ATELIER « COMBINATOIRE, RANDOMISATION, ALGORITHMES ET PROBABILITÉS »
4–8 MAI 2009

WORKSHOP "COMBINATORICS, RANDOMIZATION, ALGORITHMS AND PROBABILITY"
MAY 4–8, 2009

Matrix Completion from Fewer Entries

ANDREA MONTANARI

Department of Electrical Engineering
Stanford University
Packard Bldg
Stanford, CA 94305-9510
USA

montanari@stanford.edu

Low-rank models are frequently used in machine learning and statistics. An important domain of application is provided by collaborative filtering, whereby a low-rank matrix describes the ratings that a large set of users attribute to a large set of products. The problem is in this case to predict future ratings from a sparse subset currently available. The dataset released for the Netflix challenge provides an ideal testbed for theory and algorithms for learning low-rank matrices.

Given M, a random $n \times n$ matrix of rank r, we assume that a uniformly random subset E of its entries is observed. We describe an efficient procedure that reconstructs M from $|E| = O(rn)$ observed entries with arbitrarily small root mean square error, whenever M is satisfies an appropriate incoherence condition. If $r = O(1)$, the algorithm reconstructs M exactly from $O(n \log n)$ entries. This settles a recent open problem by Candes and Recht.

In the process of proving these statements, we obtain a generalization of a celebrated result by Friedman–Kahn–Szemeredi and Feige–Ofek on the spectrum of sparse random matrices.

This is joint work with R. H. Keshavan and S. Oh.