In Favour of Posting

香港中文大學
The Chinese University of Hong Kong

Institute of Theoretical Computer Science and Communications

ITCSC Seminar

Sparse Johnson-Lindenstrauss Transforms

By

Dr. Jelani Nelson, MIT CSAIL

June 23, 2011, Thursday

11:30am – 12:30noon

Rm. 121, 1/F., Ho Sin Hang Engineering Building, CUHK

Abstract: The Johnson-Lindenstrauss (JL) lemma (1984) states that any n points in d-dimensional Euclidean space can be embedded into k =O((log n)/eps2) dimensions so that all pairwise distances are preserved up to 1+/-eps. Furthermore, this embedding can be achieved via a linear mapping. The JL lemma is a useful tool for speeding up solutions to several high-dimensional problems: closest pair, nearest neighbor, diameter, minimum spanning tree, etc. It also speeds up some clustering and string processing algorithms, reduces the amount of storage required to store a dataset, and can be used to reduce memory required for numerical linear algebra problems such as linear regression and low rank approximation.

The original proofs of the JL lemma let the linear mapping be specified by a random dense k x d matrix (e.g. i.i.d. Gaussian entries). Thus, performing an embedding requires dense matrix-vector multiplication. There has been much recent work on developing distributions that allow for embedding vectors quickly, begun by the work of [Ailon, Chazelle, STOC 2006]. Unfortunately, these works cannot take advantage of sparsity of the vector to be embedded, and they take Omega(d) time to embed a vector even with only one non-zero entry. This feature is particularly unfortunate in streaming applications, where updates can be seen as 1-sparse vectors.

One way to speed up embeddings for sparse vectors is to develop distributions over linear mappings whose corresponding matrices themselves are sparse. In this talk, we give two JL distributions with the sparsest known matrices. In fact, these are the first distributions where every matrix in their support has only o(1) of its entries non-zero for all settings of eps and n (specifically, only an O(eps)-fraction of entries in each column are non-zero).

This is joint work with Daniel Kane (Harvard University)

Biography: Jelani Nelson just finished his PhD with the Theory of Computation group at MIT CSAIL three weeks ago, and he will next be headed to the Mathematical Sciences Research Institute in Berkeley this Fall for a semester-long program on quantitative geometry. His research interests include sketching, streaming, and pseudorandomness. At the current time, he has an unhealthy obsession with dimensionality reduction.

******* ALL ARE WELCOME *******

Hosted & Enquiries : Institute of Theoretical Computer Science and Communications Tel: 2696 1257