

Presents the Spring 2014 EECS Seminar Series

Dr. Lidan Wang
University of Illinois at Urbana Champaign

“Efficient Modeling-Based Approaches for Large-Scale Mining and Retrieval”
Wednesday, February 12, 2014 • 10:00 a.m. • HEC 450

ABSTRACT

In the era of "Big data", when we design mining and retrieval models for large data (e.g., in the order of billions of nodes/documents) to extract knowledge and patterns, the scalability and efficiency of the models are of critical importance. Intuitively, as dataset size increases, the models should not just be effective, but must be fast and scalable also (i.e., run-time speed of the models when applied to very large amount of data).

Traditional mining and retrieval models tend to get more and more costly to sustain reasonable run-time speed on large data. In this talk, I will describe cost-effective and efficient modeling-based approaches for mining and retrieval on large data. I will illustrate how to achieve highly effective mining and retrieval results on massive data, without incurring too much additional computational latency and costs. I will describe three key challenges (model, metric, and learning) for this problem and present systematic solutions to address the challenges.

An implication of this work is that now we can handle large data more elegantly -- maintaining good mining and retrieval effectiveness, without sacrificing run-time efficiency on massive data or requiring too much additional computational and energy resources, thus offering a more attractive and cost-effective solution for mining and retrieval on massive data.

BIOGRAPHY

Lidan Wang is a postdoctoral researcher in the Data Mining Research Group at the University of Illinois at Urbana Champaign, working with Professor Jiawei Han. She received a Ph.D. in Computer Science from the University of Maryland, College Park in 2012. Her current research interests are data mining, information retrieval, and machine learning, with a focus on efficient, scalable, and robust techniques for mining and search from large-scale structured and unstructured data. Her work has been recognized with a SIGIR 2012 Best Paper Honorable Mention, a CIKM 2010 Best Paper Award finalist, and an EWSN 2008 Best Paper Award.