SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

DISTINGUISHED SPEAKER SERIES

Moses Charikar

Stanford University

Friday, March 22, 2019

3:00 pm

Luddy Hall, Rm. 1106

Importance Sampling in High Dimensions via Hashing

Abstract: Locality sensitive hashing (LSH) is a popular technique for nearest neighbor search in high dimensional data sets. Recently, a new view at LSH as a biased sampling technique has been fruitful for density estimation problems in high dimensions. Given a set of points and a query point, the goal (roughly) is to estimate the density of the data set around the query. One way to formalize this is by kernel density estimation: Given a function that decays with distance and represents the "influence" of a data point at the query, sum up this influence function over the data set. Yet another way to formalize this problem is by counting the number of data points within a certain radius of the query. While these problems can easily be solved by making a linear pass over the data, this can be prohibitive for large data sets and multiple queries. Can we preprocess the data so as to answer queries efficiently? This talk will survey several recent papers that use locality sensitive hashing to design unbiased estimators for such density estimation problems and their extensions.

This talk will survey joint works with Arturs Backurs, Piotr Indyk, Vishnu Natchu, Paris Syminelakis and Xian (Carrie) Wu.

Biography: Moses Charikar is the Donald E. Knuth professor of Computer Science at Stanford University. He obtained his PhD from Stanford in 2000, spent a year in the research group at Google, and was on the faculty at Princeton from 2001-2015. He is broadly interested in approximation algorithms (especially the power of mathematical programming approaches), metric embeddings, algorithmic techniques for big data, efficient algorithms for computational problems in high-dimensional statistics and optimization problems in machine learning. He won the best paper award at FOCS 2003 for his work on the impossibility of dimension reduction, the best paper award at COLT 2017 and the 10 year best paper award at VLDB 2017. He was jointly awarded the 2012 Paris Kanellakis Theory and Practice Award for his work on locality sensitive hashing inspired by random hyperplane rounding, and was named a Simons Investigator in theoretical computer science in 2014.

INDIANA UNIVERSITY

SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING