Learning Political Embeddings from Text
Dr. Noah Smith

Abstract
Research in natural language processing is increasingly taking into account the context—physical and social—in which linguistic communication occurs. In this talk, I will present models of text data that exploit assumptions about the social world from which those data emerged. Vector representations of political actors' policy preferences, known as “ideal points,” have been an important representational tool in political science since the 1980s. I'll describe how textual evidence can help to infer these embeddings in models that capture relationships between language and politics. The first project considers high-stakes texts generated for the Supreme Court and explores the assumption that authors of amicus curiae (“friends of the court”) are rational agents seeking to maximize expected utility. The second project considers political embeddings of propositions, using social media text to infer a continuous representation of the political import of claims like “Obama is a socialist.”

The primary collaborators on this research are my former Ph.D. student David Bamman (now at UC Berkeley), my current Ph.D. student Yanchuan Sim (at CMU), and Prof. Bryan Routledge (Tepper School of Business at CMU).

Bio
Noah Smith is an Associate Professor of Computer Science & Engineering at the University of Washington. Previously, he was an Associate Professor of Language Technologies and Machine Learning in the School of Computer Science at Carnegie Mellon University. He received his Ph.D. in Computer Science from Johns Hopkins University in 2006 and his B.S. in Computer Science and B.A. in Linguistics from the University of Maryland in 2001. His research interests include statistical natural language processing, especially unsupervised methods, machine learning, and applications of natural language processing. His book, Linguistic Structure Prediction, covers many of these topics. He has served on the editorial board of the journals Computational Linguistics (2009–2011), Journal of Artificial Intelligence Research (2011–present), and Transactions of the Association for Computational Linguistics (2012–present), and as the secretary-treasurer of SIGDAT (2012–2015). Alumni of his research group, Noah's ARK, are international leaders in NLP in academia and industry. Smith's work has been recognized with a Finmeccanica career development chair at CMU (2011–2014), an NSF CAREER award (2011–2016), a Hertz Foundation graduate fellowship (2001–2006), numerous best paper nominations and awards, and coverage by NPR, BBC, CBC, New York Times, Washington Post, and Time.