

Problem Set 1: Part 2

Dan Gamarnik

April 16, 2018

Paper Critiqued: Grimmer, J. (2016). Measuring Representational Style in the House: The Tea Party, Obama, and Legislators' Changing Expressed Priorities.

1. Research Question

In this paper the researcher's primary research question is to investigate how legislators utilize communication as a way to express their representation strategies to their constituents. More specifically, he was interested in how the legislators may change these representation strategies when under the pressure of different political contexts. This paper was interested in the contexts of electoral pressure and instances in which there was a change in party control of congress. Grimmer also identified a methodological research question such that he wanted to investigate how effectively computational methodologies (specifically topic modeling) can operationalize these communication strategies.

2. Data

Grimmer operationalized the idea of legislator's communication strategies through a documentation of every House of Representatives press release from 2005-2010. He uses this data to calculate legislator's "Expressed priorities." His primary declared data method is "text as data"

3. Theory

While this paper is heavier on methodological innovations, one theory that Grimmer implicitly builds into his topic models is degrees of granularity. Grimmer nests more granular and narrow topics into more coarse and broad topics and consequently builds a hierarchy of topics which he analyzed accordingly. He defines his grouping theory as a strategy that groups topics "that use broadly similar language, or language that accomplishes a similar substantive goal," (7).

4. Classification

This paper combines elements of both a descriptive study and an identification study. Grimmer employs elements of identification when he runs simple regressions on the proportion of press releases in his created farming group on the proportion of employed constituents who work in farming to show that legislators who represent farming districts tend to discuss farming more frequently providing validity to his topic model. He also exemplifies elements of a descriptive study through his primary methodology - unsupervised learning. He merely describes the changes in the topics that the unsupervised model produces.

5. Computational Methodologies

As mentioned, Grimmer utilizes text as data and he does this through topic modeling methodology. His methodologies in this paper consisted of constructing a two-layer hierarchy of topics: coarse and granular. These layers allowed him to nest more narrow topics in broader ones and conduct more generalizable analyses. Firstly, he cleaned the text using standard topic modeling techniques. This included discarding word order, punctuation, capitalization. Grimmer then "stemmed the words, mapping words that refer to the same basic

content to a common stem.” Finally, he “removed words that occurred in less than 0.5% of the press releases, words that occurred in more than 90% of the press releases, stop words, and proper nouns that refer to specific Congressional districts, members of Congress, or American cities.” All of these steps constituted simply cleaning that he conducted before arranging the text into his hierarchies. Finally, in establishing these hierarchies Grimmer assumes that “each representative in the House of Representatives, i , divides attention over a set of K topics θ_{ik} , where θ_{ik} represents the proportion of the representative’s press release allocated to topic k .” Results: Grimmer finds that legislators are likely to change their communication strategies after changes in congressional power balances. One such example of this was Republican House members abandoning credit claiming strategies after the 2008 victory of Barack Obama while Democratic House members amplified their credit claiming strategies to appeal to constituents more effectively. However, legislators’ attention to broad, coarse topics has been found to be more stable.

6. Suggestions

There are two suggests I would offer:

1. I found the visualizations of the granular topics to be a bit underwhelming and I think that more extensive visualizations showing his hierarchy strategy might have been helpful for showing why this is a better modeling technique. He also does not elaborate on the disadvantages of this strategy as compared to other topic modeling techniques.
2. Finally, I wish that he had gone into more detail on examples of topics which were decided to be more coarse or granular and the numbers of these topics. This is the most pivotal part of the paper and felt a bit glossed over. I also wish that this paper was a bit more theoretically robust. While the methodological changes were very interesting more theory might have contextualized the implications a bit better.