Sponsored by

2017 IEEE International Conference on Big Data

Organization Committee

Conference Co-Chairs

Dr. Ricardo Baeza-Yates: NTENT, USA / Universitat Pompeu Fabra, Spain
Prof. Xiaohua Tony Hu: Drexel University, USA
Dr. Jeremy Kepner: MIT Lincoln Laboratory, USA

Program Co-Chairs

Prof. Jian-Yun Nie: University of Montreal, Canada
Prof. Zoran Obradovic: Temple University, USA
Dr. Toyotaro Suzumura: IBM T.J. Watson Research Center, USA

Industry and Government Program Committee Co-Chairs

Dr. Rumi Ghosh: BOSCH, USA
Dr. Raghunath Nambiar: CISCO, USA
Dr. Sudarsan Rachuri: Dept of Energy, USA
Dr. Chonggang Wang: InterDigital Inc., USA
Dr. Hui Zang: Huawei Research, USA

Workshop Co-Chairs

Prof. Alfredo Cuzzocrea: University of Trieste, Italy
Prof. Jian Tang: Syracuse University, USA
Prof. Masashi Toyoda: University of Tokyo, Japan

Tutorial Co-Chairs

Prof. Feng Chen: SUNY at Albany, USA
Dr. Jianliang Gao: Drexel University, USA
Dr. Mu Qiao: IBM Almaden Research Center, USA

Poster Co-Chairs

Prof. Alexandru Costan: IRISA/INSA, France
Prof. Jingrui He: Arizona State University, USA
Prof. Feng Luo: Clemson University, USA

Sponsorship Chair

Prof. Yicheng Tu: University of South Florida, USA

Registration Chair

Prof. Yuan An: Drexel University, USA

Publicity Co-Chairs

Prof. Yihua Huang: Nanjing University, China
Prof. Wookey Lee: Inha University, Korea
Dr. Dominik Slezak: Infobright, Canada / University of Warsaw, Poland

Student Travel Award Chair

Prof. Jianwu Wang: Univ of Maryland at Baltimore Country, 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
Prof. Wei Wang: University of California at Los Angle, USA

Program Committee

Main Conference Senior PC Members

Name	Organization	Country
Naoki Abe	IBM, USA	USA
Karl Aberer	EPFL	Switzerland
Gagan Agrawal	The Ohio State University	USA
Danilo Ardagna	Politecnico di Milano, Italy	Italy
Peter Baumann	Jacobs University	Germany
Elisa Bertino	Purdue University	United States
Albert Bifet	Télécom ParisTech	France
Francesco Bonchi	Yahoo! Research	USA
Ricardo J. G. B. Campello	James Cook University	Australia
Barbara Carminati	University of Insubria, Italy	Italy
Carlos Castillo	Eurecat	Spain
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
Ming-Syan Chen	National Taiwan University	Taiwan
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
Frans Coenen	University of Liverpool	United Kingdom
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
Xiaoyong DU	Renmin University of China	China
Johannes Furnkranz	TU Darmstadt	Germany
Wook-Shin Han	Postech	Korea
Howard Ho	IBM Almaden Research	USA
Jimmy Huang	School of Information Technologies, 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
Latifur Khan	The University of Texas at Dallas	United States
Taghi Khoshgoftaar	Florida Atlantic University	United States
Alex Kuo	University of Victoria	Canada
Keqin Li	State University of New York	USA
Yunyao Li	IBM Almaden Research	United States
Huan Liu	Arizona State University	United States
Tie-Yan Liu	Microsoft Research Asia, China	China
Eric Lo	Chinese University of Hong Kong	Hong Kong
Jay Lofstead	Sandia National Laboratories	USA
Bradley Malin	Vanderbilt University	USA
Wagner Meira	Universidade Federal de Minas Gerais	Brazil
John Miller	University of Georgia	United States
Shinichi Morishita	University of Tokyo	Japan
Dimitrios Nikolopoulos	Queen's University Belfast	UK
Beng Chin Ooi	National University of Singapore	Singapore
Jian Pei	School of Computing Science, Simon Fraser University	Canada
Barbara Pernici	Politecnico Milano	Italy
Evaggelia Pitoura	University of Ioannina	Greece
Lakshmish Ramaswamy	The University of Georgia	United States
Huzefa Rangwala	GEORGE MASON UNIVERSITY	United States
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
Alkis Simitsis	HPE Labs	USA
Domenico Talia	University of Calabria	Italy

Jie Tang	Tsinghua University	China
Hanghang Tong	Arizona State University	USA
Hong-Linh Truong	Vienna University of Technology	Austria
Vincent Tseng	National Cheng Kung University	Taiwan
Vassilis J. Tsotras	University of California, Riverside	United States
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
Jianliang Xu	Hong Kong Baptist University	China
Jian Yang	Macquarie University	Australia
Philip S. Yu	University of Illinois at Chicago	USA
Xin Yuan	Florida State University	USA
Xiaofang Zhou	Univ. of Queensland, Australia	Australia
Zhi-Hua Zhou	Nanjing University, China	China
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
Amy Apon	Clemson University	United States
Chris Argenta	Applied Research Associates, Inc.	USA
Antonio Badia	University of Louisville	USA
Nathalie Baracaldo	IBM Almaden Research, USA	USA
Roberto J. Bayardo	Google, USA	USA
David Belanger	Stevens Institute of Technology	USA
Boualem Benatallah	University of New South Wales	Australia
Martin Berzins	University of Utah	USA
Kanishka Bhaduri	Intuit Inc.	USA
Bishwaranjan Bhattacharjee	IBM Research	USA
Mario Bravetti	University of Bologna	Italy
Hoang Bui	Western Illinois University	USA
Ali R. Butt	Virginia Tech	USA
Surendra Byna	LBNL	USA
Cornelia Caragea	Kansas State University	USA

David Carrera	Technical University of Catalonia, USA	USA
Abhishek Chandra	University of Minnesota	USA
Lijun Chang	The University of Sydney	Australia
Lin-Ching Chang	Catholic University of America	USA
Rong Chang	IBM T.J. Watson Research Center, USA	USA
Chun-Fu (Richard) Chen	IBM T.J. Watson Research Center, USA	USA
Enhong Chen	University of Science and Technology of China	China
Shiping Chen	CSIRO, Australia	Australia
Yong Chen	Texas Tech University	USA
Zhiyuan Chen	University of Maryland, Baltimore County	USA
James Cheng	The Chinese University of Hong Kong	Hong Kong
Malolan Chetlur	IBM India,	India
Kenneth Chiu	Binghamton University	USA
Wonik Choi	Inha University	Korea
Andrea Clematis	IMATI - CNR	Italy
Ayse Coskun	Boston University	United States
Alexandru Costan	INRIA	France
Sadie Creese	University of Oxford	United Kingdom
Edward Curry	National University of Ireland, Galway	Ireland
Brian D. Davison	Lehigh University, USA	USA
Miyuru Dayarathna	WSO2 Inc.	USA
Noel De Palma	University Joseph Fourier	France
Eduard Deagut	Temple University	USA
Boris Delibasic	University of Belgrade	Serbia
Marios Dikaiakos	University of Cyprus	Cyprus
Nemanja Djuric	Uber ATG	USA
Debora Donato	StumbleUpon	Italy
Matthieu Dorier	Argonne National Laboratory	USA
Zhicheng Dou	Renmin University of China	China
Nick Duffield	Texas A&M University	USA
Roee Ebenstein	The Ohio State University	United States
Magdalini Eirinaki	San Jose State University	USA
Miki Enoki	IBM Research - Tokyo	Japan
Dick Epema	Delft University of Technology	The Netherlands
Yi Fang	Santa Clara University	USA
Dmitriy Fradkin	Siemens	USA
Ada Wai-Chee Fu	The Hong Kong University of Science and Technology	Hong Kong
Yun Fu	Northeastern University	United States
Bin Gao	Microsoft Research Asia, China	China

Dario Garcia	Barcelona Supercomputing Center	Spain
Felix Gessert	University of Hamburg	Germany
Mohamed Ghalwash	IBM T.J. Watson Research Center	USA
Harald Gjermundrod	University of Nicosia	Cyprus
Bart Goethals	University of Antwerp	Belgium
Anastasios Gounaris	Aristotle University of Thessaloniki	Greece
Jane Greenberg	Drexel University	USA
Paul Grefen	Eindhoven University of Technology	The Netherlands
Clemens Grelck	University of Amsterdam	Netherlands
Le Gruenwald	University of Oklahoma	United States
Jayant Gupchup	Microsoft	United States
Amarnath Gupta	San Diego Supercomputing Center	USA
Sandeep Gupta	Biocomplexity Institute of Virginia Tech	USA
Vijay K. Gurbani	Bell Laboratories, Alcatel-Lucent	USA
Masatoshi Hanai	Nanyang Technological University	Singapore
Mohammad Hasan	IUPUI	USA
Claudia Hauff	Delft University of Technology	The Netherlands
Bingsheng He	National University of Singapore	Singapore
Daqing He	University of Pittsburgh	USA
Mark Hedges	King's College London	UK
Hiroshi Horii	IBM Research - Tokyo	Japan
Xia Hu	Texas A&M University	USA
Xiaohua Hu	Drexel University	United States
Jun Huan	University of Kansas	USA
Ruihong Huang	Texas A&M University	USA
Yihua Huang	Nanjing University	China
Fabrice Huet	INRIA-I3S-CNRS	France
Marty Humphrey	University of Virginia	USA
Hiroshi Inoue	IBM Research - Tokyo	Japan
Kazuaki Ishizaki	IBM Research - Tokyo	Japan
Saltz Jeffrey	Syracuse University	USA
Yan Jia	National University of Defense Technology	China
Milos Jovanovic	University of Belgrade	Serbia
David Kaeli	Northeastern University	United States
Jaap Kamps	University of Amsterdam	The Netherlands
George Karypis	University of Minnesota	USA
Kiyokuni Kawachiya	IBM Research - Tokyo	Japan
Hideyuki Kawashima	University of Tsukuba	Japan
Yiping Ke	Nanyang Technological University	Singapore
Vlado Keselj	Dalhausie University	Canada

Harald Kornmayer	DHBW Mannheim	Germany
Alexander Kotov	Wayne State University, USA	USA
Eren Kursun	Columbia University	USA
Alberto Laender	Universidade Federal de Minas Gerais	Brazil
Jack Lange	University of Pittsburgh	USA
Alexey Lastovetsky	University College Dublin	Ireland
Kisung Lee	Louisiana State University	USA
Chengkai Li	University of Texas at Arlington	USA
Pan Li	Case Western Reserve University	USA
Xue Li	The University of Queensland	Australia
Zhoujun Li	BAUU	China
Hongfei Lin	Dalian Univ. of Technology, China	China
Shou-De Lin	National Taiwan University	Taiwan
Chengfei Liu	Swinburne University of Technology	Australia
Xiaohua Liu	Huawei	China
Yan Liu	University of Southern California	USA
Yiqun Liu	Tshinghua University	China
Dr. Shiyong Lu	Wayne State University	United States
Claudio Lucchese	National Research Council of Italy (CNR)	Italy
Heiko Ludwig	IBM Research - Almaden	USA
feng luo	Clemson University	USA
Qiong Luo	Hong Kong University of Science and Technology	Hong Kong
Kwan-Liu Ma	University of California,Davis	USA
Tiziana Margaria	University of Limerick and Lero	Ireland
Amirreza Masoumzadeh	SUNY at Albany	USA
George Mathew	MIT Lincoln Laboratory	USA
Satoshi Matsuoka	Tokyo Institute of Technology, Japan	Japan
Carolyn McGregor	university of Ontario Institute of Technology	CANADA
Edgar Meij	University of Amsterdam	Netherlands
Christoph Meinel	Hasso-Plattner-Institute, Germany	Germany
Wagner Meira, Jr.	UFMG	Brazil
Ningfang Mi	Northeastern University	USA
Taneli Mielikainen	Oath	USA
Christine Morin	University of Rennes, France	France
John P. Morrison	University of Cork	Ireland
Alessandro Moschitti	University of Trento	Italy
Sebastien Mosser	Universit? Nice-Sophia Antipolis	France
Abdullah Mueen	Microsoft Research	USA
Hidemoto Nakada	National Institute of Advanced Industrial Science and Technology	Japan

Wolfgang Nejdl	Institut für Verteilte Systeme	Germany
Surya Nepal	CSIRO	Australia
Alexandros Ntoulas	LinkedIn	USA
Salvatore Orlando	Universit? Ca' Foscari Venezia	Italy
Balaji Palanisamy	University of Pittsburgh	US
Dino Pedreschi	University of Pisa	Italy
Dana Petcu	West University of Timisoara, Romania	Romania
Tao Qin	Microsoft Research Asia, China	China
Baojun Qiu	Chaoda Foodmall Group	China
Christoph Quix	Fraunhofer FIT	Germany
Tilmann Rabl	TU Berlin	Germany
Vladan Radosavljevic	Uber Advanced Technology Group	USA
Milos Radovic	University of Kragujevac	Serbia
Jan Ramon	INRIA Lille	France
Andreas Rauber	Technical University of Vienna	Austria
Stephan Reiff-Marganiec	University of Leicester, UK	UK
Chiara Renso	ISTI-CNR	Italy
Abdelmounaam Rezgui	New Mexico Tech	USA
Philip Rhodes	University of Mississippi	USA
Uwe Roehm	The University of Sydney	Australia
Paolo Romano	Lisbon University/INESC-ID	Portugal
Lotfi Ben Romdhane	University of Sousse	Tunisia
Aki-Hiro Sato	Kyoto University	Japan
Martin Schulz	Lawrence Livermore National Laboratory	USA
Bruno Schulze	National Lab. for Scientific Computing, Brazil	Brazil
Matthias Schunter	Intel	USA
Assaf Schuster	Technion â€?Israel Institute of Technology, Israel)	Israel
Jangwon Seo	Google Inc.	USA
Bin Shao	Microsoft Research Asia, China	China
Haiying Shen	Clemson University	USA
Conglei Shi	Airbnb	United States
Weidong Shi	University of Houston	USA
Lidan Shou	Zhejiang University, China	China
Yogesh Simmhan	Indian Institute of Science (IISc)	INDIA
Sean Smith	Dartmouth College	USA
Fengguang Song	Indiana University-Purdue University Indianapolis	USA
Guojie Song	Peking University	China
Shaoxu Song	Tsinghua University	China
Gregor Stiglic	University of Maribor	Slovenia
Torsten Suel	Polytechnic Institute of New York University, USA	USA

Aixin Sun	Nanyang Technological University, Singapore	Singapore
Toyotaro Suzumura	IBM T.J. Watson Research Center	USA
Hassan Takabi	University of North Texas	USA
Douglas Talbert	Tennessee Technological University	USA
Pang-Ning Tan	Michigan State University, USA	USA
Wei Tan	IBM	US
Katsumi Tanaka	Graduate School of Informatics, Kyoto University	Japan
Gabriel Tanase	Graphen Inc.	United States
Jian Tang	University of Michigan	USA
Yusuke TANIMURA	National Institute of Advanced Industrial Science and Technology and University of Tsukuba	Japan
Vahid Taslimitehrani	PhysioSigns Inc.	USA
Shirish Tatikonda	IBM Almaden Research Cente	USA
Doug Thain	University of Notre Dame, USA	USA
Andrew Trotman	University of Otago	New Zealand
Dimitrios Tsoumakos	Ionian University	Greece
Mauricio Tsugawa	Microsoft	USA
Takanori Ueda	IBM Research - Tokyo	Japan
Frias-Martinez Vanessa	UMD, College Park	USA
Ana Lucia Varbanescu	University of Amsterdam	Netherlands
Carlos Varela	Rensselaer Polytechnic Institute	USA
Anthony Ventresque	University College Dublin	Ireland
Maksims Volkovs	Layer6 AI	Canada
Slobodan Vucetic	Temple University	USA
Milan Vukicevic	University of Belgrade	Serbia
Thomas Walsh	Kronos Inc.	USA
Xiaojun Wan	Peking University	China
Bin Wang	Institute of Information engineering	China
Cho-Li Wang	The University of Hong Kong, China	China
Jianwu Wang	University of Maryland, Baltimore County	USA
Jun Wang	University Central Florida, USA	USA
Shuliang Wang	Beijing Institute of Technology	USA
Ting Wang	http://x-machine.github.io/	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
Lingfei Wu	IBM T.J. Watson Research Center, USA	USA
Sai Wu	Zhejiang University	China
Yongwei Wu	Tsinghua University	China

Hui Xiong	Rutgers University	USA
Hongbo Xu	Chinese Academy of Sciences, China	China
Jian Xu	TouchPal Inc.	USA
Jun Xu	ICT, Chinese Academy of Science	China
Weijia Xu	University of Texas at Austin	USA
Feng Yan	College of William and Mary	USA
Rui Yan	Baidu	China
Haiqin Yang	Hang Seng Management College	Hong Kong
Xiaoyan Yang	ADVANCED DIGITAL SCIENCES CENTER	Singapore
Jianwei Yin	Zhejiang University	China
Peifeng Yin	IBM	United States
Yiming Ying	University of Exeter, UK	UK
Chun-Nam Yu	Cornell University, USA	USA
Qi Yu	Rochester Institute of Technology	USA
Shui Yu	Deakin University	Australia
Weikuan Yu	Florida State University	USA
Carlo Zaniolo	UCLA	USA
Chengxiang Zhai	University of Illinois at Urbana-Champaign	USA
Allan Zhang	Singapore Institute of Manufacturing Technology	Singapore
Jingyuan Zhang	Baidu Big Data Lab	USA
Kai Zhang	Temple University	USA
Kunpeng Zhang	University of Maryland, College Park	USA
Meihui Zhang	Singapore University of Technology and Design	Singapore
Min Zhang	Tsinghua University, China	China
Rui Zhang	IBM Research - Almaden	USA
Wenjie Zhang	University of New South Wales	Wales
Ya Zhang	Shanghai Jiao Tong University	China
Ming Zhao	Arizona State University	USA
Fang Zhou	Temple University	USA
Guangyou Zhou	Central China Normal University	China
Nianjun (Joe) Zhou	IBM T.J. Watson Research Center, USA	USA
Feida Zhu	Singapore Management University	Singapore
Yanmin Zhu	Shanghai Jiao Tong University, China	China

Industry and Governement Program PC members

Name	Organization	Country
Andrew Aslinger	Lockheed Martin	
Soshant Bali	Facebook	USA
Cheng Bo	Huawei Technologies	USA

Daniel Bowers	Gartner	USA
Burcin Bozkaya	Sabanci University	Turkey
Paul Cao	HP	USA
Carlos Castillo	Eurecat	Spain
Zhengzhang Chen	NEC Laboratories America	USA
ramesh Chitor	WDC	USA
alain crolotte	Teradata Corporation	United States
Mehdi Dadfarnia	NIST	USA
Akon Dey	Awake Security, Inc.	Australia
Sameh Elnikety	Microsoft Research	USA
Rumi Ghosh	Robert Bosch LLC	USA
Nancy Grady	SAIC	United States
Hansu Gu	Seagate Technology	USA
Jan-Ming Ho	Institute of Information Science, Academia Sinica	Taiwan
Bo Hu	LinkedIn	United States
Tridivesh Jena	Software AG	USA
ashok joshi	Oracle	USA
Joseph Kasten	Penn State York	USA
Balaji Krishnapuram	IBM	USA
Karthik Kulkarni	Cisco	USA
Harumi Kumo	HPE	USA
Kincho Law	Stanford University	United States
Min Li	IBM Research - Almaden	USA
Qiang Ma	Google Inc.	USA
Tariq Magdon-Ismail	VMWare	USA
Bob Marcus	ET-Strategies	USA
Kiran Mehta	MapR Technologies	USA
Xiaofeng Meng	Renmin University of China	China
Kevin Mills	NIST	US
Ye Ouyang	Verizon Wireless	USA
Pouria Pirzadeh	Microsoft	United States
Meikel Poess	Oracle	United States
Nicolas Poggi	Barcelona Supercomputing Center (BSC)	Spain
Tilmann Rabl	Bankmark	Germany
Sudarsan Rachuri	Advanced Manufacturing Office EERE, DOE	USA
Krithi Ramamritham	IIT	India
Gyan Ranjan	Symantec	USA
Lei Rao	General Motors Research Lab	USA
Amit Saha	Cisco	USA
Ramendra Sahoo	KPMG	USA

Koichi Shirahata	Fujitsu	Japan
Rekha Singhal	Tata Consultancy Services Limited	India
Srivathsan Srinivasagopalan	Visa	USA
Reza Taheri	VMWare	USA
Ilie Tanase	IBM Research	USA
Adel Taweel	Birzeit University	UK
Chii-Ren Tsai	Marriott	
Yathiraj Udupi	Cisco	USA
Chonggang Wang	InterDigital	USA
Nan Wang	LinkedIn	United States
Qingsong Wen	Alibaba Group	USA
Carl Willis-Ford	CSRA, Inc.	
Jonathan Wong	University of Toronto	Canada
Lingfei Wu	IBM Research	USA
Yinglong Xia	Huawei Research America,	USA
Han Yang	Cisco	USA
Jin Yang	Huawei Technologies Inc.	USA
Kai Yang	Tongji University	China
Roelof van Zwol	Netflix	

IEEE Big Data 2017 Program Schedule

Boston, MA, USA
December 11 - December 14, 2017

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 Copley Place, Boston located at 10 Huntington Avenue, Boston,MA.

Sunday, 10-December	
3:00 – 8:00 pm Location:	**Registration** *EssexBlrm Foyer -3rd FL*

Monday, 11-December			
7:20am-6:00 pm Location:	**Registration** *EssexBlrm Foyer -3rd FL*		
10:00-10:20 am and 3:30 – 3:50 pm Location:	**Coffee Break** *Essex Blrm Foyer, 3rd Fl, America Foyer-4th Fl and 7th Fl Foyer*		
2:00 – 6:00 pm Location:	**Poster Session (Set up only)** *Essex Blrm Foyer, 3rd Fl and America Foyer-4th Fl*		
Time	**Workshops**	**Session Chair**	**Location**
Full Day Workshops 8:00 – 6:30 pm (the starting and ending time of each workshop varies, please check the deatild workshop schedule in the Workshops Section)	6th Workshop on Scalable Cloud Data Management	Felix Gessert	Adam/Parl--7th Fl
	2nd International Workshop on Application of Big Data for Computational Social Science	Akira Ishii	Essex North-3rd Fl
	The 2nd IEEE International Workshop on Big Spatial Data (BSD 2017)	Farnoush Banaei-kashani	Essex South-3rd Fl
	IWSC17: International Workshop on Smart Cities: People, Technology, and Data	Koh Takeuchi	Staffordshire-3rd Fl
	Big Data Analytics for Internet of Things (BDA-IoT)	Levente Klein	Gloucester/Newbury 2nd FL
	First IEEE Workshop on Human-Machine Collaboration in BigData (HMData 2017)	Atsuyuki Morishima, Senjuti Basu Roy, and Lei Chen	Great Republic - 7th Fl
	The 2nd IEEE Workshop on Big Data Metadata and Management (BDMM 2017) Hackathon: 24 hours on Data Mashup (Varieties Problem) Big Data Analytics	Alex Mu-Hsing Kuo, Mahmoud Daneshmand, Wo Chang, Kathy Grise, Yinglong Xia, David Belanger	*EssexCenter - 3rd FL*

Monday, 11-December - continued			
Time	**Sessions/Workshops**	**Session Chair**	**Location**
Half Day Workshop 8:00 – 12:00 pm (the starting and ending time of each workshop varies, please check the deatild workshop schedule in the Workshops Section)	3rd International Workshop on Methodologies to Improve Big Data projects	Jeff Saltz	Courier Room - 7th Fl
	Second workshop on Real-time and stream processing in Big Data	Sabri Skhiri	Defender Room - 7th FL
	5th International Workshop on Distributed Storage Systems and Coding for Big Data	Bing Zhu	Empire Room - 7th Fl
	Big Social Media Data Management and Analysis / Natural Language Processing for Social Media	Xin Huang, Cheng-Te Li	Helicon Room - 7th Fl
	The 1st IEEE Big Data International Workshop on Policy-based Autonomic Data Governance (PADG)	Seraphin Calo	St. George A - 3rd Fl
	2nd International Workshop on Methods to Manage Heterogeneous Big Data and Polystore Databases	Vijay Gadepally	St. George B - 3rd Fl
	Fourth International Workshop on High Performance Big Graph Data Management, Analysis, and Mining (BigGraphs 2017)	Mohammad Al Hasan Kamesh Madduri Nesreen Ahmed	St. George C - 3rd Fl
	The 4th Workshop on Pattern Mining and Application of Big Data (BigPMA 2017)	Yi-Cheng Chen, Jiun-Long Huang	North Star Room - 7th Fl
Tutorial 8:15-10:15am	Tutorial 7: Game Theory for Data Science: Eliciting truthful information	Boi Faltings, Goran Radonovic	St. George D - 3rd Fl
Tutotial 10:30-12:30noon	Tutorial 4: Time Series Data Mining using the Matrix Profile: A Unifying View of Motif Discovery, Anomaly Detection, Segmentation, Classification, Clustering and Similarity Joins	Abdullah Mueen, Eamonn Keogh	St. George D - 3rd F
12:00 - 1:30 pm	**Lunch (On Own)**		
Time	**Sessions/Workshops/Tutorials**	**Session Chair**	**Location**
Half day Workshop 1:30 – 6:00 pm (the starting and ending time of each workshop varies, please check the deatild workshop schedule in the Workshops Section)	4th Workshop on Advances in Software and Hardware for Big Data Science (ASH)	Weijia Xu	Courier Room - 7th Fl
	2nd International Workshop on Enterprise Big Data Semantic and Analytics Modeling	Michael Peran	Defender Room - 7th FL
	The First International Workshop on Big Data Analytic for Cyber Crime Investigation and Prevention	Andrii Shalaginov	Empire Room - 7th Fl
	Data Science for Emergency Management	Paolo Garza	Helicon Room - 7th Fl
	Applications of Big Data Technology in the Transport Industry	Nii Attoh-Okine	North Star Room - 7th Fl
	Big Data Analytic Technology for Bioinformatics and Health Informatics (KDDBHI)	Donghui Wu	St. George A - 3rd Fl
	2nd International Workshop on Big Data Transfer Learning (BDTL) -- Automatic Knowledge Mining and Transfer for Digital Healthcare	Ming Shao	St. George B - 3rd Fl
	Big Social Media Data Management and Analysis / Natural Language Processing for Social Media	Xin Huang, Cheng-Te Li	St. George C - 3rd Fl
	International Workshop on Big Data for Financial News and Data	Quanzhi Li	St. George C - 3rd Fl
	Open Science in Big Data (OSBD) Workshop	Shannon Quinn	St. George D - 3rd Fl

Tuesday, 12-December			
7:20-6:00 pm Location:	**Registration** *Essex Blrm Foyer-3rd Fl*		
Time	**Sessions**	**Session Chair**	**Location**
8:30-08:45	Opening and Welcome	Conf and PC chairs	America N&C-4th Fl
8:45-09:45	**Keynote Session 1**: Human-in-the-loop Applied Machine Learning Prof. Carla E. Brodley	Raghunath Nambiar	America N&C-4th Fl
9:45-10:45	**Keynote Session 2**: TextScope: Enhance Human Perception via Text Mining Dr. ChengXiang Zhai, Professor, University of Illinois at Urbana-Champaign, USA	Jian-Yun Nie	America N&C-4th Fl
8:00am-6:00pm	Special Session: Intelligent Data Mining	Uraz Yavanoglu	St. George A - 3rd Fl
10:45 – 11:05 am Location:	**Coffee Break** *Essex Blrm Foyer-3rd Fl and America Foyer-4th Fl* **Poster Session (Set up)** *Essex Blrm Foyer-3rd Fl and America Foyer-4th Fl*		
11:05 am - 12:45 pm	**Sessions**	**Session Chair**	**Location**
	L1 High Performance Platforms for Big Data (I)	Balaji Palanisamy	Essex Center-3rd Fl
	L2 Big Data Applications	Geeth De Mel	Essex North-3rd Fl
	L3 Algorithms and Systems for Big Data Search (I)	Michael Gubanov	Staffordshire-3rd Fl
	L4 Novel Theoretical Models for Big Data	Liang Ma	Gloucester/Newbury 2nd Fl
	I&G-Regular 1: Big Data Analytics	Raghunath Nambiar	ADAM/PARL-7th Fl
	Manufacturing Symposium	Sudarsan Rachuri, Tina Lee, Ronay Ak, Anantha Narayanan, Rumi Ghosh	America N&C-4th Fl
12:45 – 2:00 pm Location:	**Lunch (Provided by Conference)** America South-4th Fl, Essex South-3rd Fl **Poster Session Sets Up and Displays** *Essex Blrm Foyer – 3rd Fl and America Foyer-4th Fl*		
2:00 – 4:05 pm	**Sessions**	**Session Chair**	**Location**
	L5 High Performance Platforms for Big Data (II)	Norbert Ritter	Essex Center-3rd Fl
	L6 Spatiotemporal Analytics and Traffic Applications	Abdeltawab Hendawi	Essex North-3rd Fl
	L7 Algorithms and Systems for Big Data Search (II)	Meng-Fen Chiang	Staffordshire-3rd Fl
	L8 Large-scale Recommendation Systems and Social Media Systems	Takako Hashimoto	Gloucester/Newbury 2nd Fl
	I&G –Short 1: Big Data Algorithms & Systems	Raghunath Nambiar	ADAM/PARL-7th Fl
	Manufacturing Symposium	Sudarsan Rachuri, Tina Lee, Ronay Ak, Anantha Narayanan, Rumi Ghosh	America N&C-4th Fl
	Tutorial 2: Popularity on the Web: From Estimation to Prediction	Charalampos Chelmis, Daphney-Stavroula Zois	Great Republic -7th Fl

	IEEE WORKSHOP ON BIG DATA ANALYTICS IN MANUFACTURING AND SUPPLY CHAINS	Allan Nengsheng Zhang	Empire Room-7th Fl
	Granular Computing and Big Data Special Session	Shusaku Tsumoto	St. George C - 3rd Fl
	Solar & Stellar Astronomy Big Data (SABiD) 4th Workshop on Management, Search and Mining of Massive Repositories of Solar and Stellar Astronomy Data	Rafal Angryk	St. George D - 3rd Fl
	The 2nd IEEE Workshop on Big Data Metadata and Management (BDMM 2017)/Hackathon: 24 hours on Data Mashup (Varieties Problem) Big Data Analytics	Alex Mu-Hsing Kuo, Mahmoud Daneshmand, Wo Chang, Kathy Grise, Yinglong Xia, David Belanger	St. George B - 3rd Fl
4:05 – 4:25 pm Location:	**Coffee Break** *Essex Blrm Foyer-3rd Fl and America Foyer-4th Fl* **Poster Session Sets Up and Displays** *Essex Blrm Foyer – 3rd Fl and America Foyer-4th Fl*		
4:25 -6:25 pm	Sessions	**Session Chair**	**Location**
	S1 Algorithms for Big Data (1)	Paolo Garza	Essex Center-3rd Fl
	S2 Text Mining/NLP	Lay Wai Kong	Essex North-3rd Fl
	S3 Distributed System and Software for Big Data (1)	Jinho Lee	Staffordshire-3rd Fl
	I&G-Regular 2: Big Data Applications (1)	Honggang Wang	Adam/Parl-7th Fl
	Manufacturing Symposium	Sudarsan Rachuri, Tina Lee, Ronay Ak, Anantha Narayanan, Rumi Ghosh	America N&C-4th Fl
	Tutorial 6: Industrial Big Data for Industrial Applications – Systematic Methodology	David Siege, Jay Lee, Hossein Davari, Brian Weiss	Gloucester/Newbury 2nd Fl
	IEEE WORKSHOP ON BIG DATA ANALYTICS IN MANUFACTURING AND SUPPLY CHAINS	Allan Nengsheng Zhang	Empire Room-7th Fl
	Granular Computing and Big Data Special Session	Shusaku Tsumoto	St. George C - 3rd Fl
	Solar & Stellar Astronomy Big Data (SABiD) 4th Workshop on Management, Search and Mining of Massive Repositories of Solar and Stellar Astronomy Data	Rafal Angryk	St. George D - 3rd Fl
	The 2nd IEEE Workshop on Big Data Metadata and Management (BDMM 2017)/Hackathon: 24 hours on Data Mashup (Varieties Problem) Big Data Analytics	Alex Mu-Hsing Kuo, Mahmoud Daneshmand, Wo Chang, Kathy Grise, Yinglong Xia, David Belanger	St. George B - 3rd F

Wednesday, 13-December			
7:30-6:00 pm Location:	**Registration** *Essex Blrm Foyer-3rd Fl*		
8:30 – 8:45	**Opening Remarks / Announcements**		
Time	**Sessions**	**Session Chair**	**Location**
8:45 -9:45 am	**Keynote Speech 3**: Large-scale Graph Representation Learning Dr. Jure Leskovec, Associate Professor, Stanford University, Chief Scientist at Pinterest, USA	Zoran Obradovic	America N&C-4th Fl
9:45 -10:45 am	**Keynote Speech 4**: Contextual Reinforcement Learning **Dr. John Langford**, Microsoft Research	Ricardo Baeza-Yates	America N&C-4th Fl
10:45 - 11:05am Location:	**Coffee Break** *Essex Blrm Foyer – 3rd Fl and America Foyer-4th Fl* **Poster Session Displays** *Essex Blrm Foyer – 3rd Fl and America Foyer-4th Fl*		
Time	**Sessions**	**Session Chair**	**Location**
11:05- 12:45 pm	L9 Software Systems for Big Data Computing	Nazim Madhavji	Essex Center-3rd Fl
	L10 New Computational Models for Big Data	George Mathew	Essex North-3rd Fl
	L11 Security and Privacy	Xintao Wu	Staffordshire-3rd Fl
	L12 Multimedia and Multi-structured Data - Big Variety Data	Ioannis (Giannis) Giannakopoulos	St. George AB 3rd Fl
	I&G-Regular 3: Big Data Platforms & Frameworks	Sudarsan Rachuri	Adam/Parl-7th Fl
	Manufacturing Symposium	Sudarsan Rachuri, Tina Lee, Ronay Ak, Anantha Narayanan, Rumi Ghosh	America N&C-4th Fl
	Workshop: Computational Archival Science (8:00-12:45pm)	Mark Hedges	Gloucester/Newbury
	Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD)	Zhiyuan Chen	Empire Room -7th Fl l
	3rd International Workshop on Big Data for Sustainable Development	Aki-Hiro Sato	St. George C - 3rd Fl
12:45 - 2:00 pm Location:	Lunch (Provided by conference) *America South-4th Fl, Essex South-3rd Fl* Poster Session Displays *Essex Blrm Foyer – 3rd Fl and America Foyer-4th Fl*		
Time	Sessions	**Session Chair**	**Location**
2:00 – 4:05 pm	L13 Social Web Search and Mining	Keren Ouaknine	Essex Center-3rd Fl
	L14 Stream Data Mining - Big Velocity Data	Sumit Purohit	Essex North-3rd Fl
	L15 Data and Information Quality for Big Data	Martin Koehler	Staffordshire-3rd Fl
	L16 Link and Graph Analytics (I)	Feng Chen	St. George AB
	I&G-short2 Massive Processing & Experience	Honggang Wang	Adam/Parl-7th Fl
	Manufacturing Symposium	Sudarsan Rachuri, Tina Lee, Ronay Ak, Anantha Narayanan, Rumi Ghosh	America N&C-4th Fl
	Workshop: Computational Archival Science	Mark Hedges	Gloucester/Newbury 2nd Fl

	Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD)	Zhiyuan Chen	Empire Room -7th Fl
	Tutorial 1: Enterprise Knowledge Graphs for Large Scale Analytics	Nidhi Rajshree, Nitish Aggarwal, Sumit Bhatia, Anshu Jain	Great Republic -7th Fl
	Tutorial 5: Mathematics of Big Data	Jeremy Kepner	St. George D - 3rd Fl
	3rd International Workshop on Big Data for Sustainable Development	Aki-Hiro Sato	St. George C - 3rd Fl
4:05 – 4:25 pm Location	**Coffee Break** *Essex Blrm Foyer – 3rd Fl and America Foyer-4th Fl* **Poster Session Displays** *Essex Blrm Foyer – 3rd Fl and America Foyer-4th Fl*		
Time	**Sessions**	**Session Chair**	**Location**
4:25- 6:25 pm	S4 Distributed System and Software for Big Data (2)	Yusuke Tanimura	Essex Center-3rd Fl
	S5 Algorithms for Big Data (2)	Sheng Li	Essex North-3rd Fl
	S6 Big Data Preprocessing/Visualization	Shiaofen Fang	St. George AB-3rd Fl
	I&G-regular4: Big Data Applications (2)	Ye Ouyang	Adam/Parl-7th Fl
	Manufacturing Symposium	Sudarsan Rachuri, Tina Lee, Ronay Ak, Anantha Narayanan, Rumi Ghosh	America N&C-4th Fl
	Tutorial 8: Anti-discrimination Learning: From Association to Causation	Lu Zhang, Yongkai Wu, Xintao Wu	Great Republic - 7th Fl
	Workshop: Computational Archival Science	Mark Hedges	Gloucester/Newbury *2nd Fl*
	Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD)	Zhiyuan Chen	Empire Room -7th Fl
	3rd International Workshop on Big Data for Sustainable Development	Aki-Hiro Sato	St. George C - 3rd Fl
	Panel 1 : Big Data Bias and Transparency	Ricardo Baeza-Yates	Staffordshire-3rd Fl
7:00 – 9:00 pm Location	**Banquet (Ticket required)** America N&C-4th Fl *Chair:*Conference Chairs, PC Co-chairs, I&G PC Co-chairs, 1. *Best Paper Award*, PC Co-chairs, Elsevier Representative 2. *Best Application Paper Award*, PC Co-chairs 3. *Best Industry and Government Application Paper*, I&G PC Co-chairs		

Thursday, 14-December			
07:30-6:00pm Location:	**Registration** *Essex Blrm Foyer-3rd Fl*		
Time	**Session**	**Session Chair**	**Location**
8:30 – 8:45 am	**Opening Remarks / Announcements**		
8:45 - 09:45 am	**Keynote Speech 5:** A More Open Efficient Future for AI Development and Data Dr. Alan Edelman, Professor of Applied Mathematics, MIT, USA	Jeremy Kepner	America N&C-4th Fl
9:45 - 10:45 am	**Keynote Speech 6:** Being "BYTES-oriented" in HPC leads to an Open Big Data/AI Ecosystem and Further Advances into the Post-Moore Era Dr. Satoshi Matsuoka, Professor, Tokyo Institute of Technology, Japan	Toyotaro Suzumura	America N&C-4th Fl
10:45 - 11:05 am Location:	**Coffee Break** Essex Blrm Foyer – 3rd Fl and America Foyer-4th Fl **Poster Session Displays** *Essex Blrm Foyer – 3rd Fl and America Foyer-4th Fl*		
8:00-6:00pm	4th International Workshop on Privacy and Security of Big Data (PSBD 2017)	Alfredo Cuzzocrea	Defender Room-7th FL
8:00-12:00pm	IEEE Workshop Data Science for Networking (DS4N)	Kai Yang	Great Republic - 7th Fl
Time	**Sessions/Tutorial/Workshop**	**Session Chair**	**Location**
11:05am – 1:00pm	L17 Link and Graph Analytics (II)	Michael Gubanov	Essex Center-3rd Fl
	L18 Analytics and Data Management for Big Data	Vetria Byrd	Essex North-3rd Fl
	L&G-Regular 5: Big Data Applications (3)	Sudarsan Rachuri	Staffordshire-3rd Fl
	S7 Streaming Data	John Herbert	St. George AB - 3rd Fl
	S8 Machine Learning (1)	Luca Cagliero	Adam/Parl-7th Fl l
	Panel 2: Big Data Software and Analytic Methods- What is Next?	Vijay Raghavan	America N&C-4th Fl
8:00-12:00pm	IEEE Workshop Data Science for Networking (DS4N)	Kai Yang	Great Republic - 7th Fl
1:00- 2:00 pm	Lunch (Provided by Conference) America South-4th Fl		
Time	**Sessions/Workshops**	**Session Chair**	**Location**
2:00 – 4:15	S9 Machine Learning (2)	Stephen McGough	Essex Center-3rd Fl
	S10 Big Data Applications (1)	Levente Klein	Essex North-3rd Fl
	S11 Big Data Applications (2)	Natalia Ponomareva	Staffordshire-3rd Fl
	Tutorial 3: Security and Automated Platform Development for Big Data Analytics	Jun (Luke) Huan, Sohaib Kiani, Xiaoli Li	St. George AB - 3rd Fl
	Data Quality Issues in Big Data and Machine Learning Applications: Going Beyond Data Cleaning and Transformations	Venkat Gudivada	Great Republic - 7th Fl
	Big Data Technology and Ethics Considerations in Customer Behavior and Customer Feedback Mining	Xin Deng	Adam/Parl-7th Fl
	Big Data for Economic and Business Forecasting	Wei Shang	Empire Room - 7th Fl

	International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID)	Huaglory Tianfield	St. George C - 3rd Fl
	Big Data Analytics in the Legal Industry	Jianping Zhang	St. George D - 3rd Fl
4:10 –4:30 pm	**Coffee Break** Essex Blrm Foyer – 3rd Fl and America Foyer-4th Fl		
4:30-6:30pm	Data Quality Issues in Big Data and Machine Learning Applications: Going Beyond Data Cleaning and Transformations	Venkat Gudivada	Great Republic - 7th Fl
	Big Data Technology and Ethics Considerations in Customer Behavior and Customer Feedback Mining	Xin Deng	Adam/Parl-7th Fl
	Big Data for Economic and Business Forecasting	Wei Shang	Empire Room - 7th Fl
	International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID)	Huaglory Tianfield	St. George C - 3rd Fl
	Big Data Analytics in the Legal Industry	Jianping Zhang	St. George D - 3rd Fl

Keynote Lectures

Keynote : Human-in-the-loop Applied Machine Learning

Speaker:
Dr. Carla E. Brodley, Professor and Dean, Northeastern University, USA

Abstract:
Machine learning research in academia is often conducted in vitro, divorced from motivating practical applications. As a result researchers often lose the ability to ask the question: how can my human expert's knowledge be used to best improve the machine learning outcome? In this talk, we present three motivating applications that all benefit from human-guided machine learning: systematic reviews for evidence-based medicine, generating maps of global land cover of the Earth from remotely sensed data, and finding lesions in the MRI's of treatment resistant epilepsy patients. Our machine learning contributions span active learning, both supervised and unsupervised learning, and their combination with human input. The methods we created are applicable to a wide range of applications in science, medicine and business.

Short Bio:
Carla E. Brodley is the Dean of the College of Computer and Information Science at Northeastern University. Prior to joining Northeastern, she was a professor of the Department of Computer Science and the Clinical and Translational Science Institute at Tufts University (2004-2014). Before joining Tufts she was on the faculty of the School of Electrical Engineering at Purdue University (1994-2004).
A Fellow of the ACM and AAAI, Dean Brodley's interdisciplinary machine learning research led to advances not only in computer and information science, but in many other areas including remote sensing, neuroscience, digital libraries, astrophysics, content-based image retrieval of medical images, computational biology, chemistry, evidence-based medicine, and predictive medicine. Dean Brodley's numerous leadership positions in computer science as well as her chosen research fields of machine learning and data mining include serving as program co-chair of ICML, co-chair of AAAI, and serving as associate editor of the Journal of AI Research, and the Journal of Machine Learning Research. She has previously served on the Defense Science Study Group, the board of the International Machine Learning Society, the AAAI Council and DARPA's Information Science and Technology (ISAT) Board. She is currently serving on the CRA Board of Directors, the executive committee of the Northeast Big Data Hub, and as a member-at-large of the section on Information, Computing, and Communication of AAAS.

Keynote : A More Open Efficient Future for AI Development and Data Science with an Introduction to Julia

Speaker:
Dr. Alan Edelman, Professor of Applied Mathematics, MIT, USA

Abstract:
We propose a more open, efficient, expressive, and ergonomic future for AI development, machine learning, and data science based on the Julia programming language. Our thesis is that the current tapestry of high level codes with library calls creates programmer indirections that can work well for the "one off", but can slow general progress. We provide examples from Machine Learning, Automatic Differentiation, and Data Handling Technologies.

Short Bio:
Alan Edelman is Professor of Applied Mathematics, and member of MIT's Computer Science and AI Lab. He has received many prizes for his work on mathematics and computing, and is a founder of Interactive Supercomputing, Inc. and Julia Computing, Inc. He received the B.S. and M.S. degrees in mathematics from Yale in 1984, and the Ph.D. in applied mathematics from MIT in 1989 under the direction of Lloyd N. Trefethen. Edelman's research interests include Julia, high-performance computing, numerical computation, linear algebra and random matrix theory. He has consulted for Akamai, IBM, Pixar, and NKK Japan among other corporations.

Keynote : Being "BYTES-oriented" in HPC leads to an Open Big Data/AI Ecosystem and Further Advances into the Post-Moore Era

Speaker:

Dr. Satoshi Matsuoka, Professor, Tokyo Institute of Technology, Japan

Abstract:

With rapid rise and increase of Big Data and AI as a new breed of high-performance workloads on supercomputers, we need to accommodate them at scale, traditional simulation-based HPC and BD/AI will converge. Our TSUBAME3 supercomputer at Tokyo Institute of Technology became online in Aug. 2017, and became the greenest supercomputer in the world on the Green 500 ranking at 14.11 GFlops/W; the other aspect of TSUBAME3, is to embody various Data or "BYTES-oriented" features to allow for HPC to BD/AI convergence at scale, including significant scalable horizontal bandwidth as well as support for deep memory hierarchy and capacity, along with high flops in low precision arithmetic for deep learning. Furthermore, TSUBAM3's technologies will be commoditized to construct one of the world’s largest BD/AI focused and "open-source" cloud infrastructure called ABCI (AI-Based Bridging Cloud Infrastructure), hosted by AIST-AIRC (AI Research Center), the largest public funded AI research center in Japan. The performance of the machine is slated to be several hundred AI-Petaflops for machine learning; the true nature of the machine however, is its BYTES-oriented, optimization acceleration in the memory hiearchy, I/O, the interconnect etc, for high-performance BD/AI. ABCI will be online Spring 2018 and its archiecture, software, as well as the datacenter infrastructure design itself will be made open to drive rapid adoptions and improvements by the community, unlike the concealed cloud infrastructures of today. Finally, transcending from FLOPS-centric mindset to being BYTES-oriented will be one of the key solutions to the upcoming "end-of-Moore's law" in the mind 2020s, upon which FLOPS increase will cease and BYTES-oriented advances will be the new source of performance increases over time in general for any compputing.

Short Bio:

Satoshi Matsuoka has been a Full Professor at the Global Scientifi Information and Computing Center (GSIC), a Japanese national supercomputing center hosted by the Tokyo Institute of Technology, and since 2016 a Fellow at the AI Research Center (AIRC), AIST, the largest national lab in Japan, as well as becoming the head of the joint Lab RWBC-OIL (Open Innovation Lab on Real World Big Data Computing) between the two institutions, in 2017. He received his Ph. D. from the University of Tokyo in 1993. He is the leader of the TSUBAME series of supercomputers, including TSUBAME2.0 which was the first supercomputer in Japan to exceed Petaflop performance and became the 4th fastest in the world on the Top500 in Nov. 2010, as well as the recent TSUBAME-KFC becoming #1 in the world for power efficiency for both the Green 500 and Green Graph 500 lists in Nov. 2013, and recently No.1 on the Green500 for the latest TSUBAME3 supercomputer. He is also currently leading several major supercomputing research projects, such as the MEXT Green Supercomputing, JST-CREST Extreme Big Data, and Co-PIs in several other HPC and BD/AI convergence projects. He has written over 500 articles according to Google Scholar, and chaired numerous ACM/IEEE conferences, most recently the overall Technical Program Chair at the ACM/IEEE Supercomputing Conference (SC13) in 2013. He is a fellow of the ACM and European ISC, and has won many awards, including the JSPS Prize from the Japan Society for Promotion of Science in 2006, awarded by his Highness Prince Akishino, the ACM Gordon Bell Prize in 2011, the Commendation for Science and Technology by the Minister of Education, Culture, Sports, Science and Technology in 2012, and recently the 2014 IEEE-CS Sidney Fernbach Memorial Award, the highest prestige in the field of HPC.

Keynote : TextScope: Enhance Human Perception via Text Mining

Speaker:

Dr. ChengXiang Zhai, Professor, University of Illinois at Urbana-Champaign, USA

Abstract:

Recent years have seen a dramatic growth of natural language text data (e.g., web pages, news articles, scientific literature, emails, enterprise documents, blog articles, forum posts, product reviews, and tweets). Text data contain all kinds of knowledge about the world and human opinions and preferences, thus offering great opportunities for analyzing and mining vast amounts of text data ("big text data") to support user tasks and optimize decision making in all application domains. However, computers cannot yet accurately understand unrestricted natural language; as such, involving humans in the loop of interactive text mining is essential. In this talk, I will present the vision of TextScope, an interactive software tool to enable users to perform intelligent information retrieval and text analysis in a unified task-support framework. Just as a microscope allows us to see things in the ìmicro world,î and a telescope allows us to see things far away, the envisioned TextScope would allow us to ìseeî useful hidden knowledge buried in large amounts of text data that would otherwise be unknown to us. As examples of techniques that can be used to build a TextScope, I will present some general statistical text mining algorithms that we have recently developed for joint analysis of text and non-text data to discover interesting patterns and knowledge. I will conclude the talk with a discussion of the major challenges in developing a TextScope and some important directions for future research in text data mining. .

Short Bio:

ChengXiang Zhai is a Professor of Computer Science and a Willett Faculty Scholar at the University of Illinois at Urbana-Champaign (UIUC), where he is also affiliated with School of Information Sciences, Carl R. Woese Institute for Genomic Biology, and Department of Statistics. He received a Ph.D. in Computer Science from Nanjing University in 1990, and a Ph.D. in Language and Information Technologies from Carnegie Mellon University in 2002. He worked at Clairvoyance Corp. as a Research Scientist and a Senior Research Scientist from 1997 to 2000. His research interests are in the general area of intelligent information systems, including specifically intelligent information retrieval, data mining, natural language processing, machine learning, and their applications. He has published over 200 papers in these areas and a textbook on text data management and analysis. He is the America Editor of Springerís Information Retrieval Book Series and an Associate Editor of BMC Medical Informatics and Decision Making, and previously served as an Associate Editor of ACM Transactions on Information Systems, Associate Editor of Elsevierís Information Processing and Management, Program Co-Chair of NAACL HLT 2007, ACM SIGIR 2009, and WWW 2015. He is an ACM Distinguished Scientist, and received a number of awards,such as ACM SIGIR Test of Time Paper Award (three times), the Presidential Early Career Award for Scientists and Engineers (PECASE), Alfred P. Sloan Research Fellowship, IBM Faculty Award, HP Innovation Research Award, UIUC Rose Award for Teaching Excellence, and UIUC Campus Award for Excellence in Graduate Student Mentoring.

Keynote: Large-scale Graph Representation Learning

Speaker:

Dr. Jure Leskovec, Associate Professor, Stanford University, Chief Scientist at Pinterest, USA,

Abstract:

Machine learning on graphs is an important and ubiquitous task with applications ranging from drug design to friendship recommendation in social networks. The primary challenge in this domain is finding a way to represent, or encode, graph structure so that it can be easily exploited by machine learning models. However, traditionally machine learning approaches rely on user-defined heuristics to extract features encoding structural information about a graph. In this talk I will discuss methods that automatically learn to encode graph structure into low-dimensional embeddings, using techniques based on deep learning and nonlinear dimensionality reduction. I will provide a conceptual review of key advancements in this area of representation learning on graphs, including random-walk based algorithms, and graph convolutional networks.

Short Bio:

Jure Leskovec is Associate Professor of Computer Science at Stanford University and Chief Scientist at Pinterest. Computation over massive data is at the heart of his research and has applications in computer science, social sciences, economics, marketing, and healthcare. This research has won several awards including a Lagrange Prize, Microsoft Research Faculty Fellowship, the Alfred P. Sloan Fellowship, and numerous best paper awards. Leskovec received his bachelor's degree in computer science from University of Ljubljana, Slovenia, and his PhD in in machine learning from the Carnegie Mellon University and postdoctoral training at Cornell University.

Keynote: Contextual Reinforcement Learning

Speaker:

Dr. John Langford, Principal Researcher, Microsoft Research New York, USA,

Abstract:

I will discuss a decade long research project to create the foundations of reinforcement learning with context (aka features). This research project has multiple threads including Contextual Bandits, Learning to Search, and Contextual Decision Processes. The most mature of these (Contextual Bandits) is now driving many real-world RL applications while the least mature (CDPs) is a fascinating theoretician's toy.

Short Bio:

John Langford is a machine learning research scientist, a field which he says "is shifting from an academic discipline to an industrial tool". He is the author of the weblog hunch.net and the principal developer of Vowpal Wabbit. John works at Microsoft Research New York, of which he was one of the founding members, and was previously affiliated with Yahoo! Research, Toyota Technological Institute at Chicago, and IBM's Watson Research Center. He studied Physics and Computer Science at the California Institute of Technology, earning a double bachelor's degree in 1997, and received his Ph.D. in Computer Science from Carnegie Mellon University in 2002. He was the program co-chair for the 2012 International Conference on Machine Learning.

Conference Paper Presentations

L1: High Performance Platforms for Big Data (I)	
Regular	BigD216 "Jointly Optimizing Task Granularity and Concurrency for In-Memory MapReduce Frameworks" Jonghyun Bae, Hakbeom Jang, Wenjing Jin, Jun Heo, Jaeyoung Jang, Joo-Young Hwang, Sangyeun Cho, and Jae W. Lee
Regular	BigD238 "Making Caches Work for Graph Analytics" Yunming Zhang, Vladimir Kiriansky, Charith Mendis, Matei Zaharia, and Saman Amarasinghe
Regular	BigD254 "HarpLDA+: Optimizing Latent Dirichlet Allocation for Parallel Efficiency" Bo Peng, Bingjing Zhang, Langshi Chen, Mihai Avram, Robert Henschel, Craig Stewart, Shaojuan Zhu, Emily Mccallum, Lisa Smith, Tom Zahniser, Jon Omer, and Judy Qiu
Regular	BigD643 "Hierarchical Automata Construction for Approximate Pattern Matching on Automata Processors" Xiaodong Yu, Kaixi Hou, Hao Wang, and Wu-chun Feng

L2: Big Data Applications	
Regular	BigD301 "Fast Interpolation of Grid Data at a Non-Grid Point" Hiroshi Inoue
Regular	BigD313 "Application of Big Data analytics in process safety and risk management" Pankaj Goel, Prerna Jain, Aniruddha Datta, and M.Sam Mannan
Regular	BigD343 "HealthEdge: Task Scheduling for Edge Computing with Health Emergency and Human Behavior Consideration in Smart Homes" Haoyu Wang, Jiaqi Gong, Yan Zhuang, Haiying Shen, and John Lach
Regular	BigD345 "Joint Sparse Auto-encoder: A Semi-supervised Spatio-temporal Approach in Mapping Large-scale Croplands" Xiaowei Jia, Yifan Hu, Ankush Khandelwal, Anuj Karpatne, and Vipin Kumar

L3: Algorithms and Systems for Big Data Search (I)	
Regular	BigD224 "Sampling Algorithms to Update Truncated SVD" Ichitaro Yamazaki, Stanimire Tomov, and Jack Dongarra
Regular	BigD314 "Active Learning Based News Veracity Detection with Feature Weighting and Deep-Shallow Fusion" Sreyasee Das Bhattacharjee, Ashit Talukder, and Bala Venkatram Balantrapu
Regular	BigD410 "Rectangular Hash Table: Bloom Filter and Bitmap Assisted Hash Table with High Speed" Tong Yang, Binchao Yin, Hang Li, Muhammad Shahzad, Steve Uhlig, Bin Cui, and Xiaoming Li
Regular	BigD418 "Potentiality of Healthcare Big data: Improving Search by Automatic Query Reformulation" Yueyao Wang, Qinmin Vivian Hu, Yang Song, and Liang He

L4: Novel Theoretical Models for Big Data	
Regular	BigD409 "Lifelong Multi-Task Multi-View Learning Using Latent Spaces" Xiaoli Li, Sai Nivedita Chandrasekaran, and Jun Huan
Regular	BigD448 "Compact Multi-Class Boosted Trees" Natalia Ponomareva, Thomas Colthurst, Gilbert Hendry, Salem Haykal, and Soroush Radpour
Regular	BigD375 "The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction" Eric Breck, Shanqing Cai, Eric Nielsen, Michael Salib, and D Sculley
Regular	BigD395 "LSTM for Septic Shock: Adding Unreliable Labels to Reliable Predictions" Yuan Zhang, Chen Lin, Min Chi, Julie Ivy, Muge Capan, and Jeanne M. Huddleston

L5: High Performance Platforms for Big Data (II)	
Regular	BigD333 "Scaling Up Data-Parallel Analytics Platforms: Linear Algebraic Operation Cases" Luna Xu, Seung-Hwan Lim, Min Li, Ali R. Butt, and Ramakrishnan Kannan

Regular	BigD649 "I/O Load Balancing for Big Data HPC Applications" Arnab K. Paul, Arpit Goyal, Feiyi Wang, Sarp Oral, Ali R. Butt, Michael J. Brim, and Sangeetha B. Srinivasa
Regular	BigD645 "Characterizing and Accelerating Indexing Techniques on Distributed Ordered Tables" Shashank Gugnani, Xiaoyi Lu, Houliang Qi, Li Zha, and Dhabaleswar K. Panda
Regular	BigD657 "Performance Characterization and Acceleration of Big Data Workloads on OpenPOWER System" Xiaoyi Lu, Haiyang Shi, Dipti Shankar, and Dhabaleswar K. Panda
Regular	BigD257 "Low-latency Multi-threaded Ensemble Learning for Dynamic Big Data Streams" Diego Marrón, Eduard Ayguadé, José R. Herrero, Jesse Read, and Albert Bifet

L6: Spatiotemporal Analytics and Traffic Applications	
Regular	BigD350 "Spatiotemporal Range Pattern Queries on Large-scale Co-movement Pattern Datasets" Shahab Helmi and Farnoush Banaei-Kashani
Regular	BigD462 "A Data-Driven Congestion Diffusion Model for Characterizing Traffic in Metrocity Scales" Baoxin Zhao, Cheng-Zhong Xu, and Siyuan Liu
Regular	BigD520 "BTCI: a New Framework for Identifying Congestion Cascades Using Bus Trajectory Data" Meng-Fen Chiang, Ee-Peng Lim, Wang-Chien Lee, and Agus Trisnajaya Kwee
Regular	BigD635 "Enabling Versatile Analysis of Large Scale Traffic Video Data with Deep Learning and HiveQL" Lei Huang, Weijia Xu, Si Liu, Venktesh Pandey, and Natalia Ruiz Juri
Regular	BigD365 "Automated Scalable Detection of Location-Specific Santa Ana Conditions from Weather Data using Unsupervised Learning" Mai Nguyen, Daniel Crawl, Jianxin Li, Dylan Uys, and Ilkay Altintas

L7: Algorithms and Systems for Big Data Search (II)	
Regular	BigD476 "A Scalable Model for Tracking Topical Evolution in Large Document Collections" Sheikh Motahar Naim, Arnold Boedihardjo, and M. Shahriar Hossain
Regular	BigD491 "A Fast Non-Volatile Memory aware Algorithm for Generating Random Scale-Free Networks" Cheng-Chin Tu, Mi-Yen Yeh, and Tei-Wei Kuo
Regular	BigD399 "High-Performance Geometric Algorithms for Sparse Computation in Big Data Analytics" Philipp Baumann, Dorit Hochbaum, and Quico Spaen
Regular	BigD608 "Distributed Top-N Local Outlier Detection in Big Data" Yizhou Yan, Lei Cao, and Elke Rundensteiner
Regular	BigD393 "On On-line Task Assignment in Spatial Crowdsourcing" Mohammad Asghari and Cyrus Shahabi

L8: Large-scale Recommendation Systems and Social Media Systems	
Regular	BigD303 "Hierarchical Collaborative Embedding for Context-Aware Recommendations" Lei Zheng, Bokai Cao, Vahid Noroozi, Philip S. Yu, and Nianzu Ma
Regular	BigD433 "ImWalkMF: Joint Matrix Factorization and Implicit Walk Integrative Learning for Recommendation" Chuxu Zhang, Lu Yu, Xiangliang Zhang, and Nitesh Chawla
Regular	BigD497 "A Comparative Study of Matrix Factorization and Random Walk with Restart in Recommender Systems" Haekyu Park, Jinhong Jung, and U Kang
Regular	BigD564 "Online City-scale Hyper-local Event Detection via Analysis of Social Media and Human Mobility" Jun Hu, Yuxin Wang, and Ping Li
Regular	BigD648 "Text-based Geolocation Prediction of Social Media Users with Neural Networks" Ismini Lourentzou, Alex Morales, and Chengxiang Zhai

L9: Software Systems for Big Data Computing	
Regular	BigD439 "Sanzu: A Data Science Benchmark" Alex Watson, Deepigha Vittal Babu, and Suprio Ray
Regular	BigD516 "Delphi: A multi-user, multi-method cloud based model exploration system" Kalyan Veeramachaneni, Thomas Swearingen, and Arun Ross
Regular	BigD596 "ooc_cuDNN: Accommodating Convolutional Neural Networks over GPU Memory Capacity" Yuki Ito, Ryo Matsumiya, and Toshio Endo
Regular	BigD650 "A Semantics-Aware Storage Framework for Scalable Processing of Knowledge Graphs on Hadoop" HyeongSik Kim, Padmashree Ravindra, and Kemafor Anyanwu

L10: New Computational Models for Big Data	
Regular	BigD379 "Robust Multi-Label Semi-Supervised Classification" Sheng Li and Yun Fu
Regular	BigD646 "Collective Subjective Logic: Scalable Uncertainty-based Opinion Inference" Feng Chen, Chunpai Wang, and Jin-Hee Cho
Regular	BigD394 "Sequential algorithms to split and merge ultra-high resolution 3D images" Valerie Hayot-Sasson, Yongping Gao, Yuhong Yan, and Tristan Glatard
Regular	BigD316 "Multi-step Prediction with Missing Smart Sensor Data using Multi-task Gaussian Processes" Pasan Karunaratne, Masud Moshtaghi, Shanika Karunasekera, Aaron Harwood, and Trevor Cohn

L11: Security and Privacy	
Regular	BigD457 "Shade: A Differentially-Private Wrapper For Enterprise Big Data" Alec Heifetz, Vaikkunth Mugunthan, and Lalana Kagal
Regular	BigD594 "Group Privacy-aware Disclosure of Association Graph Data" Balaji Palanisamy, Chao Li, and Prashant Krishnamurthy
Regular	BigD351 "Contaminant Removal for Malware Detection on Android" Lichao Sun, Xiaokai Wei, Jiawei Zhang, Lifang He, Philip S. Yu, and Witawas Srisa-an

L12: Multimedia and Multi-structured Data - Big Variety Data	
Regular	BigD525 "Error-Robust Multi-View Clustering" Mehrnaz Najafi, Lifang He, and Philip S. Yu
Regular	BigD604 "VIGAN: Missing View Imputation with Generative Adversarial Networks" Chao Shang, Aaron Palmer, Jiangwen Sun, Ko-Shin Chen, Jin Lu, and Jinbo Bi
Regular	BigD613 "Inverse Extreme Learning Machine for Learning with Label Proportions" Limeng Cui, Jiawei Zhang, Zhensong Chen, Yong Shi, and Philip S. Yu
Regular	BigD549 "Fast Access to Columnar, Hierarchical Data via Code Transformation" Jim Pivarski, Peter Elmer, Brian Bockelman, and Zhe Zhang

L13: Social Web Search and Mining	
Regular	BigD552 "Large-Scale Joint Topic, Sentiment & User Preference Analysis for Online Reviews" Xinli Yu, Zheng Chen, Wei-Shih Yang, Xiaohua Hu, and Erjia Yan
Regular	BigD505 "MRAttractor: Detecting Communities from Large-Scale Graphs" Nguyen Vo, Kyumin Lee, and Thanh Tran
Regular	BigD276 "Connecting Emerging Relationships from News via Tensor Factorization" Philip S. Yu, Yi Chang, Bokai Cao, Chun-Ta Lu, and Jingyuan Zhang
Regular	BigD437 "Holistic and Scalable Ranking of RDF Data" Axel-Cyrille Ngonga Ngomo, Michael Hoffmann, Ricardo Usbeck, and Kunal Jha
Regular	BigD260 "Crack Random Forest for Arbitrary Large Datasets" Alessandro Lulli, Luca Oneto, and Davide Anguita

L14: Stream Data Mining - Big Velocity Data	
Regular	BigD539 "Tiered Sampling: An Efficient Method for Approximate Counting Sparse Motifs in Massive Graph Streams" Lorenzo De Stefani, Erisa Terolli, and Eli Upfal
Regular	BigD560 "S-Isomap++: Multi Manifold Learning from Streaming Data" Suchismit Mahapatra and Varun Chandola
Regular	BigD612 "Multistream Regression with Asynchronous Concept Drift Detection" Ahsanul Haque, Bo Dong, Yifan Li, Yang Gao, Latifur Khan, and Mohammad Masud
Regular	BigD664 "Detecting Changes in Streaming Data with Information-Theoretic Windowing" Ryoya Kaneko, Kohei Miyaguchi, and Kenji Yamanishi
Regular	BigD404 "Rhea: Adaptively Sampling Authoritative Content from Social Activity Streams" Panagiotis Liakos, Alexandros Ntoulas, and Alex Delis

L15: Data and Information Quality for Big Data	
Regular	BigD327 "Quality-aware Aggregation & Predictive Analytics at the Edge" Natascha Harth and Christos Anagnostopoulos
Regular	BigD368 "Elastic Management of Cloud Applications using Adaptive Markov Models" Konstantinos Lolos, Ioannis Konstantinou, Verena Kantere, and Nectarios Koziris
Regular	BigD527 "How Fast Can One Scale Down a Distributed File System?" Nathanael Cheriere and Gabriel Antoniu
Regular	BigD252 "A Decision Tree Based Approach Towards Adaptive Modeling of Big Data Applications" Ioannis Giannakopoulos, Dimitrios Tsoumakos, and Nectarios Koziris
Regular	BigD526 "Constraint-Aware Dynamic Truth Discovery in Big Data Social Media Sensing" Daniel (Yue) Zhang, Dong Wang, and Yang Zhang

L16: Link and Graph Analytics (I)	
Regular	BigD432 "Hybrid Algorithms for Subgraph Pattern Queries in Graph Databases: An Evaluation" Foteini Katsarou, Nikos Ntarmos, and Peter Triantafillou
Regular	BigD602 "Domain-specific Hierarchical Subgraph Extraction: A Recommendation Use Case" Sarasi Lalithsena, Sujan Perera, Pavan Kapanipathi, and Amit Sheth
Regular	BigD233 "Closed Walk Sampler: An Efficient Method for Estimating the Spectral Radius of Large Graphs" Guyue Han and Harish Sethu
Regular	BigD274 "Estimation of distance-based metrics for very large graphs with MinHash Signatures" Giambattista Amati, Simone Angelini, Giorgio Gambosi, Gianluca Rossi, and Paola Vocca
Regular	BigD376 "Bias Correction in Clustering Coefficient Estimation" Roohollah Etemadi and Jianguo Lu

L17: Link and Graph Analytics (II)	
Regular	BigD571 "Towards Robust Models of Food Flows and Their Role in Invasive Species Spread" Srinivasan Venkatramanan, Sichao Wu, Bowen Shi, Achla Marathe, Madhav Marathe, Stephen Eubank, Lalit Sah, A.P. Giri, Luke Colavito, Nitin S, Sridhar V, Asokan R, Rangaswamy Muniappan, George Norton, and Abhijin Adiga
Regular	BigD402 "CoEuS: Community Detection via Seed-set Expansion on Graph Streams" Panagiotis Liakos, Alexandros Ntoulas, and Alex Delis
Regular	BigD590 "E-CLoG: Counting Edge-Centric Local Graphlets" Vachik Dave, Nesreen Ahmed, and Mohammad Hasan
Regular	BigD357 "Bayesian Multi-View Models for Member-Job Matching and Personalized Skill Recommendations" Abhinav Maurya and Rahul Telang

L18: Analytics and Data Management for Big Data	
Regular	BigD449 "Exploiting Visual and Textual Neighborhood Information to Improve Image-Tag Relevance" Chandramani Chaudhary, Poonam Goyal, and Yi-Ping Phoebe Chen
Regular	BigD512 "Drum: A Rhythmic Approach to Interactive Analytics on Large Data" Jianfeng Jia, Chen Li, and Michael Carey
Regular	BigD425 "QuAD: A Quorum Protocol for Adaptive Data Management in the Cloud" Ilir Fetai, Alexander Stiemer, and Heiko Schuldt

S1: Algorithms for Big Data (1)	
short	BigD309 “Distributed Decision Tree v.2.0” Ankit Desai and Sanjay Chaudhary
short	BigD239 “Entropic Determinants” Diego Granziol and Stephen Roberts
short	BigD294 “Judicious Setting of Dynamic Time Warping’s Warping Window Width Allows More Accurate Classification of Time Series” Hoang Anh Dau, Diego Furtado Silva, FranÃ§ois Petitjean, Germain Forestier, Anthony Bagnall, and Eamonn Keogh
short	BigD323 “Discrimination Detection by Causal Effect Estimation” Jiuyong Li, Jixue Liu, Lin Liu, Thuc Le, Saisai Ma, and Yizhao Han
short	BigD581 “Setting the threshold for high throughput detectors: A mathematical approach for ensembles of dynamic, heterogeneous, probabilistic anomaly detectors” Robert Bridges, Jessie Jamieson, and Joel Reed
short	BigD335 “Iterative Matrix Correlation for Bisection Clustering” Byron Gao, Robert Tung, and Yong Yang
short	BigD665 “NVMD: Non-Volatile Memory Assisted Design for Accelerating MapReduce and DAG Execution Frameworks on HPC Systems” Md Wasi-ur- Rahman, Nusrat Islam, Xiaoyi Lu, and Dhabaleswar Panda

S2: Text Mining/NLP	
short	BigD662 “Product Function Need Recognition via Semi-supervised Attention Network” Hu Xu, Sihong Xie, Lei Shu, and Philip S. Yu
short	BigD506 “Reliable Fake Review Detection via Modeling Temporal and Behavioral Patterns” Xian Wu, Yuxiao Dong, Jun Tao, Chao Huang, and Nitesh Chawla
short	BigD501 “WEAC: Word Embeddings for Anomaly Classification from Event Logs” Amit Pande and Vishal Ahuja
short	BigD557 “Scalable Document Similarity Using Linear-Complexity Word Mover's Distance” Kubilay Atasu, Thomas Parnell, Celestine Duenner, Manolis Sifalakis, Haralampos Pozidis, Vasileios Vasileiadis, Michail Vlachos, Cesar Berrospi, and Abdel LAbbi
short	BigD331 “Bringing Semantic Structures to User Intent Detection in Online Medical Queries” Chenwei Zhang, Nan Du, Wei Fan, Yaliang Li, Chun-Ta Lu, and Philip S. Yu

short	BigD622 “Energy Efficient Stochastic-Based Deep Spiking Neural Networks for Sparse Datasets” Mohammed Alawad, Hong-Jun Yoon, and Georgia Tourassi
short	BigD662 “Product Function Need Recognition via Semi-supervised Attention Network” Hu Xu, Sihong Xie, Lei Shu, and Philip S. Yu
short	BigD231 “Analysis of the Term 'Big Data': Usage in Biomedical Publications” Allard Jan-Jaap van Altena, Perry D Moerland, Aeilko H Zwinderman, and SÃlvia Delgado Olabarriaga

S3: Distributed System and Software for Big Data (1)	
short	BigD236 “Big Data and HPC collocation: Using HPC Idle Resources for Big Data Analytics” Michael Mercier, David Glesser, Yiannis Georgiou, and Olivier Richard
short	BigD318 “CStorage: An Efficient Classification-based Image Storage System in Cloud Datacenters” Heng Zhou and Haiying Shen
short	BigD559 “A Distributed k-Core Decomposition Algorithm on Spark” Aritra Mandal and Mohammad Al Hasan
short	BigD374 “Multi-objective Optimization of Scheduling Dataflows on Heterogeneous Cloud Resources” Ilia Pietri, Yannis Chronis, and Yannis Ioannidis
short	BigD373 “Universal Distant Reading through Metadata Proxies with ArchiveSpark” Helge Holzmann, Vinay Goel, and Emily Novak Gustainis
short	BigD618 “Fast Graph Scan Statistics Optimization Using Algebraic Fingerprints” Jose Cadena, Saliya Ekanayake, and Anil Vullikanti
short	BigD415 “Towards Memory and Computation Efficient Graph Processing on Spark” Xinhui Tian, Yuanqing Guo, and Jianfeng Zhan
short	BigD442 “Understanding and Optimizing the Performance of Distributed Machine Learning” Applications on Apache Spark Celestine DÃnner, Thomas Parnell, Kubilay Atasu, Manolis Sifalakis, and Haralampos Pozidis

S4: Distributed System and Software for Big Data (2)	
short	BigD219 “An Open-Source Tool For The Transcription of Paper-Spreadsheet Data” Mohammad Ghassemi, Willow Jarvis, Tuka Alhanai, Emery Brown, Roger Mark, and M. Brandon Westover
short	BigD591 “eTRIKS Analytical Environment: A Modular High Performance Framework for Medical Data Analysis” Axel Oehmichen, Florian Guitton, Kai Sun, Jean Grizet, Thomas Heinis, and Yike Guo
short	BigD680 “Standardizing Big Earth Datacubes” Peter Baumann
short	BigD493 “On the Usability of Hadoop MapReduce, Apache Spark & Apache Flink for Data Science”

	Bilal Akil and Uwe Roehm
short	BigD475 “External Memory Pipelining Made Easy With TPIE” Lars Arge, Mathias Rav, Svend C. Svendsen, and Jakob Truelsen
short	BigD392 “Optimal Reducer Placement to Minimize Data Transfer in MapReduce-Style Processing” Xiao Meng and Lukasz Golab
short	BigD362 “A comparative analysis of state-of-the-art SQL-on-Hadoop systems for interactive analytics” Ashish Tapdiya and Daniel Fabbri
short	BigD473 “Dione: Profiling Spark Applications Exploiting Graph Similarity” Nikos Zacheilas, Stathis Maroulis, and Vana Kalogeraki

S2: Algorithms for Big Data (2)	
short	BigD446 “Two-level Clustering Fast Betweenness Centrality Computation for Requirement-driven Approximation” Angelo Furno, Nour Eddin El Faouzi, Rajesh Sharma, and Eugenio Zimeo
short	BigD369 “Compressed Domain-Specific Data Processing and Analysis” Dapeng Dong and John Herbert
short	BigD213 “In-Depth Exploration of Single-Snapshot Lossy Compression Techniques for N-Body Simulations” Dingwen Tao, Sheng Di, Zizhong Chen, and Franck Cappello
short	BigD511 “Discovering Co-occurrence Patterns of Heterogeneous Events from Unevenly-distributed Spatiotemporal Data” Hung Tran-The and Koji Zettsu
short	BigD349 “Cellular Network Configuration via Online Learning and Joint Optimization” Xueying Guo, George Trimponias, Xiaoxiao Wang, Zhitang Chen, Yanhui Geng, and Xin Liu
short	BigD445 “Large-scale Point-of-Interest Category Prediction Using Natural Language Processing Models” Daniel (Yue) Zhang, Dong Wang, Hao Zheng, Xin Mu, Qi Li, and Yang Zhang
short	BigD441 “Mining Pros and Cons of Actions from Social Media for Decision Support” Ebad Ahmadzadeh and Philip Chan

S6: Big Data Preprocessing/Visualization	
Short	BigD682 “Enhancing Data Quality by Cleaning Inconsistent Big RDF Data” Salima Benbernou and Mourad Ouziri
Short	BigD683 “Data Context Informed Data Wrangling” Martin Koehler, Alex Bogatu, Cristina Civili, Nikolaos Konstantinou, Edward Abel, Alvaro A A Fernandes, John Keane, Leonid Libkin, and Norman W. Paton
Short	BigD684 “Micro-Clustering by Data Polishing” Takeaki Uno, Hiroki Maegawa, Takanobu Nakahara,

	Yukinobu Hamuro, Ryo Yoshinaka, and Makoto Tatsuta
Short	BigD310 “A Distributed Rough Set Theory based Algorithm for an Efficient Big Data Pre-processing under the Spark Framework” Zaineb Chelly Dagdia, Christine Zarges, GaÃ«l Beck, and Mustapha Lebbah
Short	BigD686 “Visual Analytics with Unparalleled Variety Scaling for Big Earth Data” Lina Yu, Michael Rilee, Yu Pan, Feiyu Zhu, Kwo-Sen Kuo, and Hongfeng Yu
Short	BigD687 “Toward Granular Knowledge Analytics for Data Intelligence” Alexander Denzler and Michael Kaufmann
Short	BigD469 “Seq2Img: A Sequence-to-Image based Approach Towards IP Traffic Classification using Convolutional Neural Networks” Zhitang Chen, Ke He, Jian Li, and Yanhui Geng

S7: Streaming Data	
Short	BigD633 “AnyFI: An Anytime Frequent Itemset Mining Algorithm for Data Streams” Poonam Goyal, Jagat Sesh Challa, Shivin Shrivastava, and Navneet Goyal
Short	BigD547 “Queryable Compression on Streaming Social Networks” Michael Nelson, Sridhar Radhakrishnan, Amlan Chatterjee, and Chandra Sekharan
Short	BigD428 “Efficient Diversified Set Monitoring for Mobile Sensor Stream Environments” Masahiro Yokoyama, Takahiro Hara, and Sanjay Madria
Short	BigD471 “Fishing in the Stream: Similarity Search over Endless Data” Naama Kraus, David Carmel, and Idit Keidar
Short	BigD430 “Event Pattern Discovery by Keywords in Graph Streams” Mohammad Hossein Namaki, Peng Lin, and Yinghui Wu
Short	BigD663 “Big Data Transfer Optimization Based on Offline Knowledge Discovery and Adaptive Real-time Sampling” MD S Q Zulkar Nine, Kemal Guner, Ziyun Huang, Xiangyu Wang, Jinhui Xu, and Tevfik Kosar
Short	BigD579 “RePAIR: Recommend Political Actors In Real-time From News Websites” Mohiuddin Solaimani, Sayeed Salam, and Latifur Khan

S8: Machine Learning (1)	
Short	BigD679 “Quality-Efficiency Trade-offs in Machine Learning for Text Processing” Ricardo Baeza-Yates and Zeinab Liaghat
Short	BigD541 “Automated Industry Classification with Deep Learning” Sam Wood, Rohit Muthyala, Yi Jin, Hua Gao, Yixing Qin, Amit Rai, and Nilaj Rukadikar
Short	BigD515 “Discovering Potential Traffic Risk in Japan using Supervised Learning Approach” Tatsuru Kobayashi, Shin Matsushima, Lee Taito, and Kenji Yamanishi

Short	BigD400 “Distributed Bayesian Piecewise Sparse Linear Models” Masato Asahara and Ryohei Fujimaki
Short	BigD585 “Forecasting the Rise and Fall of Volatile Point-of-Interests” Xinjiang Lu, Zhiwen Yu, Chuanren Liu, Yanchi Liu, Hui Xiong, and Bin Guo
Short	BigD429 “A Closed-loop Deep Learning Architecture for Robust Activity Recognition using Wearable Sensors” Ramyar Saeedi, Skyler Norgaard, and Assefaw Gebremedhin
Short	BigD463 “Semi-Supervised Convolutional Neural Networks for Human Activity Recognition” Ming Zeng, Tong Yu, Xiao Wang, Le T. Nguyen, Ole J. Mengshoel, Ian Lane, and Joy Zhang
Short	BigD641 “Big Active Learning” Er-Chen Huang, Hsing-Kuo Pao, and Yuh-Jye Lee

S9: Machine Learning (2)	
Short	BigD450 “Differentially Private Query Learning: from Data Publishing to Model Publishing” Tianqing Zhu, Ping Xiong, Gang Li, Wanlei Zhou, and Philip S. Yu
Short	BigD385 “Multi-View Graph Learning with Adaptive Label Propagation” Sheng Li, Hongfu Liu, Zhiqiang Tao, and Yun Fu
Short	BigD387 “Graphical Approach for Influence Maximization in Social Networks Under Generic Threshold-based Non-submodular Model” Liang Ma, Guohong Cao, and Lance Kaplan
Short	BigD677 “Sandpiper: Scaling Probabilistic Inferencing to Large Scale Graphical Models” Alexander Ulanov, Manish Marwah, Mijung Kim, Roshan Dathathri, Carlos Zubieta, and Jun Li
Short	BigD638 “Application-Specific Graph Sampling for Frequent Subgraph Mining and Community Detection” Sumit Purohit, Lawrence Holder, and Sutanay Choudhury
Short	BigD230 “Model-Driven Reverse Engineering of NoSQL Property Graph Databases” Isabelle Comyn-Wattiau and Jacky Akoka
Short	BigD237 “Exponential Random Graph Models with Big Networks: Maximum Pseudolikelihood Estimation and the Parametric Bootstrap” Christian Schmid and Bruce Desmarais
Short	BigD391 “A novel approach to optimization of iterative machine learning algorithms: over heap structure” Hasan Kurban and Mehmet Dalkilic

S10: Big Data Applications (1)	
Short	BigD220 “A Data-Driven Approach to Predict NOx-Emissions of Gas Turbines” Giuseppe Cuccu, Somayeh Danafar, Philippe CudrÃ©-Mauroux, Martin Gassner, Stefano Bernero, and Krzysztof Kryszczuk
Short	BigD317 “Predicting Treatment Repetitions in the Implant Denture Therapy Process” Marzieh Bakhshandeh, Dennis M.M. Schunselaar, Henrik Leopold, and Hajo A. Reijers

Short	BigD678 “Predicting regional economic indices using big data of individual bank card transactions” Emanuele Massaro, Stanislav Sobolevsky, Iva Bojic, Juan Murillo Arias, and Carlo Ratti
Short	BigD403 “OTPS: A Decision Support Service for Optimal Airfare Ticket Purchase” Yuchang Xu and Jian Cao
Short	BigD234 “A Single-Node Datastore for High-Velocity Multidimensional Sensor Data” Juan Colmenares, Reza Dorrigiv, and Daniel Waddington
Short	BigD531 “Exploring the Dynamics of Surge Pricing in Mobility-on-Demand Taxi Services” Wenbo Zhang, Dheeraj Kumar, and Satish Ukkusuri
Short	BigD264 “Identifying and Quantifying Nonlinear Structured Relationships in Complex Manufactural Systems” Tingyang Xu, Tan Yan, Dongjin Song, Wei Cheng, Haifeng Chen, Geoff Jiang, and Jinbo Bi
Short	BigD266 “Detecting Unmetered Taxi Rides from Trajectory Data” Xibo Zhou, Ye Ding, Fengchao Peng, Qiong Luo, and Lionel M. Ni
Short	BigD496 “T-BMIRT: Estimating Representations of Student Knowledge and Educational Components in Online Education” Jiankun Huang and Wenjun Wu

S11: Big Data Applications (2)	
Short	BigD321 “Weatherman: Exposing Weather-based Privacy Threats in Big Energy Data” Dong Chen and David Irwin
Short	BigD454 “Personalized Travel Mode Detection with Smartphone Sensors” Xing Su, Yuan Yao, Qing He, Jie Lu, and Hanghang Tong
Short	BigD416 “Privacy-protected Place of Activity Mining on Big Location Data” Shuo Wang, Richard Sinnott, and Surya Nepal
Short	BigD356 “Event-Based Non-Parametric Clustering of Team Sport Trajectories” Fengchao Peng, Yudian Ji, Qiong Luo, and Lionel M. Ni
Short	BigD361 “Personalized Flight Recommendations via Paired Choice Modeling” Jian Cao, Fangzhou Yang, and Yuchang Xu
Short	BigD593 “OReONet: Deep Convolutional Network for Oil Reservoir Optimization” Chung Ming Cheung, Palash Goyal, Viktor K. Prasanna, and Arash Saber Tehrani
Short	BigD417 “Sensitive Gazetteer Discovery and Protection for Mobile Social Media Users” Shuo Wang, Richard Sinnott, and Surya Nepal
Short	BigD455 “Travel Purpose Inference with GPS Trajectories, POIs, and Geo-tagged Social Media Data” Chuishi Meng, Yu Cui, Qing He, Lu Su, and Jing Gao
Short	BigD661 "Discovering Scientific Influence using Cross-Domain Dynamic Topic Modeling" Jennifer Sleeman, Milton Halem, Tim Finin, and Mark Cane

Industry and Government Paper Presentations

I&G-regular1: Big Data Analytics	
Regular	N214-*What is Skipped: Finding Desirable Items in E-Commerce Search by Discovering the Worst Title Tokens* Ishita Khan, Prathyusha Senthil Kumar, Daniel Miranda, and David Goldberg
Regular	*N223-Dependency Analysis of Cloud Applications for Performance Monitoring using Recurrent Neural Networks* Syed Yousaf Shah, Zengwen Yuan, Songwu Lu, and Petros Zerfos
Regular	*N221-Predicting Over-Indebtedness on Batch and Streaming Data* Jacob Montiel, Albert Bifet, and Talel Abdessalem
Regular	N233- *Topic Models for RFID Data Modeling and Localization* Timothy Kennedy, Robert Provence, James Broyan, Patrick Fink, Phong Ngo, and Lazaro Rodriguez
Regular	N243- *Representativeness of Latent Dirichlet Allocation Topics Estimated from Data Samples with Application to Common Crawl* Yuheng Du, Alexander Herzog, Andre Luckow, Ramu Nerella, and Amy Apon

I&G-regular2: Big Data Applications (1)	
Regular	*N209-Application of Dynamic Logistic Regression with Unscented Kalman Filter in Predictive Coding* Yihua Astle, Xuning Tang, and Craig Freeman
Regular	*N210-RAVEN: Web-based Smart Home Exploration System Through Interactive Pattern Discovery* Mansurul Bhuiyan and Mohammad Hasan
Regular	N217- *Trendi: Tracking Stories in News and Microblogs via Emerging, Evolving and Fading Topics* Xuchao Zhang, Liang Zhao, Zhiqian Chen, Arnold Boedihardjo, Dai Jing, and Chang-Tien Lu
Regular	N231- *A data-driven approach for multivariate contextualized anomaly detection: industry use case* Nenad Stojanovic, Marko Dinic, and Ljiljana Stojanovic
Regular	N241- *Fast Botnet Detection From Streaming Logs Using Online Lanczos Method* Zheng Chen, Xinli Yu, Chi Zhang, Jin Zhang, Cui Lin, Xiaohua Hu, Erjia Yan, and Wei-Shih Yang

I&G-short1: Big Data Algorithms & Systems	
Short	N224- *On Event-Driven Leaning of Knowledge in Smart Factories: The Case of Siemens* Martin Ringsquandl, Steffen Lamparter, Evgeny Kharlamov, Raffaello Lepratti, Daria Stepanova, Peer Kröger, and Ian Horrocks
Short	N232- *TRACES: Generating Twitter Stories via Shared Subspace and Temporal Smoothness* Xuchao Zhang, Zhiqian Chen, Liang Zhao, Arnold Boedihardjo, and Chang-Tien Lu
Short	N245- *Demystifying Dark Matter for Online Experimentation* Nirupama Appikatala, Miao Chen, Michael Natkovich, and Joshua Walters
Short	N247- *Faster Online Experimentation by Eliminating Traditional A/A Validation* Russell Chen, Miao Chen, Mahendrasinh Ramsinh Jadav, Joonsuk Bae, and Don Matheson,
Short	N249- *Linking Many Unusual Co-Incidences* Kevin Pratt
Short	N255- *Detecting and Summarizing Emergent Events in Microblogs and Social Media Streams by Dynamic Centralities* Neela Avudaiappan, Alexander Herzog, Sneha Kadam, Yuheng Du, Jason Thatcher, and Ilya Safro
Short	N262- *Connected Health: Opportunities and Challenges* Ankita Nambiar, Nikitha Reddy, and Debojyoti Dutta
Short	N213- *Real Time Semantic Enrichment of Broadcast Content in the Big Data Age* Maurizio Montagnuolo, Alberto Messina, Nicolò Bidotti, Paolo Platter, and Alessio Bosca
Short	N219- *Tracking and Predicting the Evolution of Research Topics in Scientific Literature* Christine Balili and Aviv Segev

Short	N222- *On the Improvement of Classifying EEG Recordings Using Neural Nerworks* Yiran Zhao, Shuochao Yao, Shaohan Hu, Shiyu Chang, Raghu Ganti, Mudhakar Srivatsa, Shen Li, and Tarek Abdelzaher
Short	*N235-Towards a Semantic Keyword Search over Industrial Knowledge Graphs (Extended Abstract)* Gong Cheng and Evgeny Kharlamov
Short	N246- *Designing a High Performance Cluster for Large-Scale SQL-on-Hadoop Analytics* Ajay Dholakia, Prasad Venkatachar, Kshitij Doshi, Ravikanth Durgavajhala, Stewart Tate, Berni Schiefer, Matthew Sheard, and Ramnath Sai Sagar

I&G-regular3: Big Data Platforms & Frameworks	
Regular	*N215-Scalable Time-Versioning Support for Property Graph Databases* Warut D. Vijitbenjaronk, Jinho Lee, and Toyotaro Suzumura
Regular	N220- *Integrated Access to Big Data Polystores through a Knowledge-driven Framework* Justin McHugh, Paul Cuddihy, Jenny Williams, Kareem Aggour, Vijay Kumar, and Varish Mulwad
Regular	N226- *Flux: Groupon's automated, scalable machine learning platform* Derrick Spell, Xiao-Han Zeng, Jae-Young Chung, Bahador Nooraei, Ricki Shomer, Ling-Yong Wang, James Gibson, and Daniel Kirsche
Regular	*N205-Performance Optimization In Scale-out Storage Using Design Of Experiment As Heuristic* Lay Wai Kong
Regular	N229- *A Gamma-based Regression for Winning Price Estimation in Real-Time Bidding Advertising* Wen-Yuan Zhu, Wen-Yueh Shih, Ying-Hsuan Lee, Wen-Chih Peng, and Jiun-Long Huang

I&G-short2: Massive Processing & Experience	
Short	*N206-Architectural Considerations for Highly Scalable Computing to Support On-demand Video Analytics* George Mathew
Short	N234- *Scalable Distributed Change Detection and its Application to Maritime Traffic* Leonardo Maria Millefiori, Paolo Braca, and Gianfranco Arcieri
Short	N211- *Knowledge extraction from maritime spatiotemporal data: An evaluation of clustering algorithms on Big Data* Giannis Spiliopoulos, Konstantinos Chatzikokolakis, Dimitrios Zissis, Evmorfia Biliri, Dimitrios Papaspyros, and Giannis Tsapelas
Short	N257- *BBC: A DSL for Designing Cloud-based Heterogeneous Bigdata Pipelines* Ferosh Jacob, Ilamgumaran Karunanithi, Pramod Salian, and Ravi Sambhu
Short	*N258-Predictive Edge Computing for Time Series of Industrial IoT and Large Scale Critical Infrastructure based on Open-source Software Analytics of Big Data,* Emmanuel Oyekanlu

I&G-regular4: Big Data Applications (2)	
Regular	N242- *Reuters Tracer: Toward Automated News Production Using Large Scale Social Media Data* Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Sameena Shah, Robert Martin, and John Duprey
Regular	N218- *SMART: Sponsored Mobile App RecommendaTion by Balancing App Downloads and Appstore Profit* Zhiwei Zhang, Ning Chen, Jun Wang, and Luo Si
Regular	N236- *APP-SON: Application Characteristics-Driven SON to Optimize 4G/5G Network Performance and Quality of Experience* Ye Ouyang and Zhongyuan Li
Regular	*N239-A Configurable, Big Data System for On-Demand Healthcare Cost Prediction* Karthikeyan Natesan Ramamurthy, Dennis Wei, Emily Ray, Moninder Singh, Vijay Iyengar, Dmitriy Katz-Rogozhnikov, Jingwei Yang, Kevin Tran, and Gigi Yuen-Reed

Regular	N240- *Empirical Evaluations of Active Learning Strategies in Legal Document Review* Nathaniel Huber-Fliflet, Jianping Zhang, Haozhen Zhao, Robert Keeling, and Rishi Chhatwal

I&G-regular5: Big Data Applications (3)	
Regular	*N250-Implementing Scalable Structured Machine Learning for Big Data in the SAKE Project* Simon Bin, Patrick Westphal, Jens Lehmann, and Axel-Cyrille Ngonga Ngomo
Regular	*N260-A Study on Intelligent Personalized Push Notification with User History* Hyunjong Lee, Youngin Jo, Sanghyuk Chun, and Gwangseop Gim
Regular	N251-*A Cognitive Assistant for Risk Identification and Modeling* Vijil Chenthamarakshan, Dharmashankar Subramanian, Debarun Bhattachrajya, Ruben Torrado, Jeff Kephart, and Jesus Rios
Regular	N252- *Ranking the Importance of Ontology Concepts Using Document Summarization Techniques* Youngho Kim, Petros Zerfos, Vadim Sheinin, and Nancy Greco
Regular	N263- *Help Me Find a Job: A Graph-based Approach for Job Recommendation at Scale* Walid Shalaby, BahaaEddin AlAila, Mohammed Korayem, Layla Pournajaf, Khalifeh Aljadda, Shannon Quinn, and Wlodek Zadrozny

Tutorials

TUTORIAL 1: Enterprise Knowledge Graphs for Large Scale Analytics

Presenters:

Nidhi Rajshree (Contact Author)
IBM Watson San Jose, USA
Email: nidhi.rajshree@us.ibm.com
Nitish Aggarwal
IBM Watson San Jose, USA
Email: nitish.aggarwal@ibm.com
Sumit Bhatia
IBM Research New Delhi, India
Email: sumitbhatia@in.ibm.com
Anshu Jain
IBM Research Almaden San Jose, USA
Email: anshu.n.jain@us.ibm.com

Abstract:

In recent years, there have been lot of efforts in facilitating an user-friendly access to vast amounts of heterogeneous text data, ranging from news articles, social media post, scientific publications, associated with various domains such as corporate reports, legal acts, patient history, advertisements and security). Transforming such massive variety of unstructured text into an actionable knowledge, is a grand challenge to the research community. Through the proposed tutorial, we aim to present a comprehensive catalog of the best practices of building such large scale enterprise knowledge graphs, and enabling them to provide a user-friendly access to large amount of unstructured text data through various analytic applications. We will share our experiences of various challenges in construction of Knowledge Graph in IBM Watson Discovery Services and its applications in life sciences and intelligence domains.

TUTORIAL 2: Popularity on the Web: From Estimation to Prediction

Presenters:

Charalampos Chelmis, Assistant Professor (Contact Author)
University at Albany - SUNY
Email: cchelmis@albany.edu
Daphney-Stavroula Zois, Assistant Professor
University at Albany - SUNY
Email: dzois@albany.edu

Abstract:

There sharing of content on the Web has become an important mechanism by which people promote themselves, as well as discover and consume information, services, and products online. In certain instances, a product, a photo, a news article, or other piece of information may get reshared multiple times (i.e., a user shares content with her set of friends, several of whom share it with their respective sets of friends, and so on, such that the content potentially reaches a large number of people), liked or "pined" (e.g., on a content sharing service such as Pinterest), highly reviewed (e.g., on Amazon), or cited (e.g., academic publications in Google Scholar). A growing body of research has focused on characterizing such aspects of "popularity", identifying its characteristics, and estimating and predicting its dynamics in these domains. Popularity estimation and prediction are problems of particular interest with multiple applications, including facilitating beŠtter provision of resources, marketing and monetization, and blocking of illegal content. ŒThe goal of this tutorial is to (1) perform an in-depth study of the fundamental properties and similarities of popularity estimation and prediction with an emphasis on the algorithmic techniques and key ideas developed to derive e• fficient solutions; (2) identify the universal challenges associated with approaching the estimation and prediction tasks regardless of domain; and (3) summarize the most promising paths for future research.

TUTORIAL 3: Security and Automated Platform Development for Big Data Analytics

Presenters:

- **Jun (Luke) Huan**, Professor (Contact Author)
  University of Kansas
  Email: jhuan@ittc.ku.edu
- **Sohaib Kiani**, Ph.D. Candidate
  University of Kansas
  Email: kiani@ittc.ku.edu
- **Xiaoli Li**, Ph.D. Candidate
  University of Kansas

Abstract:

Data science is penetrating virtually every aspect of our society. However, data science systems—including data acquisition and processing pipelines and analytical techniques, such as deep learning—are becoming increasingly complex. Many data analytics and predictive analytics algorithms and systems are not transparent to the end-user. For example, how the underlying models work and when such models may fail, are not clear. Many approaches, especially those that apply to human subjects, may learn and reinforce pre-existing biases leading, for example, to unfair treatment of minority sections of a population. To enable widespread adoption of data science approaches requires assurances that the system will operate safely and securely, in a controlled and transparent manner. However, current research in this area is very limited. In this tutorial, we plan to cover a set of theories behind secure data analytics. We review recent efforts in developing algorithms to achieve data science safety using different techniques based on various evaluation metrics. We use several real-world applications of safe data science to further illustrate the importance of the topic. We also review efforts to provide open analytics platform. We conclude the tutorial by pointing out challenges, issues in current research of safe data science and future research directions.

TUTORIAL 4: Time Series Data Mining using the Matrix Profile: A Unifying View of Motif Discovery, Anomaly Detection, Segmentation, Classification, Clustering and Similarity Joins

Presenters:

Abdullah Mueen, Assistant Professor (Contact Author)
University of New Mexico
Email: mueen@unm.edu
Eamonn Keogh, Professor
University of California Riverside
Email: eamonn@cs.ucr.edu

Abstract:

Time series data mining is a perennially popular research topic in ACM SIGKDD, due to the ubiquity of time series in medical, financial, industrial, and scientific domains. There are about a dozen major time series data mining tasks, including:

- Time Series Motif Discovery
- Time Series Joins
- Time Series Classification (shapelet discovery)
- Time Series Density Estimation
- Time Series Semantic Segmentation
- Time Series Visualization
- Time Series Clustering
- Time Series Similarity Search (indexing)
- Time Series Monitoring (complex event processing)

In 2016, an international group of researchers introduced the Matrix Profile, with the following two surprising claims. Firstly, if you have the Matrix Profile computed, then all time series data mining tasks are easy or trivial, and secondly, computing the Matrix Profile is unexpectedly scalable, and is completely free of the curse of dimensionality. Given these

two facts, the Matrix Profile is poised to become an incredibly useful and ubiquitous primitive for time series data mining. It is difficult to overstate the scalability of the Matrix Profile computation, it has been used to perform ten exact quadrillion pairwise comparisons of a single time series during a self-join, surely the largest exact self-join ever attempted. In this tutorial, two of the inventers of the Matrix Profile will explain how to use it efficiently to solve problems in time series analytics. The tutorial will be illustrated with case studies from domains as diverse as entomology, oil-and-gas production, music, bioinformatics. medicine, seismology and human behavior understanding. All attendees will be given free access to a Matlab toolbox that will allow them to immediately leverage the power of the Matrix Profile, and start building their own novel applications and extensions.

TUTORIAL 5: Mathematics of Big Data

Presenters:

Kepner (Contact Author)
MIT Lincoln Laboratory Supercomputing Center
Email: kepner@ll.mit.edu

Abstract:

Big Data describes a new era in the digital age in which the volume, velocity, and variety of data created across a wide range of fields (e.g., internet search, healthcare, finance, social media, defense, ...) are increasing at a rate well beyond our ability to analyze the data. Tools such as spreadsheets, databases, matrices, and graphs have been developed to address these challenges. The common theme amongst these tools is the need to store and operate on data as whole sets instead of as individual data elements. This tutorial provides hands-on programming examples that illustrate the common mathematical foundations of these data sets (associative arrays) that apply across many applications and technologies. Associative arrays unify and simplify data, leading to rapid solutions to volume, velocity, and variety problems. Understanding the mathematical underpinnings of big data allows the student to see past the differences that lie on the surface of these tools and to leverage their mathematical similarities to solve the hardest data big challenges. Specifically, understanding associative arrays (1) reduces the effort required to pass data between steps in a data processing system, (2) allows steps to be interchanged with full confidence that the results will be unchanged, and (3) makes it possible to recognize when steps can be simplified or eliminated.

TUTORIAL 6: Industrial Big Data for Industrial Applications – Systematic Methodology

Presenters:

David Siege (Contact Author)
Predictronics Corp
Email: siegel@predictronics.com
Jay Lee, Professor
University of Cincinnati
Email: Jay.lee@uc.edu
Hossein Davari, Post-doctoral Fellow
University of Cincinnati
Email: davarihn@ucmail.uc.edu
Brian Weiss
National Institute of Standards and Technology
Email: brian.weiss@nist.gov

Abstract:

Industrial big data presents significant opportunities for organizations to improve their operation, reduce maintenance cost, and have higher productivity. These potential benefits can only be properly harnessed if one can extract actionable information and value from these large industrial data sets. This tutorial will first highlight the differences between industrial big data and other big data applications, including the structure of the data, the data quality, the volume of data, and the balance of the data classes. The tutorial will then focus on the data analysis methodology for predictive monitoring, including pre-processing and data quality checks, feature engineering, health index and anomaly detection algorithms, diagnosis and prognostic methods. In addition to the data analysis methodology, test methods, verification and validation approaches will also be included and discussed. Industrial case studies in manufacturing, transportation, and energy domains will be shown

to highlight the methodology and the successful use cases in industry. Lastly, some concluding remarks on the future direction for industrial big data and the unmet challenges will be discussed.

TUTORIAL 7: Game Theory for Data Science: Eliciting truthful information

Presenters:

Boi Faltings, Professor (Contact Author)
Swiss Federal Institute of Technology (EPFL)
Email: boi.faltings@epfl.ch
Goran Radanovic, Post-doctoral Researcher
Harvard University
Email: gradanovic@g.harvard.edu

Abstract:

As Big Data is increasingly used as a basis for decision making, it becomes important to ensure its quality. Often, data is provided by other agents, for example in sensor networks, user-contributed content, or crowdsourcing. Providing accurate and relevant data requires costly effort that agents may not always be willing to provide. Thus, it becomes important both to verify the correctness of data, but also to provide incentives so that agents that provide high-quality data are rewarded while those that do not are discouraged by low rewards. We will show how game theory makes such rewards possible. We will cover different settings and the assumptions they admit, including sensing, human computation, peer grading, reviews and predictions. We will survey different incentive mechanisms, including proper scoring rules, prediction markets and peer prediction, Bayesian Truth Serum, Peer Truth Serum, and the settings where each of them would be suitable. As an alternative, we also consider reputation mechanisms. We complement the game-theoretic analysis with practical examples of applications in prediction platforms, community sensing and peer grading.

TUTORIAL 8: Anti-discrimination Learning: From Association to Causation

Presenters:

Lu Zhang, Post-doctoral Fellow(Contact Author)
University of Arkansas
Email: lz006@uark.edu
Yongkai Wu, Ph.D. student
University of Arkansas
Email: yw009@uark.edu
Xintao Wu, Professor
University of Arkansas
Email: xintaowu@uark.edu

Abstract:

Anti-discrimination learning is an increasingly important task in data mining and machine learning fields. Discrimination discovery is the problem of unveiling discriminatory practices by analyzing a dataset of historical decision records, and discrimination prevention aims to remove discrimination by modifying the biased data and/or the predictive algorithms. Discrimination is causal, which means that to prove discrimination one needs to derive a causal relationship rather than an association relationship. Although it is well-known that association does not mean causation, the gap between association and causation is not paid enough attention by many researchers. The aim of this tutorial is to survey existing association-based approaches and point out their limitations, introduce a causal modeling-based framework and cover the very latest research on causal modeling-based fairness aware learning, and finally suggest potential future research directions.

Panel

Panel 1: Big Data Bias and Transparency

In January of this year, the ACM US policy committee published a statement of the seven characteristics that algorithms should have to be transparent and accountable:

1. Awareness
2. Access and redress
3. Accountability
4. Explanation
5. Data Provenance
6. Auditability
7. Validation and Testing

These characteristics change completely how we design algorithms as they cannot be black boxes any longer. One of the main motivations behind this initiative is certainly big data bias in all its forms, from Internet usage to financial transactions. How the algorithmic design process must change to address these concerns? How we can incorporate ethics and social values, yet legality, in our software systems? How we can make sure that at least bias is not being amplified by machine-learning based solutions? Will our privacy be affected even more?

These questions and others will be answered by:

- Cynthia Dwork, Harvard University
- John Langford, Microsoft Research
- Jure Leskovec, Stanford University & Pinterest
- Jeanna Matthews, Clarkson University
- Ricardo Baeza-Yates, NTENT (Moderator)

Panel 2: Big Data Software and Analytic Methods- What is Next?

Big data software and analytic methods are among the hottest IT themes in both academics and industry worldwide. It is a stark reality that computing will not get faster as it did in the past. We are witnessing today that smart phones/tablets are not visibly faster than they were two years back. Given that HPC and the hardware community have started thinking about what's beyond Moore's era, what is its impact for the big data community? Many a times, it is not only the data complexities that drive the big data analytic methods, but also the computing platform supporting the computations that helps us in overcoming formidable challenges.

Even though there are some recent big data software advances, such as SparkSQL

BigQuery, NoSQL databases and Hadoop, these are designed to address enterprise data that consist mostly of unstructured text and structured databases. Also, these restrict the programming environment provided and not all machine learning algorithms can perform well in these environments. Furthermore, big data over rich media such as video, audio, photos are reality today. The consumption of these big data on commercial hardware is becoming very difficult. For eg., 2.5TB of videos/images for competition are reality today. Just loading these data on to machine's memory will be 4-10 orders of magnitude slower than if equivalent content is consumed as text or as structured data. But the data pipelines to ingest multimedia big data to analytic methods have not been evolved adequately and remain a forte of the top internet companies.

In this panel, the panelists will present their point of view on pressing next challenges for Big Data Software and Analytic Methods. The discussion will leverage a diverse set of experiences and viewpoints, since the panel includes participants from both the leadership of R & D labs in industrial settings and from research groups conducting high-profile, Big Data research projects in academia.

Panelists may share their controversial points of view and provocative positions on issues/ questions, listed here, during panel presentation and discussion.

The following structure will be followed:

1. Welcome, Panel mechanics for discussion and Q & A, Introduction of panel members,
2. Presentations from Panelists (10- 15 min. each, including any quick questions/ comments)
3. Moderator-directed Panel Q & A
4. Questions from the Audience and open discussion

Issues/ Questions:

- What are the computational expectations of future big data platforms – distributed, micro parallelism, custom architecture such as FPGA's/neuromorphic computing? Which direction will drive the growth - standardized open platforms vs proprietary
- Which of the Big Data analytics/ technology areas (e.g. Descriptive, Inquisitive, Predictive, Prescriptive and Pre-emptive Analytics) are most important to your organization, and can you perhaps give some example research projects that address such challenges?
- What kind of BI tools do you use today, and what are the limitations of these tools- such as in dealing with unstructured data sets, streaming data, interactive visualization, etc.?
- What type of unstructured datasets (text, video, audio, photos) do you (or your organization) deal with today and which of those pose greater challenges?
- What are the biggest trends you see that will emerge, in the next several years, for data-driven decision making from your perspective – in terms of how your stake holders will make decisions by exploiting knowledge extracted from data

Panelists:
Peter Bauman, Jacobs University, Germany
Alan Edelman, MIT, USA
Ronald D. Hagan, BAE Systems, USA
Satoshi Matsuoka, Tokyo Institute of Technology, Japan
Dominik Slezak, University of Warsaw, Poland

Moderator:
Vijay Raghavan, University of Louisiana at Lafayette, USA

Workshops

Computational Archival Science *Workshop Chairs: Mark Hedges, Victoria Lemieux, Richard Marciano*		
Time	**Title**	**Presenter/Author**
9:00am-9:15am	Welcome	Mark Hedges, Victoria Lemieux, Richard Marciano
9:15am-10:35am	**Paper Session I: Exploring Archival Data**	
	Building new knowledge from distributed scientific corpus; HERBADROP & EUROPEANA: two concrete case studies for exploring big archival data	Pascal Dugenie, Nuno Freire, Daan Broeder
	An Infrastructure and Application of Computational Archival Science to Enrich and Integrate Big Digital Archival Data: Using Taiwan Indigenous Peoples Open Research Data (TIPD) as Example	Ji-Ping Lin
	Computational Curation of a Digitized Record Series of WWII Japanese-American Internment	William Underwood, Richard Marciano, Sandra Laib, Carl Apgar, Luis Beteta, Waleed Falak, Marisa Gilman, Riss Hardcastle, Keona Holden, Yun Huang, David Baasch, Brittni Ballard, Tricia Glaser, Adam Gray, Leigh Plummer, Zeynep Diker, Mayanka Jha, Aakanksha Singh, Namrata Walanj
	The Cybernetics Thought Collective Project: Using Computational Methods to Reveal Intellectual Context in Archival Material	Bethany Anderson, Christopher Prom, Kevin Hamilton, James Hutchinson, Mark Sammons, Alex Dolski
10:35am-10:45am	**Questions and Discussion**	
10:45am-11:05am	**Coffee Break**	
11:05am-12:25pm	**Paper Session II: Curation and Appraisal**	
	Towards Automated Quality Curation of Video Collections from a Realistic Perspective	Todd Goodall, Maria Esteva, Sandra Sweat, Alan Bovik
	Line Detection in Binary Document Scans: A Case Study with the International Tracing Service Archives	Benjamin Lee
	Auto-Categorization & Future Access to Digital Archives	Nathaniel Payne, Jason R. Baron
	Heuristics for Assessing Computational Archival Science (CAS) Research: The Case of the Human Face of Big Data Project	Myeong Lee, Yuheng Zhang, Shiyun Chen, Edel Spencer, Jhon Dela Cruz, Hyeonggi Hong, Richard Marciano
12:25pm-12:45pm	**Paper Session III: CAS Methods**	
	What Can a Knowledge Complexity Approach Reveal About Big Data and Archival Practice?	Nicola Horsley
12:45pm-2:00pm	**Lunch Break**	
2:00pm-3:00pm	**Paper Session III: CAS Methods (Cont.)**	
	Protecting Privacy in the Archives: Preliminary Explorations of Topic Modeling for Born-Digital Collections	Tim Hutchinson
	Identifying Epochs in Text Archives	Tobias Blanke, Jon Wilson
	GraphQL for Archival Metadata: An Overview of the EHRI GraphQL API	Mike Bryant
3:00pm-3:40pm	**Paper Session IV: Creation and Management of Current Records**	
	The Blockchain Litmus Test	Tyler Smith
	A Typology of Blockchain Recordkeeping Solutions and Some Reflections on their Implications for the Future of Archival Preservation	Victoria Lemieux
3:40pm-4:05pm	**Questions and Discussion**	
4:05pm-4:25pm	**Coffee Break**	
4:25pm-4:55pm	**Demo Session**	

	ArchiveSpark: Efficient Web Archive Access, Extraction, and Derivation of smaller datasets	Helge Holzmann
	Digital Repository At Scale That Invites Computation (DRAS-TIC)	Greg Jansen
4:55pm-5:15pm	**Student Session**	
	Graduate students from U. Maryland discuss their responses to incorporating CAS into Master's programmes in Library and Information Science.	
5:15pm-5:45pm	**Closing Remarks**	

3rd International Workshop on Methodologies to Improve Big Data projects

WorkshopChair: Jeff Saltz

Time	Title	Presenter/Author
8:00am – 8:30am	Saving Costs with a Big Data Strategy	Alan Serrano
8:30am – 9:00am	Towards a Requirements Engineering Artefact Model	Darlan Arruda / Nazim Madhavji
9:00am – 9:30am	Does Pair Programming work in a Data Science Context	Ivan Shamshurin
9:30am – 10:00am	Predicting Outcomes for Big Data Projects: Big Data Project Dynamics	David Becker
10:00am – 10:20am	**Coffee Break**	
10:20am – 10:50am	The Ambiguity of Data Science Team Roles and the Need for a Data Science Workforce Framework	Jeff Saltz
10:50am – 11:20am	Make Accumulated Data in Companies Eloquent by SQL Statement Constructors	Toshiyuki Shimono
11:20am – 11:50am	Agile Big Data Analytics: AnalyticsOps for Data Science	Jason Payne
11:50 – 12:00pm	**Closing Remarks**	

Second Workshop on Real-time and Stream Analytics in Big Data

WorkshopChairs: Sabri SKHIRI, EURA NOVA
Albert Bifet, Télécom Paris Tech
Alessandro Margara, Politecnico di Milano, IT

Time	Title	Presenter/Author
8:30	Workshop Keynote 1: The rise of Stream Processing for data management & micro service Architecture	Sabri SKHIRI, EURA NOVA8:45
8:45	Workshop Keynote 2: Apache Kafka – New features	Viktor Gamov, CONFLUENT
9:15	**"ABC: a Practicable Sketch Framework for Non-uniform Multisets"**	Junzhi Gong, Tong Yang, Yang Zhou, Dongsheng Yang, Shigang Chen, Bin Cui, and Xiaoming Li
9:25	**"Online Mining for Association Rules and Collective Anomalies in Data Streams"**	Shaaban Abbady, Cheng-Yuan Ke, Jennifer Lavergne, Jian Chen, Vijay Raghavan, and Ryan Benton
9:50	**"A Study of a Video Analysis Framework Using Kafka and Spark Streaming"**	Ayae Ichinose, Atsuko Takefusa, Hidemoto NAKADA, and Masato Oguchi
10:10	**Coffee Break**	
10:30	**"RASP: Real-time Network Analytics with Distributed NoSQL Stream Processing"**	Georgios Touloupas, Ioannis Konstantinou, and Nectarios Koziris
10:50	**"Towards a Unified Storage and Ingestion Architecture for Stream Processing"**	Ovidiu-Cristian Marcu, Alexandru Costan, Gabriel Antoniu, Maria Perez, Radu Tudoran, Stefano Bortoli, and Bogdan Nicolae
11:10	**"Smart Distributed Query Execution over Data Streams"**	Salman Ahmed Shaikh and Hiroyuki Kitagawa
11:30	**"Harnessing the Power of Hashtags in Tweet Analytics"**	Vibhuti Gupta and Rattikorn Hewett
11:50	**"Predicting Concept Drift via Dynamic Naïve Bayes"**	Qian Zhao, Chris Klaue, and Chih Lai

12:10	**Closing Remarks**

6th Workshop on Scalable Cloud Data Management *Workshop Chairs: Norbert Ritter, Felix Gessert*		
Time	**Title**	**Presenter/Author**
8:25- 8:30pm	**Opening Remarks**	Norbert Ritter, Felix Gessert (University of Hamburg, Germany)
	Session I: Data Management	
8:30 - 10:00	Uncovering the Evolution History of Data Lakes	Meike Klettke (University of Rostock, Germany)
	Dynamic Data Transformation for Low Latency Querying in Big Data Systems	Leandro Ordonez-Ante (Ghent University, Belgium)
	Improving user interaction in mobile-cloud database query processing	Chenxiao Wang (University of Oklahoma, USA)
10:00 - 10:20	**Coffee Break**	
10:20 - 11:50	**Session II: Cloud Databases and Systems**	
	Icarus: Towards a Multistore Database System	Marco Vogt (University of Basel, Switzerland)
	Trilogy: Data Placement to Improve Performance and Robustness of Cloud Computing	Chin-Jung Hsu (North Carolina State University, USA)
	Leveraging Distributed Big Data Storage Support in CLAaaS for WINGS Workflow Management System	Hadeel Alghamdi (Queen's University, Canada)
11:50 - 1:50	**Coffee Break**	
1:50 - 3:20	**Session III: Big Data Infrastructure**	
	Highly Consolidated Servers with Container-based Virtualization	Joichiro Kon (Kogakuin University, Japan)
	Understanding and Improving Disk-based Intermediate Data Caching in Spark	Kaihui Zhang (University of Tsukuba, Japan)
	Online Machine Learning for Cloud Resource Provisioning of Microservice Backend Systems	Hanieh Alipour (Concordia University, Canada)
3:20 - 3:50	**Coffee Break**	
	Session IV: Big Data Applications	
3:50 - 4:50	Closing the Loop – Finding Lung Cancer Patients using NLP	Bipin Karunakaran (Geisinger, USA)
	Discovery of Action Rules at Lowest Cost in Spark	Angelina Tzacheva (University of North Carolina at Charlotte, USA)
4:50 - 4:55	**Closing Remarks**	

Workshop on Solar & Stellar Astronomy Big Data *WorkshopChairs: Rafal A. Angryk, Piet C. Martens, Russel J. White*		
Time	**Title**	**Presenter/Author**
14:00-14:25	Parallel Computation of Magnetic Field Parameters from HMI Active Region Patches	Sunitha Basodi et al.
14:25-14:50	Accelerating Scientific Algorithms in Array Databases with GPUs	Simon Marcin et al.
14:50-15:15	Improving Functionality of Tamura Directionality on SDO AIA Images	Azim Ahmadzadeh et al.
15:15-15:40	Multi-wavelength Solar Event Detection using Faster R-CNN	Ahmet Kucuk et al.

15:40-16:05	Improving Expectation Maximization Algorithm over Stellar Data	Hasan Kurban et al.
16:05-16:25	**Coffee Break**	
16:25-16:50	On the Prediction of >100 MeV Solar Energetic Particle Events Using GOES Satellites Data	Soukaina Fiali Boubrahimi et al.
16:50-17:15	A Time Series Classification-based Approach for Solar Flare Prediction	Shah Muhammad Hamdi et al.
17:15-17.40	Solar Flare Prediction using Multivariate Time Series Decision Trees	Ruizhe Ma et al.
17:40-18:05	**Closing Remarks**	

4th Workshop on Advances in Software and Hardware for Big Data to Knowledge Discovery (ASH)

WorkshopChairs: Weijia Xu, Hui Zhang, Hongfeng Yu

Time	Title	Presenter/Author
1:30 pm- 2:00 pm	Cloud Big Data Decision Support System for Machine Learning on AWS	Alex Kaplunovich
2:00 pm – 2:30 pm	Big Data Machine Learning Using Apache Spark MLlib	Mehdi Assefi
2:30pm – 3:00pm	Map-Scan Node Accelerator for Big Data	Mihaela Malita
3:00pm – 3:30pm	Divide-and-Conquer Strategies for Large-scale Simulations in R	Juan Lin
	Coffee Break	
3:50pm – 4:20pm	The Sampling Peak Criterion Method for Large Data	Sergiy Peredrity
4:20pm – 4:50pm	An Online Spatio-Temporal Model for Inference and Predictions of Taxi Demand	Jian Zou
4:50pm – 5: 20pm	Ranked Time Series Matching by Interleaving Similarity Distance	Charles Lovering
4:20pm – 5: 50pm	Return of Experience on the Mean-shift Clustering for Heterogeneous Architecture Use Case	Christophe Cerin
	Closing Remarks	

Data Quality Issues in Big Data and Machine Learning Applications: Going Beyond Data Cleaning and Transformations

WorkshopChairs: Venkat Gudivada, Junhua Ding, and Srividya Bansal

Time	Title	Presenter/Author
2:00 – 2:10 PM	*Session Overview*	Junhua Ding
2:10 – 2:40 PM	*Generative Adversarial Networks for Increasing the Veracity of Big Data*	Matthew Dering
2:40 – 3:10 PM	*Augmentation and Evaluation of Training Data for Deep Learning*	Junhua Ding
3:10 – 3:40 PM	*Towards high precision gender categorization*	Daniel Müller
3:40 – 4:10 PM	*Is Data Quality Enough for a Clinical Decision? Apply Machine Learning and Avoid Bias*	Hee Kim
4:10 – 4:30 PM	**Coffee Break**	
4:30 – 5:00 PM	*Data Quality Challenges with Missing Values and Mixed Types in Joint Sequence Analysis*	Alina Lazar
5:00 – 5:30 PM	*Collapsing Corporate Confusion*	Tim Marple
5:30 – 6:00 PM	*Toward Data Quality Analytics in Signature Verification Using a Convolutional Neural Network*	Shahab Tayeb
6:00 – 6:25 PM	*Identifying and Mitigating Risks to the Quality of Open Data in the Post-truth Era*	Adrienne Colborne
6:25 – 6:30 PM	**Closing Remarks**	

5th International Workshop on Distributed Storage Systems and Coding for Big Data

Time	Title	Presenter/Author

8:00-8:25 am	The Architecture of Distributed Storage System Under Mimic Defense Theory	Hui Li, Jiawei Hu, Huajun Ma, and Ting Huang
8:25-8:50 am	MDFS: A Mimic Defense Theory based Architecture for Distributed File System	Zhili Lin, Hanxu Hou, Xin Yang, Kedan Li, and Hui Li
8:50-9:15 am	On the Implementation of Efficient BRS Codes in Ceph	Jiyang Zhang, Hanxu Hou, Kedan Li, and Hui Li
9:15-9:40 am	A Scheduling Strategy Based on Multi-Queues of Cassandra	HaoPeng Li and Hui Li
9:40-10:05 am	An improved P2P File System Scheme based on IPFS and Blockchain	Yongle Chen, Hui Li, Kejiao Li, and Jiyang Zhang
10:05-10:20 am	**Coffee Break**	
	Closing Remarks	

BSMDMA-SocialNLP Workshop

BSMDMA Workshop Chairs: Xin Huang, Rui Chen, Xuan Song, and Bolei Zhou
SocialNLP Workshop Chairs: Cheng-Te Li, and Lun-Wei Ku

Time	Title	Presenter/Author
8:30-9:30	Keynote Talk	Evimaria Terzi
9:30-9:45	*Identifying emergency stages in Facebook posts of police departments with convolutional and recurrent neural networks and support vector machines*	Nicolai Pogrebnyakov and Edgar Maldonado
9:45-10:00	*Characterization of daily tourism behaviors based on place sequence analysis from photo sharing websites*	Thomas-Joseph Loiseau, Sonia Djebali, Thomas Raimbault, Bérengère Branchet, and Gaël Chareyron
10:00-10:15	*Ticket-Purchase behavior under the Effects of Marketing Campaigns on Facebook Fan Pages*	Hsiao-Wei Hu, Ching-Han Cheng, Yun-Chu Chung, and Chia-Yu Lee
10:15-10:30	*Outbound Behavior Analysis Through Social Network Data: a case study of Chinese people in Japan*	TIANQI XIA, Xuan Song, Dou Huang, Satoshi Miyazawa, Zipei Fan, Renhe Jiang, and Ryosuke Shibasaki
10:30-10:45	*PSEISMIC: A Personalized Self-Exciting Point Process Model for Predicting Tweet Popularity*	Hsin-Yu Chen and Cheng-Te Li
10:45-11:05	**Coffee Break**	
11:05-11:20	*Detection of Profile Injection Attacks in Social Recommender Systems Using Outlier Analysis*	Anahita Davoudi and Mainak Chatterjee
11:20-11:35	*Evaluating the Quality of Graph Embeddings via Topological Feature Reconstruction*	Stephen Bonner, John Brennan, Ibad Kureshi, Georgios Theodoropoulos, Stephen McGough, and Boguslaw Obara
11:35-11:50	*Topic Life Cycle Extraction from Big Twitter Data based on Community Detection in Bipartite Networks*	Takako Hashimoto, Hiroshi Okamoto, Tetsuji Kuboyama, and Kilho Shin
11:50-12:05	*Using Sentiment Analysis to Explore the Degree of Risk in Sharing Economy*	Wei-Lun Chang
12:05-12:20	*Big Social Data Analytics for Public Health: Comparative Methods Study and Performance Indicators of Health Care Content on Facebook*	Nadiya Straton, Raghava Rao Mukkamala, and Ravi Vatrapu
12:20-12:35	*A Big Social Media Data Study of the 2017 German Federal Election based on Social Set Analysis of Political Party Facebook Pages with SoSeVi*	Benjamin Flesch, Ravi Vatrapu, and Raghava Rao Mukkamala
12:35-12:50	*Digital Content Recommendation System Using Implicit Feedback Data*	Saayan Mitra, Viswanathan Swaminathan, Ratnesh Kumar, and Gang Wu
12:50-14:00	**Lunch Break**	
16:00-16:20	**Coffee Break**	
16:20-16:35	*Detecting Polarization in Ratings: An Automated Pipeline and a Preliminary Quantification on Several Benchmark Data Sets*	Mahsa Badami, olfa Nasraoui, Wenlong Sun, and Patrick Shafto
16:35-16:50	*Language Identification in Multilingual, Short and Noisy Texts using Common N-Grams*	Dijana Kosmajac and Vlado Keselj
16:50-17:05	*Characterizing Online Community Practices with Orthographic Variation*	Ian Stewart, Stevie Chancellor, Munmun De Choudhury, and Jacob Eisenstein
17:05-17:20	*Using an Asset Price Bubble Model in Tweet Analytics*	K.M. George
17:20-17:35	*An Entity Disambiguation Method Based on LeaderRank*	Bingjing Jia, Bin Wu, Jinna Lv, Pengpeng Zhou, Yao Bu, and Ying Xing

17:35-17:50	*Improving Arabic Sentiment Analysis with Sentiment-Specific Embeddings*	A. Aziz Altowayan and Ashraf Elnagar
17:50-18:05	*Topic Modelling enriched LSTM Models for the Detection of Novel and Emerging Named Entities from Social Media*	Patrick Jansson and Shuhua Liu
18:05-18:20	*Differences in Emoji Sentiment Perception between Readers and Writers*	Jose Berengueres and Dani Castro
	Closing Remarks	

Workshop on Benchmarking, Performance Tuning and Optimization for Big Data Applications (BPOD)

WorkshopChairs: Zhiyuan Chen & Jianwu Wang

Time	Title	Presenter/Author
11:05am-11:15am	Welcome, opening remark	Zhiyuan Chen & Jianwu Wang
11:15am-11:30am	S10214: Imbalance in the Cloud: an Analysis on Alibaba Cluster Trace	Chengzhi Lu, Kejiang Ye, Guoyao Xu, Cheng-Zhong Xu, and Tongxin Bai
11:30am-11:45am	S10208: A Performance Study of Big Data Analytics Platforms	Pouria Pirzadeh, Michael Carey, and Till Westmann
11:45am-12pm	S10205: Plug and Play Bench : Simplifying Big Data Benchmarking Using Containers	Sheriffo Ceesay, Dr. Adam Barker, and Dr. Blesson Varghese
12pm-12:15pm	S10219: Quantifying Volume, Velocity, and Variety to support (Big) Data-Intensive Application Development	Rustem Dautov and Salvatore Distefano
12:15pm-12:30pm	S10215: A Comparison of Big Data Application Programming Approaches: A Travel Companion Case Study	Pei Guo, Jianwu Wang, and Zhiyuan Chen
12:30pm-12:45pm	BigD566: A Performance Study of AsterixDB	Keren Ouaknine and Michael Carey
12:45pm-2pm	**Lunch**	
2pm-2:15pm	BigD522: A Novel Compression Algorithm Decision Method for Spark Shuffle Process	Huang Shanshan, Liao Husheng, Xu Jungang, and Liu Renfeng
2:15-2:30pm	S10211: Tula: A Disk Latency Aware Balancing and Block Placement Strategy for Hadoop	Janakiram Dharanipragada, Srikant Padala, Balaji Kammili, and Vikram Kumar
2:30pm-2:45pm	S10212: ECL-Watch: A Big Data Application Performance Tuning Tool in the HPCC Systems Platform	Lili Xu, Edin Muharemagic, Flavio Villanustre, and Amy Apon
2:45pm-3pm	S10206: Towards Online Graph Processing with Spark Streaming	Tariq Abughofa and Farhana Zulkernine
3pm-3:15pm	S10213: Schema Design Support for Semi-Structured Data: Finding the Sweet Spot between NF and De-NF	Vincent Reniers, Dimitri Van Landuyt, Ansar Rafique, and Wouter Joosen
3:15pm-3:30pm	S10210: Enhancing the MapReduce Training of BP Neural Networks Based on Local Weight Matrix Evolution	Wanghu Chen, Xintian Li, Jing Li, and Jianwu Wang
3:30pm-3:45pm	BigD451: Finding the Best Box-Cox Transformation from Massive Datasets on Spark	Baijian Yang, Tonglin Zhang, and Huayi Fang
3:45pm-4:05pm	S10204: SUDS: System for Uncertainty Decision Support	Maaike de Boer, Barry Nouwt, and Michael van Bekkum
4:05pm-4:25pm	**Coffee Break**	
4:25pm-4:40pm	S10216: Adaptive Scalable Pipelines for Political Event Data Generation	Yan Liang, Andy Halterman, Phanindra Jalla, Solaimani Mohiuddin, Manar Landis, Jill Irvine, and Christan Grant
4:40pm-4:55pm	S10217: Scaling Point Set Registration in 3D across Thread Counts on Multicore and Hardware Accelerator Platforms through Autotuning for Large Scale Analysis of Scientific Point Clouds	Piotr Luszczek, Jakub Kurzak, Ichitaro Yamazaki, David Keffer, and Jack Dongarra
4:55pm-5:10pm	BigD281: Big Data processing: Is there a framework suitable for Economist and Statisticians	Giuseppe Bruno, Demetrio Condello, Alberto Falzone, and Andrea Luciani
5:10pm-5:25pm	BigD381: CloudEC: A MapReduce-based Algorithm for Correcting Errors in Next-generation Sequencing Big Data	Wei-Chun Chung, Jan-Ming Ho, Chung-Yen Lin, and D. T. Lee
5:25pm-5:40pm	BigD540: Efficient Incremental Data Analytics with Apache Spark	Sina Gholamian, Wojciech Golab, and Paul A. S. Ward
5:40pm-5:55pm	S10207: Performance Evaluation of Multiple Sports Player Tracking System Based on Graph Optimization	Yuri Nishikawa, Hitoshi Sato, and Jun Ozawa
5:55pm-6:25pm	Discussion of future plan, closing remarks	

1st IEEE Big Data International Workshop on Policy-based Autonomic Data Governance (PADG) **Workshop Chairs: Seraphin Calo, Elisa Bertino, Dinesh Verma**		
Time	Title	Presenter/Author
8:15am –8:30am	**Opening Remarks**	**Elisa Bertino**
8:30am –9:10am	**Keynote Address**	
	My Fair (Big) Data	**Tiziana Catarci**
9:10am –10:00am	**Session 1**	Chair: **Dinesh Verma**
	Improving Data Sharing in Data Rich Environments	**Erisa Karafili**, Emil Lupu, Alan Cullen, Bill Williams, Saritha Arunkumar, Seraphin Calo
	Community-based Self Generation of Policies and Processes for Assets: Concepts and Research Directions	**Geeth de Mel**, Elisa Bertino, Alessandra Russo, Seraphin Calo, Dinesh Verma
10:00am –10:15am	**Coffee Break**	
10:15am –11:05am	**Session 1 (continued)**	
	LightSpy: Optical Eavesdropping on Displays Using Light Sensors on Mobile Devices	**Supriyo Chakraborty**, Wentao Ouyang, Mani Srivastava
	Identifying Sensor Accesses from Service Descriptions	**Raghu Ganti**, Antara Palit, Mudhakar Srivatsa, Christopher Simpkin
11:05am –11:45am	**Session 2**	Chair: **Seraphin Calo**
	Combining Semantic Web and IoT to Reason with Health and Safety Policies	**Emre Göynügür**, Murat Şensoy, Geeth de Mel
	Edge Computing Architecture for applying AI to IoT	**Maroun Touma**, Dinesh Verma, Alan Cullen, Seraphin Calo
11:45am –12:15am	**Session 3**	Chair: **Elisa Bertino**
	Policy Enabled Caching for Distributed AI	**Dinesh Verma**, Graham Bent
	Research Challenges in Dynamic Policy-Based Autonomous Security	**Seraphin Calo**, Elisa Bertino, Emil Lupu, Saritha Arunkumar, Alan Cullen, Gregory Cirincione, Brian Rivera
12:15am –12:30am	**Closing Remarks**	

Big Data Metadata Management 2017 *WorkshopChairs: Wo Chang* December 11, 2017		
Time	**Title**	**Presenter/Author**
08:00 – 08:10	Welcome	Wo Chang
08:10 – 08:20	Opening Remark	David Belanger
08:20 – 10:00	Briefing about the use case, datasets, challenges, Q/As	Wo Chang
10:00 – 08:00 (next day)	Solving hackathon challenges	Hackathon Participants
10:45 – 11:05	**Coffee Break**	
16:05 – 16:25	**Coffee Break**	

Big Data Metadata Management 2017 *WorkshopChairs: Wo Chang* December 12, 2017		
Time	**Title**	**Presenter/Author**
08:00 – 09:00	Hackathon Evaluation	Hackathon Participants Evaluation Team: Wo Chang, David Belanger, Mahmoud Daneshmand, Kathy Grise, Cherry Tom, Robby Robson
14:00 – 14:10	Welcome	Wo Chang
14:10 – 14:30	Opening Remark	David Belanger
14:30 – 15:00	Keynote Speech: ***Digital Object Architecture***	Larry Lannom
15:00 – 15:20	Invited Talk: ***Managing Big Time Series & Text Data for Unsupervised Feature Representation Learning***	Linqfei Wu
15:20 – 15:50	Invited Talk: TBD	Yu Luo
15:50 – 16:05	***Why-Diff: Explaining Differences amongst Similar Workflow Runs by exploiting Scientific Metadata***	Priyaa Thavasimani, Jacek Cala, and Paolo Missie
16:05 – 16:25	**Coffee Break**	
16:25 – 16:40	***Case: Big Geosciences Data Validation Challenges and Achievements***	Hussain Alajmi,
16:40 – 16:55	***Deep Learning for Big Data Analytics: A Review from Fog and Edge Computing Perspective***	Swarnava Dey and Arijit Mukherjee
16:55 – 17:20	Hackathon Ceremony	David Belanger and Kathy Grise
17:20 – 17:30	Closing Remarks	Wo Chang

2nd International Workshop on Application of Big Data for Computational Social Science *Workshop Chairs: Akira Ishii, Fujio Toriumi, Hiroki Takikawa, Kazutoshi Sasahara*		
Time	**Title**	**Presenter/Author**
8:30-8:40	Opening	
8:40-9:00	Bias reduction of peer influence effects with latent coordinates and community membership	Daniel Rajchwald, Natasha Markuzon, and Edoardo Airoldi
9:00-9:20	Evaluating Funding Programs through Network Centrality Measures of Co-Author Networks of Technical Papers	Masanori Fujita, Hiroto Inoue, and Takao Terano
9:20-9:40	Using Machine Learning Methods to Identify Atrocity Perpetrators	Benjamin E. Bagozzi and Ore Koren
9:40-10:00	Inference of Personal Attributes from Tweets Using Machine Learning	Take Yo and Kazutoshi Sasahara
10:00-10:20	**Coffee Break**	
10:20-10:40	Detecting two types of seasonal words using simple autocorrelation analysis	Kenta Yamada
10:40-11:00	Analyzing Regional Characteristics of Living Activities of Elderly People from Large Survey Data with Probabilistic Latent Spatial Semantic Structure Modeling	Ayae Ide, Kazuya Yamashita, Yoichi Motomura, and Takao Terano
11:00-11:20	A statistical analysis of behavioral bursts occurring in a social networking game	Mitsuki Murase, Masanori Takano, Reiji Suzuki, and Takaya Arita

11:20-11:40	Analysis of the Changes in Listening Trends of a Music Streaming Service	Masanori Takano, Hiroki Mizukami, Fujio Toriumi, Makoto Takeuchi, Kazuya Wada, Masahiro Yasuda, and Ichiro Fukuda
11:40-12:00	Analysis of EXILE TRIBE in the Music Scene Using Mathematical Model of Hit Phenomenon	Toshimichi Wakabayashi, Yasuko Kawahata, and Akira Ishii
12:00-13:30	**Lunch Time**	
13:30-13:50	Political Polarization in Social Media: Analysis of the "Twitter Political Field" in Japan	Hiroki Takikawa and Kikuko Nagayoshi
13:50-14:10	Analysis of Twitter Messages about the Osaka Metropolis Plan in Japan	Kouki Hayashi, Eiichi Umehara, and Yuuki Ogawa
14:10-14:30	Cross-National Measurement of Polarization in Political Discourse: Analyzing floor debate in the U.S. and the Japanese legislatures	Takuto Sakamoto and Hiroki Takikawa
14:30-14:50	"Fake News" Drives out Real: Analyzing Posts and Links shared on Public Facebook Pages During 2016 US Presidential Election	King-wa Fu
14:50-15:10	Facebook and Public Health: A Survey Study to Better Understand Clustering and Supervised Learning Findings about Facebook Post Performance of 153 Health and Care Organisations	Nadiya Straton, Ravi Vatrapu, and Raghava Rao Mukkamala
15:10-15:30	An Overview of Social Media Analysis for Disasters Management -- Toward Leveraging Social Media Data for Community Recovery	Yuya Shibuya
15:30 -15:50	**Coffee Break**	
15:50-16:10	Position-sensitive propagation of information on social media using social physics approach	Akira Ishii, Takayuki Mizuno, and Yasuko Kawahata
16:10-16:30	Comparison between Spatial Distributions of Tweet Base and Population in Japan	Shouji Fujimoto, Atushi Ishikawa, and Takayuki Mizuno
16:30-16:50	When Do Users Change Their Profile Information on Twitter?	Jinsei Shima, Mitsuo Yoshida, and Kyoji Umemura
16:50-17:10	Relationships between market impact characteristics and order book properties	Kenta Yamada and Takayuki Mizuno
17:10-17:30	Time Dependent Analysis of Financial Networks using Supervised Latent Feature Relational Models	Shotaro Ito and Koji Eguchi
17:30-17:50	Develop Method to Predict the Increase in the Nikkei VI index	Hirohiko Suwa, Yuuki Ogawa, Eiichi Umehara, Kento Kakigi, Tatsuo Yamashita, and Kota Tsubouchi
	Closing Remarks	

METHODS TO MANAGE HETEROGENEOUS BIG DATA AND POLYSTORE DATABASES *WorkshopChairs: Vijay Gadepally, Timothy Mattson, Michael Stonebraker*		
Time	**Title**	**Presenter/Author**

0800-0810	Welcome/Introduction	
0810-0830	Enabling Query Processing across Heterogeneous Data Models: A Survey	Ran Tan, Rada Chirkova, Vijay Gadepally, Tim Mattson
0830-0850	Polystore Mathematics of Relational Algebra	Hayden Jananthan, Ziqi Zhou, Vijay Gadepally, Dylan Hutchison, Suna Kim, Jeremy Kepner
0850-0910	Querying Web Polystores	Yasar Khan, Antoine Zimmerman, Alokkumar Jha, Dietrich Rebholz-Schuhmann, Ratnesh Sahay
0910-0930	Demo: AWESOME Polystore	Amarnath Gupta, Subashish Das
0930-0950	A novel object placement protocol for minimizing the average response time of get operations in distributed key-value stores	Antonios Makris, Konstantinos Tserpes, and Dimosthenis Anagnostopoulos
1000-1020	**Coffee Break**	
1020-1050	Keynote	Andy Palmer
1050-1110	A novel object placement protocol for minimizing the average response time of get operations in distributed key-value stores	Antonios Makris, Konstantinos Tserpes, and Dimosthenis Anagnostopoulos
1110-1130	Demo: SciDB	Marilyn Matz
1130-1150	Demo: TileDB	Stavros Papadopoulos, Jake Bolewski
1150-1210	An Apache Calcite-based Polystore Variation for Federated Querying of Heterogeneous Healthcare Sources	Ashwin Kumar Vajantri, Kunwar Deep Singh Toor, Edmon Begoli, Jack Bates
	Closing Remarks	

IEEE WORKSHOP ON BIG DATA ANALYTICS IN MANUFACTURING AND SUPPLY CHAINS

WorkshopChairs: Dr Allan Zhang and Dr Gurdal Ertek

Time	Title	Presenter/Author
2:00 pm - 2:05 pm	**Opening remarks**	
2:05 pm - 2:25 pm	Text Mining Analysis of Wind Turbine Accidents: An Ontology-Based Framework	Gurdal Ertek, et al.
2:25 pm - 2:45 pm	A Detection Mechanism with Text Mining Cross Correlation Approach (CCA)	JOSE LUIS GUERRERO CUSUMANO
2:45 pm – 3:05 pm	Performing literature review using text mining, Part I: Retrieving technology infrastructure using Google Scholar and APIs	Dazhi Yang, et al.
3:05 pm – 3:25 pm	Performing literature review using text mining, Part II: Expanding domain knowledge with abbreviation identification	Dazhi Yang and Jihoog Hong
3:25pm - 3:45pm	Forecast and analysis of food donations using support vector regression	Nigel Pugh and Lauren Davis
3:45pm- 4.05pm	Adaptive Spatio-temporal Mining for Route Planning and Travel Time Estimation	Rong Wen and Wenjing Yan
4:05 pm - 4:25 pm	**Coffee Break**	
4:25 pm - 4:45 pm	Streaming Analytics Processing in Manufacturing Performance Monitoring and Prediction	Yi-Hsin Wu, et al.

4:45 pm – 5:05 pm	Application of Deep Neural Network and Generative Adversarial Network to Industrial Maintenance: A Case Study of Induction Motor Fault Detection	Yong Oh Lee, et al.
5:05 pm - 5:25 pm	Learning Automata Based Method for Solving Demand and Supply Problem with Periodic Behaviours	Haoye Lu, et al.
5:25 pm - 6:25 pm	**Poster and Network**	
	Association Analysis of Supply Chain Risk and Company Sales	Murat Tunc, et al.
	A model for analysing a disrupted supply chain's time-to-recovery under uncertainty	Jie Liang Aloysious Lee, et al.
	Closing remarks	

The 2nd International Workshop on Big Spatial Data (BSD 2017)

Workshop Chairs: Farnoush Banaei Kashani, Siyuan Lu, Chengyang Zhang, Abdeltawab Hendawi

Time	Title	Presenter/Author
8:00 – 8:30	All in One: Encoding Spatio-Temporal Big Data in XML, JSON, and RDF without Information Loss	Peter Baumann, Eric Hirschorn, Joan Maso, Vlad Merticariu, and Dimitar Misev
8:30 – 9:00	Spaten: a Spatio-temporal and Textual Big Data Generator	Thaleia Dimitra Doudali, Ioannis Konstantinou, and Nectarios Koziris
9:00 – 9:30	SQL versus NoSQL Databases for Geospatial Applications	Elena Baralis, Andrea Dalla Valle, Paolo Garza, Claudio Rossi, and Francesco Scullino
9:30 – 9:45	Scalable Parallel Data Loading in SciDB (Short Paper)	Sangchul Kim, Junhee Lee, Taehoon Kim, and Bongki Moon
9:45 – 10:00	Towards development of spark based agricultural information system including Geo-spatial data (Short Paper)	Purnima Shah, Deepak Hiremath, and Sanjay Chaudhary
10:00 – 10:20	**Coffee Break**	
10:20 11:20	Keynote #1	Dr. Hendrik F. Hamann, IBM T.J. Watson Research Center, Yorktown Heights, NY
11:20 11:50	Discovering Dynamic Patterns of Urban Space via Semi-Nonnegative Matrix Factorization	Zhicheng Liu, Jun Cao, Sannyuya Liu Junyan Yang, and Qiao Wang
11:50 – 12:05	Multiscale Graph Theoretical Tools Reveal Subtle Patterns in Big Geospatial Data (Short Paper)	Ronald Hagan, Charles Phillips, Michael Langston, and Bradley Rhodes
12:05 – 12:20	Challenges and Trends about Smart Big Geospatial Data: A Position Paper (Short Paper)	Victor Saquicela, Luis Vilches, and Andrés Tello
12:20 – 13:30	**Lunch**	
13:30 – 14:00	Identifying Coherent Anomalies in Multi-Scale Spatio-Temporal Data using Markov Random Fields	Adway Mitra
14:00 – 14:30	A Map-Based Visual Analysis Method for Patterns Discovery of Mobile Learning in Education with Big Data	Dongbo ZhouHao Li, Sannyuya Liu, Bo Song, and Xiaohua Hu
14:30 – 15:00	Techniques for Efficient Detection of Rapid Weather Changes and Analysis of their Impacts on a Highway Network	Adil Alim, Aparna Joshi, Feng Chen, Catherine T. Lawson
15:00 – 15:30	Road Map Extraction from Satellite Imagery Using Connected Component Analysis and Landscape Metrics	Kulsawasd Jitkajornwanich, Peerapon Vateekul, Teerapong Panboonyuen, Siam Lawawirojwong, and Siwapon Srisonphan
15:30 – 15:50	**Coffee Break**	
15:50 – 16:50	Keynote #2	Michael Whitby, Digital Globe
16:50 – 17:20	Optimal Viewpoint Finding for 3D Visualization of Spatio-Temporal Vehicle Trajectories on Caution Crossroads Detected from Vehicle Recorder Big Data	Masahiko Itoh, Daisaku Yokoyama, Masashi Toyoda, and Masaru Kitsuregawa
17:20 – 17:50	Spatiotemporal Visualization of Traffic Paths Using Color Space Time Curve	Savitha Baskaran, Shiaofen Fang, and Shenghui Jiang
17:50 – 18:20	A Tale of Two Cities: Analyzing Road Accidents with Big Spatial Data	Rene Richard, and Suprio Ray
18:20 – 18:30	**Closing Remarks**	

2nd International Workshop on Enterprise Big Data Semantic and Analytics Modeling *Workshop Chair: Michael Peran; December 11, 2017*		
Time	**Title**	**Presenter**/Author(s)
13:00 - 13:05	Opening remarks and introduction	
13:05 - 13:30	Innovation in Big Data Analytics (Invited Talk 1)	**Eva K. Lee**
13:30 - 13:35	Q&A	
13:35 - 13:55	Hitting your number or not? A Robust & Intelligent Sales Forecast System	**Xin Xu**, Lei Tang, and Venkat Rangan
13:55 - 14:00	Q&A	
14:00 - 14:20	Artificial Intelligence Applied to Challenges in the Fields of Operations and Customer Support	**Ravi Santosh Arvapally**, Hasan Hicsasmaz, and Wally Lo Faro
14:20 - 14:25	Q&A	
14:25 - 14:45	Estimating Skill Fungibility and Forecasting Services Labor Demand	**Brian Johnston**, Benjamin Zweig, Michael Peran, Charlie Wang, and Rachel Rosenfeld
14:45 - 14:50	Q&A	
14:50 - 15:10	A Hybrid Bipartite Graph based Recommendation algorithm for Mobile Games	Yong Cai, Shaorong Liu, **Jinlong Hu**, Guihong Bai, and Shoubin Dong
15:10 - 15:15	Q&A	
15:15 - 15:50	**Coffee Break**	
15:50 - 16:15	Semantic Search (Invited Talk 2)	**Ricardo Baeza-Yates**
16:15 - 16:20	Q&A	
16:20 - 16:40	Machine learning approach for early detection of autism using a parental questionnaire and home video screening	Halim Abbas, **Ford Garberson**, Eric Glover, and Dennis Wall
16:40 - 16:45	Q&A	
16:45 - 17:05	A comparative sequence analysis of career paths among knowledge workers in a multinational bank	**Paul Squires**, Harold Kaufman, Julian Togelius, and Catalina Maria Jaramillo
17:05 - 17:10	Q&A	
17:10 - 17:30	Automated Knowledge Extraction from the Federal Acquisition Regulations System (FARS)	**Srishty Saha**, Karuna Joshi, Renee Frank, Michael Aebig, and Jiayong Lin
17:30 - 17:35	Q&A	
17:35 - 17:55	Governance Framework for Enterprise Analytics and Data	Atsushi Yamada and **Michael Peran**
17:55 - 18:00	Q&A	
18:00 - 18:20	Artificial Intelligence(AI), Automation, and its Impact on Data Science (Short paper)	**Richard Boire**
18:20 - 18:40	**Closing Remarks, discussion, and Q&A**	

The 1st International Workshop on Big Data Analytic for Cyber Crime Investigation and Prevention *Workshop Chairs: Andrii Shalaginov, Katrin Franke, Jan William Johnsen; Norwegian University of Science and Technology*		
Time	**Title**	**Presenter/Author**
13:30-13:40	**Opening Remarks and Welcome**	Andrii Shalaginov
13:40-14:20	Computational Forensics	**Invited Speaker:** Katrin Franke NTNU Digital Forensics Group
14:20-14:40	*"Neural Reputation Models learned from Passive DNS Data"*	Pierre Lison and Vasileios Mavroeidis
14:40-15:00	*"Extracting Cyber Threat Intelligence From Hacker Forums: Support Vector Machines versus Convolutional Neural Networks"*	Isuf Deliu, Katrin Franke, and Carl Leichter
15:00-15:20	*"Forensics Analysis of Wi-Fi Communication Traces in Mobile Devices"*	Anja Evelyn Amundsen and Kenneth M. Ovens
15:20-15:30	*"Topical Behavior Prediction from Massive Logs"*	Shih-Chieh Su
15:30-15:50	**Coffee Break**	
15:50-16:10	*"Identifying Extremism in Social Media with Multi-view Context-Aware Subset Optimization"*	Sreyasee Das Bhattacharjee, Bala Venkatram Balantrapu, William Tolone, and Ashit Talukder
16:10-16:30	*"Introducing DeepBalance: Random Deep Belief Network Ensembles to Address Class Imbalance"*	Peter Xenopoulos
16:30-16:50	*"Forensic Database Reconstruction"*	Joshua Sablatura, Bing Zhou
16:50-17:10	*"A First Estimation of the Proportion of Cybercriminal Entities in the Bitcoin Ecosystem using Supervised Learning"*	Haohua Sun Yin and Ravi Vatrapu
17:10-17:30	*"Exploratory studies into forensic logs for criminal investigation using case studies in industrial control systems in the power sector"*	Asif Iqbal
17.30-17:40	**Closing Remarks**	

Data Science for Emergency Management *Workshop Chairs: Elena Baralis, Paolo Garza, Laura Rusu, and Gandhi Sivakumar*		
Time	**Title**	**Presenter/Author**
1:30pm – 2:15pm	**Keynote**: Social Media and Digital Volunteering in Disaster Management	Carlos Castillo
2:15pm – 2:35pm	The Role of Unstructured Data in Real-Time Disaster-related Social Media Monitoring	Francesco Tarasconi, Michela Farina, Alessio Bosca, and Antonio Mazzei
2:35pm – 2:55pm	A Language-agnostic Approach to Exact Informative Tweets during Emergency Situations	Jacopo Longhini, Claudio Rossi, Claudio Casetti, and Federico Angaramo
2:55pm – 3:10pm	All in a twitter: self-tuning strategies for a deeper understanding of a crisis tweet collection	Evelina Di Corso, Francesco Ventura, and Tania Cerquitelli
3:10pm – 3:30pm	A Comparison of Classification Models for Natural Disaster and Critical Event Detection from News	Tim Nugent, Fabio Petroni, Natraj Raman, Lucas Carstens, and Jochen L. Leidner
3:30pm – 3:50pm	**Coffee Break**	
3:50pm – 4:05pm	Analyzing spatial data from Twitter during a disaster	Luca Venturini and Evelina Di Corso
4:05pm – 4:25pm	Summarization of emergency news articles driven by relevance feedback	Luca Cagliero
4:25pm – 4:40pm	Gamified Crowdsourcing for Disaster Risk Management	Antonella Frisiello, Quynh Nhu Nguyen, Claudio Rossi, and Fabrizio Dominici
4:40pm – 5:00pm	Coupling Early Warning Services, Crowdsourcing, and Modelling for Improved Decision Support and Wildfire Emergency Management	Conrad Bielski, Victoria O'Brien, Ceri Whitmore, Kaisa Ylinen, Ilkka Juga, Pertti Nurmi, Juha Kilpinen, Ignasi Porras, Josep Maria Sole, Pedro Gamez, Maria Navarro, Azra Alikadic, Andrea Gobbi, Cesare Furlanello, Gunter Zeug, M. Weirather, Jesus Martinez, Raquel Yuste, S Castro, Victoria Moreno, Tonny Vellin, and Claudio Rossi
5:00pm – 5:15pm	A Heat Wave Forecast System for Europe	Andrea Gobbi, Azra Alikadic, Cesare Furlanello, Ylinen Kaisa, and Federico Angaramo

5:15pm – 5:30pm	River segmentation for flood monitoring	Laura Lopez-Fuentes, Claudio Rossi, and Harald Skinnemoen
5:30pm – 5:45pm	Optimal Geospatial Volunteer Allocation Needs Realistic Distances	Jasmin Pielorz, Matthias Prandtstetter, Markus Straub, and Christoph H. Lampert
5:45pm – 6:00pm	Crowd Control and Evacuation Guidance Based on Simulations	Tomoichi Takahashi and Katsuki Ichinose

Applications of Big Data Technology in the Transport Industry *Workshop Chairs: Nii Attoh-Okine and John Easton*		
Time	**Title**	**Presenter/Author**
13:30 – 14:00	*Exploring Pavement Texture and Surface Friction Using Soft Computing Techniques*	Joshua Qiang Li, Guangwei Yang, You ZHAN, and Kelvin Wang
14:00 – 14:30	*Application of Machine Learning for Fuel Consumption Modelling of Trucks*	Federico Perrotta, Tony Parry, and Luis Neves
14:30 – 15:00	*Understanding data quality - Ensuring data quality by design in the rail industry*	Qian Fu and John Easton
15:00 – 15:30	*Edge Computing for Traffic Scene Analytics*	Yaw Adu-Gyamfi
15:30 – 16:00	**Coffee Break**	
16:00 – 16:30	*Track Geometry Big Data Analysis: A Machine Learning Approach*	Emmanuel Martey, Ahmed Lasisi, and Nii Attoh-Okine
16:30 – 17:00	*Privacy-Preserving Trajectory Classification of Driving Trip Data Based on Pattern Discovery Techniques*	Gene P. K. Wu and Keith C. C. Chan
17:00 – 17:30	*Comparison of Different Driving Style Analysis Approaches based on Trip Segmentation over GPS Information*	Marco Brambilla, Paolo Mascetti, and Andrea Mauri
17:30 – 17:50	*Round Table Discussion – Managing Data in the Multimodal Transport System*	John Easton
17:50 – 18:00	**Closing Remarks**	

4th KDDBHI Workshop: Big Data Analytic Technology for Bioinformatics and Health Informatics *Workshop Chairs: Donghui Wu and Xin Deng*		
Time	**Title**	**Presenter/Author**
1:30pm – 1:50pm	Toward Predicting Medical Conditions Using k-Nearest Neighbors	Shahab Tayeb
1:50pm – 2:10pm	bigNN: an open-source big data toolkit focused on biomedical sentence classification	Ahmad P. Tafti
2:10pm - 2:30pm	A Multi-task Machine Learning Approach for Comorbid Patient Prioritization	Goutam Mylavarapu
2:30pm - 2:50pm	A Medical Price Prediction System using Hierarchical Decision Trees	sanket.tavarageri
2:50pm -3:10pm	Explainable Data-Driven Modeling of Patient Satisfaction Survey Data	Ning Liu
3:10pm – 3:30pm	Mining Accompanying Relationships between Diseases from Patient Records	Arbee Chen
3:30pm -3:55pm	**Coffee Break**	
4:00pm - 4 :20pm	High Dimensional Data Processing for Fetal Activity Evaluation	Denis Kouame
4:20pm – 4:40pm	,iVAR: Interactive Visual Analytics of Radiomics Features from Medical Images	Lina Yu
4:40pm – 5:00pm	A Multimedia Big Data Retrieval Framework to Detect Dyslexia Among Children	Elham Hassanain
5:00pm – 5:20pm	Patient-individual Morphological Anomaly Detection in Multi-lead Electrocardiography Data Streams	Alexander Acker
5:20pm – 5:40pm	Visualization of Non-metric Relationships by Adaptive Learning Multiple Maps t-SNE Regularization	Xianjun Shen
5:40pm – 6:00pm	Predicting Efficacy of Therapeutic Services for Autism Spectrum Disorder Using Scientific Workflows	Fahima Bhuyan
	Closing Remarks	

Open Science in Big Data (OSBD) *Workshop Chairs: Shannon Quinn, Suchi Bhandarkar, John Miller*		
Time	**Title**	**Presenter/Author**
1:40 – 2:15	*Deep Learning Enabled National Cancer Surveillance*	Dr. Georgia Tourassi
2:15 – 2:50	*Unifying the Open Big Data World: The Possibilities* of Apache BEAM*	Holden Karau
2:50 – 3:25	*New Data Paradigms: From the Crowd and Back*	Dr. Rumi Chunara
3:30 – 3:50	**Coffee Break**	
3:50 – 4:15	*iEnvironment: A Software Platform for Integrated Environmental Monitoring and Modeling of Surface Water*	Paulo Alencar, Donald Cowan, Doug Mulholland, Bruce MacVicar, Simon Courtenay, Stephen Murphy, and Fred McGarry
4:15 – 4:40	*Preparing Data Managers to Support Open Ocean Science: Required Competencies, Assessed Gaps, and the Role of Experiential Learning*	Lee Wilson, Adrienne Colborne, and Michael Smit
4:40 – 5:05	*Modeling Multiple Subskills by Extending Knowledge Tracing Model Using Logistic Regression*	Xuan Zhou, Wenjun Wu, and Yong Han
5:05 – 5:30	*An Intelligent Update of SDN Forwarding Tables*	Manal Taoufiki, Nour Gritli, and Omar Cherkaoui
5:30 – 5:55	*MSRM ：A Novel Model to Retrieve Meaningful Opinion Sentences for New Products*	Nana Du
5:55 – 6:00	**Closing Remarks**	

2nd International Workshop on Big Data Transfer Learning (BDTL) -- Automatic Knowledge Mining and Transfer for Digital Healthcare *WorkshopChairs: Yun Fu, Honggang Wang, Yu Cao, and Ming Shao*		
Time	**Title**	**Presenter/Author**
1:30PM	Invited Talk: CNN with Small Data	Dr. Sarah Ostadabbas
2:20PM	Automatic Topic Discovery of Online Hospital Reviews Using an Improved LDA with Variational Gibbs Sampling	Richard de Groof and Haiping Xu
2:50PM	Invited Talk: CNN based Person Recognition for Uncontrolled Clinical Scenarios in Contactless Population Monitoring	Dr. Haibo Wang
3:40PM	Fragrance to Vector as Scent Technology	Noriaki Koide and Yu Ichifuji
4:05PM	**Coffee Break**	
4:25PM	Invited Talk: Recent Advances in Transfer Learning and Applications	Dr. Ming Shao
5:15PM	Cross-Database Mammographic Image Analysis through Unsupervised Domain Adaptation	Deepak Kumar, Chetan Kumar and Ming Shao
5:45PM	**Closing Remarks**	

Big Data for Economic and Business Forecasting December 14, 2017 *WorkshopChairs: Shouyang Wang, Wei Shang*		
Time	**Title**	**Presenter/Author**
2:00-2:20pm	*Towards Building a Hybrid Model for Predicting Stock Indexes*	Farhana Zulkernine
2:20-2:40pm	*Insurance premium optimization using motor insurance policies - a business growth classification approach*	Daniel Meuller
2:40-3:00pm	*A new time series prediction method based on complex network theory*	Minggang Wang
3:00-3:20pm	*Predicting Business Performance through Patent Applications*	Daniel Meuller

3:20-3:40pm	*Stock Price Forecasting Using Support Vector Regression: Based on Network Behavior Data*	Quan Jin
3:40-4:00pm	*Integrating heterogeneous data sources for traffic flow prediction through extreme learning machine*	Wei Dai
4:00-4:20pm	*Forecasting Tourist Arrivals with Machine Learning and Internet Search Index*	Yunjie Wei
4:20 –4:30	**Coffee Break**	
4:30-4:50pm	*An enhanced LGSA-SVM for S&P 500 index forecast*	Wei Shang
4:50-5:10pm	*The Construction and Application of Expectations Index on Monetary Policy*	Guihuan Zheng
5:10-5:30pm	*Impress Backers at First Sight: An Image-Oriented Analytics Approach for Predicting Success in Crowdfunding Platforms*	Wei Xu
5:30-5:50pm	*Agglomeration, Network and Urban Development: A Study on Media Connection Network Index of Cities*	Xiaolan Yang
5:50-6:10pm	*Can search data help forecast inflation? Evidence from a 13-country panel*	Xun Zhang
6:10-6:20pm	*An Augmented Fama and French Three-Factor Model Using Social Interaction*	Lin Huo
	Closing Remarks	

3rd International Workshop on Big Data for Sustainable Development

WorkshopChairs: Aki-Hiro Sato, Chu-Hua Kuei, and Antoaneta Serguieva

Time	Title	Presenter/Author
11:05-11:15	**Opening Remarks**	Dr. Aki-Hiro Sato Dr. Antoaneta Serguieva
11:15-11:40	Micro-sensoring: Antibodies and aptamer-based Micro-ELISA as performing offline/online tool for allergens and mycotoxins detection in foods	Rob Dolci
11:40-12:05	Cluster-Overlap Algorithm for Evaluating Relevance and Processing Choices in Environmental Sustainability Contexts with Multiple Dependent Attributes	Anne Denton and Arighna Roy
12:05-12:30	Big Data processing: Is there a framework suitable for Economists and Statisticians?	Giuseppe Bruno, Demetrio Condello, Alberto Falzone, and Andrea Luciani
12:30-2:00	**Lunch Break**	
2:00-2:25	World Grid Square Codes: Definition and an example of world grid square data	Aki-Hiro Sato, Shoki Nishimura, and Hiroe Tsubaki
2:25-2:50	Characterization of Cities Based on World Grid Square Statistics about Specific Properties	Aki-Hiro Sato
2:50-3:15	Statistical Analysis of Hotel Plan Popularity in Regional Tourist Areas	Hiroshi Tsuda, Masakazu Ando, and Yu Ichifuji
3:15-3:40	Developing Sustainable Trading Strategies Using Directional Changes with High Frequency Data	Ailun Ye, V L Raju Chinthalapati, Antoaneta Serguieva, and Edward Tsang
3:40-4:05	Sustainable Blockchain-Enabled Services: Smart Contracts	Craig Wright and Antoaneta Serguieva
4:05-4:25	**Coffee Break**	
4:25-4:50	Critical Enablers of Sustainable Water Management (SWM): Text Evidences from 10 Countries	Chu-hua Kuei, Christian N. Madu, and Picheng Lee
4:50-5:00	**Closing Remarks**	Dr. Aki-Hiro Sato Dr. Antoaneta Serguieva

The First IEEE Workshop on Human-Machine Collaboration in BigData (HMData 2017) *Workshop Chairs: Atsuyuki Morishima (University of Tsukuba), Senjuti Basu Roy (New Jersey Institute of Technology), Lei Chen (HKUST)* *The detailed program is available at http://humanmachinedata.org*		
Time	**Title**	**Presenter/Author**
8:45	**Opening**	
8:50	Formalizing Interruptible Algorithms for Human over-the-loop Analytics	Austin Graham, Yan Liang, Le Greunwald, and Christan Grant
9:20	Active Preference Learning for Generative Adversarial Networks	Masahiro Kazama and Viviane Takahashi
9:30	Improving Classification Accuracy in Crowdsourcing through Hierarchical Reorganization	Xiaoni Duan and Keishi Tajima
9:40	A Crowd-in-the-Loop Approach for Generating Conference Programs with Microtasks	Naoki Kobayashi, Masaki Matsubara, Keishi Tajima, and Atsuyuki Morishima
9:50	Clarifying the Transition of Workload for Victims Life Reconstruction Support Programs in Affected Local Governments Using the Victims Master Database -Comparison between the 2007 Chuetsu-oki Earthquake and the 2016 Kumamoto Earthquake-	Munenari Inoguchi, Keiko Tamura, Kei Horie, and Haruo Hayashi
10:00	**Coffee Break**	
10:20	Implicit Order Join: Joining Log Data with Property Data by Discovering Implicit Order-oriented Keys with Human Assistance	Takahiro Komamizu, Toshiyuki Amagasa, and Hiroyuki Kitagawa
10:50	Crossing the Streams: Fuzz testing with user input	Joseph Cottam, Leslie Blaha, Dimitri Zarzhitsky, Mathew Thomas, and Elliott Skomski
11:00	Crowd-based Best-effort Number Estimation	Yuzuki Furuhashi, Masaki Matsubara, and Atsuyuki Morishima
11:10	Conceptual design for comprehensive research support platform	Mamiko Matsubayashi and Keiko Kurata
11:20	A Method to Generate Disaster-Damage Map by Using 3D photometry and Crowd Sourcing	Koyo Kobayashi, Hidehiko Shishido, Yoshinari Kameda, and Itaru Kitahara
11:30	Participants Self-Introduction Session	All Participants
12:00	**Lunch**	
13:30	Keynote: Targeted Crowdsourcing with a Billion (Potential) Users	Panos Ipeirotis
14:30	Collaborative Filtering and Rating Aggregation Based on Multicriteria Rating	Hiroki Morise, Satoshi Oyama, and Masahito Kurihara
15:00	Towards Predicting Task Performance from EEG Signals	Michalis Papakostas, Konstantinos Tsiakas, Theodoros Giannakopoulos, and Fillia Makedonn
15:10	Proactive Preservation of World Heritage by Crowdsourcing and 3D Reconstruction Technology	Hidehiko Shishido, Yutaka Ito, Youhei Kawamura, Toshiya Matsui, Atsuyuki Morishima, and Itaru Kitahara
15:20	“DEKATSU” Activity of Data and Service Collaboration among Private Companies and Academic Institutions for Tokyo Metropolitan Resilience Project	Keiko Tamura and Naoshi Hirata
15:30	**Coffee Break (with Posters)**	
16:10	A Trade-off between Estimation Accuracy of Worker Quality and Task Complexity	Yoshitaka Matsuda, Yu Suzuki, and Satoshi Nakamura
16:40	Using categorized web browsing history to estimate the user's latent interests for web advertisement recommendation	Panote Siriaraya, Yuriko Yamaguchi, Mimpei Morishita, Yoichi Inagaki, Reyn Nakamoto, Jianwei Zhang, Junichi Aoi, and Shinsuke Nakajima
17:10	**Closing Remarks**	

3rd International Workshop on Smart Cities: People, Technology, and Data *WorkshopChairs: Frederico Lopes (UFRN, Brazil), Koh Takeuchi (NTT, Japan)*		
Time	**Title**	**Presenter/Author**
8:45-9:00	**Opening remarks: co-organizer Frederico Lopes**	
	Session 1: City management (Chair: Koh Takeuchi)	
9:00-9:20	Road Marking Blurs Detection with Drive Recorder	Makoto Kawano
9:20-9:40	Self-Adaptive and Resilient Urban Networking Infrastructure for Disasters and Smart City Services	Paul Flikkema
9:40-10:00	Reliability Analysis of an IoT-Based Smart Parking Application for Smart Cities	Gustavo Girao
10:00-10:20	**Coffee Break**	
	Session 2: First International Workshop on Big Data in Smart Cities and Smart Buildings (Chair: Gustavo Girão)	
10:20-10:40	Proposing an Access Gate to Facilitate Knowledge Exchange for Smart City Services	Viviana Angely Bastidas Melo
10:40-11:00	A Model for the Socially Smart City	Paulo Alencar
11:00-11:20	A Whole Building Fault Detection Using Weather Based Pattern Matching and Feature Based PCA Method	Yimin Chen
11:20-14:30	**Lunch**	
	Session 3: City event detection (Chair: Takuro Yonezawa)	
14:30-14:50	Datafying city: detecting and accumulating spatio-temporal events by vehicle-mounted sensors	Yasue Kishino
10:20-10:40	Using Social Media Photos to Identify Tourism Preferences in Smart Tourism Destination	Frederico Lopes
15:10-15:30	Data Analysis on Train Transportation Data with Nonnegative Matrix Factorization	Kyoichi Ito
15:30-15:50	**Coffee Break**	
	Session 4: City applications (Chair: Futoshi Naya)	
15:50-16:10	Analytical Toolbox for Smart City Applications: Garbage Collection Log Use Case	Takahiro Komamizu
16:10-16:30	GuideMe: Route Coordination of Participating Agents in Mobile Crowd Sensing Platforms	Christine Bassem
16:30-16:50	MM360: GPS-assisted 360° Video Sharing System for Participatory Events	Naoya Shibahara
16:50-17:00	**Short Break**	
17:00-17:45	Panel Discussion	Frederico Lopes
17:45-18:00	**Closing Remarks: co-organizer Koh Takeuchi**	

Big Data Analytics and Internet of Things Chair: Levente Klein, IBM TJ Watson Research Center		
Time	**Title**	**Presenter/Author**
8:00-8:15	Introduction and workshop aims	
8:15-9.20	**Invited Presentation "Better Cities through Imaging"**	Greg Dobler, NYU, USA
9:20-10:00	Session 1: Scalable data analytics and Data Fusion	
9:20-9:40	Using Big Data Analytics and IoT Principles to Keep an Eye on Underground Infrastructure	Joshua Lieberman,Tumbling Walls, USA
9:40-10:00	Data driven modelling for energy consumption prediction on smart buildings	Aurora González-Vidal,University of Murcia, Spain
10:00-10.20	Coffee Break	
10.20-12.00	Session 2: Industry specific big data analytics for IoT	
10:20-10:40	Machine Learning and Air Quality Modeling	Christoph Keller, NASA Global Modeling and Assimilation Office / USRA, USA
10.40-11:00	Event Clustering & Event Series Characterization on Expected Frequency	Conrad Albrecht, IBM, USA
11:00-11:20	Understanding the Impact of Lossy Compressions on IoT Smart Farm Analytics	Aekyeung Moon, Electronics and Telecommunications Research Institute, Korea
11:20-12:00	A low maintenance particle pollution sensing system using the Minimum Airflow Particle Counter (MAPC)	Ted van Kessel, and Ramachandran Muralidhar, IBM, USA
12:00-13.30	Lunch break (Lunch on you own)	
13:30-14.30	Invited Presentation: "Big Data Challenges for Industry"	Matt Nielsen, GE,USA
14:30-15.30 pm	Session 3: Edge computing and Edge Data Informatics	
14.30-14:50	Source characterization of airborne emissions using a sensor network: examining the impact of sensor quality, quantity, and wind climatology	Xiaochi Zhou, Vinicius Amaral, and John Albertson,Cornell University, USA
14.50-15.10	'Petroleum Analytics Learning Machine' For Optimizing the Internet of Things Of Today's Digital Oil Field To Refinery Petroleum System	Roger Anderson, Columbia University, USA
15.10-15:30	Wireless Sensor Network for fugitive methane gas emission monitoring	Levente Klein, IBM, USA
15:30-15.50	Coffee Break	
15.50-16.30	Session 4: Advanced analytics	
15.50-16.10	Developing an edge computing platform for real-time descriptive analytics	Hung Cao and Monica Wachowicz,U. of New Brunswick, Canada
16.10-16:30	Energy Efficiency Driven by a Storage Model and Analytics on a Multi-System Semantic Integration	Domitille Couloumb, Schneider Electric, USA
16.30-16.50	Measures of Network Centricity for Edge Deployment of IoT Applications	Dinesh Verma, IBM, USA
16:50-18.10	Panel discussion "Will IoT change fundamentally our life?"	Andreas Oloffson, Hon Pak, Jono Anderson, Matt Nielsen, Greg Dobler
18:10-18:15	Closing Remarks	

4th International Workshop on Privacy and Security of Big Data (PSBD 2017)		
Time	**Title**	**Presenter/Author**

8:00am – 8.25am	**Session PSBD17_1: Opening** **Chair: Alfredo Cuzzocrea**	
8:25am – 9.25am	**Session PSBD17_2: Invited Talk – Jacob Whitehill, "Climbing the Kaggle Leaderboard by Exploiting the Log-Loss Oracle"** **Chair: Alfredo Cuzzocrea**	
9:25am – 10.45am	**Session PSBD17_3: Security Analysis Models and Algorithms for Big Data** **Chair: Alfredo Cuzzocrea**	
9:25am – 9.45am	Modeling User Communities for Identifying Security Risks in an Organization	Anirban Das, Min-Yi Shen, Jisheng Wang
9:45am – 10.05am	Impact of Security Awareness Training on Phishing Click-Through Rates	Antohny Carella, Murat Kotsoev, Traian Marius Truta
10:05am – 10.25am	Link Before You Share: Managing Privacy Policies through Blockchain	Agniva Banerjee, Karuna Pande Joshi
10:25am – 10.45am	Tor Traffic Analysis and Detection via Machine Learning Techniques	Alfredo Cuzzocrea, Fabio Martinelli, Francesco Mercaldo, Gianni Vercelli
10:45am - 11:05am	**Coffee Break**	
11:05am – 12.45am	**Session PSBD17_4: Advanced Applications Embedding Big Data Privacy and Security** **Chair: Dan Chia-Tien Lo**	
11:05am – 11.25am	Securing the Positioning Signals of Autonomous Vehicles	Shahab Tayeb, Gabriel Esguerra, Kimiya Ghobadi, Jimson Huang, Robin Hill, Derwin Lawson, Stone Li, Tiffany Zhan
11:25am – 11.45am	Collaborative Caching Techniques for Privacy- Preserving Location-based Services in Peer-to-Peer Environments	Kangsoo Jung, Seog Park
11:45am – 12.05am	Automated Big Security Text Pruning and Classification	Khudran Alzhrani, Ethan Rudd, Edward Chow, Terrance Boult
12:05am – 12.25am	Automated Microsoft Office Macro Malware Detection using Machine Learning	Ruth Bearden, Dan Chia-Tien Lo
12:25am – 12.45am	Secure Power Scheduling Auction for Smart Grids Using Homomorphic Encryption	Haya Shajaiah, Ahmed Abdelhadi, Charles Clancy
12:45am - 2:00pm	**Lunch**	
2:00pm – 4.05pm	**Session PSBD17_5: Privacy-Preserving Big Data Management and Cloud Applications** **Chair: Traian Truta**	
2:00pm – 2.20pm	A Top-Down k-Anonymization Implementation for Apache Spark	Ugur Sopaoglu, Osman Abul
2:20pm – 2.40pm	Efficient and Private Approximations of Distributed Databases Calculations	Philip Derbeko, Shlomi Dolev, Ehud Gudes, Jeffrey Ullman
2:40pm – 3.00pm	Data Masking Techniques for NoSQL Database Security: A Systematic Review	Alfredo Cuzzocrea, Hossain Shahriar
3:00pm – 3.20pm	User-Profile-Based Analytics for Detecting Cloud Security Breaches	Trishita Tiwari, Ata Turk, Alina Oprea, Katzalin Olcoz Herrero, Ayse Coskun
3:20pm – 4.05pm	**Session PSBD17_6: Panel: "Fighting Fake News Spread in Online Social Networks: Actual Trends and Future Research Directions"** **Chair: Alfredo Cuzzocrea**	
4:05pm - 4:25pm	**Coffee Break**	

Workshop on Big Data Technology and Ethics Considerations in Customer Behavior and Customer Feedback Mining (BEBF BigData 2017)

WorkshopChairs: Xin Deng, Ross Smith

Time	Title	Presenter/Author
2:00-2:20pm	**Open Remark**: Big Data Technology and Ethics Considerations in Customer Behavior and Customer Feedback Mining	Xin Deng
2:20-2:45pm	Towards an Ethical Application of Customer Feedback Data	Ross Smith

2:45-3:10pm	Customer Churn Prediction in an Internet Service Provider	Phuong Vo.T.H
3:10-3:35pm	A Big Data Analytics Framework for Forecasting Rare Customer Complaints	Donghui Wu
3:35-4:10pm	Dynamic Bayesian Predictive Model for Box Office Forecasting	Wutao Wei
4:10-4:30pm	**Coffee Break**	
4:30-4:55pm	Training on the Poles for Review Sentiment Polarity Classification	Michael Kranzlein
4:55-5:20pm	Understanding Rating Behavior based on Moral Foundations: The case of Yelp Reviews	Pegah Nokhiz
5:20-5:45pm	Heterogeneous Knowledge Transfer via Domain Regularization for Improving Cross-Domain Collaborative Filtering	yizhou zang
5:45-6:10pm	A Scalable Sequential Principal Component Analysis Algorithm (SeqPCA) with Application to User Access Control Analysis	Yixuan Qiu and Wutao Wei
6:10-6:15pm	**Closing Remarks**	

International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID) *WorkshopChairs:* Huaglory Tianfield		
Time	**Title**	**Presenter/Author**
14:00-14:20	Network Intrusion Detection using Word Embeddings	Xiaoyan Zhuo, Jialing Zhang, and Seung Woo Son
14:20-14:40	DNS Graph Mining For Malicious Domain Detection	Hau Tran Xuan, An Nguyen Thanh, Phuong Vo.T.H, and Tu Vu Anh
14:40-15:00	Deriving Cyber Use Cases from Graph Projections of Cyber Data Represented as Bipartite Graphs	Mohammed Eslami, George Zheng, Hamed Eramian, and Georgiy Levchuk
15:00-15:20	Improving Cyber-Attack Predictions Through Information Foraging	Adam Dalton, Bonnie Dorr, Leon Liang, and Kristy Hollingshead
15:20-15:40	Towards a Definition of Cyberspace Tactics, Techniques and Procedures	Fernando Maymi, Robert Bixler, Randolph Jones, and Scott Lathrop
15:40-16:00	Binary Malware Image Classification using Machine Learning with Local Binary Pattern	Jhu-sin Luo and Dan Chia-Tien Lo
16:00-16:10	Workshop discussion on cybersecurity issues	All participants
16:10 –16:30	**Coffee Break**	
16:30-16:50	Twitter-Enhanced Android Malware Detection	Jordan DeLoach and Doina Caragea
16:50-17:10	Detection of Hacking Behaviors and Communication Patterns on Social Media	Olga Babko-Malaya, Rebecca Cathey, Steve Hinton, David Maimon, and Taissa Gladkova
17:10-17:30	On the Relevance of Social Media Platforms in Predicting The Volume and Patterns of Web Defacement Attacks	David Maimon, Andrew Fukuda, Olga Babko-Malaya, Rebecca Cathey, and Steve Hinton
17:30-17:50	Sentiment Analysis via Multi-Layer Perceptron Trained by Meta-Heuristic Optimisation	Dabiah Ahmed Alboaneen, Huaglory Tianfield, and Yan Zhang
17:50-18:00	Workshop discussion on social media issues	All participants

Big Data Analytics in the Legal Industry *Workshop Chairs: Jianping Zhang, Robert Keeling, Rishi Chhatwal, Peter Gronvall, Nathaniel Huber-Fliflet*		
Time	**Title**	**Presenter**
2:15-3:00pm	Introduction and Roundtable Session	Nathaniel Huber-Fliflet

3:00-3:25pm	An Empirical Study of the Application of Machine Learning and Keyword Terms to Privilege Document Review	Peter Gronvall
3:25-3:50pm	Predictive Analytics for Litigation Case Management	Jerzy Bala and Michael Kellar
3:50-4:15pm	A feasibility experiment on the application of predictive coding to instant messaging corpora	Thanasis Schoinas
4:15-4:35pm	**Coffee Break**	
4:35-5:00pm	Glyph Analytics: A New Tool Set for Managing Big Unstructured Data Collections	John Martin
5:00-5:25pm	Using Google Analytics to Support Cybersecurity Forensics	Han Qin
5:25-5:50pm	Word Embeddings for Text Categorization with Application in Legal Text Analytics	Haozhen Zhao
5:50-6:15pm	Empirical Evaluations of Active Learning Strategies in Legal Document Review	Nathaniel Huber-Fliflet

International Workshop on Big Data for Financial News and Data *Workshop Chairs: Quanzhi Li, Sameena Shah*		
Time	**Title**	**Presenter/Author**
2:00pm-2:30pm	*Predicting Stock Movement Direction with Machine Learning: an Extensive Study on S&P 500 Stocks*	Yang Jiao, Jérémie Jakubowicz
2:30pm-3:00pm	*Credit Decision Tool using Mobile Application Data for Microfinance in Agriculture*	Naomi Simumba, Suguru Okami, Naohiko Kohtake
3:00pm-3:30pm	*Building Industry Network Based on Business Text: Corporate Disclosures and News*	Sung Whan Jeon, Hye Jin Lee, Sungzoon Cho
3:30pm-4:00pm	Natural Language Processing R&D at Alibaba Group	Quanzhi Li

Symposium on Data Analytics for Advanced Manufacturing

Tuesday, December 12, 2017

Time	**Event**
8:45 – 9:45	**Conference Keynote Speech:** *Human-in-the-loop Applied Machine Learning* **Prof. Carla E. Brodley, Northeastern University, USA**

9:45 – 10:45	**Conference Keynote Speech:** *TextScope: Enhance Human Perception via Text Mining* **Dr. ChengXiang Zhai, Professor, University of Illinois at Urbana-Champaign, USA**
10:45 – 11:05	*Coffee Break*
11:05 – 11:15	**Opening Remarks: Sudarsan Rachuri, DOE**
11:15 – 12:00	**Symposium Keynote Speech:** *IoT: Opportunities and Challenges* **Dr. Sanjay Sarma - Vice President for Open Learning, MIT**
12:00 – 12:45	**Symposium Keynote Speech:** *Federal Internet of Things Initiatives for Industrie 4.0 (I4.0)* **Dr. Regine Gernert and Matthias Kuom – Program Managers, German Aerospace Center (DLR-PT)**
12:45– 14:00	*Lunch*
14:00 – 16:05	**Tutorial:** *Building and Deploying Predictive Analytics Models Using PMML Standard* **Svetlana Levitan, Advisory Software Engineer, IBM Hybrid Cloud, IBM Corp**
16:05 – 16:25	*Coffee Break*
16:25 – 18:05	**Technical Paper Session 1** *(Session Chair: Dr. Ronay Ak)*
16:25 – 16:50	*Statistically-substantiated density characterizations of additively manufactured steel alloys through verification, validation, and uncertainty quantification* **Heather Reed, Corbin Robeck, Richard P. Vinci, Trevor Verdonik, Christina Viau Haden, Michael Pires, Maria Castro, and Wojciech Misiolek**
16:50 – 17:15	*A Data-Driven Approach for Improving Sustainability Assessment in Advanced Manufacturing* **Yunpeng Li, Heng Zhang, Utpal Roy, and Yung-Tsun Tina Lee**
17:15 – 17:40	*Issues in Synthetic Data Generation for Advanced Manufacturing* **Don Libes, David Lechevalier, and Sanjay Jain**
17:40	Adjourn

Wednesday, December 13, 2017

Time	
8:45 – 9:45	**Conference Keynote Speech:** *Large-scale Graph Representation Learning* **Dr. Jure Leskovec, Associate Professor, Stanford University, Chief Scientist at Pinterest, USA**
9:45 – 10:45	**Conference Keynote Speech:** *Contextual Reinforcement Learning* **Dr. John Langford, Microsoft Research**
10:45 – 11:05	*Coffee Break*
11:05	**Technical Paper Session 2** *(Session Chair: Dr. Anantha Narayanan)*
11:05 – 11:30	*Hybrid Datafication of Maintenance Logs from AI-Assisted Human Tags* **Thurston Sexton, Michael Brundage, Michael Hoffman, and KC Morris**
11:30 – 11:55	*Automatic Localization of Casting Defects with Convolutional Neural Networks* **Max Ferguson, Ronay Ak, Yung-Tsun Tina Lee, and Kincho Law**
11:55 – 12:20	*Estimation of online tool wear in turning processes using recurrence quantification analysis (RQA)* **Srinivasan Radhakrishnan, Yung-Tsun Tina Lee, and Sagar Kamarthi**
12:20 – 12:45	*Manufacturing and Contract Service Networks: Composition, Optimization and Tradeoff Analysis based on a Reusable*

	Alexander Brodsky, Mohan Krishnamoorthy, M. Omar Nachawati, William Z. Bernstein, and Daniel A. Menasce
12:45 – 14:00	*Lunch*
14:00-14:05	*Session Chair: Y. Tina Lee*
14:05 – 14:45	**Symposium Keynote Speech:** *Guided Deep Reinforcement Learning for Additive Manufacturing Control Applications* **Dr. Kishore K. Reddy – Research Scientist, United Technologies Research Center**
14:45 – 15:25	**Symposium Keynote Speech:** *Creating Data-driven Advanced Manufacturing Collaborative Communities* **Douglas Ramsey, Vice President, Business Development, Citrine Informatics**
15:25 – 16:05	**Symposium Keynote Speech:** *Clean Energy Smart Manufacturing Innovation Institute-Vison and Roadmap* **Jim Wetzel – CEO, CESMII**
16:05 – 16:25	*Coffee Break*
16:25 – 17:30	**Panel: Big Data Analytics and IoT for Advanced Manufacturing: Challenges and opportunities** **Panelists:** Regine Gernert (German Aerospace Center-DLR-PT), Sagar Kamarthi (Northeastern Univ.), Kincho Law (Stanford Univ.), Douglas Ramsey (Citrine Informatics), Kishore K. Reddy (UTRC), Jim Wetzel (CESMII), Valri Lightner (AMO, EERE, DOE) **Panel Moderator:** Dr. Sudarsan Rachuri, Department of Energy
17:30	Adjourn

Keynote Speeches

Keynote 1:

Title: IoT: Opportunities and Challenges

Speaker: Dr. Sanjay Sarma - Vice President for Open Learning, MIT

Abstract:

The Internet of Things has garnered a lot of attention recently in terms of its potential. However, many questions remain: how does one start? What will its eventual impact be? What are the deeper research questions? Do we need more fundamental science and math to deal with it? What to do with all the data? Where do adjacent technologies such as machine learning and big data fit? I will present a view that IoT and cyber-physical systems require a fundamental rethinking of how we build, secure, manage, operate and monetize these systems ranging from factories to cars to homes, and that the incremental approach of adding IoT capabilities without thinking the issues may lead to irreparable problems in the future. In particular, I will describe concepts such as the cognitive firewall, data proxies, the use of the cloud, and the underlying concepts from control theoretic, computing architecture, networking and security perspectives.

Bio:

Sanjay Sarma is the Vice President for Open Learning. He also leads the Office of Digital Learning, which oversees MIT OpenCourseWare and supports the development and use of digital technology for on-campus teaching and massive open online courses (MOOCs). He is also the Fred Fort Flowers (1941) and Daniel Fort Flowers (1941) Professor of Mechanical Engineering at MIT.

A co-founder of the Auto-ID Center at MIT, Sarma developed many of the key technologies behind the EPC suite of RFID standards now used worldwide. Currently, Sarma serves on the boards of GS1, EPCglobal, several startup companies including Senaya and

ESSESS, and edX, the not-for-profit company set up by MIT and Harvard to create and promulgate an open-source platform for the distribution of free online education worldwide.

Author of more than 75 academic papers in computational geometry, sensing, RFID, automation, and CAD, Sarma is the recipient of numerous awards for teaching and research, including the MacVicar Fellowship, the Business Week eBiz Award, and InformationWeek's Innovators and Influencers Award. He received his bachelor's degree from the Indian Institute of Technology, his master's degree from Carnegie Mellon University, and his PhD from the University of California at Berkeley.

Keynote 2:

Title: Federal Internet of Things Initiatives for Industrie 4.0 (I4.0)

Speakers: Dr. Regine Gernert and Matthias Kuom – Project Managers, German Aerospace Center (DLR-PT)

Abstract:

The presentation gives an insight about the Federal Industrial Internet of Things initiatives of Germany including Platform Industrie 4.0 (I4.0). The talk will also integrate a European and global perspective with respect to collaboration and standardization aspects. For instance, the Alliance for the Internet of Things Innovation and the Industrial Internet Consortium are of interest. If companies want to provide their data for digital services in a distributed and decentralized manner, the upcoming reference architecture of the Industrial Data Space will be of interest to them. The Platform Industrie 4.0 is in the process of standardizing its reference architecture RAMI at a European and international level.

There is a long tradition in the digitization of industry in Germany, also in supporting this with national funding programs. With the launch of the "Smart Data" initiative four projects in the industrial application area started to develop and elaborate new technologies that enable big data to be used in a secure and legally compliant manner and they will give an impact to data analytics in manufacturing, mainly in post-production maintenance. In manufacturing, there was a special initiative called "Autonomics" oriented on exploring the Internet of Things. In the subsequent initiative "PAiCE" IIoT pioneering technology fields, such as secure industrial communication, are addressed. The flagship project "Industrial Communication for Factories" specifically targets future industrial applications and, accordingly, a communication reference architecture and an easy-to-use modular technology toolkit will be developed. In particular, the goal is a modular approach to enable a flexible composition of components including key technologies like 5G or Edge Cloud Computing. Thus, starting with the networking of smart objects for industrial applications (the Internet of Things), organizing the data management and using analytics platforms (the Internet of Data), new web-based knowledge infrastructures are possible that pave the way for new electronic services (the Internet of Services). With the initiative "Smart Service World" all these levels are addressed. In a "Smart Services World" different digital user areas are connected using a targeted, secure combination of open service platforms, data management technologies, and Internet of Things. The presentation shows best practice examples as well as new approaches.

Bios:

Regine Gernert is employed at the Project Management Agency at German Aerospace Center (DLR-PT). She supports with her working group at DLR-PT in Berlin the European and internationalization strategy of the Unit "Digital Technologies" of the German Federal Ministry for Economic Affairs and Energy (BMWi). In addition to the monitoring of international developments and the initiation of collaborations, one of the objectives of the last years was the support of innovations in electronic services and out of data (Internet of Services, Internet of Data).

Previously Mrs. Gernert worked from 2006 to 2013 as project manager for research funding activities in the field of information and communication technologies (ICT) in the BMWi. Dr. Gernert studied computer science and received her degree Dr.-Ing. from the Technical University of Berlin in the field of Industrial Information Technology.

Matthias Kuom joined the DLR-PT in 2002. Mr. Kuom is responsible for the implementation of ICT R&D funding programmes, including the evaluation of funding concepts, professional project tracking and financial control. Currently he supports the German Federal Ministry for Economic Affairs and Energy (BMWi) in the conceptual and organisational design of cross-border research funding as well as in the internationalisation of programme activities (esp. Industrie 4.0 and Autonomous Systems).

Prior to joining DLR, Mr. Kuom managed the implementation of e-health solutions at T-Systems – including portal technologies, interfaces with an electronic health record and pervasive end-user assistance systems. He designed an R&D project (in cooperation with well-known hospitals and clinics in Germany) for a tele-medicine solution which provided support for stroke sufferers during their rehabilitation process. Earlier in his career, Mr. Kuom was a scientist in the ICT Media and Communication Group at the Institute for Futures Studies and Technology Assessment.

Keynote 3:

Title: Guided Deep Reinforcement Learning for Additive Manufacturing Control Applications

Speaker: Dr. Kishore K. Reddy, United Technologies Research Center

Abstract:

Additive manufacturing (e.g., 3D printing, cold spray and powder bed manufacturing) encompasses a wide range of tasks that commonly involve complex trajectory traversal by a robotic agent to meet multifaceted objectives such as surface quality, material properties etc. Traditional control or reinforcement learning can be inefficient in handling such a rich motion range with limited scalability to perception in a natural environment. In this talk, we will present a deep neural network based guided policy search (GPS) framework for optimizing trajectory policy of nozzle dynamics in cold spray application.

Bio:

Kishore K. Reddy is a Research Scientist at the United Technologies Research Center (UTRC) working in the area of computer vision, human machine interaction (HMI) and machine learning. He is currently leading the Digital Initiative at UTRC, primarily focusing on Deep Learning application in aerospace and building systems to perform anomaly detection, multi-modal sensor fusion, data compression, and design space exploration. He has published over 20 papers, and he is the co-author on 7 pending patents. Kishore earned his Ph.D. in 2012 from University of Central Florida, where he developed advanced video and image analysis algorithms for multiple contracts funded by DARPA, IARPA and NIH.

Keynote 4:

Title: Creating Data-driven Advanced Manufacturing Collaborative Communities

Speaker: Douglas Ramsey, Citrine Informatics

Abstract:

There are several key technologies that are driving a revolution in manufacturing across the world that will reshape how we source, design, and manufacturing everything from aircraft to mobile phones. These four drivers are artificial intelligence (AI), additive manufacturing (AM), collaborative robotics, and smart manufacturing. These technologies are not isolated geographically and will require increased levels of industrial cooperation. The challenge before industry today is how to best coordinate coalitions of innovators that include industry, academia, and government organizations. The US, Germany, China, Japan, and South Korea are all leaders in organizing their technology and innovations ecosystems domestically. However, there are challenges to establishing industrial collaborations due to IP, trade secret, and competitive intelligence concerns. New multisided platforms are allowing for these ecosystems to cross borders and build collaborative partnerships that balance between sharing innovation and developing competitive advantage. International collaborations in all four areas have been successfully demonstrated in North America, Asia, and Europe. However, there continue to be pitfalls and challenges related to conflicting patent regulations and concerns for protecting intellectual property. Coalitions of corporate collaborators can generally navigate these challenges, but that tends to exclude participation of federal research organizations that often have much to offer in terms of technical expertise and pre-competitive research and innovation. Commercially driven multi-sided platforms have the power and promise to act as honest brokers between these parties and creative collaborative test-beds that best leverage the full range of available partners and expertise. The 4th Industrial Revolution is eliminating traditional borders and driving changes that will force changes across both traditional manufacturers and new market entrants. A failure to think broadly about borderless collaboration models will harm not only individual sectors, but also more widely confound the next wave of manufacturing innovation.

Bio:

Douglas Ramsey is the Vice President for Business Development with Citrine Informatics. Citrine is the world leader in Artificial Intelligence (AI) and Materials Informatics (MI) for materials discovery, product design, and manufacturing. Mr. Ramsey has over 25 years of experience working across many manufacturing industries including primary metals, automotive, aerospace, energy, defense, and consumer products. Mr. Ramsey served as the Chairman of the Industrial Control Board for LIFT in Detroit and as a manufacturing technology contributor to the White House as part of the Advanced Manufacturing Partnership (AMP 2.0). He has also held advisory roles with many manufacturing institutes including America Makes, CESMII, and IACME. Mr. Ramsey also served as the Alcoa-Oak Ridge National Laboratory Technologist-In-Resident (TIR) with a focus on innovations in metals manufacturing technology. More recently, Mr. Ramsey helped lead the team that secured a $250m award from the US Department of Defense to establish a new national robotics institute (ARM Institute) in Pittsburgh, PA. Mr. Ramsey holds a MScEcon in Strategic Studies from the National University of Wales, Aberystwyth.

Keynote 5:

Title: Clean Energy Smart Manufacturing Innovation Institute- Transforming Manufacturing, the vision and roadmap

Speaker: Jim Wetzel, CESMII

Abstract:

This session will introduce the efforts of the Clean Energy Smart Manufacturing Innovation Institute (CESMII) operated by the Smart Manufacturing Leadership Coalition sponsored by the Department of Energy. The institute brings together industry, academia and government partners within a growing network of advanced manufacturing institutes, called Manufacturing USA, to increase U.S. manufacturing competitiveness. Smart Manufacturing (SM) is the business, technology, infrastructure, and workforce practice of optimizing manufacturing using engineered systems that integrate operational technologies and information technologies (OT/IT). The CESMII Roadmap will address R&D challenges and knowledge gaps related to the integration of manufacturing OT/IT, including: hardware, software, and security requirements; sensor technologies, multi-sensor data fusion, and sensor-actuator-human interfaces; process model (e.g., physics-based, empirical, data-driven, cognitive, and quantitative) verification, validation, and uncertainty quantification; data structures, contextualization, configuration, and management; and reference architectures and platform for process technology digitization.

Bio:

Jim Wetzel is currently the interim CEO of the Clean Energy Smart Manufacturing Innovation Institute (CESMII). This is a National Institute sponsored by the Department of Energy, and is one of 14 Institutes of Manufacturing USA. CESMII aims to radically accelerate the development and adoption of Smart Manufacturing —including advanced sensors, controls, platforms, and models—to help companies of all sizes across a variety of manufacturing industries to realize the benefits of Smart Manufacturing.

After 32 years at General Mills Inc, Mr. Wetzel retired in August 2017 as Director Engineering–Global Reliability. In this role, he was responsible for improving the existing asset base in GMI Manufacturing Plants across the Globe. This function was responsible for technology, standardization, reliability and maintenance and energy reduction. In addition, he was responsible for the technical mastery, learning and development for all of engineering.

Mr. Wetzel holds a BS Mechanical Engineering and MBA, both from the University of Minnesota.

PANEL- Big Data Analytics and IoT for Advanced Manufacturing: Challenges and opportunities

Moderator: Dr. Sudarsan Rachuri, Federal Program Officer and Technology Manager, Advanced Manufacturing Office, Department of Energy

Panelists:

Dr. Regine Gernert, German Aerospace Center (DLR-PT)

Prof. Sagar Kamarthi, Northeastern University

Prof. Kincho Law, Stanford University

Valri Lightner, Acting Deputy Director, AMO, DOE

Douglas Ramsey, Citrine Informatics

Dr. Kishore K. Reddy, United Technologies Research Center

Jim Wetzel, CESMII

Bios:

Sagar Kamarthi is Professor in the Dept. of Industrial and Mechanical Engineering at Northeastern University, Boston. He is the founding director of MS in Data Analytics Engineering program at Northeastern. He teaches courses in manufacturing, data mining, and machine learning. Prof. Kamarthi received his PhD and MS degrees in Industrial Engineering from The Pennsylvania State University and a BS in Chemical Engineering from Sri Venkateswara University, India.

His research interests are in smart and sustainable manufacturing, predictive analytics for engineering and healthcare applications, and engineering education research. He has published more than 190 articles in internationally reputed journal and conference proceedings and has secured several grants from the National Science Foundation (NSF) and other federal agencies. Through his NSF funded education research grants he co-pioneered Engineering Based Learning (EBL) model (a structured version of project based learning), "Transform" curriculum model to train non-STEM graduates for manufacturing careers, and Mass Customized Instruction (MCI) model to enable personalized learning. Data analytics in engineering education is one of his current interests.

Kincho H. Law is currently Professor of Civil and Environmental Engineering at Stanford University. He obtained his BS in Civil Engineering, and BA in Mathematics from University of Hawaii in 1976, and his MS and PhD in Civil Engineering in 1979 and 1981, respectively, from Carnegie Mellon University. Prof. Law's research has been focused on innovative use of computational and information science in engineering. His work has dealt with various aspects of data analytics and machine learning, smart infrastructures, smart manufacturing, wireless sensing, monitoring and control, high performance computing, engineering and legal information management, Internet and cloud computing. He has authored and co-authored over 400 articles in journals and conference proceedings and has received best research paper awards from ASCE, ASME, IEEE and Digital Government Society. He was the recipient of the ASCE Computing in Civil Engineering Award in 2011. Prof. Law has been elected as a Distinguished Member of the American Society of Civil Engineers and as a Fellow of the American Society of Mechanical Engineering in 2017.

Valri Lightner has 30 years of experience managing technology development for the federal government. She has been with the Department of Energy for 20 years. For the last 3 years she has worked in the Loan Programs Office providing technical management of a $30 billion portfolio including advanced technology vehicle manufacturing, transmission, fossil, nuclear, efficiency and renewable energy projects.

Previously, Mrs. Lightner worked within the DOE's Office of Energy Efficiency and Renewable Energy where she led public-private partnerships in biomass program deployment including biorefinery and infrastructure activities, fuel cell technology development for transportation applications, and pulp and paper energy efficiency development. Mrs. Lightner has a Bachelor's Degree in Chemical Engineering from Villanova University.

Special Sessions

3rd SPECIAL SESSION ON INTELLIGENT DATA MINING

Session Organizer: Uraz YAVANOGLU, PhD

Schedule-12 December 2017 Tuesday		
Time	**Title**	**Presenter/Author**

07:00am-08:00am	**Registration**	
08:00am-08:30am	**Session Keynote Speech** Uraz Yavanoglu, PhD	
08:30am-08:40am	Modeling Self-Service Machine-Learning Agents for Distributed Stream Processing	Philipp Zehnder and Dominik Riemer
08:40am-08:50am	Using Meta-learning for Model Type Selection in Predictive Big Data Analytics	Mustafa Nural, Hao Peng, and John Miller
08:50am-09:00am	A Distributed Proximal Gradient Descent Method for Tensor Completion	Thomas Papastergiou and Vasilis Megalooikonmou
09:00am-09:10am	Mined Semantic Analysis: A New Concept Space Model for Semantic Representation of Textual Data	Walid Shalaby and Wlodek Zadrozny
09:10am-09:20am	Predicting High Taxi Demand Regions Using Social Media Check-ins	Xuefeng Peng, Yiming Pan, and Jiebo Luo
09:20am-09:30am	Sleep-deprived Fatigue Pattern Analysis using Large-Scale Selfies from Social Media	Xuefeng Peng, Jiebo Luo, Catherine Glenn, Li-Kai Chi, and Jingyao Zhan
09:30am-09:40am	One-shot Learning for Fine-grained Relation Extraction via Convolutional Siamese Neural Network	Jianbo Yuan, Han Guo, Zhiwei Jin, Hongxia Jin, Xianchao Zhang, and Jiebo Luo
09:40am-09:50am	Understanding What affects Career Progression Using LinkedIn and Twitter Data	Yiming Pan, Xuefeng Peng, Tianran Hu, and Jiebo Luo
09:50am-10:00am	A Filter-based Feature Selection Model for Anomaly-based Intrusion Detection Systems	Imtiaz Ullah and Qusay Mahmoud
10:00am-10:10am	A Hybrid Model for Anomaly-based Intrusion Detection in SCADA Networks	Imtiaz Ullah and Qusay Mahmoud
10:10am-10:20am	Big Data Impact on Stability and Reliability Improvement of Smart Grid	Shady Khalil, Amira Mohamed, and Haitham Abu-Rub
10:20am-10:30am	Event Detection from Time-Series Streams Using Directional Change and Dynamic Thresholds	Nora Alkhamees and Maria Fasli
10:30am-10:40am	A Review on Cyber Security Datasets for Machine Learning Algorithms	Ozlem Yavanoglu and Murat Aydos
10:40am-11:00am	**Break**	
11:00am-11:15am	ClusTop: A Clustering-based Topic Modelling Algorithm for Twitter using Word Networks	Kwan Hui Lim, Shanika Karunasekera, and Aaron Harwood
11:15am-11:30am	Graph-based Information Exploration over Structured and Unstructured Data	Giannis Koumoutsos, Maria Fasli, Ian Lewin, and David Milward
11:30am-11:45am	A Natural Language Normalization Approach to Enhance Social Media Text Reasoning	Long Nguyen, Andrew Salopek, Liang Zhao, and Fang Jin
11:45am-12:00pm	Recovering Loss to Followup Information Using Denoising Autoencoders	Lovedeep Gondara and Ke Wang
12:00pm-12:15pm	What's Trending Tomorrow, Today: Using Early Adopters to Discover Popular Posts on Tumblr	Daniel Xie, Jiejun Xu, and Tsai-Ching Lu
12:15pm-12:30pm	Towards MapReduce based Bayesian Deep Learning Network for Monitoring Big Data Applications	Omair Shafiq

12:30pm-12:45pm	Effects of Language Processing in Turkish Authorship Attribution	Hayri Volkan Agun, Sibel Yilmazel, and Ozgur Yilmazel
12:45pm-02:00pm	**Lunch Break**	
02:00pm-02:15pm	Genomic Variant Analysis with Big Data Technologies	Tuğçe Döngel and Yasemin Timar
02:15pm-02:30pm	A Recommender Model Based on Trust Value and Time Decay	Muhittin Isık and Hasan Dag
02:30pm-02:45pm	Mathematical Programming for Social Network Analysis	Harun Pirim
02:45pm-03:00pm	Harvey Flooding Rescue in Social Media	Zhou Yang, Long Nguyen, Joshua Stuve, Guofeng Cao, and Fang Jin
03:00pm-03:15pm	Weather Data Analysis and Sensor Fault Detection Using An Extended IoT Framework with Semantics, Big Data, and Machine Learning	Aras Can Onal, Omer Berat Sezer, Ahmet Murat Ozbayoglu, and Erdogan Dogdu,
03:15pm-03:30pm	SpEnD Portal: Linked Data Discovery using SPARQL Endpoints	Semih Yumusak, Riza Emre Aras, Elif Uysal, Erdogan Dogdu, Halife Kodaz, and Kasim Oztoprak
03:30pm-03:45pm	Estimation of Parameters for Free-form Machining with Deep Neural Network	Gokberk Serin, Mehmet Ugur Gudelek, Ahmet Murat Ozbayoglu, and Hakkı Ozgur Unver,
03:45pm-04:00pm	Real-Time Lexicon-Based Sentiment Analysis Experiments On Twitter With A Mild (More Information, Less Data) Approach	Yusuf Arslan, Aysenur Birturk, Bekjan Djumabaev, and Dilek Küçük
04:00pm-04:30pm	**Break**	
04:30pm-04:40pm	A Comparative Study on Learning to Rank with Computational Methods	İnci Batmaz, Pinar Karagoz, and Gulsah Serdar
04:40pm-04:50pm	Semi-supervised learning and Social Media Text Analysis towards multi-labeling categorization	Billal Belainine, Alexsandro Fonseca, Fatiha Sadat, and Hakim Lounis
04:50pm-05:00pm	A deep learning model for air quality prediction in smart cities	Ibrahim Kok, Mehmet Ulvi Simsek, and Suat Ozdemir
05:00pm-05:10pm	A Data-Driven Approach to Help Understanding the Preferences of Public Transport Users	Vasco Furtado, Elizabeth Furtado, Carlos Caminha, André Lopes, Caio Pontes, Victor Dantas, and Sofia Cavalcante
05:10pm-05:20pm	Focus Location Extraction from Political News Reports with Bias Correction	Maryam Bahojb Imani, Swarup Chandra, Samuel Ma, Latifur Khan, and Bhavani Thuraisingham
05:20pm-05:30pm	DxNAT - Deep Neural Networks for Explaining Non-Recurring Traffic Congestion	Fangzhou Sun, Abhishek Dubey, and Jules White
05:30pm-05:40pm	Convolutional Neural Network for Clinical Narrative Categorization	Paula Lauren, Guangzhi Qu, and Paul Watta
05:40pm-05:50pm	Unsupervised Deep Learning for Subspace Clustering	Ali Sekmen, Ahmet Bugra Koku, Mustafa Parlaktuna, Ayad Abdul-Malek, and Nagendrababu Vanamala
05:50pm-06:00pm	Principal Coordinate Clustering	Ali Sekmen, Akram Aldroubi, Ahmet Bugra Koku, and Keaton Hamm
06:00pm-06:10pm	Augmenting Word Embeddings through External Knowledge-base for Biomedical Application	Kishlay Jha, Guangxu Xun, Vishrawas Gopalakrishnan, and Aidong Zhang
06:10pm-06:20pm	Online Video Ad Measurement for Political Science Research	Adisak Sukul, Baskar Gopalakrishnan, Wallapak Tavanapong, and David A.M. Peterson

Special Session on Information Granulation in Data Science and Scalable Computing *Special Session Organizers: Shusaku Tsumoto, Dominik Slezak, Tzung-Pei Hong and Shyue-Liang Wang*		
Time	**Title**	**Presenter/Author**
14:00-14:10	Introduction	Shusaku Tsumto
14:10-14:30	Secure Information Flow and File Movements	T.Y.Lin
14:30-14:50	Scalable Cyber-Security Analytics with a New Summary-based Approximate Query Engine	Dominik Slezak, Agnieszka Chadzynska-Krasowska, Joel Holland, Piotr Synak, Rick Glick, and Marcin Perkowski
14:50-15:10	Big-Data-Enabled Modelling and Optimization of Granular Speed-based Vessel Schedule Recovery Problem	Fatemeh Cheraghchi, Ibrahim Abualhaol, Rafael Falcon, Rami Abielmona, Bijan Raahemi, and Emil Petriu
15:10-15:30	A Preliminary Study on Deep Learning for Predicting Social Insurance Payment Behavior	Jia-Ching Ying, Po-Yu Huang, Chih-Kai Chang, and Don-Lin Yang
15:30-15:50	Unsupervised Deep Embedding for Novel Class Detection over Data Stream	Ahmad Mustafa and Latifur Khan
	Coffee Break	
16:20-16:40	Improving Text Classification with Word Embedding	Lihao Ge and Teng Moh
16:40-17:00	Mining Text for Disease Diagnosis in Hospital Information System	Shusaku Tsumoto, Shoji Hirano, and Tomohiro Kimura
17:00-17:20	Quasi-Erasable Itemset Mining	Tzung-Pei Hong, Lu-Hung Chen, Shyue-Liang Wang, Chun-Wei Lin, and Bay Vo
17:20-17:40	Noise Self-Filtering k-Nearest Neighbors Algorithms	Shuyin Xia, Guoyin Wang, Yunsheng Liu, Qun Liu, and Hong Yu
17:40-18:00	On the Role of Feature Space Granulation in Feature Selection Processes	Marek Grzegorowski, Andrzej Janusz, Dominik Slezak, and Marcin Szczuka
18:00-18:20	Data treatment from the viewpoint of granular computing	Akinori Abe and Yuki Hayashi
	Closing Remarks	

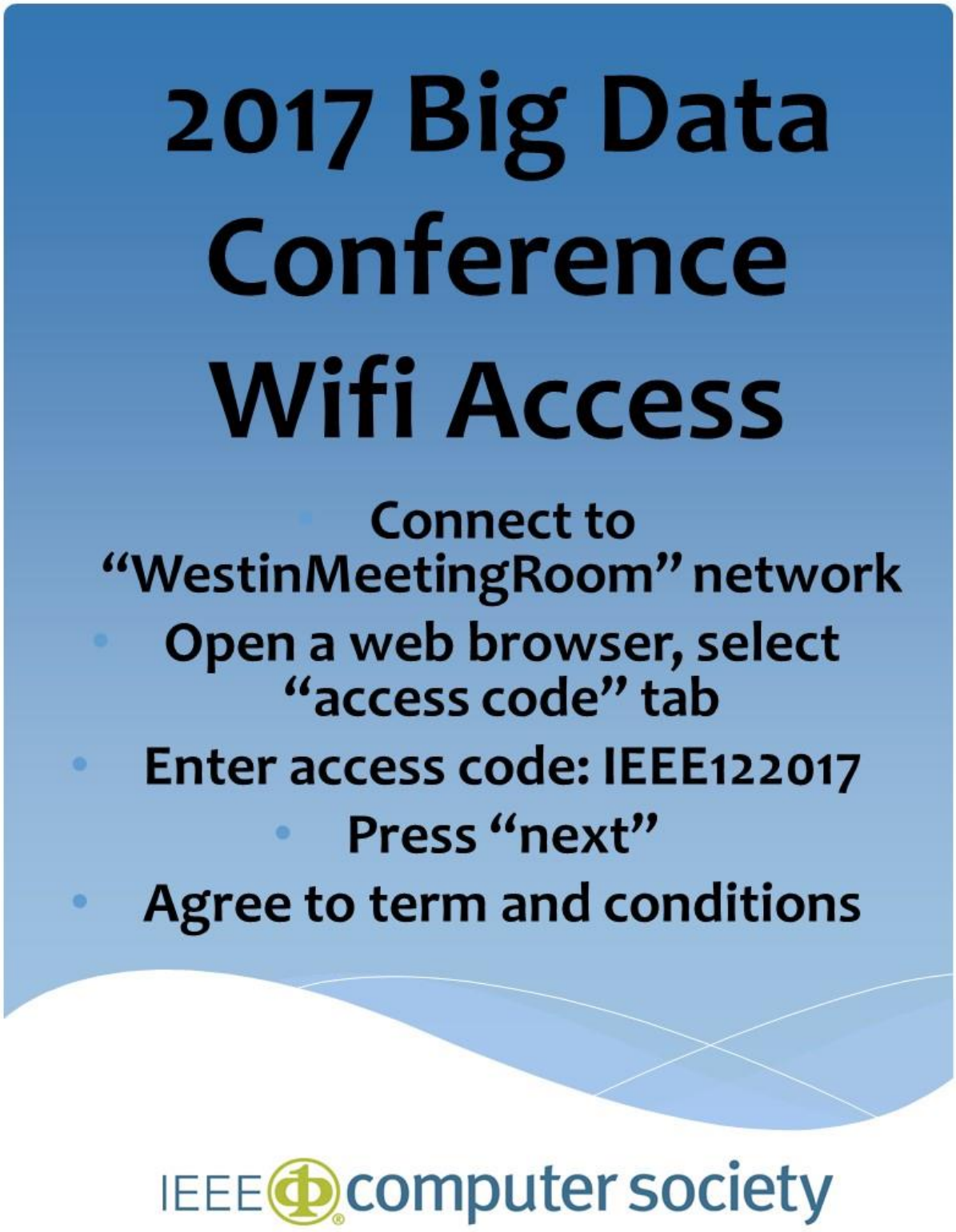
Posters

Paper ID	Accept Posters
P201	Monika Nawrocka and Marcin Łukowski, *Biofeedback EEG Data Integration and Visualization Analytics for Endurance Exercise Practices*
P203	Shilpa Balan, Nishant Shristiraj, Vrunda Shah, and Anusha Manjappa, *Big Data Analysis of Youth Tobacco Smoking Trends in the United States*
P204	Binyam Zemede and Byron Gao, *Personalized Search with Editable Profiles*
P208	Jonathan Wang, Alex Sim, Kesheng Wu, and Seongwook Hwangbo, *Accurate Signal Timing from High Frequency Streaming Data*

P214	Mohammed Eslami, George Zheng, Hamed Eramian, and Georgiy Levchuk, *Anomaly Detection on Bipartite Graphs for Cyber Situational Awareness and Threat Detection*
P215	Haiyan Yu, Kun Xiang, and Jiang Yu, *Understanding a Moderating Effect of Physicians' Endorsement to Online Workload*
P220	Kwan Hui Lim, Shanika Karunasekera, Aaron Harwood, and Lucia Falzon, *Spatial-based Topic Modelling using Wikidata Knowledge Base*
P221	Yin Zhang and Jiming Hu, *Discovering the Interdisciplinary Nature of Big Data Research*
P223	Hyun-Chul Lee, Tong-Il Jang, and Kwangsu Moon, *Anticipating Human Errors from Periodic Big Survey Data in Nuclear Power Plants*
P225	Mauri Kaipainen and Olli Pitkänen, *User-directed stepwise cluster analysis method*
P227	Akira Umayabara and Hayato Yamana, *MCMalloc: A Scalable Memory Allocator for Multithreaded Applications on a Many-Core Shared-Memory Machine*
P228	Yoshiko Yasumura, Hiroki Imabayashi, and Hayato Yamana, *Attribute-based Proxy Re-encryption Method for Revocation in Cloud Data Storage*
P229	Iwao Fujino, *Extracting Route Patterns of Vessels from AIS Data by Using Topic Model*
P230	Chaochao Chen, Xinxing Yang, Li Wang, Jun Zhou, and Xiaolong Li, *Large Scale App Recommendation in Ant Financial*
P232	Tayfun Pay and Stephen Lucci, *Automatic Keyword Extraction An Ensemble Method*
P233	Javier Mata, Ignacio de Miguel, Ramón J. Durán, Juan Carlos Aguado, Noemí Merayo, Lidia Ruiz, Patricia Fernández, Rubén M. Lorenzo, and Evaristo J. Abril, *A SVM Approach for Lightpath QoT Estimation in Optical Transport Networks*
P235	Frank Greguska, Thomas Huang, Brian Wilson, Nga Quach, and Joe Jacob, *Analyzing Big Ocean Science Data with NEXUS*
P236	Li-Chen Cheng, *Applied deep learning in fake review detection and dataset generation*
P237	Kenji Nakashima, Joichiro Kon, Gil Lee, Jose Fortes, and Saneyasu Yamaguchi, *A Study on Big Data I/O Performance with Modern Storage Systems*
P239	Alec Parise, Kaine Black, and Monica Wachowicz, *Using Bipartite Graphs to Cluster Complex Networks*
P240	Shaunak Bopardikar and George Ekladious, *Towards Scalable Kernel Machines For Streaming Data Analytics*
P242	Michel Généreux, Bryor Snejfella, and Marta Maslej, *Big data in psychology: using word embeddings to study theory-of-mind*
P243	Iulia Popescu, Kurt Portelli, Christos Anagnostopoulos, and Nikos Ntarmos, *The Case for Graph-based Recommendations*
P245	Chen Li, Annisa Annisa, Asif Zaman, and Yasuhiko Morimoto, *MapReduce-Based Computation of Area Skyline Query for Selecting Good Locations in a Map*
P246	Kasumi Kato, Atsuko Takefusa, Hidemoto Nakada, and Masato Oguchi, *Consideration of Parallel Data Processing over an Apache Spark Cluster*
P247	Nat Pavasant, Hiroshi Furutani, Masayuki Numao, and Ken-ichi Fukui, *ART-2b: Adapted ART-2a for large scale data clustering on PM2.5 mass spectra*
P248	Mohammed Elshambakey, Mohamed Khalefa, William J. Tolone, Sreyasee Das Bhattacharjee, Huikyo Lee, Luca Cinquini, Shannon Schlueter, Isaac Cho, Wenwen Dou, and Daniel J. Crichton, *Towards a Distributed Infrastructure for Data-Driven Discoveries & Analysis*
P249	Yasuko Kawahata, Yukari Moriyama, Shinichirou Yamada, Mingyi Sun, and Taketo Kawamura, *Analytical the Large-scale Collection of Data on the Results of the Guides for Foreigners Visiting Japan*
P250	Sebastian Trinks and Carsten Felden, *Real Time Analytics - State of the art*
P252	Lixin Liu and Jun Chen, *The influences of deep-sea vision data quality on observational analysis*
P253	Shohei Shirataki and Saneyasu Yamaguchi, *A Study on Interpretability of Decision of Machine Learning*
P255	Mark Simmons, Daniel Armstrong, Dylan Soderman, and Michael Gubanov, *Hybrid.media: High Velocity Video Ingestion in an In-Memory Scalable Analytical Polystore*
P256	Philipp Zehnder and Dominik Riemer, *Towards Automatic Infrastructure Provisioning for Highly Dynamic Streaming Applications*
P257	Amin Majd and Elena Troubitsyna, *Data-Driven Approach to Ensuring Fault Tolerance and Efficiency of Swarm Systems*

P259	Santiago Villasenor, Tom Nguyen, Anusha Kola, Sean Soderman, and Michael Gubanov, *Scalable Spam Classifier for Web Tables*
P260	Anusha Kola, Harshal More, Sean Soderman, and Michael Gubanov, *Generating Unified Famous Objects (UFOs) from the Classified Object Tables*
P262	Masanori Ajito, Yasuko Kawahata, and Akira Ishii, *Analysis of National Election Using Mathematical Model of Hit Phenomenon*
P264	Yifang Wei and Lisa Singh, *Understanding the Impact of Sampling and Noise on Detecting Events Using Twitter*
P265	Yiheng Liang and Prathyusharani Merla, *Data Analysis using Hadoop MapReduce Environment*
P266	Bharath K. Samanthula, *Privacy-Preserving Outsourced Collaborative Frequent Itemset Mining in the Cloud*
P267	Tai Yeon Ku, Wan-Ki Park, and Hoon Choi, *Energy information collection mechanism Using big data correlation map*
P268	Takuya Yonezawa, Ismail Arai, Toyokazu Akiyama, and Kazutoshi Fujikawa, *Proposal of classification method of bus operation states using sensor data*
P270	Jason Radford, Luke Horgan, and David Lazer, *Baselines for Demographic Inference on a New Gold Standard Twitter Corpus*
P272	Paul Le Noac'h, Alexandru Costan, and Luc Bougé, *A Performance Evaluation of Apache Kafka in Support of Big Data Streaming Applications*
P273	Jason Radford, *Piloting A Theory-based Approach to Inferring Gender in Big Data*
P275	S Khanna, Y Sethi, and A Nambiar, *iSkin Specialist – A Big Data Based Expert System for Dermatology*
P276	Steven Ortiz, Caner Enbatan, Maksim Podkorytov, Dylan Soderman, and Michael Gubanov, *Hybrid.JSON: High-velocity Parallel In-Memory Polystore JSON Ingest*
P277	Daisaku Yokoyama and Masashi Toyoda, *Towards Constructing a Driver Management System Based on Large-scale Driving Operation Records*
P278	Ziwei Zhu, Weijia Xu, and Wei He, *Big Data System for Information Aggregation and Model Comparison for Precision Medicine*
P279	Masashi Toyoda, Daisaku Yokoyama, Junpei Komiyama, and Masahiko Itoh, *Road Safety Estimation Utilizing Big and Heterogeneous Vehicle Recorder Data>*
P280	Darlan Arruda and Nazim H. Madhavji, *Towards a Big Data Requirements Engineering Artefact Model in the Context of Big Data Software Development Projects: Poster Extended Abstract*
P281	Abdeltawab Hendawi, Aqeel Rustum, Mohamed Ali, and John Stankovic, *Turning Big Spatial Data Into Smart Routing*
P282	Tsumugi Tairaku, Akihiro Nakao, Shu Yamamoto, Saneyasu Yamaguchi, and Masato Oguchi, *Application Specific Traffic Control in Large-Scale Disasters*
P285	Thomas Kitson, Paula Olaya, Elizabeth Racca, Michael Wyatt, Mario Guevara, Rodrigo Vargas, and Michela Taufer, *Data Analytics for Modeling Soil Moisture Patterns across United States Ecoclimatic Domains*
P286	Lisa Singh and Raghu Pemmaraju, *EOS: A Multilingual Text Archive of International Newspaper & Blog Articles*
P287	Ranjeet Devarakonda, Michael Giansiracusa, Jitendra Kumar, and Harold Shanafield, *Social Media Based NPL System to Find and Retrieve ARM Data: Concept Paper*

Conference Wifi Access



Westin Copley Place, Boston Floor Plan

FLOOR PLANS
CAPACITY CHARTS

THE WESTIN
COPLEY PLACE
BOSTON

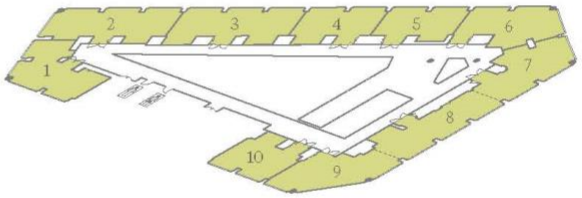
MEETING SPACE
OVERVIEW, PART ONE

10 Huntington Avenue
Boston, Massachusetts 02116
T 617.262.9600 F 617.424.7483
E westincopley@westin.com
westincopleyplaceboston.com

THE WESTIN
COPLEY PLACE
BOSTON

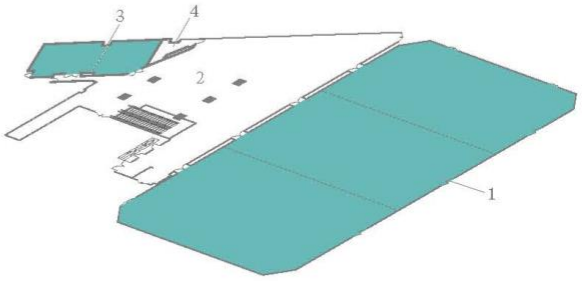
SEVENTH FLOOR

1. DEFENDER
2. EMPIRE
3. GREAT REPUBLIC
4. HELICON
5. MASTIFF
6. NORTHSTAR
7. PARLIAMENT
8. ADAMS
9. BALTIC
10. COURIER



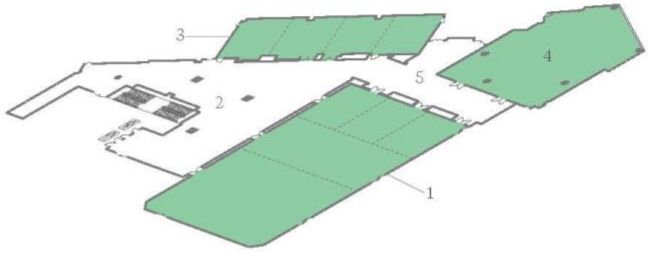
FOURTH FLOOR

1. AMERICA BALLROOM
2. AMERICA BALLROOM FOYER
3. INDEPENDENCE
4. CONVENTION OFFICE



THIRD FLOOR

1. ESSEX BALLROOM
2. ESSEX BALLROOM FOYER
3. ST. GEORGE
4. STAFFORDSHIRE
5. STAFFORDSHIRE FOYER



ROOMS AT A GLANCE

TOTAL GUEST ROOMS	803
TOTAL MEETING ROOMS	32
LARGEST MEETING ROOM CAPACITY	2,000
LARGEST MEETING ROOM SIZE	1,425 m^2 / 15,337 ft^2

Additional spaces for meetings and events, not displayed here, may also be available. Contact your hotel representative for more information.

This document contains approximate measurements and square footage that are for illustrative purposes only. We cannot guarantee the floor plan accuracy or completeness, therefore encourage you to review the space to make sure it is suitable for your event.

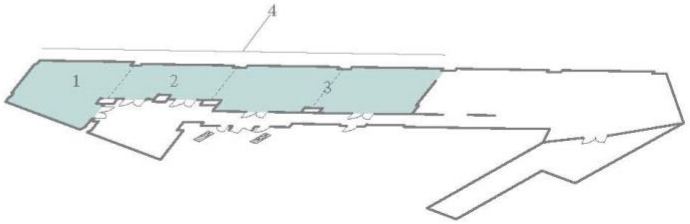
MEETING SPACE OVERVIEW, PART TWO

10 Huntington Avenue
Boston, Massachusetts 02116
T 617.262.9600 F 617.424.7483
E westincopley@westin.com
westincopleyplaceboston.com

THE WESTIN
COPLEY PLACE
BOSTON

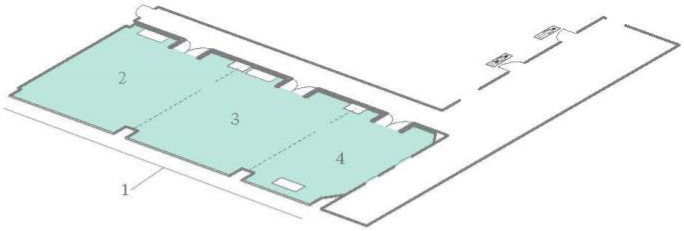
SECOND FLOOR

1. NEWBURY
2. GLOUCESTER
3. HUNTINGTON
4. HUNTINGTON HALL



GROUND FLOOR

1. HARBOUR/IPSWICH/ROCKPORT
2. HARBOUR
3. IPSWICH
4. ROCKPORT



ROOMS AT A GLANCE

TOTAL GUEST ROOMS	803
TOTAL MEETING ROOMS	32
LARGEST MEETING ROOM CAPACITY	2,000
LARGEST MEETING ROOM SIZE	1,425 m^2 / 15,337 ft^2

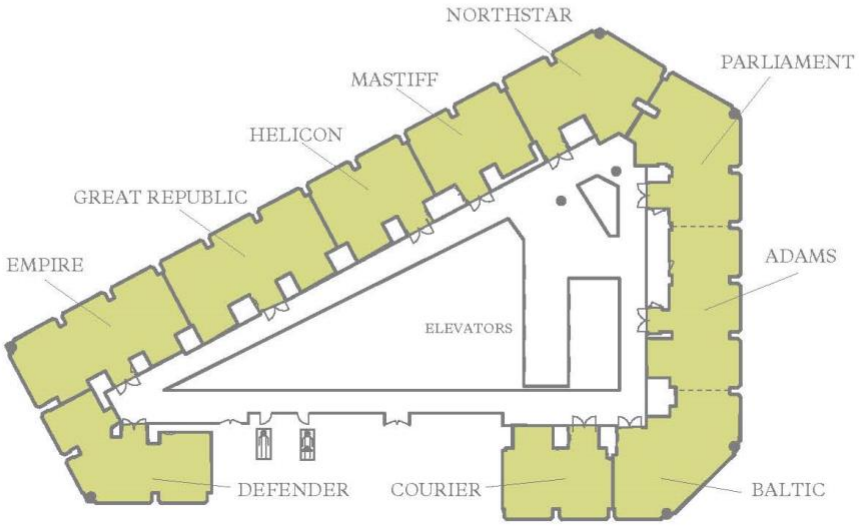
Additional spaces for meetings and events, not displayed here, may also be available. Contact your hotel representative for more information.

This document contains approximate measurements and square footage that are for illustrative purposes only. We cannot guarantee the floor plan accuracy or completeness, therefore encourage you to review the space to make sure it is suitable for your event.

MEETING SPACE
SEVENTH FLOOR

10 Huntington Avenue
Boston, Massachusetts 02116
T 617.262.9600 F 617.424.7483
E westincopley@westin.com
westincopleyplaceboston.com

THE WESTIN
COPLEY PLACE
BOSTON



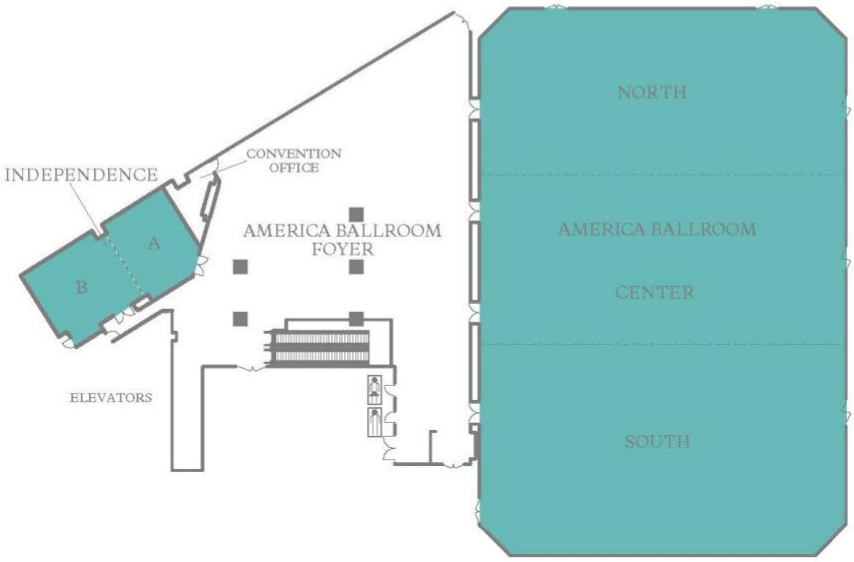
	DIMENSIONS	AREA (SQ. FT.)	MAX CEILING HEIGHT	THEATER	THEATER CHEVRON	CLASSROOM	CLASSROOM CHEVRON	BANQUET	BANQUET CHEVRON	CONFERENCE	U-SHAPE	HOLLOW SQ.	RECEPTION
COURIER	24' X 28'	615	8.3'	41	—	26	—	41	—	20	20	22	41
DEFENDER	30' X 45'	797	8.2'	59	—	41	—	59	—	28	23	28	59
EMPIRE	23' X 45'	902	8.2'	64	—	49	—	64	—	40	40	40	64
GREAT REPUBLIC	24' X 43'	939	8.2'	64	—	46	—	64	—	38	38	30	64
HELICON	24' X 28'	617	8.3'	43	—	26	—	43	—	16	22	22	43
MASTIFF	24' X 28'	606	8.2'	—	—	—	—	43	—	14	—	—	63
NORTH STAR	29' X 40'	813	8.2'	56	—	30	—	56	—	28	24	24	56
PARLIAMENT/ADAMS/BALTIC	34' X 117'	2,529	8.3'	—	—	—	—	90	—	90	87	84	175
ADAMS	24' X 43'	891	8.3	63	40	42	30	40	50	36	33	30	64
ADAMS/BALTIC	34' X 77'	1,697	8.3'	102	—	66	—	70	—	54	45	48	119
BALTIC	33' X 34'	806	8.3'	56	—	39	—	60	60	18	15	22	55
PARLIAMENT	30' X 40'	833	8.3'	51	—	27	—	30	—	24	21	24	56
PARLIAMENT/ADAMS	30' X 83'	1,723	8.3'	120	—	60	—	70	—	66	69	66	120

This document contains approximate measurements and square footage that are for illustrative purposes only. We cannot guarantee the floor plan accuracy or completeness, therefore encourage you to review the space to make sure it is suitable for your event.

MEETING SPACE
FOURTH FLOOR

10 Huntington Avenue
Boston, Massachusetts 02116
T 617.262.9600 F 617.424.7483
E westincopley@westin.com
westincopleyplaceboston.com

THE WESTIN
COPLEY PLACE
BOSTON



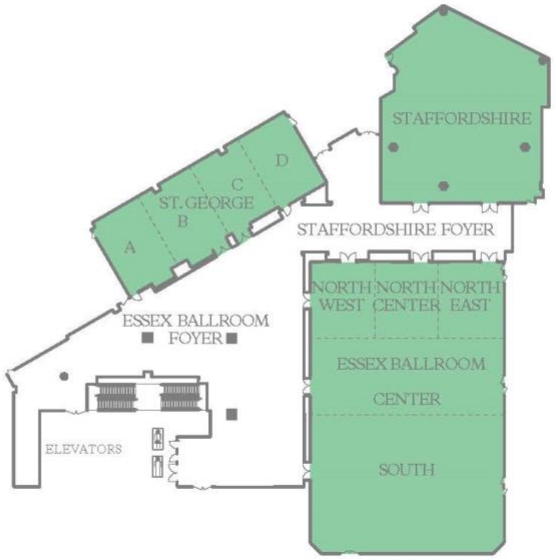
	DIMENSIONS	AREA (SQ. FT.)	MAX. CEILING HEIGHT	THEATER	THEATER CHEVRON	CLASSROOM	CLASSROOM CHEVRON	BANQUET	BANQUET CHEVRON	CONFERENCE	U-SHAPE	HOLLOW SQ.	RECEPTION
▪ AMERICA BALLROOM	102' X 152'	15,329	16.3'	1,800	—	970	—	1,300	—	—	—	—	2,000
AMERICA BALLROOM FOYER	128' X 106'	6,668	12'	—	—	—	—	—	—	—	—	—	600
▪ AMERICA CENTER	102' X 46'	4,710	16.3'	450	—	250	—	320	—	88	70	80	500
▪ AMERICA CENTER & NORTH	102' X 93'	9,398	16.3'	1,012	950	609	540	710	800	90	120	150	989
▪ AMERICA NORTH	102' X 46'	4,662	16.3'	500	—	252	—	360	—	90	70	80	500
▪ AMERICA SOUTH	102' X 59'	5,922	16.3'	500	—	306	—	450	—	90	78	90	600
▪ AMERICA SOUTH & CENTER	102' X 106'	10,668	16.3'	1,332	1,102	711	588	900	1,250	90	126	162	1,122
▪ CONVENTION OFFICE	16' X 22'	176	9.6'	—	—	—	—	—	—	—	—	—	43
▪ INDEPENDENCE	28' X 50'	1,308	10'	90	—	70	—	80	—	32	30	36	120
▪ INDEPENDENCE A	28' X 23'	608	10'	60	—	40	—	50	—	26	24	28	70
▪ INDEPENDENCE B	27' X 27'	700	10'	50	—	30	—	40	—	24	22	28	50

This document contains approximate measurements and square footage that are for illustrative purposes only. We cannot guarantee the floor plan accuracy or completeness, therefore encourage you to review the space to make sure it is suitable for your event.

MEETING SPACE
THIRD FLOOR

10 Huntington Avenue
Boston, Massachusetts 02116
T 617.262.9600 F 617.424.7483
E westincopley@westin.com
westincopleyplaceboston.com

THE WESTIN
COPLEY PLACE
BOSTON



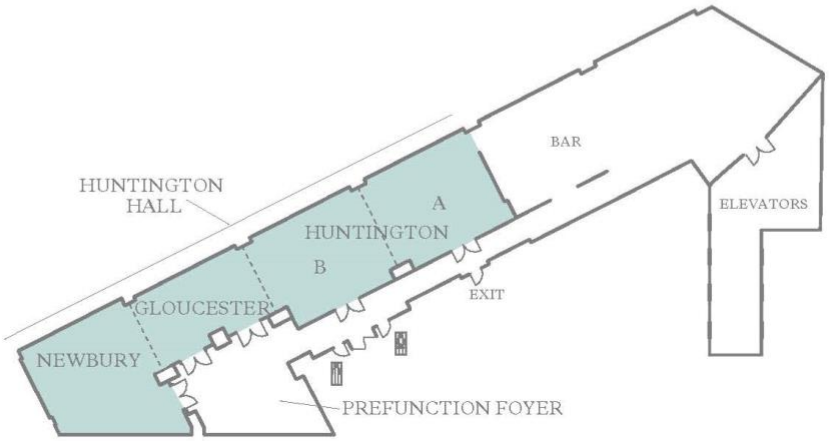
	DIMENSIONS	AREA (SQ. FT.)	MAX. CEILING HEIGHT	THEATER	THEATER CHEVRON	CLASSROOM	CLASSROOM CHEVRON	BANQUET	BANQUET CHEVRON	CONFERENCE	U-SHAPE	HOLLOW SQ.	RECEPTION
ESSEX BALLROOM	74' X 111'	8,136	14.3'	850	—	428	—	590	—	—	—	—	1,030
ESSEX BALLROOM FOYER	110' X 116'	5,931	11.2'	—	—	—	—	—	—	—	—	—	600
ESSEX CENTER	74' X 29'	2,135	14.3'	200	—	100	—	170	—	66	60	76	270
ESSEX CENTER & NORTH	74' X 58'	4,262	14.3'	459	364	273	240	300	320	60	75	90	448
ESSEX CENTER & SOUTH	74' X 82'	6,009	14.3'	638	544	390	300	420	490	66	87	114	632
ESSEX NORTH	74' X 29'	2,127	14.3'	216	144	126	90	140	200	66	60	72	270
ESSEX NORTH CENTER	24' X 29'	706	14.3'	50	—	40	—	50	—	24	28	32	80
ESSEX NORTH CENTER & NORTH EAST	49' X 29'	1,422	14.3'	110	88	72	63	80	90	42	42	48	149
ESSEX NORTH CENTER & NORTH WEST	49' X 29'	1,411	14.3'	144	—	72	—	80	100	42	45	48	148
ESSEX NORTH EAST	25' X 29'	715	14.3'	50	—	40	—	50	—	24	28	30	80
ESSEX NORTH WEST	25' X 29'	707	14.3'	50	—	40	—	50	—	24	28	30	80
ESSEX SOUTH	74' X 53'	3,874	14.3'	380	—	200	—	240	—	64	60	78	490
ST. GEORGE	34' X 87'	2,721	11.1'	200	—	140	—	220	—	72	69	72	300
ST. GEORGE A	34' X 22'	718	11.1'	50	—	40	—	50	—	24	28	32	80
ST. GEORGE A - B	34' X 45'	1,435	11.1'	100	—	80	—	110	—	36	38	48	160
ST. GEORGE A - C	34' X 66'	2,063	11.1'	210	152	72	72	100	170	48	54	54	217
ST. GEORGE B	34' X 23'	717	11.1'	50	—	40	—	50	—	24	28	32	80
ST. GEORGE B - C	34' X 43'	1,345	11.1'	100	—	75	—	110	—	36	37	48	160
ST. GEORGE B- D	34' X 64'	2,003	11.1'	180	140	72	81	180	150	54	54	54	210
ST. GEORGE C	32' X 21'	632	11.1'	49	—	30	—	40	40	24	24	24	75
ST. GEORGE C - D	32' X 42'	1,289	11.1'	100	—	75	—	110	—	36	37	46	160
ST. GEORGE D	32' X 21'	658	11.1'	50	—	40	—	50	—	24	28	30	80
STAFFORDSHIRE	75' X 64'	3,937	13.2'	270	—	165	—	300	—	54	62	78	400
STAFFORDSHIRE FOYER	50' X 81'	2,111	10.5'	—	—	—	—	—	—	—	—	—	222

This document contains approximate measurements and square footage that are for illustrative purposes only. We cannot guarantee the floor plan accuracy or completeness, therefore encourage you to review the space to make sure it is suitable for your event.

MEETING SPACE
SECOND FLOOR

10 Huntington Avenue
Boston, Massachusetts 02116
T 617.262.9600 F 617.424.7483
E westincopley@westin.com
westincopleyplaceboston.com

THE WESTIN
COPLEY PLACE
BOSTON



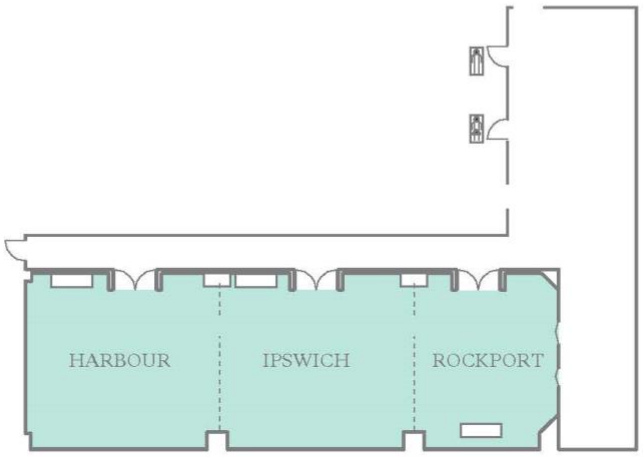
	DIMENSIONS	AREA (SQ. FT.)	MAX CEILING HEIGHT	THEATER	THEATER CHEVRON	CLASSROOM	CLASSROOM CHEVRON	BANQUET	BANQUET CHEVRON	CONFERENCE	U-SHAPE	HOLLOW SQ.	RECEPTION
HUNTINGTON HALL (NEWBURY/GLOUCESTER/HUNTINGTON)	35' X 116'	2,764	12'	—	—	108	—	180	190	84	—	—	271
GLOUCESTER	20' X 29'	506	12'	48	—	24	—	48	—	16	14	16	48
GLOUCESTER/HUNTINGTON	25' X 87'	1,925	12'	110	—	—	—	120	—	—	—	—	140
HUNTINGTON	25' X 59'	1,374	12'	100	—	—	—	72	—	—	—	—	100
HUNTINGTON A	25' X 29'	686	12'	50	—	36	—	36	—	18	18	22	50
HUNTINGTON B	25' X 29'	688	12'	50	—	36	—	36	—	18	18	22	50
NEWBURY	35' X 28'	768	12'	48	—	36	—	50	—	20	24	24	48
NEWBURY/GLOUCESTER	35' X 57'	1,262	12'	90	—	50	—	112	—	40	40	40	112
PREFUNCTION FOYER	31' X 102'	1,136	12'	—	—	—	—	—	—	—	—	—	122

This document contains approximate measurements and square footage that are for illustrative purposes only. We cannot guarantee the floor plan accuracy or completeness, therefore encourage you to review the space to make sure it is suitable for your event.

MEETING SPACE
GROUND FLOOR

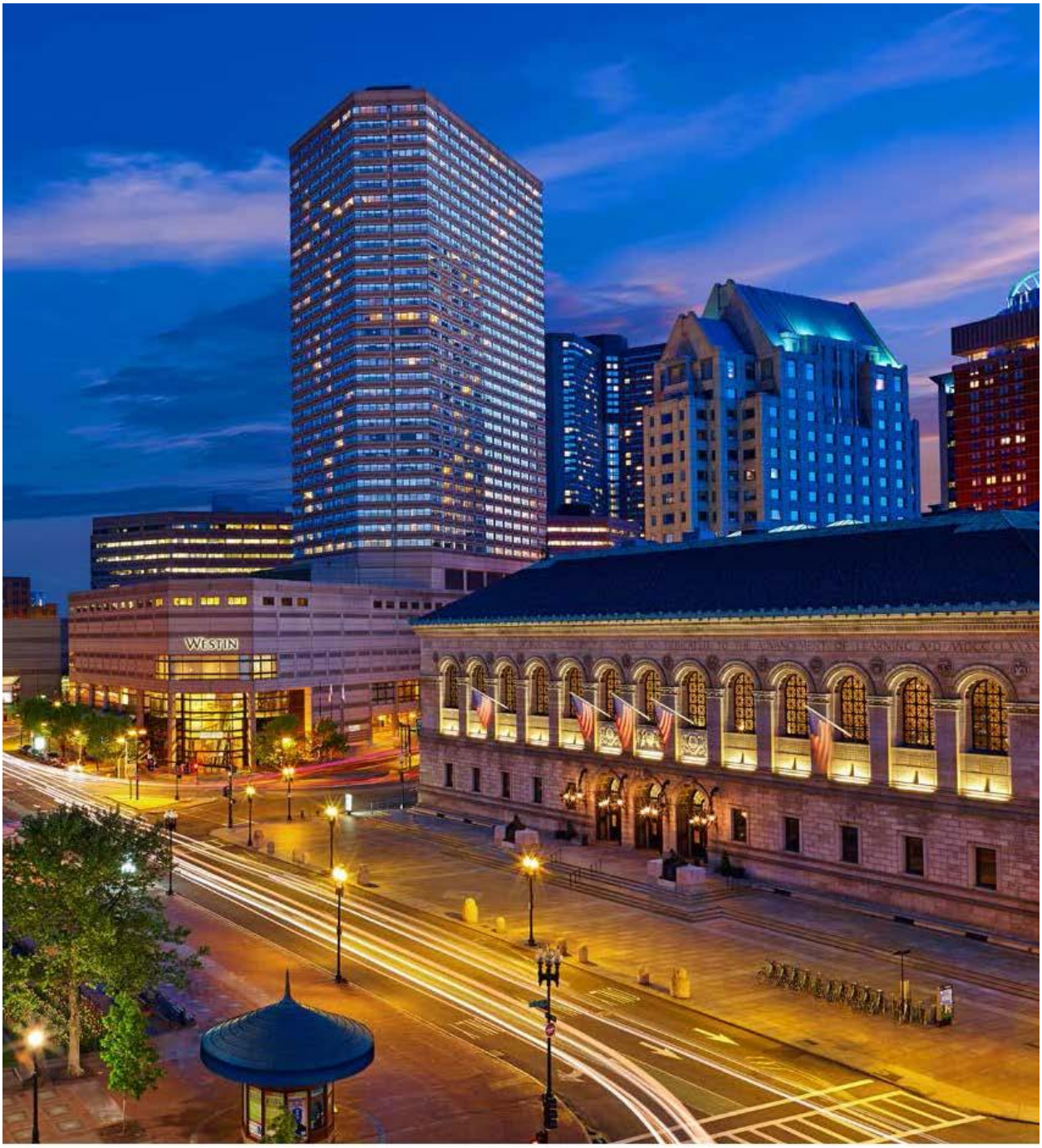
10 Huntington Avenue
Boston, Massachusetts 02116
T 617.262.9600 F 617.424.7483
E westincopley@westin.com
westincopleyplaceboston.com

THE WESTIN
COPLEY PLACE
BOSTON



	DIMENSIONS	AREA (SQ. FT.)	MAX CEILING HEIGHT	THEATER	THEATER CHEVRON	CLASSROOM	CLASSROOM CHEVRON	BANQUET	BANQUET CHEVRON	CONFERENCE	U-SHAPE	HOLLOW SQ.	RECEPTION
HARBOUR/IPSWICH/ROCKPORT	26' X 79'	1,880	10'	—	—	—	—	130	—	—	—	—	278
HARBOUR	26' X 28'	688	10'	55	—	30	—	50	40	22	22	20	106
IPSWICH	26' X 29'	705	10'	50	—	30	—	50	40	20	20	18	102
IPSWICH/HARBOUR	26' X 57'	1,393	10'	117	78	66	42	100	90	42	48	48	146
ROCKPORT	26' X 21'	486	10'	30	—	20	—	30	—	16	18	18	50
ROCKPORT/IPSWICH	26' X 50'	1,192	10'	108	—	54	—	80	80	42	36	42	125

This document contains approximate measurements and square footage that are for illustrative purposes only. We cannot guarantee the floor plan accuracy or completeness, therefore encourage you to review the space to make sure it is suitable for your event.



10 Huntington Avenue
Boston, Massachusetts 02116
T 617.262.9600 F 617.424.7483
E westincopley@westin.com
westincopleyplaceboston.com

THE WESTIN
COPLEY PLACE
BOSTON