SCMS&SDS Seminar

AN ASYMPTOTIC ANALYSIS OF NONPARAMETRIC DIVIDE-AND-CONQUER METHODS

Speaker：Dr. Botond Szabo
Leiden University, The Netherlands

Time: 10:00-11:30 a.m., Tuesday, June 13, 2017
Venue: Room 2201, East Guanghua Tower (Main), Fudan University

Abstract: In the era of "big data" computer scientists and statisticians are facing new challenges to handle the large amount of available data. Datasets have become so large that it becomes unfeasible, or computationally undesirable, to carry out the analysis on a single machine. This gave rise to divide-and-conquer algorithms where the data is distributed over several "local" machines and the computations are done on these machines parallel to each other. Then the outcome of the local computations is somehow aggregated to a global result in a central machine.

Over the years, various divide-and-conquer algorithms were proposed, many of them with limited theoretical underpinning. First, we compare the theoretical properties of a (not complete) list of proposed methods on the benchmark nonparametric signal-in-white-noise model. Most of the investigated algorithms use information on aspects of the underlying true signal (for instance regularity), which is usually not available in practice. A central question is whether one can tune the algorithms in a data-driven way, without using any additional knowledge about the signal. We show that (a list of) standard data-driven techniques (both Bayesian and frequentist) cannot recover the underlying signal with the minimax rate. This, however, does not imply the absence of an adaptive distributed method.

To address the theoretical limitations of data-driven divide-and-conquer algorithms we consider a setting where the amount of information sent between the local and central machines is expensive and limited. We show that it is not possible to construct data-driven methods which adapt to the unknown regularity of the underlying signal and at the same time communicates the optimal amount of information between the machines.

This is a joint work with Harry van Zanten.

.