"Deep Learning for Political Economy"

Melissa Dell

Abstract

Deep learning provides a robust method for learning a mapping between unstructured data (e.g., text, images, audio) and computable representations that can power downstream analyses. These methods, which have already transformed a variety of disciplines, allow us to process traditional data sources at an unprecedented scale and to bring completely new types of data into political economy analyses. Yet taking existing methods off-the-shelf often has significant limitations - particularly for historical applications or those in non-Western societies – given the domain shift from the pre-training corpora that power much of deep learning. This talk will provide an overview of work developing novel datasets and methods for using deep learning to examine social science questions. These include a series of user-friendly open-source packages for deep learning-powered document layout analysis, OCR, record linkage, and other data wrangling tasks, designed to be highly extensible to a diversity of societies. I will also introduce massive-scale open-source text datasets that we curated by applying deep learning to historical newspapers. These are useful both for large-scale pre-training and for social science research. Finally, I will discuss deep learning methods designed to examine the influence of historical media.