingli Song*		
Time	**Title**	**Presenter/Author**
13:30-13:50	Watermarking Based Data Spoofing Detection Against Speech Synthesis and Impersonation with Spectral Noise Perturbation	Xuping Huang
13:50-14:10	Improve Cross-Domain Face Recognition with IBN-block	Yangchun Qing, Yafei Zhao, Yongjie Shi, Dan Chen, Yining Lin, and Yao Peng

14:10-14:30	Image Quality Assessment Based on BIQI with Gray Beep	Weipeng Wu, Chaoqun Hong, Yong Xie, and Liang Chen
14:30-14:50	Salient Object Detection with Convex Hull Overlap	Yongqing Liang
14:50-15:10	Adaptive Query Re-ranking Based on ImageGraph for Image Retrieval	Haonan Fan, Hai-Miao Hu, Rong Wang, and Yugui Zhang
15:10-15:30	Semi-automatic Data Annotation Tool for Person Re-identification Across Multi Cameras	Tianyi Zhao, Shengcai Liao, and Zhen Lei
15:30-15:50	**Coffee Break**	
15:50-16:10	Evenly Cascaded Convolutional Networks	Chengxi Ye, Chinmaya Devaraj, Michael Maynord, Cornelia Fermüller, and Yiannis Aloimonos
16:10-16:30	AutoHighlight : Automatic Highlights Detection and Segmentation in Soccer Matches	Kaiyu Tang, Yixin Bao, Zhijian Zhao, Liang Zhu, Yining Lin, and Yao Peng
16:30-16:50	PepAls: Performance Prediction and Algorithm Selection Framework for Data Mining Applications	Mingyu You, Xuanhui Xu, and Zheng Wang
16:50-17:10	Improving the Optical Flow Accuracy Based on the Total Variation of Local-Global method	Yugui Zhang, Haonan Fan, Jin Zheng, and Chi Zhang
17:10-17:30	Is Re-ranking Useful for Open-set Person Re-identification?	Hongsheng Wang, Shengcai Liao, Zhen Lei, and Yang Yang
17:30-17:50	Semi-Supervised Dictionary Learning Based on Atom Graph Regularization	Xiaoqin Zhang, Qianqian Liu, Di Wang, Jie Hu, Nannan Gu, and Tianhao Wang
17:50-18:00	**Closing Remarks**	

The 2nd International Workshop on Big Data for Financial News and Data

WorkshopChairs: Quanzhi Li, Sameena Shah, Xiaozhong Liu

Time	Title	Presenter/Author
9:00am- 9:20am	*Predicting the Effects of News Sentiments on the Stock Market*	**Dev Shah**, Haruna Isah, and Farhana Zulkernine
9:20am- 9:40am	*Subscription and Redemption Prediction in Mutual Funds Using Machine Learning Techniques*	**Morteza Mashayekhi**, Iman Rezaeian, Annie Z. Zhang, and Jonathan Anders
9:40am- 10:00am	*Applied attention-based LSTM neural networks in stock prediction*	Mu-En Wu, Yu-Hsiang Huang, and **Li-Chen Cheng**
10:00am- 10:20am	**Coffee Break**	
10:20am- 10:40am	*Financial Networks: A Study of the Toronto Stock Exchange*	**Dhanya Jothimani**, Can Kavaklioglu, and Ayse Basar,
10:40am- 11:00am	*FinanViz: Visualizing Emerging Topics in Financial News*	Ngan V.T. Nguyen, Vinh T. Nguyen, **Tommy Dang**, and Vung Pham

11:00am- 11:20am	*Classification of "Hot News" for Financial Forecast Using NLP Techniques*	Savas Yildirim, **Dhanya Jothimani**, Can Kavaklioglu, and Ayse Basar
11:20am- 11:40am	*Hierarchical Bayesian Modeling for Clustering Sparse Sequences in the Context of User Profiling in Customer Loyalty Program*	**Ishani Chakraborty**
11:40am	**Closing Remarks**	

Workshop on Energy-Efficient Big Data Analytics *WorkshopChairs: Mohammed Alawad*		
Time	**Title**	**Presenter/Author**
8:30-9:00	*Parallel Sampling-Pipeline for Indefinite Stream of Heterogeneous Graphs using OpenCL for FPGAs*	Muhammad Usman Tariq and Fahad Saeed
9:00-9:30	*FRLDM: Empowering K-nearest Neighbor (KNN) through FPGA-based Reduced-rank Local Distance Metric*	Ashkan Samiee, Yinjie Huang, and Yu Bai
9:30-10:00	*Optimizing Radial Basis Function Kernel on OpenCL FPGA Platform*	Zheming Jin and Hal Finkel
10:00-10:20	**Coffee Break**	
10:20-10:50	*Bob Jenkins Lookup3 Hash Function on OpenCL FPGA Platform*	Zheming Jin and Hal Finkel
10:50-11:20	*A study on modeling using big data and deep learning method for failure diagnosis of system*	chung-ki seo, Jun-Ha Kim, and Sun-Youl Kwon
11:20-11:50	*CompactNet: High Accuracy Deep Neural Network Optimized for On-Chip Implementation*	Abhinav Goel, Zeye Liu, and Ronald Blanton
	Closing Remarks	

3rd International Workshop on Big Data Transfer Learning *Workshop Chairs: Ming Shao, Tongliang Liu, Zhengming Ding, Yun Fu*		
Time	**Title**	**Presenter/Author**
8:30AM	Causal Domain Adaptation	Invited Speaker: Dr. Mingming Gong
9:20AM	PROPS: Probabilistic Personalization of Black-box Sequence Models	Michael Wojnowicz and Xuan Zhao
	Coffee Break	
10:05AM	Object Detection at Scale as Cloud Services	Invited Speaker: Dr. Lei Zhang
10:55AM	Virtual Touch-Point: Trans-Domain Behavioral Targeting via Transfer Learning	Mori Kurokawa, Hao Niu, Kei Yonekawa, Arei Kobayashi, Daichi Amagata, Takuya Maekawa, and Takahiro Hara
11:20AM	Curriculum Domain Adaptation	Invited Speaker: Dr. Boqing Gong

12:10AM	K-nearest Neighbor Search by Random Projection Forests	Donghui Yan, Yingjie Wang, Jin Wang, Honggang Wang, and Zhenpeng Li
	Closing Remarks	

BDMM Workshop *WorkshopChairs: Wo Chang*		
Time	**Title**	**Presenter/Author**
Dec. 10	**Day-1**	
09:00 – 09:10	Welcome	Wo Chang
09:10 – 09:20	Opening Remark	David Belanger
09:20 – 11:00 (in Parallel)	Hackathon Briefing on use case, datasets, challenges, Q/As	Elizabeth Chang, David Ziegler, Seth Elkin-Frankston
	Coffee Break	
11:00 – till next day 09:00	Solving hackathon challenges	Hackathon Participants
Dec. 11	**Day-2**	
09:00 – 12:00	Hackathon Presentation and Evaluation	Hackathon Evaluation Team
12:00 – 13:30	**Lunch**	
13:30 – 13:40	Welcome	Wo Chang
13:40 – 14:00	Opening Remark	David Belanger
14:00 – 14:40	Keynote Speaker: NCI Cancer Research Data Commons	Allen Dearry
14:40 – 15:10	Invited Speaker: Aggregating and Sharing De-Identified Clinical, Imaging, and Genomic Data from the VA to External Repositories for the APOLLO Network	Luis E. Selva
15:10 – 15:40	**Coffee Break**	
15:40 – 16:10	Invited Speaker on Big Data Metadata Management	TBD
16:10 – 16:40	Invited Speaker: Modeling visual cortex through the lens of interpretable machine learning and biophysics	Reza Abbasi-Asl
16:40 – 16:55	Paper Presentation #1: Efficient Query Answering On Uncertain Big RDF Data	Mourad Ouziri and Salima Benbernou
16:55 - 17:10	Paper Presentation #2: A Path to Big Data Readiness	Claire C. Austin
17:10 – 17:25	Hackathon Ceremony	David Belanger and Team
17:25 – 17:30	Announcement for next BDGMM Event	Wo Chang
	Closing Remarks	

International Workshop on Conversational Agents and Chatbots with Machine Learning (ChatbotML 2018)

WorkshopChairs: Huaglory Tianfield

Time	Title	Presenter/Author
2:00 pm – 2:30 pm	Affective Natural Language Generation by Phrase Insertion	Tomasz Dryjański, Paweł Bujnowski, Hyungtak Choi, Katarzyna Podlaska, Kamil Michalski, Katarzyna Beksa, and Paweł Kubik
2:30 pm – 3:00 pm	Domain-specific Topic Model for Knowledge Discovery through Conversational Agents in Data Intensive Scientific Communities	Yuanxun Zhang, Prasad Calyam, Trupti Joshi, Satish Nair, and Dong Xu
3:00 pm – 3:30 pm	Fuzzy-Based Conversational Recommender for Data Intensive Science Gateway Applications	Arjun Ankathatti Chandrashekara, Radha Talluri, Sai Swathi Sivarathri, Reshmi Mitra, Prasad Calyam, Kerk Kee, and Satish Nair
3:30 pm – 3:50 pm	**Coffee Break**	
3:50 pm –4:20 pm	Applications of Sequence to Sequence Models for Technical Support Automation	Santosh Aditham, Ghodrat Aalipour, Pranav Kumar, Trung Nguyen, and Aditya Sood
4:20 pm – 4:50 pm	Towards Edge-Cloud Computing	Huaglory Tianfield
4:50 pm – 5:20 pm	Chatbots and Cloud Computing	Workshop Discussion
5:20 pm – 5:30 pm	**Closing Remarks**	

The Second International Workshop on Automation in Machine Learning and Big Data

Workshop Chairs: Tao Wang, Patrick Koch, Brett Wujek

Time	Title	Presenter/Author
10:05-10:15am	**Opening and Welcome**	**Dr. Tao Wang**
10:15-11am	**Keynote: Data and Evaluation Challenges in Social Media Mining**	**Dr. Huan Liu**
11-11:20am	S38205: Classification of Various Daily Activities using Convolution Neural Network and Smartwatch	Min-Cheol Kwon
11:20-11:40am	S38207: An Approach to Automatically Extract Predictive Properties from Nominal Attributes in Relational Databases	Valentin Kassarnig
11:40-12pm	S38208: Resource Optimization for Circuit Simulation using Machine Learning	Gangotree Chakma
12-1:20pm	**Lunch**	
1:20-1:40pm	S38209: American Sign Language Recognition using Deep Learning and Computer Vision	Kshitij Bantupalli
1:40-2pm	S38210: Toward Efficient Automation of Interpretable Machine Learning	Boris Kovalerchuk
2-2:20pm	S38211: Automatic Hyperparameter Tuning of Machine Learning Models under Time Constraints	Zhen Wang
2:20-2:40pm	S38212: A new approach for automated feature selection	Andreas Gocht
2:40-3pm	S38213: Unified Analytical Framework for Trustable Machine Learning and Automation Running with Blockchain	Tao Wang

3-3:20pm	S38214: CTE: Continuous Training Engine for Hyperparameter Optimization	Samarth Tripathi
3:20-3:40pm	S38215: Algorithm Selection for Classification Problems via Cluster-based Meta-features	Daren Ler
3:40-4pm	S38217: Energy Anomaly Detection with Forecasting and Deep Learning	Keith Hollingsworth
4-4:20pm	S38218: New Profile Recommendation Approach Based on Multi-Criteria Algorithm	Tarek Menouer
4:20-4:40pm	BigD743: MCDD: Multi-class Distribution Model for Large Scale Classification	Mayanka Chandrashekar
4:40-5pm	S38220: Learning and Multi-Objective Optimization for Linking Virtual Identities	Leila Jalali
5-5:10pm	**Closing Remarks and Best Paper Award**	**Dr. Patrick Koch**

International Workshop on Big Data Analytics for Cyber Threat Hunting (CyberHunt 2018)

Monday, December 10, 2018 – Room: Vashon II, Floor 3

Workshop Chairs: Vasileios Mavroeidis, Kamer Vishi, Thirimachos Bourlai

Time (PM)	Title	Presenter/Author
1:30	*Opening remarks*	Vasileios Mavroeidis
1:30 - 2:00	**KEYNOTE SPEECH 1: The 21st Century SOC**	**Simon Pope**
2:00 - 2:20	An Evaluation of DGA Classifiers	Martine De Cock
2:20 - 2:40	ECG-based Human Authentication using High-level Spectro-temporal Signal Features	Sara Abdeldayem
2:40 - 3:00	Defining a Metric Space of Host Logs and Operational Use Cases	Miki Verma
3:00 - 3:20	A Hierarchical Framework to Detect Targeted Attacks using Deep Neural Network	Nahid Farhady Ghalaty
3:20 - 3:40	AIL - The Design and Implementation of an Analysis Information Leak Framework	Sami Mokaddem
3:40 - 3:50	*Coffee Break (GRAND FOYER: Floor 4)*	--
3:50 - 4:20	**KEYNOTE SPEECH 2: Oslo Analytics Cyber Security Research**	**Audun Jøsang**
4:20 - 4:40	Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR	Kamer Vishi
4:40 - 5:00	Collecting Cyber Threat Intelligence from Hacker Forums via a Two-Stage, Hybrid Process using Support Vector Machines and Latent Dirichlet Allocation	Isuf Deliu
5:00 - 5:20	Corpus and Deep Learning Classifier for Collection of Cyber Threat Indicators in Twitter Stream	Avishek Bose
5:20 - 5:40	High Performance Attack Estimation in Large-Scale Network Flows	Christopher Freas
5:40 - 6:00	Towards a Data-driven Behavioral Approach to Prediction of Insider Threat	Subhasree Basu
6:00 - 6:20	Predicting Malicious Insider Threat Scenarios Using Organizational Data and a Heterogeneous Stack-Classifier	Adam Hall

-	MPMPA: A Mitigation and Prevention Model for Social Engineering Based Phishing attacks on Facebook *(*Video presentation will be shared electronically)*	Abid Jamil
6:20 - ∞	*Closing Remarks* *and* *Social Gathering (Optional)*	Vasileios Mavroeidis

PSBD 2018 Detailed Program Schedule		
Time	**Title**	**Presenter/Author**
9:00am – 9.25am	**Session PSBD18_1: Opening** **Chair: Alfredo Cuzzocrea**	
9:25am – 10.45am	**Session PSBD17_2: Security of Big Data: Models and Algorithms** **Chair: Alfredo Cuzzocrea**	
9:25am – 9.45am	Parallel Mining of Correlated Heavy Hitters on Distributed and Shared-Memory Architectures	Marco Pulimeno, Italo Epicoco, Massimo Cafaro, Catiuscia Melle, Giovanni Aloisio
9:45am – 10.05am	Fake Account Identification in Social Networks	Loredana Caruccio, Domenico Desiato, Giuseppe Polese
10:05am – 10.25am	Improving Machine Learning Tools with Embeddings: Applications to Big Data Security	Alfredo Cuzzocrea, Francesco Mercaldo, Fabio Martinelli
10:25am – 10.45am	Discussion time among participants	
10:45am - 11:05am	**Coffee Break**	
11:05am – 12.45am	**Session PSBD18_2: Privacy of Big Data: Models and Algorithm** **Chair: TBA**	
11:05am – 11.25am	Privacy-Preserving Frequent Pattern Mining from Uncertain Data	Carson Leung, Calvin Hoi, Adam Pazdor, Bryan Wodi, Alfredo Cuzzocrea
11:25am – 11.45am	Secure Your Specific Confidential Columns In Big Data Using SAS Enterprise Guide	Kaiqing Fan
11:45am – 12.05am	A Comparative Analysis of Data Hiding Techniques in Multimedia Data	Muazzam Ali Khan
12:05am – 12.45am	**Session PSBD18_3: Panel: "Privacy-Preserving Big Data Management and Analytics Models, Methods and Techniques in Specific Domains: Static and Dynamic Distributed Environments"** **Chair: Alfredo Cuzzocrea**	
12:45am - 2:00pm	**Lunch**	

First International Workshop on the Internet of Things Data Analytics		
Time	**Title**	**Presenter/Author**

10:05 am – 10:15 am	**Opening Remarks and Welcome**	
10:15 am– 10:50 am	**Invited Speaker: Dr. Di Wang (Microsoft Research)**	
11:00 am– 11:20 am	(S41212) An IoT Analytics Embodied Agent Model based on Context-Aware Machine Learning	Nathalia Nascimento, Paulo Alencar, Carlos Lucena, and Donald Cowan
11:20 am– 11:40 am	(S41210) Efficient Data Compression for IoT Devices Using Huffman Coding Based Techniques	Amlan Chatterjee, Rushabh Jitendrakumar Shah, and Khondker Hasan
11:40 am– 12:00 pm	(S41208) Governance in Adaptive Normative Multiagent Systems for the Internet of Smart Things: Challenges and Future Directions	Marx Viana, Lauro Caetano, Francisco Cunha, Paulo Alencar, and Carlos Lucena
12:10 pm – 1:30 pm	**Lunch**	
1:40 pm– 2:00 pm	(S41217) Internet of Things Big Data Analytics: The Case of Noise Level Measurements at the Roskilde Music Festival	Tor Morten Groenli, Benjamin Flesch, Raghava Mukkamala, Ravi Vatrapu, Sindre
2:00 pm– 2:20 pm	(S41204) Metric Indexing for Efficient Data Access in the Internet of Things	Christian Beecks, Alexander Grass, and Shreekantha Devasya
2:20 pm– 2:40 pm	(S41206) Continuous Location Statistics Sharing Algorithm with Local Differential Privacy	Fatima Zahra Errounda and Yan Liu
2:40 pm– 3:00 pm	(S41207) Scheduling Stream Processing Tasks on Geo-Distributed Heterogeneous Resources	Gerrit Janßen, Ilya Verbitskiy, Thomas Renner, and Lauritz Thamsen
3:00 pm– 3:20 pm	(S41209) File-system Front-end for Seamless Job Management in Sensitive Data e-Infrastructures and Cloud Federation	Abdulrahman Azab, Hein Meling, Eivind Hovig, and Antti Pursula
3:20 pm– 3:40 pm	(S41208) Governance in Adaptive Normative Multiagent Systems for the Internet of Smart Things: Challenges and Future Directions	Marx Viana, Lauro Caetano, Francisco Cunha, Paulo Alencar, and Carlos Lucena
3:40 pm– 4:00 pm	(S41211) Enhancing the Microservices Architecture for the Internet of Things	Eyhab Al-Masri
4:10 pm – 4:30 pm	**Coffee Break GRAND FOYER (Floor 4)**	
4:40 pm– 5:00 pm	(S41203) IoT Devices Recognition Through Network Traffic Analysis	Mustafizur Rahman Shahid, Hervé Debar, Gregory Blanc, and Zonghua Zhang
5:00 pm– 5:20 pm	(S41216) Utilizing Twitter Data for Early Flood Warning in Thailand	Kulsawasd Jitkajornwanich, Chanwit Kongthong, Nattaya
5:20 pm– 5:40 pm	(S41213) Audio IoT Analytics for Home Automation Safety	Sayed Khushal Shah, Zeenat Tariq, and Yugyung Lee
5:40 pm– 6:00 pm	(S41202) A Framework for IoT Data Acquisition and Forensics Analysis	Hongmei Chi, Temilola aderibigbe, and Bobby Granville
6:00 pm– 6:20 pm	(S41205) Intelligence Retrieval from a Centralized IoT Network	Dave Poortvliet and Xinli Wang
6:20 pm – 6:30 pm	**Closing Remarks**	

Special Symposiums

Bench 2018 Program

Day 1 (Mon, Dec 10th, 2018)
8:30-8:35 Opening Remarks (Dr. Chen Zheng)
8:35-8:40 Sponsor Remarks (Prof. Xiaohua Hu)
8:40-9:05 **Prof. Jianfeng Zhan, Institute of Computing Technology, CAS**
Benchmarking Opportunities and Challenges: present and future of BenchCouncil
9:05-10:05 ***Keynote I:***
Prof. Geoffrey Fox, Indiana University, APS and ACM Fellow.
10:00-10:20 coffee Break
10:20-10:40 **Session 1: Cloud I**
10:40-12:10 ***BenchCouncil: Benchmarking proposal***
10:40-11:10 **Dr. Wanling Gao, ICT, CAS**
DataMotif: A Benchmark Proposal for Big Data and AI
11:10-11:40 **Prof. Xiaoyi Lu, The Ohio State University**
A Benchmark proposal for Deep Learning Benchmarks
11:40-12:10 **Discussion**
12:00 Lunch
13:30-14:50 ***BenchCouncil: Benchmarking proposal***
13:30-14:00 **Dr. Chen Zheng, ICT, CAS**
A Benchmark proposal for Datacenter Computing
14:00-14:30 **Prof. Weining Qian, East China Normal University**
PeakBench: A Benchmark Proposal for Scalable Transaction Processing
14:30-14:50 **Discussion**
14:50 **Session 2: Best Paper Session I**
15:30 Coffee Break
15:50 **Session 3: Best Paper Session II**
16:40 **Session 4: Big Data**
17:40 **Session 5: Modeling and Prediction**
18:20 End

Day 2 (Tue, Dec 11th, 2018)
8:35 Opening and Welcome
8:45-9:45 ***Keynote 2:***
Prof. Vijay Janapa Reddi, The University of Texas at Austin
9:45-10:05 coffee Break
10:05-10:35 **Invited talk: Dr. Arne Berre, SINTEF Digital**
Benchmarking for Digital Platforms with Big Data, IoT, AI, Cloud, HPC and CyberSecurity
10:35-11:55 ***BenchCouncil: Benchmarking proposal***
10:35-11:05 **Prof. Yueguo Chen, Renmin University of China**
TS-benchmark: a benchmark proposal for time series databases
11:05-11:35 **Prof. Zhiyuan Chen, Prof. Jianwu Wang, University of Maryland, Baltimore County**
A Benchmark proposal for large-scale and high-speed spatiotemporal data processing and analytic
11:35-11:55 **Discussion**
11:55-12:20 **BenchCouncil Open Meeting: Formation of working groups**
12:20 Lunch
13:30-14:10 **Session 6: Algorithms and Implementations**
14:10-15:10 **Session 7: Cloud II**
15:10-17:30 **Tutorial: BigDataBench Tutorial: a scalable and unified Big Data and AI benchmark suite**
15:10 **Introduction of BigDataBench 4.0 & Benchmarking Methodology**
16:10 Coffee Break

16:30 **How to use BigDataBench 4.0**
17:00 **Big data and AI proxy benchmarks for simulation**
17:30 **Best Paper Award and Closing the symposium**
17:45 End

Program Details
Length of presentations (including Q&A):
Keynotes: 60 minutes
Benchmark Proposal: 30 minutes
Invited talk: 30 minutes
Best Paper Candidates: 25 minutes
Regular Papers: 20 minutes

Session 1: Cloud I
Monday (Dec 10th) 9:40-10:40 (Including a Coffee Break)
Benchmarking VM Startup Time in the Cloud

Session 2: Best Paper Session I
Monday (Dec 10th), 14:30-15:30
Session Chair:
DCMIX: Generating Mixed Workloads for the Cloud Data Center
EC-Bench: Benchmarking Onload and Offload Erasure Coders on Modern Hardware Architectures

Session 3: Best Paper Session II
Monday (Dec 10th), 15:50-16:40
Testing Raft-replicated Database Systems
Machine-Learning Based Spark and Hadoop Workload Classification Using Container Performance Patterns

Session 4: Big Data
Monday (Dec 10th), 16:40-17:40
Benchmarking for Transaction Processing Database Systems in Big Data Era
UMDISW: A Universal Multi-Domain Intelligent Scientific Workflow Framework for the Whole Life Cycle of Scientific Data
IBDB: A Benchmark Suite for Industrial Big Data System

Session 5: Modeling and Prediction
Monday (Dec 10th), 17:40-18:20
Power Characterization of Memory Intensive Applications: Analysis and Implications
Multi-USVs coordinated detection in marine environment based on deep reinforcement learning

Session 6: Algorithms and Implementations
Tuesday (Dec 11th), 13:30-14:10
Benchmarking SpMV on Many-core Platforms
Benchmarking Parallel Implementations of K-Means Cloud Type Clustering from Satellite Data

Session 7: Cloud II
Tuesday (Dec 11th), 14:10-15:10
An Open Source Cloud-based NoSQL and NewSQL Database Benchmarking Platform for IoT Data
Scalability Evaluation of Big Data Processing Services in Clouds
PAIE: A Personal Activity Intelligence Estimator in the Cloud

Special Sessions

4th Special Session on Intelligent Data Mining		
Time	**Title**	**Presenter/Author**
07:00am-08:00am	**Registration**	
08:00am-08:30am	**Session Keynote Speech** Uraz Yavanoglu, PhD	
08:30am-08:40am	SP01202	Countermeasure of Statistical Inference in Database Security
08:40am-08:50am	SP01203	How to Become Instagram Famous: Post Popularity Prediction with Dual-Attention
08:50am-09:00am	SP01204	You Type a Few Words and We Do the Rest: Image Recommendation for Social Multimedia Posts
09:00am-09:10am	SP01220	A unified scheme of text localization and structured data extraction for joint OCR and data mining
09:10am-09:20am	SP01229	Probabilistic Relational Supervised Topic Modelling using Word Embeddings
09:20am-09:30am	SP01209	Matrix factorization for co-training algorithm to classify human rights abuses
09:30am-09:40am	SP01211	Learning Patterns from Imbalanced Evolving Data Streams
09:40am-09:50am	SP01214	MIS-IoT: Modular Intelligent Server Based Internet of Things Framework with Big Data and Machine Learning
09:50am-10:00am	SP01216	Evaluation of Classification Algorithms, Linear Discriminant Analysis and a New Hybrid Feature Selection Methodology for Coronary Artery Disease Diagnosis
10:00am-10:10am	SP01227	Classification of TrashNet Dataset Based on Deep Learning Models
10:10am-10:20am	SP01231	Transfer Learning Effects on Image Steganalysis with Pre-Trained Deep Residual Neural Network Model
10:20am-10:30am	SP01232	Evaluation of Distributed Machine Learning Algorithms for Anomaly Detection from Large-Scale System Logs: A Case Study
10:30am-10:40am	SP01233	Complex Event Analysis of Urban Environmental Data based on Deep CNN of Spatiotemporal Raster Images
10:40am-11:00am	**Break**	
11:00am-11:10am	SP01213	A Simple Method to Remove Reviews against Guideline for Online Review Services
11:10am-11:20am	SP01215	Customer Lifetime Value in Video Games Using Deep Learning and Parametric Models
11:20am-11:30am	SP01218	StaTIX — Statistical Type Inference on Linked Data
11:30am-11:40am	SP01221	An In-depth Comparison of Group Betweenness Centrality Estimation Algorithms
11:40am-11:50am	SP01222	DyBED: An Efficient Algorithm for Updating Betweenness Centrality in Directed Dynamic Graphs
11:50am-12:00pm	SP01225	Evaluating the EEG and Eye Movements for Autism Spectrum Disorder
12:00pm-12:10pm	SP01226	Ordinal Hyperplane Loss
12:10pm-12:20pm	SP01228	Influence Propagation for Social Graph-based Recommendations
12:20pm-12:30pm	SP01208	OPOSSAM: Online Prediction of Stream Data Using Self-adaptive Memory

12:30pm-12:40pm	SP01230	From Big Data to Knowledge: Issues of Provenance, Trust, and Scientific Computing Integrity
12:40pm-12:50pm	SP01235	Conditioning Neural Networks: A Case Study of Electricity Load Forecasting
12:40pm-02:00pm	**Lunch Break**	
02:00pm-02:10pm	SP01234	Deep Neural Networks for Social Media Word Segmentation of Asian Languages
02:10pm-02:20pm	N202	Forecasting and Anomaly Detection on Application Metrics using LSTM
02:20pm-02:30pm	BigD230	A Statistical Approach to Inferring Business Locations Based on Purchase Behavior
02:30pm-02:40pm	BigD270	Tracking the Evolution of Words with Time-reflective Text Representations
02:40pm-02:50pm	BigD272	One-Class Learning Time-Series Shapelets
02:50pm-03:00pm	BigD273	Object Detections by a Super-Resolution Method and Convolution Neural Networks
03:00pm-03:10pm	BigD283	User-centered Information Retrieval using Semantic Multimedia Big Data
03:10pm-03:20pm	BigD320	LaHiIO: Accelerating Persistent Big Data Machine Learning via Latency Hiding IOs
03:20pm-03:30pm	BigD355	Comparative Study of CNN and LSTM based Attention Neural Networks for Aspect-Level Opinion Mining
03:30pm-03:40pm	BigD503	Ensemble Machine Learning Systems for the Estimation of Steel Quality Control
03:40pm-03:50pm	BigD551	Performance Prediction using Neural Network and Confidence Intervals: a Gas Turbine application.
03:50pm-04:00pm	BigD584	Multi-layer Embedding Neural Architecture with External Memory for Large-Scale Text Categorization
04:00pm-04:30pm	**Break**	
04:30pm-04:40pm	BigD591	A Density-based Preprocessing Technique to Scale Out Clustering
04:40pm-04:50pm	BigD606	Towards a New Approach to Empower Periodic Pattern Mining for Massive Data using Map-Reduce
04:50pm-05:00pm	BigD660	Context Aware Flow Prediction of Bike Sharing Systems
05:00pm-05:10pm	BigD742	DeepMove: Learning Place Representations through Large Scale Movement Data
05:10pm-05:20pm	SP01224	Twitter Sentiment Analysis: 3-Way Classification Positive, Negative or Neutral?
05:20pm-05:30pm	SP01236	Distributed Big Data Mining Platform for Smart Grid
05:30pm-05:40pm	Discussions	Discussions
05:40pm-05:50pm	Session Closing	Session Closing

IEEE Big Data 2018 - 1th Special Session on HealthCare Data

Special Session Chairs: Ozgun Pinarer

Time	Title	Presenter/Author
14:00-14:10	Welcome	
14:10-14:30	HL7 Data Acquisition & Integration: Challenges and Best Practices	Shweta Sinha

14:30-14:50	The Role of Selfies in Creating the Next Generation Computer Vision Infused Outpatient Data Driven Electronic Health Records (EHR)	Chandrasekar Vuppalapati
14:50-15:10	A Machine Learning Based Natural Language Question and Answering System for Healthcare Data Search using Complex Queries	Hangu Yeo
15:10-15:30	Multi-Database Monitoring Tool for the E-Health Services	Igor Kotsiuba
15:30-15:50	Privacy-Preserving Scoring of Tree Ensembles: A Novel Framework for AI in Healthcare	Keerthanaa Saminathan
15:50-16:10	Distributed Rough Set Based Feature Selection Approach to Analyse Deep and Hand-crafted Features for Mammography Mass Classification	Zaineb Chelly Dagdia
16:10-16:30	**Coffee Break**	
16:30-16:50	Web Service Solution for Adverse Drug Events and Medication Errors	Ozgun Pinarer
16:50-17:10	Rule Discovery from Breast Cancer Risk Factors using Association Rule Mining	Md Faisal Kabir
17:10-18:30	ROUND TABLE	
	Closing Remarks	

Special Session on Information Granulation in Data Science and Scalable Computing

Special Session Chairs: Shusaku Tsumoto, Dominik Slezak, Tzung-Pei Hong and Shuye-Liang Wang

Time	Title	Presenter/Author
	Session I: Granular Computing Theory	
2:30-2:50	A Distributed Rough Set Theory Algorithm based on Locality Sensitive Hashing for an Efficient Big Data Pre-processing	Zaineb Chelly Dagdia, Christine Zarges, Gael Beck, Hanene Azzag, and Mustapha Lebbah
2:50-3:10	Deep Similarity-Enhanced K Nearest Neighbors	Linh Le, Ying Xie, and Vijay Raghavan
3:10-3:30	CoUPM: Correlated Utility-based Pattern Mining	Wensheng Gan, Jerry Chun-Wei Lin, Han-Chieh Chao, Tzung-Pei Hong, and Philip S. Yu
3:30-3:42	A Multi-Granular Relative Density Model for Class Noise Detection	Xiao Liang, Shuyin Xia, Qun Liu, Yunsheng Liu, Baiyun Chen, and Guoyin Wang
3:42-3:54	Reducing Database Scan in Maintaining Erasable Itemsets from Product Deletion	Tzung-Pei Hong, Chia-Che Li, Shyue-Liang Wang, and Jerry Chun-Wei Lin
3:54-4:06	Improving Database Security with Pixel-based Granular Encryption	Ahmet Aydogan and Bing Zhou
	Coffee Break	
	Session II: Survey and Applications	
4:30-4:50	Privacy Preserving Utility Mining: A Survey	Wensheng Gan, Jerry Chun-Wei Lin, Han-Chieh Chao, Shyue-Liang Wang, and Philip S. Yu
4:50-5:10	R2-D2: ColoR-inspired Convolutional NeuRal Network (CNN)-based AndroiD Malware Detections	TonTon Hsien-De Huang and Hung-Yu Kao
5:10-5:30	Data-Driven Vessel Service Time Forecasting using Long Short-Term Memory Recurrent Neural Networks	Ibrahim AbuAlhaol, Rafael Falcon, Rami Abielmona, and Emil Petriu

5:30-5:50	From Hospital Big Data to Clinical Process: A Granular Computing Approach	Shusaku Tsumoto, Shoji Hirano, Tomohiro Kimura, and Haruko Iwata
5:50-6:02	A Soft Sensing Prediction Model of Superheat Degree in the Aluminum Electrolysis Production	Hong Yu, Jisen Yang, Xiaofang Chen, Zhong Zou, Guoyin Wang, and Tao Sang
6:02-6:14	Similarity-based Detection of Fertile Days at OvuFriend	Lukasz Sosnowski, Wojciech Chaber, Lukasz Milobedzki, Tomasz Penza, Jadwiga Sosnowska, Karol Zaleski, Joanna Fedorowicz, Iwona Szymusik, and Dominik Slezak
6:14-6:26	Toward Machine Learning on Granulated Data -- a Case of Compact Autoencoder-based Representations of Satellite Images	Mateusz Przyborowski, Tomasz Tajmajer, Lukasz Grad, Andrzej Janusz, Piotr Biczyk, and Dominik Slezak
	Closing Remarks	

BigData Cup Challenges

Road Damage Detection and Classification Challenges *Chairs: Hiroya Maeda, Yoshihide Sekimoto, Takehiro Kashiyama, Toshikazu Seto, Hiroshi Omata*		
Time	**Title**	**Presenter/Author**
9:00	Opening Remarks	Yoshihide Sekimoto
9:20	Road Damage Detection and Classification with Faster R-CNN	Wenzhe Wang, Bin Wu, Sixiong Yang, Zhixiang Wang
9:40	A Deep Learning Approach for Road Damage Detection from Smartphone Images	Abdullah Alfarrarjeh, Dweep Trivedi, Seon Ho Kim, Cyrus Shahabi
10:00	Road Damage Detection Using RetinaNet	Laha Ale, Ning Zhang, Longzhuang Li
10:20	**Coffee Break**	
10:40	Region-based Cycle-Consistent Data Augmentation for Object Detection	Florian Kluger, Christoph Reinders, Kevin Raetz, Philipp Schelske, Bastian Wandt, Hanno Ackermann∗ and Bodo Rosenhahn
11:00	Deep Proposal and Detection Networks for Road Damage Detection and Classification	Yanbo J. Wang1, Ming Ding, Shichao Kan, Shifeng Zhang, Chenyue Lu
11:20	Automated Road Crack Detection Using Deep Convolutional Neural Networks	Vishal Mandal, Lan Uong, Yaw Adu-Gyamfi
11:40	**Closing Remarks & Awards ceremony**	

IEEE Big Data Cup competitions - FEMH Voice Data Challenge (FEMH Voice Data Challenge) **Workshop Chair: Yu Tsao**		
Time	**Title**	**Presenter/Author**
10:00-10:20	FEMH Voice Data Challenge Opening	Yu Tsao
10:20-10:40	Pathological Voice Classification Using Mel-Cepstrum Vectors and Support Vector Machine	Maryam Pishgar, Fazle Karim, Somshubra Majumdar, and Houshang Darabi
10:40-11:00	IEEE FEMH Voice Data Challenge 2018	Chandrasekar Vuppalapati, Archana Ramalingam, and Sharat Kedari
11:00-11:20	Parameterization of Sequence of MFCCs for DNN-based voice disorder detection	Tomasz Grzywalski, Adam Maciaszek, Adam Biniakowski, Jan Orwat, Szymon Drgas, Mateusz Piecuch, Riccardo Belluzzo, Krzysztof Joachimiak, Dawid Niemiec, Jakub Ptaszynski, and Krzysztof Szarzynski
11:20-11:40	DNN-based Approach to Detect and Classify Pathological Voice	Zong-Ying Chuang, Xiao-Tong Yu, Ji-Ying Chen, Yi-Te Hsu, Zhe-Zhuang Xu, Chi-Te Wang, Feng-Chuan Lin, and Shih-Hau Fang

11:40-12:00	ByoVoz Automatic Voice Condition AnalysisSystem for the 2018 FEMH Challenge	J.D Arias-Londoño, J.A Gómez-García, L. Moro-Velázquez, and J.I Godino-Llorente
Lunch Break		
2:00-2:20	A Multi-Representation Ensemble Approach to Classifying Vocal Diseases	Mingxuan Ju, Zhengkai Jiang, Yufan Chen, and Soumya Ray
2:20-2:40	FEMH Voice Data Challenge: Voice disorder Detection and Classification using Acoustic Descriptors	Chitralekha Bhat and Sunilkumar Kopparapu,
2:40-3:00	The UCD System for the 2018 FEMH Voice Data Challenge	Kevin Degila, Rahhal Errattahi, and Asmaa El Hannani,
3:00-3:20	A Transfer Learning Approach for the 2018 FEMH Voice Data Challenge	Kazi Aminul Islam, Daniel Perez, Yuzhong Shen, and Jiang Li,
3:20-3:40	Diagnosing Voice Disorder with Machine Learning	Minh Pham, Jing Lin, and Yanjia Zhang,
3:40-4:00	Discussion and Closing Remarks	Yu Tsao

Posters

Poster ID	Accept Posters
P201	Thuan Nguyen, *A Framework for Five Big V's of Big Data and Organizational Culture in Firms*
P203	Byron J. Gao and Jose A. Lopez, *LIGHT: Enabling Instant Communication for Web Surfers with Momentary Needs*
P205	Eyhab Al-Masri, *Detecting ECG Heartbeat Abnormalities using Artificial Neural Networks*
P206	Mark Lokanan, *"Methodological Problems with Big Data When Conducting Financial Crime*
P207	Vyacheslav Romanov and Jeffrey Hawk, *Mapping non-linear influence of alloying elements on tensile strength of martensitic steel*
P208	Akash Nambiar and Yuvraj Sethi, *iSkin Specialist – An Artificial Intelligence Aided Diagnostic Support System for Dermatology*
P209	Edouard Ngor SARR, Ousmane SALL, Aminata MAIGA, Lamine FATY, and Reine Marie Ndela MARONE, *Automatic Segmentation and tagging of facts in French for automated fact-checking*
P210	Seung-Shik Kang and Minhaeng Lee, *Automatic Construction of Sentiment Lexicon by Analyzing SMS Bigdata*
P211	Seok Won Chang and Jin Woo Lee, *Detailed Configuration of Spatial Hadoop-based Spatial Big Data System and Main Service Status*
P213	Mark Lokanan, *Methodological Problems with Big Data When Conducting Financial Crime Research*
P214	RANJEET DEVARAKONDA, *Machine Learning and Social Media to Mine and Disseminate Big Scientific Data*
P215	Vijayarangan Natarajan, Swaminathan Meenakshisundaram, Gautham Balasubramanian, and Shubham Sinha, *Flight delay prediction model for Airlines*
P216	Kavya Guntupally and Ranjeet Devarakonda, *Spring Boot based REST API to Improve Data Quality Report Generation for Big Scientific Data: ARM Data Center Example*
P217	Zihao Xu and Mariam Salloum, *Deep Neural Networks for Object Enumeration*
P218	Tayfun Pay, James L. Cox, and Stephen Lucci, *Another Perspective on Ensemble Methods for Automatic Keyword Extraction*
P219	Manish Puri, Aparna Varde, and Boxiang Dong, *Pragmatics and Semantics to Connect Specific Local Laws with Public Reactions*
P220	Chen Li, Minjia He, Mahboob Qaosar, Saleh Ahmed, and Yasuhiko Morimoto, *Capturing Temporal Dynamics of Users' Preferences from Purchase History Big Data for Recommendation System*
P221	Kazuyoshi Ootani and Hayato Yamana, *External Content-dependent Features for Web Credibility Evaluation*
P222	Jose Antonio Martinez Torres and Byron J. Gao, *Investigating Comparative Evaluation for Large Data*
P223	Seiki Miyamoto, Takumi Zamami, and Hayato Yamana, *Improving Recommendation Diversity across users by reducing Frequently Recommended Items*
P224	Alina Lazar, Kesheng Wu, and Alex Sim, *Predicting Network Traffic Using TCP Anomalies*
P225	Abhijit Suresh, Tamara Sumner, Isabella Huang, Jennifer Jacobs, Bill Foland, and Wayne Ward, *Using deep learning to automatically detect talk moves in teachers' mathematics lessons*
P226	Ibrahim Almubark, Lin-Ching Chang, Rahsaan Holley, iian Black, Ji Chen, Alexander Dromerick, and Peter S. Lum, *Machine Learning Approaches to Predict Functional Upper Extremity Use in Individuals with Stroke*
P227	Kasumi Kato, Atsuko Takefusa, Hidemoto Nakada, and Masato Oguchi, *A Study of a Scalable Distributed Stream Processing Infrastructure Using Ray and Apache Kafka*
P228	Misbah Khan, Narayanan Krishnamoorthy, Leila Jalali, and Rahul Biswas, *Adobe Identity Graph*
P229	Peng Xu, Dalin Chen, Xu Liu, and Jonathan Loo, *Image-based Dietary Assessment System for Chinese Children*
P230	Kangsoo Jung, Jaewon Kim, Youngjun Kim, and Seog Park, *DRAKO: Differentially pRivate Algorithm to meet K-anonymity for Online portal service*
P231	Kazuma Kusu and Kenji Hatano, *Combining Two Types of Database System for Managing Property Graph Data*
P232	Sihyun Jeong and Chong-kwon Kim, *Online Spammer Detection using User-Neighbor Relationship*
P233	Jianliang Gao, Chuqi Lei, Yuan Ling, and Bo Song, *Distributed Top-k Subgraph Search over Big Graphs*
P234	Eyhab Al-Masri and Misba Momin, *Detecting Heart Rate Variability using Millimeter-Wave Radar Technology*

P235	Chae-Soo Kim and Seung-Beom Son, *A Study on Big Data Cluster in Smart Factory using Raspberry-Pi*
P236	Rituparna Khan and Michael Gubanov, *Nested Dolls: Towards Unsupervised Clustering of Web Tables*
P237	Maksim Podkorytov and Michael Gubanov, *Hybrid.Poly: Performance Evaluation of Linear Algebra Analytical Extensions*
P238	Anisha Agarwal, Rafael Dowsley, Nicholas D. McKinney, Dongrui Wu, Chin-Teng Lin, Martine De Cock, and Anderson Nascimento, *Privacy-Preserving Linear Regression for Brain-Computer Interface Applications*
P239	Kanika Binzani and Jin Soung Yoo, *Spark-based Spatial Association Mining*
P240	Sanchya Bhagat, Keerthanaa Saminathan, Anisha Agarwal, Rafael Dowsley, Martine De Cock, and Anderson Nascimento, *Privacy-Preserving User Profiling with FB Likes*
P241	Gautam Pal, Gangmin Li, and Katie Atkinson, *Big Data Ingestion and Lifelong Learning Architecture*
P242	Sotirios Tasoulis, Aristidis Vrahatis, Spiros Georgakopoulos, and Vassilis Plagianakos, *Visualizing High-dimensional single-cell RNA-sequencing data through multiple Random Projections*
P243	Shi Dong, Zlatan Feric, Leiming Yu, David Kaeli, John Meeker, Ingrid Y. Padilla, Jose Cordero, Carmen Velez Vega, Zaira Rosario, and Akram Alshawabkeh, *An Efficient Data Management Framework for Puerto Rico Testsite for Exploring Contamination Threats (PROTECT)*
P245	Fausto Fasano, Fabio Martinelli, Francesco Mercaldo, and Antonella Santone, *Measuring Mobile Applications Quality and Security in Higher Education*
P246	Giovanni Capobianco, Umberto Di Giacomo, Francesco Mercaldo, and Antonella Santone, *A Formal Methodology for Notational Analysis and Real-Time Decision Support in Sport Environment*
P248	Maryem AIT EL HADJ, Ahmed Khoumsi, Yahya Benkaouz, and Mohammed Erradi, *Validation and Correction of Large Security Policies: A Clustering and Access Log Based Approach*
P250	SHILPA BALAN, TEJAS AGARA CHANDRAKUMAR, and SOHONG CHAKRABORTY, *A Time Series Analysis of the IT Stock Market during the 2007 – 2009 Recession*
P251	Tan Tran, Lin-Ching Chang, Ibrahim Almubark, Elaine M. Bochniewicz, Lqii Shu, Peter S. Lum, and Alexander Dromerick, *Robust Classification of Functional and Nonfunctional Arm Movement after Stroke Using a Single Wrist-Worn Sensor Device*
P252	Antonio Lopardo and Marco Brambilla, *Analyzing and Predicting the US Midterm Elections on Twitter with Recurrent Neural Networks*
P253	Jo ão Cardoso, Tomasz Miksa, and Jos éBorbinha, *Debunking Active Data Management Plans*
P254	Navyasree Petluri and Eyhab Al-Masri, *Web Traffic Prediction of Wikipedia Pages*
P255	Jessica Wojtkiewicz, Satya Katragadda, and Raju Gottumukkala, *A Concept-Drift Based Predictive-Analytics Framework: Application for Real-Time Solar Irradiance Forecasting*
P256	Takuya Ueoka and Akira Ishii, *Consideration on TV audience rating and influence of social media*
P257	Jian Wu, Bharath Kandimalla, Shaurya Rohatgi, Athar Sefid, Jianyu Mao, and C. Lee Giles, *CiteSeerX-2018: A Cleansed Multidisciplinary Scholarly Big Dataset*
P258	Yingchi Liu, Quanzhi Li, Xiaozhong Liu, and Luo Si, *Document Information Assisted Event Trigger Detection*
P259	Tao Li, Xudong Liu, and Shihan Su, *Semi-supervised Text Regression with Conditional Generative Adversarial Networks*
P260	Naoki Mizusawa, Joichiro Kon, Yuya Seki, Jian Tao, and Saneyasu Yamaguchi, *Improving I/O Performance in Container with OverlayFS*
P261	Tao Li, Minsoo Choi, Yuntao Guo, and Lei Lin, *Opinion Mining at Scale: A Case Study of the First Self-driving Car Fatality*
P262	Masafumi Oyamada, *Accelerating Feature Engineering with Adaptive Partial Aggregation Tree*
P263	Hayato Nakashima, Ismail Arai, and Kazutoshi Fujikawa, *Proposal of a Method for Estimating the Number of Passengers with Using Drive Recorder and Sensors Equipped in Buses*
P264	Li Chin Ho and Kyong Jin Shim, *Data Mining Approach to the Identification of At-Risk Students*
P265	Yawei Hui, Rizwan Ashraf, Byung-Hoon Park, and Christian Engelmann, *Real-Time Assessment of Supercomputer Status by a Comprehensive Informative Metric through Streaming Processing*
P266	Antonio Candelieri, Wenbin Zhang, Enza Messina, and Francesco Archetti, *Automated Rehabilitation Exercises Assessment in Wearable Sensor Data Streams*
P267	Jane Seah and Kyong Jin Shim, *Data Mining Approach to the Detection of Suicide in Social Media: A Case Study of Singapore*
P269	Julia Hockett, Yaguang Liu, Yifang Wei, Lisa Singh, and Nathan Schneider, *Detecting and Using Buzz from Newspapers to Understand Patterns of Movement*
P270	Tiffany Hyun-Jin Kim and Joshua Lampkins, *BRICS: Blockchain-based Information Control Systems*
P272	Colin Tay and Kyong Jin Shim, *A Cloud-Based Data Gathering and Processing System for Intelligent Demand Forecasting*

P275	Eyhab Al-Masri, Ibrahim Diabate, Richa Jain, Ming Hoi Lam, and Swetha Reddy Nathala, *Recycle.io: An IoT-Enabled Framework for Urban Waste Management*
P276	Eyhab Al-Masri and Lingwei Meng, *A Quality-Driven Recommender System for IaaS Cloud Services*

Conference Wifi Access

2018 Big Data Conference Wifi Access

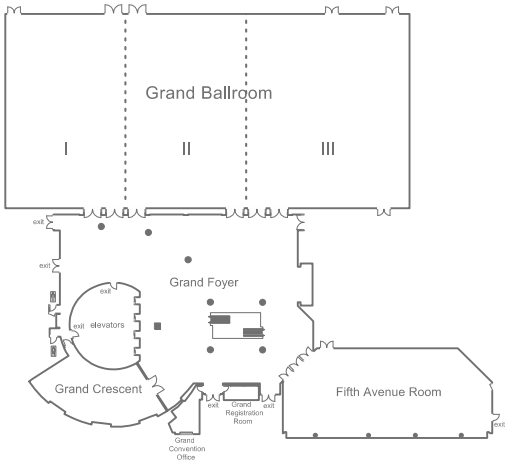
Connect to

“Westin Meeting” network

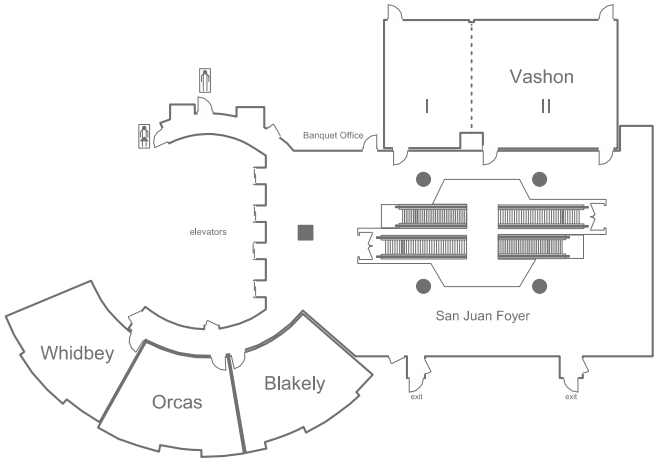
Enter (case sensitive): IEEE2018

Westin Seattle Floor Plan

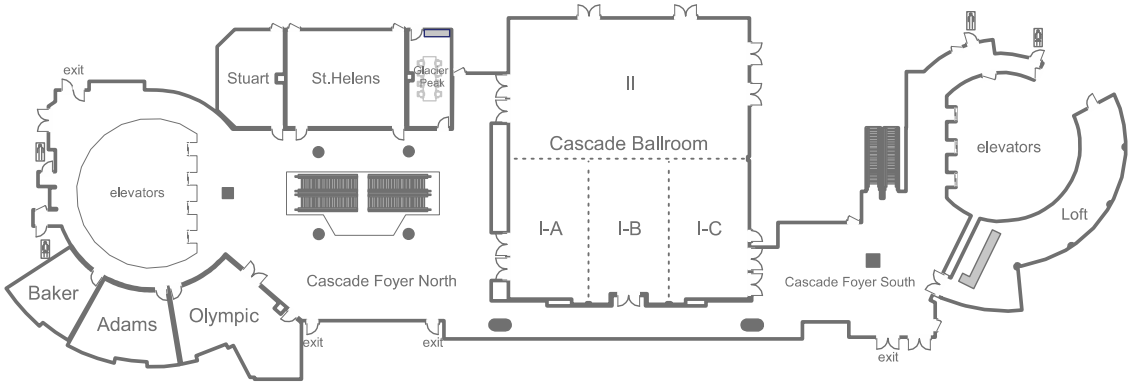
Floor Plans



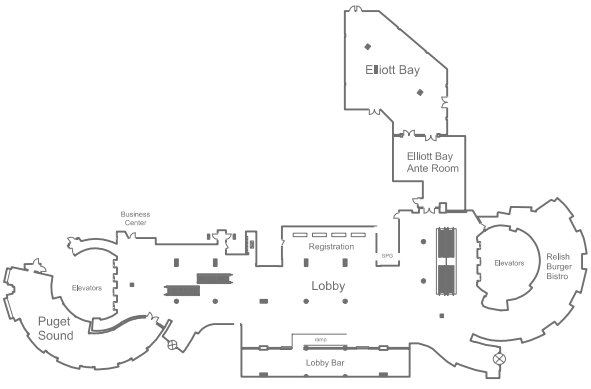
GRAND LEVEL (FOURTH FLOOR)



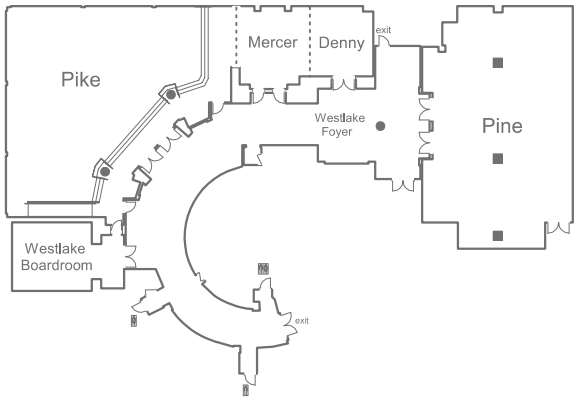
SAN JUAN LEVEL (THIRD FLOOR)



MEZZANINE LEVEL (SECOND FLOOR)



LOBBY LEVEL (FIRST FLOOR)



WESTLAKE LEVEL (LOWER LEVEL)

IEEE BIGDATA 2019

December 9-12, 2019, Los Angeles, CA, USA

The IEEE Big Data conference series is a leading forum for disseminating the latest advances in big data research , development and application. We solicit high-quality original research papers (including significant work-in-progress) in any aspect of Big Data with emphasis on 5Vs (Volume, Velocity, Variety, Value and Veracity): big data science and foundations, big data infrastructure, big data management, big data searching and mining, big data privacy/security, and big data applications. Relevant topics include but are not limited to:

1. Big Data Science and Foundations

a. Novel Theoretical Models for Big Data
b. New Computational Models for Big Data
c. Data and Information Quality for Big Data
d. New Data Standards

2. Big Data Infrastructure

a. Cloud/Grid/Stream Computing for Big Data
b. High Performance/Parallel Computing Platforms for Big Data
c. Autonomic Computing and Cyber-infrastructure, System Architectures, Design and Deployment
d. Energy-efficient Computing for Big Data
e. Programming Models and Environments for Cluster, Cloud, and Grid Computing to Support Big Data
f. Software Techniques and Architectures in Cloud/Grid/Stream Computing
g. Big Data Open Platforms
h. New Programming Models for Big Data beyond Hadoop/MapReduce, STORM
i. Software Systems to Support Big Data Computing

3. Big Data Management

a. Advanced database and Web Applications
b. Novel Data Model and Databases for Emerging Hardware
c. Data Preservation
d. Data Provenance
e. Interfaces to Database Systems and Analytics Software Systems
f. Data Protection, Integrity and Privacy Standards and Policies
g. Information Integration and Heterogeneous and Multi-structured Data Integration
h. Data management for Mobile and Pervasive Computing
i. Data Management in the Social Web
j. Crowdsourcing
k. Spatiotemporal and Stream Data Management
l. Scientific Data Management
m. Workflow Optimization
n. Database Management Challenges: Architecture, Storage, User Interfaces

4. Big Data Search and Mining

a. Social Web Search and Mining
b. Web Search
c. Algorithms and Systems for Big Data Search
d. Distributed, and Peer-to-peer Search
e. Big Data Search Architectures, Scalability and Efficiency
f. Data Acquisition, Integration, Cleaning, and Best Practice
g. Visualization Analytics for Big Data
h. Computational Modeling and Data Integration
i. Large-scale Recommendation Systems and Social Media Systems
j. Cloud/Grid/Stream Data Mining- Big Velocity Data
k. Link and Graph Mining
l. Semantic-based Data Mining and Data Pre-processing
m. Mobility and Big Data
n. Multimedia and Multi-structured Data- Big Variety Data

5. Big Data Security & Privacy

a. Intrusion Detection for Gigabit Networks
b. Anomaly and APT Detection in Very Large Scale Systems
c. High Performance Cryptography
d. Visualizing Large Scale Security Data
e. Threat Detection using Big Data Analytics
f. Privacy Threats of Big Data
g. Privacy Preserving Big Data Collection/Analytics
h. HCI Challenges for Big Data Security & Privacy
i. User Studies for any of the above
j. Sociological Aspects of Big Data Privacy

6. Big Data Applications

a. Complex Big Data Applications in Science, Engineering, Medicine, Healthcare, Finance, Business, Law, Education, Transportation, Retailing, Telecommunication
b. Big Data Analytics in Small Business Enterprises (SMEs),
c. Big Data Analytics in Government, Public Sector and Society in General
d. Real-life Case Studies of Value Creation through Big Data Analytics
e. Big Data as a Service
f. Big Data Industry Standards
g. Experiences with Big Data Project Deployments

INDUSTRIAL and GOVERNMENT Track

The Industrial and government Track solicits papers describing implementations of Big Data solutions relevant to industrial settings. The focus of industry track is on papers that address the practical, applied, or pragmatic or new research challenge issues related to the use of Big Data in industry. We accept full papers (up to 10 pages) and extended abstracts (2-4 pages).


IEEE COMPUTER SOCIETY
IEEE
expedia group
Baidu 百度
松鼠AI·智适应 Squirrel AI Learning
Open Access Journals by MDPI
information
algorithms
big data and cognitive computing
future internet
ankura
Springer