

MANAGEMENT SCIENCES SEMINAR SERIES

The Predictive Power of Massive Data about our Fine-Grained Behavior

Dr. Foster Provost
Professor

Stern School of Business, New York University

February 24, 2017

2:30-3:30 pm

C125 Pappajohn Business Building

Abstract

What really is it about “big data” that makes it different from traditional data? In this talk I illustrate one important aspect: massive ultra-fine-grained data on individuals' behaviors holds remarkable predictive power. I examine several applications to marketing-related tasks, showing how machine learning methods can extract the predictive power and how the value of the data "asset" seems different from the value of traditional data used for predictive modeling. I then dig deeper into explaining the predictions made from massive numbers of fine-grained behaviors by applying a counter-factual framework for explaining model behavior based on treating the individual behaviors as evidence that is combined by the model. This analysis shows that the fine-grained behavior data incorporate various sorts of information that we traditionally have sought to capture by other means. For example, for marketing modeling the behavior data effectively incorporate demographics, psychographics, category interest, and purchase intent. Finally, I discuss the flip side of the coin: the remarkable predictive power based on fine-grained information on individuals raises new privacy concerns. In particular, I discuss privacy concerns based on inferences drawn about us (in contrast to privacy concerns stemming from violations to data confidentiality). The evidence-counterfactual approach used to explain the predictions also can be used to provide online consumers with transparency into the reasons why inferences are drawn about them. In addition, it offers the possibility to design novel solutions such as a privacy-friendly "cloaking device" to inhibit inferences from being drawn based on particular behaviors.

Bio of Prof. Foster Provost

Foster Provost is Professor of Data Science, Professor of Information Systems, and Andre Meyer Faculty Fellow at New York University's Stern School of Business. He is coauthor of the best-selling data science book, *Data Science for Business*. His research focuses on modeling behavior data, modeling (social) network data, crowd-sourcing for data science, aligning data science with application goals, and privacy-friendly methods. His research has won many awards, including the INFORMS Design Science Award, the best paper award in the journal *Information Systems Research*, and best paper awards at the ACM SIGKDD Conference across three decades. He cofounded several data-science-oriented companies based on his research, including *Dstillery* and *Integral Ad Science*. Foster previously was Editor-in-Chief of the journal *Machine Learning*. His latest music album, *Mean Reversion*, will be released later this year.

Acknowledgement: This seminar is funded by the Clarence W. Tow Fund.