RUTGERS
School of Arts and Sciences

Department of Statistics and Biostatistics
Rutgers, The State University of New Jersey
110 Frelinghuysen Road
Piscataway, NJ 08854

Web: www.stat.rutgers.edu
E-mail: office@stat.rutgers.edu
Phone: (848) 445-2690
Fax: (732) 445-3428

RUTGERS UNIVERSITY

Department of Statistics, Department of Computer Science

www.stat.rutgers.edu

Workshop on Big Data: Theoretical Foundation of Big Data

Speaker: **Dr. Eric Xing**
Carnegie Mellon University

Title: **On Data Parallelism and Model Parallelism for Large Scale Machine Learning**

Time: **3:15pm – 4:10pm, *Thursday*, October 16, 2014**

Place: ***CoRE Auditorium***

Abstract

In many modern applications built on massive data and using high-dimensional models, such as web-scale content extraction via topic models, genome-wide association mapping via sparse regression, and image understanding via deep neural networks, one needs to handle BIG machine learning problems that threaten to exceed the limit of current infrastructures and algorithms. A hallmark of machine learning programs that distinguishes them from a conventional computing program such as database search or keyword extraction is that they are intrinsically an optimization problem defined by an explicit objective function to be optimized over data, rather than being a set of operations whose very high-fidelity execution itself marks the goal of the program. Furthermore, a clear dichotomy between data (which is conditionally independent and persistent throughout the process of training) and model (which is internally coupled, and is transient before converging to an optimum), and the needs for an iterative-convenient procedure to learn the model from the data opens a unique need and opportunity of adopting an ML-first principle for designing a new algorithmic and system interface of BIG ML. In this talk, I discuss elements of this interface for Data Parallelism and Model Parallelism inspired by this insight; why it is preferred over the conventional database-style consistency and fault-tolerance driven system architectures; and how it can result in orders of magnitude of improvement in speed and scale without comprising correctness empirically and theoretically. I present Petuum -- a general-purpose framework we built on such a principle for distributed machine learning, an demonstrate how innovations in scalable ML algorithms and distributed systems design work in concert to achieve dramatic scalability for a wide range of large scale problems in social network (mixed-membership inference on 100M node) and personalized genome medicine (sparse regression on 100M dimensions), even with a modest cluster.

RUTGERS

BIO:

Dr. Eric Xing is a Professor of Machine Learning in the School of Computer Science at Carnegie Mellon University. His principal research interests lie in the development of machine learning and statistical methodology; especially for solving problems involving automated learning, reasoning, and decision-making in high-dimensional, multimodal, and dynamic possible worlds in social and biological systems. Professor Xing received a Ph.D. in Molecular Biology from Rutgers University, and another Ph.D. in Computer Science from UC Berkeley. His current work involves, 1) foundations of statistical learning, including theory and algorithms for estimating time/space varying-coefficient models, sparse structured input/output models, and nonparametric Bayesian models; 2) computational and statistical analysis of gene regulation, genetic variation, and disease associations; and 3) large-scale systems for machine learning. Professor Xing has published over 200 peer-reviewed papers, and is an associate editor of the Annals of Applied Statistics (AOAS), the Journal of American Statistical Association (JASA), the IEEE Transaction of Pattern Analysis and Machine Intelligence (PAMI), the PLoS Journal of Computational Biology, and an Action Editor of the Machine Learning Journal (MLJ), the Journal of Machine Learning Research (JMLR). He is a member of the DARPA Information Science and Technology (ISAT) Advisory Group, a recipient of the NSF Career Award, the Sloan Fellowship, the United States Air Force Young Investigator Award, the IBM Open Collaborative Research Award, and best paper awards in a number of premier conferences including UAI, ACL, SDM, and ISMB. He is the Program Chair of ICML 2014.