

NYU Stern School of Business
Department of Information, Operations & Management Sciences
INFORMATION SYSTEMS RESEARCH SEMINAR

TOPIC: Scaling Up Event and Pattern Detection to Big Data

SPEAKER: Daniel B. Neill (Carnegie Mellon University)

DATE: Thursday, April 3rd, 2014

TIME: 4:00PM-5:30PM

PLACE: KMC 4-80

ABSTRACT

This talk will present a novel framework for accurate and computationally efficient pattern detection in massive, complex, and high-dimensional datasets. Our "fast subset scan" approach treats pattern detection as a search over subsets of data records and attributes, finding the subsets which maximize some score function (e.g., a likelihood ratio statistic). Many commonly used functions can be shown to satisfy the "linear-time subset scanning" property, enabling exact and efficient optimization over subsets. We can then use this unconstrained optimization step as a building block for scalable pattern detection, enabling us to incorporate a variety of real-world constraints (such as spatial proximity and graph connectivity), use prior information (for example, we expect that a dynamic pattern will evolve smoothly over time), and integrate information from many data streams. This framework has enabled more timely and more accurate detection of emerging events and other patterns across a variety of application domains, such as disease outbreak detection, container shipment monitoring, tracking contamination in water networks, crime prediction, and preventing rodent infestations. This talk will review the basic fast subset scan framework (Neill, 2012) and describe our many recent extensions and applications. Finally, I will discuss our ongoing work on scaling up pattern detection to even larger problems (where even linear-time algorithms may be insufficient) by incorporating sampling, hierarchy, and parallelism.

BIO

Daniel B. Neill is the Dean's Career Development Professor and Associate Professor of Information Systems at Carnegie Mellon University's Heinz College, where he directs the Event and Pattern Detection Laboratory and the Joint Ph.D. Program in Machine Learning and Policy. He holds courtesy appointments in Machine Learning and Robotics at CMU and is an adjunct professor in the University of Pittsburgh's Department of Biomedical Informatics. He received his M.Phil. from Cambridge University and his M.S. and Ph.D. in Computer Science from CMU. His research focuses on machine learning and event detection in massive datasets, with

applications ranging from medicine and public health to law enforcement and security. His detection methods have been incorporated into deployed disease surveillance systems throughout the world, and his "CityScan" software is in day-to-day operational use by police to predict and prevent emerging hot spots of violent crime. Dr. Neill was the recipient of an NSF CAREER award and an NSF Graduate Research Fellowship, and was recently named one of the "top ten artificial intelligence researchers to watch" by IEEE Intelligent Systems.