

DEPARTMENT OF INDUSTRIAL AND ENTERPRISE SYSTEMS ENGINEERING

GE/IE 590 SEMINAR

Learning from the Wisdom of the Crowd: Efficient Algorithms and Fundamental Limits

Sewoong Oh

Postdoctoral Associate
LIDS, Massachusetts Institute of Technology

Abstract

This talk is on designing extremely efficient and provably order-optimal algorithms to extract meaningful information from societal data, the kind of data that comes from crowdsourcing platforms like Amazon Mechanical Turk, or recommendation systems like the Netflix Challenge dataset.

Crowdsourcing systems, like Amazon Mechanical Turk, provide platforms where large-scale projects are broken into small tasks that are electronically distributed to numerous on-demand contributors. Because these low-paid workers can be unreliable, we need to devise schemes to increase confidence in our answers, typically by assigning each task multiple times and combining the answers in some way. I will present the first rigorous treatment of this problem, and provide both an optimal task assignment scheme (using a random graph) and an optimal inference algorithm (based on low-rank matrix approximation and belief propagation) for that task assignment. This approach significantly outperforms previous approaches and, in fact, is asymptotically order-optimal, which is established through comparisons to an oracle estimator.

Another important problem in learning from the wisdom of the crowd is how to make product recommendations based on past user ratings. A common and effective way to model these user ratings datasets is to use low-rank matrices. In order to make recommendations, we need to predict the unknown entries of a ratings matrix. A natural approach is to find a low-rank matrix that best explains the observed entries. Motivated by this recommendation problem, my approach is to provide a general framework for recovering a low-rank matrix from partial observations. I will introduce a novel, efficient and provably order-optimal algorithm for this matrix completion problem. The optimality of this algorithm is established through a comparison to a minimax lower bound on what the best algorithm can do.

Biography

Sewoong Oh is a postdoctoral associate at LIDS, MIT working with professor Devavrat Shah and Professor David R. Karger. He finished his Ph.D. in 2010 at Stanford University under the supervision of professor Andrea Montanari working in the area of statistical inference and graphical models. He received his Master's degree in Electrical Engineering from Stanford University in June 2007. He graduated from Seoul National University, Korea, with a Bachelor of Science in Electrical Engineering in 2002. He was awarded the Kenneth C. Sevcik outstanding student paper award at the Sigmetrics 2010.

Location: 101 Transportation Building
Date: Thursday, February 2, 2012
Time: 4:00 p.m.