ENGINEERING IN THE CONTEXT OF BIG DATA
Session co-chairs: Jennifer Dy, Northeastern University, and Sriram Raghavan, IBM-India

The advancement of technology has enabled the growth of data acquisition at an enormous rate. Sensors and memory devices are now cost-effective that engineers and scientists can continuously collect data for health monitoring, climate prediction, optimization of manufacturing processes, management and control of complex systems (such as electric energy grid, transportation, communication networks, water resources and the environment). New devices for sensing at various scales are constantly developed; examples include microarrays, optical imaging, telescopes, high resolution cameras and video, and mobile phones. Many disciplines have become increasingly exploratory that large-scale data collection has become prevalent, e.g., genome sequencing, surveying the universe. In fact, data collection has now become ubiquitous in today's world, from genetic sequencing, to health and environment monitoring, to social networks, to data that is collectively stored in the internet. Coupled with this data collection is the hope that new discoveries or knowledge can be learned from this massive data.

For example, in biomedicine, one hopes to be able to combine information from genomic, molecular, cellular, clinical and environmental data to make better understanding of diseases and better diagnosis and treatment planning for each individual. In climate science, one hopes that continuous monitoring of multi-scale climate and environmental variables will enable better prediction of extremes.

Indeed, various engineering disciplines have benefitted from this data explosion. In this session, we invited speakers who have worked on big data applied to different engineering and science areas: manufacturing, equipment health monitoring, maintenance and optimization, healthcare and drug discovery, and space exploration, to provide examples of how new discoveries are enabled with learning from data.

The first speaker is Athulan Vijayaraghavan of System Insights. He will talk about "The Internet of Manufacturing Things," which will focus on how shop-floor data (e.g., from sensors and actuators) can be collected, managed, and stored, and calls for a new set of algorithms and data analytic methods to improve the capabilities of classical solutions to manufacturing.

The second speaker is Pankaj Dayama of IBM Research. His talk is on "Predictive Analytics for Industrial Applications," where he would talk about how predictive analytics techniques can be utilized to facilitate monitoring, maintenance, and optimization of industrial assets for better availability, utilization, and performance.

The third speaker is Nirmal Keshava at AstraZeneca. His presentation is "At the Intersection of Healthcare, Drug Discovery, and Big Data." The delivery of health care in the United States is being transformed by the role data is increasingly playing in enabling decision making at all levels. This talk will examine how several data streams are poised to impact drug discovery at different points, as well as the rapidly evolving technologies that are being introduced to support these changes, influenced strongly by the emergence of engineering principles and leadership.

The final speaker is Kiri Wagstaff of Jet Propulsion Laboratory, California Institute of Technology. Her talk is on “Facilitating Discovery in Big Data Sets,” where she presents how knowledge discovery can be facilitated given a large volume of data, by prioritizing observations that are most likely to inspire new discoveries first. She demonstrates the utility of these methods on stellar time series data from Kepler’s search for starts with exoplanets, Martian rock and soil emission spectra data from the Mars Science Laboratory’s ChemCam laser spectrometer, and atmospheric aerosol observations from the Multi-angle Imaging Spectroradiometer in Earth orbit.

This information explosion introduces interesting challenges as well. The session will end with a panel discussion, where the panel of speakers will discuss the challenges involved in dealing with massive data and in how to make data analytics a reality, such as what are the data acquisition challenges, what is involved in data preparation, curation and cleaning, how to deal with noise, how to combine domain knowledge or engineering models with data, how to scale algorithms to massive data, and how to extract information from data and interpret the results.