September 2010

MADALGO seminar by Jeff M. Phillips, University of Utah

Comparing Distributions and Shapes with the Kernel Distance

Abstract:

The kernel distance is the metric formed using a using a kernel (or similarity function). Specifically, given a positive definite K: R^d x $R^d \rightarrow R$, then for two points sets P,Q in R^d the kernel distance is defined:

$$D_K(P,Q) = \sqrt{K(P,P) + K(Q,Q) - 2K(P,Q)}$$

Where

$$K(P,Q) = \sum_{p \in P} \sum_{q \in Q} K(p,q).$$

This definition generalizes naturally to shapes (curves, surfaces), distributions, clusters, graphs, and trees. In the past 5 years or so, a flurry of work in medical imaging (where D_K is called the current distance) and machine learning (where D_K is called maximum mean discrepancy or MMD) has shown the practicality of this measure as well as its favorable relation to more classic measures such as EMD.

In this talk, I will provide the first rigorous algorithmic analysis of the kernel distance. I first will reduce the kernel distance on smooth shapes and distributions with bounded error to the kernel distance on finite point sets. Then I will produce near-linear algorithms on point sets that preserves the same bounded error, beating the naive quadratic runtime bound.

Joint work with:

Sarang Joshi, Raj Varma Kommaraju, and Suresh Venkatasubramanian