香港科技大學
THE HONG KONG UNIVERSITY OF
SCIENCE AND TECHNOLOGY

JOINT SEMINAR

Nonlinear Embedding as a Tool for Big Data Algorithm Design

By

Prof Le Song
Associate Professor of Department of Computational Science and
Engineering, College of Computing
Associate Director of the Center for Machine Learning
Georgia Institute of Technology

****

Date: 21 March 2017 (Tuesday)
Time: 11:15am
Venue: LT-G (Lift 25-26)

Abstract
Many large scale data analytics problems are intrinsically hard and complex, making the design of effective and scalable algorithms very challenging. Domain experts have to perform extensive research, and experiment with many trial-and-errors, in order to craft approximation or heuristic schemes that meet the dual goals of effectiveness and scalability. Very often, restricted assumptions about the data, which are likely to be violated by real world scenarios, need to be made in order for the designed algorithms to work and have performance guarantees. Furthermore, previous algorithm design paradigms seldom systematically exploit a common trait of real-world algorithms: instances of the same type of problem are solved repeatedly on a regular basis using the same algorithm, maintaining the same problem structure, but differing mainly in their data.

In this talk, he will present a framework for scalable algorithm design based on the idea of embedding algorithm steps into nonlinear spaces, and learn these embedded algorithms from data via direct supervision or reinforcement learning. In contrast to traditional algorithm design where every steps in an algorithm is prescribed by experts, the embedding design will delegate some difficult algorithm choices to nonlinear learning models so as to avoid large memory requirement, restricted assumption on the data, or limited design space exploration. I will illustrate the benefit of this embedding framework using three examples: a materials discovery problem, a recommendation system problem over dynamic social networks, and a problem of learning combinatorial algorithms over graphs. In some cases, the learned algorithms require thousands of times less memory to run, finish hundreds of times faster, and obtain drastic improvement in predictive performance.

Speaker's Profile

Le Song is an Associate Professor in the Department of Computational Science and Engineering, College of Computing, and an Associate Director of the Center for Machine Learning, Georgia Institute of Technology. He received his Ph.D. in Machine Learning from University of Sydney and NICTA in 2008, and then conducted his post-doctoral research in the Department of Machine Learning, Carnegie Mellon University, between 2008 and 2011. Before he joined Georgia Institute of Technology in 2011, he was a research scientist at Google briefly. His principal research direction is machine learning, especially nonlinear models, such as kernel methods and deep learning, and probabilistic graphical models for large scale and complex problems, arising from artificial intelligence, network analysis, computational biology and other interdisciplinary domains. He is the recipient of the Recsys'16 Deep Learning Workshop Best Paper Award, AISTATS'16 Best Student Paper Award, IPDPS'15 Best Paper Award, NSF CAREER Award'14, NIPS'13 Outstanding Paper Award, and ICML'10 Best Paper Award. He has also served as the area chair or senior program committee for many leading machine learning and AI conferences such as ICML, NIPS, AISTATS and AAAI, and the action editor for JMLR.

All are Welcome!

Jointly hosted by Big Data Institute and

Department of Computer Science & Engineering