Learning-Based Sampling and Streaming

Abstract:
Classical algorithms typically provide “one size fits all” performance, and do not leverage properties or patterns in their inputs. A recent line of work aims to address this issue by developing algorithms that use machine learning predictions to improve their performance. In this talk I will present two examples of this type, in the context of streaming and sampling algorithms. In particular, I will show how to use machine learning predictions to improve the performance of (a) low-memory streaming algorithms for frequency estimation (ICLR19), and (b) sampling algorithms for estimating the support size of a distribution (ICLR21). Both algorithms use an ML-based predictor that, given a data item, estimates the number of times the item occurs in the input data set.

The talk will cover material from papers co-authored with T Eden, CY Hsu, D Katabi, S Narayanan, R Rubinfeld, S Silwal, T Wagner and A Vakilian.