



Association for
Computing Machinery

Advancing Computing as a Science & Profession



NEWS RELEASE

CONTACT: Jim Ormond
212-626-0505
ormond@hq.acm.org

BREAKTHROUGH INNOVATIONS IN DATA SCIENCE AND MACHINE LEARNING TO BE SHOWCASED AT ACM KDD 2018

Record Attendance Expected for Premier Data Mining Conference

New York, NY, August 1, 2018 – The Association for Computing Machinery (ACM) Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD) will hold its flagship annual conference, KDD, in London, UK, August 19-23. Started in 1989, KDD is the world’s oldest and largest data mining conference and has been the venue where concepts such as big data, data science, predictive analytics and crowdsourcing were first introduced. Continuing this tradition, KDD 2018 will showcase leading-edge research papers in data science and feature exciting keynote addresses by industry leaders, as well as informative tutorials and workshops.

“Not only has data science come into its own as a field, but its growth is intertwined with several other areas that are transforming technology and society—including artificial intelligence, machine learning and supercomputing,” said Yike Guo, of Imperial College, London and General Co-Chair for KDD 2018.

“For the first time this year, we’re excited to announce that KDD is sponsoring a Deep Learning Day and a Health Day,” added KDD General Co-Chair Faisal Farooq of IBM. “With these thematically-focused events, our goal is to provide a clear, wide overview of recent developments in these areas. We’ve assembled experts from world-class research institutions and leading companies to serve as speakers. For Health Day, we will be exploring the myriad ways large-scale data mining might impact public health issues globally in the coming years.”

The KDD 2018 program is the largest in the history of the conference, and will feature four keynote speakers, nine applied data science invited talks, 27 workshops, eight hands-on tutorials and 29 conventional tutorials. A partial listing of highlights follows. [The full KDD 2018 program is available here.](#)

Keynote Speakers

David Hand, Senior Research Investigator Emeritus Professor of Mathematics, Imperial College, London

“Data Science for Financial Applications”

Hand argues that there is considerable untapped potential for applying modern data analytic ideas to financial applications. Such applications come in three broad areas: actuarial and insurance, consumer banking, and investment banking. New model types and new sources of data are leading to a rich

opportunity for significant developments. Hand will also address other data science issues, such as data quality, ethics, and security, along with the need to understand the limitations of models.

Alvin E. Roth, Professor of Economics, Stanford University

“Market Design and Computerized Marketplaces”

In recent years, marketplaces have become computerized. Together with the introduction of smart phones, this also makes them ubiquitous. We can order car rides to the airport, plane rides to London, and hotel rooms for when we arrive, all on our smartphones. And as we do so we leave a data trail that is easily combined with other streams of data. This is changing not only how we interact with markets, but also how we manage and regard privacy. Roth will discuss recent developments in computerized markets and speculate about some still to come.

Yee Whye Teh, Professor of Statistical Machine Learning, University of Oxford and Research Scientist, Deepmind

“On Big Data Learning for Small Data Problems”

A question has arisen recently of whether machine learning systems need large amounts of data to solve a task well. An exciting recent development (under the banners of meta-learning, lifelong learning, learning to learn, multitask learning etc.) is that often there is heterogeneity within the data sets at hand, and in fact a large data set can be viewed more productively as many smaller data sets, each pertaining to a different task. In this talk, Teh will describe a view of this problem from probabilistic and deep learning perspectives, and describe a number of efforts that he has recently been involved in.

Jeannette M. Wing, Avanesians Director of the Data Sciences Institute, Columbia University

“Data for Good”

Wing uses the tagline “Data for Good” to convey how she feels the computing community should be promoting data science, especially in training future generations of data scientists. First, she believes data science should be used for the *good* of humanity and society. Second, she argues society should use data in a *good* manner. She employs the acronym FATES to emphasize fairness, accountability, transparency, ethics, and safety and security.

Tutorials (Partial List)

“Large-Scale Graph Algorithmics: Theory and Practice”

Silvio Lattanzi and Vahab Mirrokni (Google)

The presenters will discuss how to design and implement algorithms based on traditional MapReduce architecture, as well as various basic graph theoretic problems such as computing connected components, maximum matching, MST, counting triangle, and overlapping or non-overlapping clustering. They will explore the possibility of employing other distributed graph processing frameworks.

“Deep Learning for Computational Healthcare”

Edward Choi (Georgia Tech); Cao Xiao (IBM Research); Jimeng Sun (Georgia Tech)

The presenters introduce deep learning methods and their applications in computational healthcare, specifically focusing on representation learning and predictive modeling. This tutorial is intended for students, engineers and researchers who are interested in applying deep learning methods to healthcare, and prerequisite knowledge will be minimal. The presenters will introduce the nature of EHR data, basic deep learning methods and their application in healthcare, and then will focus on challenges specific to computational healthcare, introducing advanced deep learning methods to address them.

“Causal Inference and Counterfactual Reasoning”

Emre Kiciman and Amit Sharma (Microsoft Research)

As computing systems are more frequently and more actively intervening to improve people's work and daily lives, it is critical to correctly predict and understand the causal effects of these interventions. Conventional machine learning methods, built on pattern recognition and correlational analyses, are insufficient for causal analysis. This tutorial will introduce participants to concepts in causal inference and counterfactual reasoning, drawing from a broad literature on the topic from statistics, social sciences and machine learning.

Workshops (Partial List)

Data Science In Fintech

A range of topics will be covered, including how modern data-driven methods can optimize and personalize decisions in lending, insurance, estate planning, tax planning, fraud detection, transaction security and overall investing.

Workshop on Mining and Learning with Graphs

An exciting set of talks and papers will cover a broad range of topics from fundamental methods and insights on large graphs to interdisciplinary applications on computational social sciences and ecology to building product knowledge graphs.

Data Science, Journalism & Digital Media

This workshop will bring together a community of researchers and practitioners who study problems at the intersection of data science, journalism, and digital media, ranging from data-driven content creation, analysis, and dissemination to the consequences of social media on the perception of news.

epiDAMIK: Epidemiology Meets Data Mining and Knowledge discovery

The impact of Zika, MERS, and Ebola outbreaks over the past decade strongly illustrates our enormous vulnerability to emerging infectious diseases. This workshop outlines the role data mining and knowledge discovery can play in combating epidemics.

Applied Data Science Papers (Partial List)

“Active Remediation: The Search for Lead Pipes in Flint, Michigan”

Jacob Abernethy (Georgia Institute of Technology); Alex Chojnacki (University of Michigan); Arya Farahi (University of Michigan); Eric Schwartz (University of Michigan); Jared Webb (Brigham Young University)

The authors detail their ongoing work in Flint, Michigan to detect pipes made of lead and other hazardous metals. After elevated levels of lead were detected in residents' drinking water, followed by an increase in blood lead levels in area children, the state and federal governments directed over \$125 million to replace water service lines, the pipes connecting each home to the water system. In the absence of accurate records, and with the high cost of determining buried pipe materials, the authors put forth a number of predictive and procedural tools to aid in the search and removal of lead infrastructure.

“WattHome: Identifying Energy-Inefficient Homes at City-scale”

Srinivasan Iyengar, Stephen Lee, David Irwin, Prashant Shenoy, and Benjamin Weil (University of Massachusetts Amherst)

Buildings consume over 40 percent of the total energy in modern societies, and improving their energy efficiency can significantly reduce society's energy footprint. The authors present WattHome, a data-driven approach to identify the least energy-efficient buildings from a large population of buildings in a city or a region.

“Career Transitions and Trajectories: A Case Study in Computing”

Tara Safavi (University of Michigan); Maryam Davoodi (Purdue University); Danai Koutra (University of Michigan)

What do computing careers reveal about the evolution of computing research? Which institutions were and are the most important in this field, and for what reasons? Can insights into computing career trajectories help predict employer retention? The authors analyze several decades of post-PhD computing careers using a large new data set rich with professional information, and propose a versatile career network model, R 3, that captures temporal career dynamics.

Research Papers (Partial List)

“Opinion Dynamics with Varying Susceptibility to Persuasion”

Rediet Abebe (Cornell University); Jon Kleinberg (Cornell University); David Parkes (Harvard University); Charalampos Tsourakis (Boston University)

A long line of work in social psychology has studied variations in people's susceptibility to persuasion—the extent to which they are willing to modify their opinions on a topic. This body of literature suggests that in addition to considering interventions that directly modify people's intrinsic opinions, it is also natural to consider those that modify people's susceptibility to persuasion. The authors adopt a popular model for social opinion dynamics, and formalize the opinion maximization and minimization problems where interventions happen at the level of susceptibility.

“You Are How You Drive: Peer and Temporal-Aware Representation Learning for Driving Behavior Analysis”

Pengyang Wang (Missouri University of Science and Technology); Yanjie Fu (Missouri University of Science and Technology); Jiawei Zhang (Florida State University); Pengfei Wang (CNIC, Chinese Academy of Sciences); Yu Zheng (JD Finance); Charu Aggarwal (IBM)

Analyzing driving behavior can help us assess driver performances, improve traffic safety, and, ultimately, promote the development of intelligent and resilient transportation systems. Existing methods to analyze driving behavior can be improved via representation learning by jointly exploring the peer and temporal dependencies of driving behavior. To that end, the authors developed a Peer and Temporal-Aware Representation Learning based framework for driving behavior analysis with GPS trajectory data.

“Algorithms for Hiring and Outsourcing in the Online Labor Market”

Aris Anagnostopoulos (Sapienza University of Rome); Carlos Castillo (Universitat Pompeu Fabra); Adriano Fazzone (Sapienza University of Rome); Stefano Leonardi (Sapienza University of Rome); Evimaria Terzi (Boston University)

Although freelance work has grown substantially in recent years—in part facilitated by a number of online labor marketplaces, (e.g., Guru, Freelancer, Amazon Mechanical Turk)—traditional forms of “insourcing” work continue to be the dominant form of employment in most companies. This means that, at least for the time being, freelancing and salaried employment will continue to co-exist. The authors provide algorithms for outsourcing and hiring workers in a general setting.

Additional information about KDD 2018, including a full program and schedule of events, may be found at <http://www.kdd.org/kdd2018>.

About SIGKDD

ACM SIGKDD, The ACM Special Interest Group for Knowledge Discovery from Data (SIGKDD), is a professional society comprising of world-renowned data scientists from industry and academia. KDD is the premier international conference that brings together researchers and practitioners from both academia and industry to deep-dive into novel ideas, latest research results and share in-the-trenches experiences and innovations. More details can be found at kdd.org.

About ACM

ACM, the Association for Computing Machinery (www.acm.org), is the world's largest educational and scientific computing society, uniting computing educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. ACM strengthens the computing profession's collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking.

###