

Empirical Research using Big Data in Canadian Political Science

Mickael Temporão, PhD

November 19, 2019

The quantity of data generated by contemporary digital systems has been growing exponentially in the past years, creating extraordinarily large and complex datasets known as "Big Data" (Groves, 2011; Keller et al., 2017; Lazer and Radford, 2017). These datasets, combined with new research methods, allow scholars to study vast amounts of texts, images, video, and audio to address their research questions (LeCun, Bengio and Hinton, 2015). The past decade has seen Canada pioneering empirical research in the field of Artificial Intelligence fostering advances in Machine Learning, and more specifically, establishing Deep Learning, a multilayered form of machine learning, as a standard to discover patterns in massive data sets.

But, can Big Data be leveraged to conduct valid Canadian empirical political science research? To date, Canadian political scientists have not fully embraced these datasets. This can partly be explained by the fact that the knowledge required to leverage these large datasets is broader than that typically included as part of classical political science training. However, there are a number of reasons why political scientists might want to advance in this area.

- Ability to examine new phenomena. While Big Data alone can't solve many of the puzzles that occupy the study of political science, they can provide opportunities to explore new phenomena, many of which were previously invisible to datasets on a smaller scale (Grimmer, 2015). Although limited, there is some evidence of this as Canadian political scientists have begun to employ this new breadth of data in their research in order to test theories that were previously impossible to test empirically due to the lack of data (see Rheault and Cochrane, 2019; Temporão et al., 2018).
- Alternative to survey research. Due to concerns about representativeness, many researchers are seeking alternatives to survey datasets. Big Data can be such an alternative. It is important to note, however, that Big Data are nearly universally nonrandom, which bring new methodological challenges to the table in order to make generalizable claims (Temporão et al., 2019).

For Big Data to advance empirical research in political science, scholars need to acknowledge that the field is evolving into something broader than political science, including machine learning frameworks, that borrow from recent advances in statistics and computer science. This new interdisciplinary research field that is emerging is called computational political science. This field of study combines recent advances in computer science, statistics, linguistics, information technology, and many others, to develop new and scalable methods that facilitate the extraction of valuable insights from Big Data and new opportunities for empirical research.

References

- Grimmer, Justin. 2015. “We are all social scientists now: How big data, machine learning, and causal inference work together.” *PS: Political Science & Politics* 48(1):80–83.
- Groves, Robert M. 2011. “Three eras of survey research.” *Public Opinion Quarterly* 75(5):861–871.
- Keller, Sallie, Gizem Korkmaz, Mark Orr, Aaron Schroeder and Stephanie Shipp. 2017. “The evolution of data quality: Understanding the transdisciplinary origins of data quality concepts and approaches.”
- Lazer, David and Jason Radford. 2017. “Data ex machina: Introduction to big data.” *Annual Review of Sociology* 43:19–39.
- LeCun, Yann, Yoshua Bengio and Geoffrey Hinton. 2015. “Deep learning.” *Nature* 521(7553):436–444.
- Rheault, Ludovic and Christopher Cochrane. 2019. “Word Embeddings for the Analysis of Ideological Placement in Parliamentary Corpora.” *Political Analysis* pp. 1–22.
- Temporão, Mickael, Corentin Vande Kerckhove, Clifton van der Linden, Yannick Dufresne and Julien M. Hendrickx. 2018. “Ideological Scaling of Social Media Users: A Dynamic Lexicon Approach.” *Political Analysis* 26(4):457–473.
- Temporão, Mickael, Yannick Dufresne, Justin Savoie and Clifton van der Linden. 2019. “Crowdsourcing the vote: New horizons in citizen forecasting.” *International Journal of Forecasting* 35(1):1–10.