

ML SEMINAR

MACHINE LEARNING
DEPARTMENT

Simple and Deterministic Matrix Sketches

Abstract:

A sketch of a matrix $\$A\$$ is another matrix $\$B\$$ which is significantly smaller than $\$A\$$, but still approximates it well. Finding such sketches efficiently is an important building block in modern algorithms for approximating, for example, the PCA of massive matrices. This task is made more challenging in the streaming model, where each row of the input matrix can be processed only once and storage is severely limited. In this paper, we adapt a well known streaming algorithm for approximating item frequencies to the matrix sketching setting. The algorithm receives $\$n\$$ rows of a large matrix $\$A\$$ in $R^{n \times m}$ one after the other, in a streaming fashion. It maintains a sketch $\$B\$$ in $R^{\ell \times m}$ containing only ℓ rows but still guarantees that $\$A^T A \approx B^T B\$$. More accurately, $\$ \|Ax - Bx\|_2 \leq \|A\|_F \sqrt{\ell} \$$ for all x , $|x|=1$. This algorithm's error decays proportionally to $\sqrt{\ell}$ using $\$O(m\ell)$ space. In comparison, random-projection, hashing or sampling based algorithms produce convergence bounds proportional to $\$1/\sqrt{\ell}\$$. Sketch updates per row in $\$A\$$ require amortized $\$O(m\ell)$ operations and the algorithm is perfectly parallelizable. Our experiments corroborate the algorithm's scalability and improved convergence rate. The presented algorithm also stands out in that it is deterministic, simple to implement, and elementary to prove.

Bio:

Edo received his B.Sc in Physics and Computer Science from Tel Aviv university and his Ph.D in Computer Science from Yale University, under the supervision of Steven Zucker. During his PhD he spent time at both UCLA and Google as an engineer and a researcher. After that, he joined the Program in Applied Mathematics at Yale as a Post-Doctoral fellow. In 2009 he joined Yahoo! Labs in Israel. He recently moved to New York to lead the scalable machine learning group which focuses on the theory and practice of (very) large scale data mining and machine learning. In Particular, theoretical foundations of machine learning, optimizations, scalable scientific computing, and machine learning systems and platforms.

Speaker:

Edo Liberty

Senior Research Scientist, Yahoo! Labs



Date: Thur., Feb 27, 2014

Time: 10:00am

Location: 8102 Gates Building