

Data Science Seminar
Hosted by Department of Mathematical Sciences

Interdisciplinary Dean's Speaker Series in Data Science

RSVP at <http://bit.ly/DS-TAE-RSVP>.

- Date: Tuesday, November 19, 2019
- Time: 10:00 - 11:30 am
- Room: UJW-325
- Speaker: Arthur Spirling (Professor of Politics and Data Science at New York University)
- Title: Word Embeddings: What works, what doesn't, and how to tell the difference for applied research

Abstract

We consider the properties and performance of word embeddings techniques in the context of political science research. In particular, we explore key parameter choices—including context window length, embedding vector dimensions and the use of pre-trained vs locally fit variants — with respect to 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. Encouragingly, we show that 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.

Bio: Spirling is Professor of Politics and Data Science at New York University. He is the Deputy Director and the Director of Graduate Studies (MSDS) at the Center for Data Science, and Chair of the Executive Committee of the Moore-Sloan Data Science Environment. Spirling specializes in political methodology and legislative behavior, with an interest in the application of text-as-data/NLP, Bayesian statistics, machine learning, item response theory and generalized linear models in political science. His substantive field is comparative politics, and he focuses primarily on the United Kingdom. Spirling received his PhD from the University of Rochester, Department of Political Science, in 2008. From 2008 to 2015, he was an Assistant Professor and then the John L. Loeb Associate Professor of the Social Sciences in the Department of Government at Harvard University. He is the faculty coordinator for the NYU Text-as-Data speaker series.

Here is the paper Arthur Spirling is going to present.

https://github.com/ArthurSpirling/EmbeddingsPaper/blob/master/Paper/Embeddings_SpirlingRodriguez.pdf

f

Here is a FAQ that he suggested could be useful.

https://github.com/ArthurSpirling/EmbeddingsPaper/blob/master/Project_FAQ/faq.md

The Interdisciplinary Dean's Speaker Series in Data Sciences is supported by the:

- Dean's Office of Harpur College of Arts and Sciences
- Department of Biological Sciences
- Department of Mathematical Sciences
- Department of Political Science
- Department of Systems Science and Industrial Engineering
- Data Science Transdisciplinary Area of Excellence

For questions, contact David Clark (dclark@binghamton.edu) or Xingye Qiao (qiao@math.binghamton.edu).

From:

<http://www2.math.binghamton.edu/> - **Department of Mathematics and Statistics,
Binghamton University**

Permanent link:

<http://www2.math.binghamton.edu/p/seminars/datasci/191119>

Last update: **2019/11/06 14:45**

