



The Center for Social Science Research Presents:

Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research

Thursday, February 20, 2020

12:00-1:30 p.m.

1030 Jenkins Nanovic Halls



Featuring

Arthur Spirling, PhD

Professor of Politics & Data Science
New York University

Word embeddings are becoming popular for political science research, yet we know little about their properties and performance. To help scholars seeking to use these techniques, we explore the effects of key parameter choices---including context window length, embedding vector dimensions and pre-trained vs locally fit variants---on the efficiency and quality of inferences possible with these models. Reassuringly we show that results are generally robust to such choices for political corpora of various sizes and in various languages. Beyond reporting extensive technical findings, we provide a novel crowdsourced "Turing test"-style method for examining the relative performance of any two models that produce substantive, text-based outputs. Our results are encouraging: popular, easily available pre-trained embeddings perform at a level close to---or surpassing---both human coders and more complicated locally-fit models. For completeness, we provide best practice advice for cases where local fitting is required.

For more information: cssr.nd.edu/news-events/

