CCIS Distinguished Lecturer Series

Spring 2016

Guilt-Free Interactive Data Analysis: The Reusable Holdout

Omer Reingold
Samsung Research America

Tuesday
January 26, 2016
240 Dockser
12:00pm

Abstract

A great deal of effort has been made to reduce the risk of spurious scientific discoveries, from the use of holdout sets and sophisticated cross-validation techniques, to procedures for controlling the false discovery rate in multiple hypothesis testing. However, there is a fundamental disconnect between the theoretical results and the practice of science: the theory mostly assumes a fixed collection of hypotheses to be tested, or learning algorithms to be applied, selected non-adaptively before the data are gathered, whereas science is by definition an adaptive process, in which data are shared and re-used, and hypotheses and new studies are generated on the basis of data exploration and previous outcomes.

Surprisingly, the challenges of adaptivity can be addressed using insights from differential privacy, a field of study supporting a definition of privacy tailored to private data analysis. As a corollary we show how to safely reuse a holdout set a great many times without undermining its power of "correctness protection," even when hypotheses and computations are chosen adaptively. Armed with this technique, the analyst is free to explore the data ad libitum, generating and evaluating hypotheses, verifying results on the holdout, and backtracking as needed.

Joint work with Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toni Pitassi and Aaron Roth.

Biography

Omer Reingold is a Principle Research Engineer at SRA working in the Computing Science Innovation Center. Past positons include the Weizmann Institute of Science, Microsoft Research, the Institute for Advanced Study in Princeton, NJ, and AT&T Labs (together with shorter visiting appointments at Harvard University and at Stanford University). His research is in the Foundations of Computer Science and most notably in Computational Complexity and the Foundations of Cryptography with emphasis on randomness, derandomization and explicit combinatorial constructions. He is an ACM Fellow and among his distinctions are the 2005 Grace Murray Hopper Award in the 2009 Gödel Prize.

Northeastern University
College of Computer and Information Science