Colloquium

When Big Data Meets BRAIN Initiative: Large-Scale Structured Sparse Learning with Applications in Imaging Genomics

Heng Huang
University of Texas at Arlington

Abstract: Sparsity is one of the intrinsic properties of real-world data, thus sparse machine learning has recently emerged as powerful tool to obtain models of high-dimensional data with high degree of interpretability at low computational cost, and provide great opportunities to analyze the big, complex, and diverse datasets. By enforcing properly designed structured sparsity, we can integrate the specific data structures and domain knowledge into the machine learning models to simplify data models and discover predictive patterns in big data analytics. Big Data research is accelerating the translation of biological and biomedical data to advance the detection, diagnosis, treatment and prevention of diseases, including the recently announced BRAIN (Brain Research through Advancing Innovative Neurotechnologies) Initiative. To address the challenging problems in current big data mining, we proposed several novel large-scale structured sparse learning models for multi-dimensional data integration, heterogeneous multi-task learning, group/graph structured data analysis, and longitudinal feature learning. We applied our new structured sparse learning models to analyze the multi-modal neuroimaging and genome-wide array data in Imaging Genomics and discover the phenotypic and genotypic biomarkers to characterize the neurodegenerative process in the progression of Alzheimers disease and other complex brain disorders. We also utilized our new machine learning models to analyze the Electronic Medical Records for predicting the heart failure patients readmission and drug side effects, detect the multi-dimensional biomarkers in The Cancer Genome Atlas (TCGA) research, and identify the brain circuitry patterns in Human Connectome.

Monday, March 3, 2014, 4:00 pm
Mathematics and Science Center: W303

Mathematics and Computer Science
Emory University