Politics as a Full-Contact Sport:
A Neural Topic Modeling Approach to Gauging Polarization in Congressional Speeches

Rowan Cassius
University of California, Berkeley
rocassius@berkeley.edu

Maurice Williams
University of California, Berkeley
reese565@berkeley.edu

Abstract

Ideological polarization has been a defining characteristic of today's political landscape. Previous studies have successfully identified patterns in polarized speech between partisans. However, we study the phenomenon of polarized political speech in the United States Congress using a neural topic modeling approach. We define polarization as the divergence in learned topic distributions between the two major political parties. This semi-supervised technique let's us measure the trajectory of political polarization across time from 1873 to 2011, and the result is consistent with prior quantitative research on polarization. Our technique also reveals learned topics that suggest political speeches contain a semantic quality similar to that of gamesmanship observed between competing teams in athletic sports.

1 Introduction

Efforts to measure the history of political polarization have evolved in recent decades from a heavy reliance on the opinions of pundits to approaches which use language as a window into individuals' personal beliefs and biases. The former inherently suffers an obvious shortcoming: it is difficult for a pundit to gauge political polarization uninfluenced by her personal biases, which inevitably fall on one side of the political divide which she attempts to calibrate. For this reason, researchers have instead turned to more computational means of detecting and measuring partisan behaviors. We contribute to these methods by offering a neural network based approach for detecting differences in speech between partisans, and an unsupervised diagnostic of the substantive topics driving them.

2 Objective

Our goal is to contribute to the literature on polarization in the following ways. We wish to expand the tool-kit used to detect partisan differences in speech. This work is most closely related to that done by Gentzkow and Shapiro, where the approach taken is one of estimating a difference in choice distributions between groups, with the choice being the phrase chosen by a speaker (2019[4]). We build on this, framing the choice as one of topic impressions encoded in a sample span of speech.

We refer to topics that our neural topic model will learn as *topic impressions* because some of them will not necessarily represent substantive subjects, such as *health* or *money*. Instead, some of the topics could be better described as attitudes or tonal qualities, such as *volatility* or *quietness*. We find the phrase *topic impression* to adequately encompass both of these topic varieties.

We employ a neural network based approach which enables the use of real-valued vector representations of speech that contain inlaid semantic information. These semantic adhering representations subsequently permit us to infer the topics dividing these groups by training a network to determine which topic impressions are best able to reproduce the represented speech. This topic modeling piece can be contrasted with a procedure where one manually selects the topics and examines the relative differences in the frequency of their occurrence between partisans. Using this design, we hope to gain insights into the more granular issues driving at the partisan wedge over time.

3 Congressional Speech Data

For our task, we use a cleaned dataset of congressional speeches published by Gentzkow et al (2017[3]). We make use of the parsed congressional speeches labeled with speaker identifiers from all congressional sessions along with the speaker metadata characteristics, which include full name, gender, party, state, chamber, district, and speakers' voting privileges.

3.1 Identifying Documents in Congressional Speech

We aim to develop a topic model that is capable of reading a snippet of congressional speech and inferring its membership distribution over over a set of learned topics. In order to identify documents in congressional speech for training the model, we ran a script over all congressional texts which extracted documents likely to contain opinionated speech, .e.g., *"raising taxes on the middle class is abominable"*, as opposed to routine procedural speech, .e.g., *"Mr. Speaker, I yield back the balance of my time."*

The script filters out speeches with fewer than 200 tokens and omits the final 100 tokens of speeches that are not filtered. These choices increase the likelihood that the documents used to train the model are composed mostly of substantive commentary on important subjects of legislation, because both shorter speeches and the endings of long speeches tend to be procedural in nature.

To identify documents for training, the script then splits the remainder of the speech into discrete chunks of 200 words, and each chunk constitutes a document. We use a document size of exactly 200 words because this size is used by Iyyer et al (2016 [5]). After running the script, the dataset contains $3,913,325$ documents, $1,697,124$ of which belong to Republicans, while $2,175,473$ belong to Democrats.

4 Model of Speech

We assume a span of speech can be represented as a semantic adhering representation. This representation is a real-valued vector, where the differing dimensions encode semantic knowledge associated with the speech. This speech representation can then be characterized as a convex combination of topic impressions whose weights are varied by the speaker to create the desired speech substance. We thus hypothesize the following model:

$$y_i = \mathbf{R} p_i \tag{1}$$

where y_i is the vector representation for speech document i, $\mathbf{R}$ is a matrix whose columns are topic impressions, and p_i is the weighting distribution ($p_{i_k} \in (0, 1)\ \forall\ k$, and $\sum_k p_{i_k} = 1$) over the topic impressions selected by the speaker to produce document i.

5 The Neural Topic Model

We base our neural method on work done by Iyyer et al. (Iyyer, et al., 2016 [5]), who implemented their Relationship Modeling Network (RMN) to examine character-character relationship trajectories in literary works. Our variation on this neural model differs with the absence of the recurrent connection in the architecture they use. We justify this omission by assuming that congressional members are not bound by the same narrative constraints that are imposed on literary characters, where a plot must evolve in a somewhat steady manner.

We choose to adapt the RMN as a topic model instead of using a more common token based model, like Latent Dirichlet Allocation (Blei et al., 2003 [1]) for two prominent reasons. Since an RMN architecture exploits the rich semantic information contained in word embeddings, the topics it learns are not restricted to being represented through the potentially narrow and especially formal set of words which tend to make up congressional language. Secondly, the use of word vectors are also critical for handling variation in word choices and phrases over time. The congressional speech data straddles 3 different centuries, and the use of embeddings allows for distinct words with a common meaning to occupy a similar region in a word embedding space and thus be represented similarly.

5.1 Model Description

Let $(w_1, w_2, ..., w_n)$ be the set of words constituting document d_i which the model will learn from, and let $(v_1, v_2, ..., v_n)$ be the words' associated GloVe vectors in $\mathbb{R}^d$ ([8]). Iyyer et al choose to represent each document i in $\mathbb{R}^d$ by taking an element-wise average of $v_1, ..., v_n$ ($y_i = 1/n \oplus_j v_j$). However, our model performed best when we scaled each word vector by its inverse document frequency score:

$$y_i = \frac{1}{n} \sum_{j=1}^{n} v_j \times \mathrm{idf}(w_j)$$

Applying inverse document frequency weighting causes the document representation to give more weight to low-frequency words in the corpus. Our logic is that low-frequency words are more indicative of a document's meaning than high-freqency words.

Additionally, to ensure that the vector representation of a document is most representative of the subjects it discusses, we set the word vectors of stopwords to be zero ($v_j = \mathbf{0}$ if $j \in$ stopwords). For this task, we define the set stopwords to be the union of the following sets of words:

- NLTK's list of standard English stopwords.
- Most frequent 100 words in the corpus.
- Names of all US states.
- First, middle and last names of every congress member or president in history.
- Stopwords manually identified by Gentzkow et al. (e.g., *hereabout, therein*)
- Stopwords manually identified by us (e.g., *ms, speaker, madam*)

For each document d_i, y_i is only one of 7 inputs to the model. We include inputs for each piece of important metadata that is associated with document d_i. This means that we include 50-dimensional embeddings for the following pieces of metadata:

- *Speaker*, a unique identifier for the member who authored the speech snippet.
- *Party*, the party of the speaker.
- *Session*, the congressional session in which the speech was delivered.
- *State*, the state which the speaker represented at the time of the speech.
- *Chamber*, the chamber of the speaker (House or Senate).
- *Gender*, the gender of the speaker.

Iyyer et al. use both book and character embeddings to capture immutable aspects of characters which may be relevant to character-character relationships. By a similar logic, we include embeddings for all of the metadata characteristics listed above because each one could conceivably inform a speaker's word, subject, tonal choices. Most obviously, a speaker's political party is likely to be relevant to what they say. In addition, each speaker is likely to have their own linguistic idiosyncrasies, which can be captured by the speaker embedding.

For each document i and all of its associated metadata, the RMN computes a discrete distribution p_i over K d-dimensional topic vectors. The matrix of topic vectors is referred to by Iyyer et al. as the descriptor matrix, $\mathbf{R} \in \mathbb{R}^{K \times d}$, so the reconstructed document span is $\hat{y}_i = \mathbf{R} p_i$. Like Iyyer et al, we use a hinge loss function to train the reconstruction of documents:

$$J(\theta) = \frac{1}{|B|} \sum_{i \in B} \max(0, 1 - y_i^T \hat{y}_i)$$

(where B denotes the indices of a particular batch of documents) and we penalize overly similar topics:

$$X(\theta) = \left\| \mathbf{R}\mathbf{R}^T - \mathbf{I} \right\|$$

In the initial stages of training, the model tended to become overfit by reconstructing the document vectors using only one of the topics, (i.e., if $K = 3$, $p_i = (0.02, 0.97, 0.01)^T$ for nearly all i). In response, we added penalties to the loss function which encourage topic distributions to be unique (distribution similarity penalty) and encourage documents to be reconstructed using multiple (but not too many) topics (high entropy penalty). The distribution similarity penalty is:

$$S(\theta) = \sum_{\substack{i,j \in B \\ i \neq j}} p_i^T p_j$$

and the entropy penalty is:

$$E(\theta) = \frac{1}{|B|} \sum_{i \in B} H(p_i)$$

where $H(\cdot)$ is the Shannon entropy of a discrete probability distribution. Thus, our complete loss function can be expressed as:

$$L(\theta) = J(\theta) + \lambda X(\theta) + \gamma S(\theta) + \omega E(\theta)$$

5.2 Model Training Spec

We trained our variant of the RMN using 100-dimensional GloVe word vectors, and embedded each piece of metadata using 50-dimensional vectors. We found success when training a model with 100 topics ($K = 100$) with the following regularization parameters: $\lambda = 0.001, \gamma = 0.1, \omega = 0.01$, and a random word dropout of 0.5 (Iyyer et al, 2016 [5]). Finally, we optimize the model's parameters using Adam (Kingma and Ba, 2014 [7]) with a learning rate of 0.001 for 1 epoch using all 3,913,325 documents.

6 Learned Topics

According to our analysis, the RMN learned nearly all 100 topic impressions quite well. Each topic $t_k \in \mathbb{R}^d$ can be understood by looking at the nearest word vectors surrounding it the space of GloVe word embeddings. We judge the coherence of a given topic according to its nearest neighbors. Intuitively, if the nearest neighbors of t_k all appear relevant to one another, this is a sign that t_k represents a coherent idea.

However, coherent nearest neighbors are not necessarily a sign of well-learned topics, since any vector in $\mathbb{R}^d$, even a randomly chosen vector, will have nearest neighbors which may conceivably appear relevant to one another. In effect, the seeming coherence of our topics may not in fact be a property of a good model fit, but instead a property of well-trained word embeddings. As such, we measure the coherence of our learned topics formally so that we can eventually compare their coherence to the coherence of randomly chosen word vectors. We measure the coherence of topic t_k using the average pairwise cosine similarity of its nearest neighbors:

$$\mathrm{coh}(t_k) = \frac{1}{(|N|)_2} \sum_{\substack{i,j \in N \\ i \neq j}} \cos(v_i, v_j) \qquad (2)$$

where N is the set of nearest neighbors for topic t_k, and cos denotes cosine similarity.

Learned Topics			
Rank	coh	Label	Nearest Neighbors
1	0.98	years	*1899, 1902, 1903, 1907, 1909*
25	0.69	regulation	*stricter, stringent, strict, regulations, regulatory*
50	0.62	crime	*trafficking, crime, prostitution, offenses, narcotics*
75	0.51	mourning	*solemn, glow, somber, mournful, glowing*
100	0.22	religious doubt	*doubtful, jehovah, abbot, regard, heretics, hades*

Figure 1: Coherence rank, coherence score, label and nearest neighbors of a sample of 5 learned topics at the 100%, 75%, 50%, 25% and 0% percentiles of coherence.

The coherence scores across all 100 topics have a mean of 0.606 and a standard deviation of 0.126. Figure 1 displays a sample of the learned topics with a wide variety of coherence scores and their nearest neighbors.

In order to validate the topics' coherence scores as a result of well-learned topics and not merely a consequence of the topics existing in a well-learned word embedding space, we drew 1000 random vectors from $\mathbb{R}^d$ and estimated the distribution of the coherence score of a randomly drawn vector. That is, we sampled from the distribution $\boldsymbol{U}(-1,1)^d$, and computed the associated coherence scores of the random draws as if considering them to be random topics. Appendix A shows a sample of the random topics, their coherence scores, and their associated sets of nearest neighbors.

A glance at the nearest neighbors of random topics in Appendix A shows that they tend to be overwhelmingly less comprehensible than learned topics are. The overall distribution of the coherence scores of 1000 random vectors is right skewed with a mean of 0.275 (0.331 less than the mean coherence of the learned topics) and standard deviation of 0.098. Its 95% quantile is 0.456, which is lower than *90 out of 100* of the learned topics' coherence scores. This serves as a strong indication that the model's impressive topic coherence scores are meaningful.

Lastly, We examined the relationship between how often topics are used and how coherent they are according to the coherence metric described in (2). A topic's use is defined as the mean amount of probability assigned to it across all documents' topic distributions. Topic use ranges from 0.4% to 1.7% across the model's 100 topics. There is a highly statistically significant positive linear relationship between topic use and topic coherence (see appendix B). This fact strongly affirms the value of the topic coherence measure we put forward, because one would intuitively expect that the most frequently used topics are also the most coherent.

7 Measuring Polarization

Given the model of speech laid out in (1), we propose that partisan differences in a set of speech documents can be understood by examining the differences in the topic use distributions between partisan groups. Suppose you draw a Republican document at random from a particular session s. We refer to the expected value of the random topic distribution vector that will result as the Republican topic use distribution in session s, our notation for which is $p^{(R)}$, and similar for the Democrat topic use distribution ($p^{(D)}$). Since $p^{(R)} = \mathbb{E}(p_i)$ if i is drawn from the set of Republican documents, $p^{(R)}$ is estimated as follows:

$$\hat{p}^{(R)} = \frac{1}{|R \cap S|} \bigoplus_{i \in R \cap S} p_i$$

where R is the set of document belonging to Republicans, S is the set of documents belonging to session s, and $\oplus$ denotes an element-wise sum.

With both the Republican and Democrat topic use distributions, we may then derive an estimate of the amount of polarization present in session s by computing the divergence between the two topic use distributions:

$$Div_s = JS(\hat{p}^{(R)}, \hat{p}^{(D)}) \tag{3}$$

$JS(\cdot,\cdot)$ is the Jensen-Shannon divergence of two probability distributions and is bounded between 0 and 1. We use Jensen-Shannon divergence as a measure of dissimilarity between topic use distributions because it is symmetric and has previously been used by other researchers to understand differences in individuals' language use (Corritore et al., 2019, [2]).

Alone, the measure Div_s may not be sufficient to gauge polarization in a given session. Even if there are no fundamental differences between the the partisan groups being compared, it is highly unlikely that the groups' topic use distributions will be exactly the same. Instead, it is likely that they will differ slightly due to chance, and the Jensen-Shannon divergence between them will be trivially greater than zero.

To ensure that we use a polarization metric that is robust against the possibility that the observed divergence is due to chance, we derive an adjusted polarization score for session s. We first estimate the the divergence that would arise due to chance if there were no systematic difference between Republicans and Democrats by randomly reassigning party labels and computing the metric in (3):

$$Div_{s,\text{random}} = JS(\hat{p}^{(R^*)}, \hat{p}^{(D^*)}) \tag{4}$$

where $\hat{p}^{(R^*)}$, and $\hat{p}^{(D^*)}$ are the topic use distributions after party labels have been taken from the documents in S and then randomly reassigned. Because $Div_{s,\text{random}}$ is itself a random variable, we reassign party labels and recompute it 100 times, and then take an average so that we get a confident estimate of $\mathbb{E}(Div_{s,\text{random}})$.

Finally, our operative polarization score, which we refer to as *adjusted divergence*, is the difference between the observed (true) divergence between the Republicans' and Democrats' topic use distributions and the estimated divergence that would arise due to chance (random divergence):

$$Pol_s = Div_s - \overline{Div}_{s,\text{random}} \tag{5}$$

By doing this subtraction, we are ensuring that the polarization score consists of the divergence between Republicans and Democrats that is due to meaningful differences between them and not due to chance party assignment.

The adjusted divergence series $\{Pol_s\}$ is displayed in figure 2, along with a smoothing spline and a 95% confidence band for the conditional mean. The series has a distinctive u-shape, and the smooth increase in the adjusted

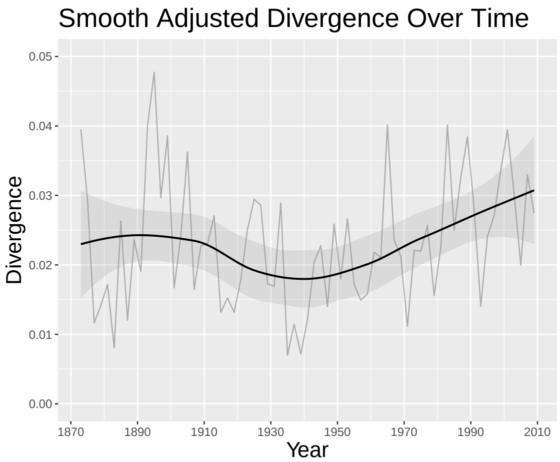


Figure 2: Adjusted divergence time series with a smoothing spline and 95% confidence interval for the conditional mean

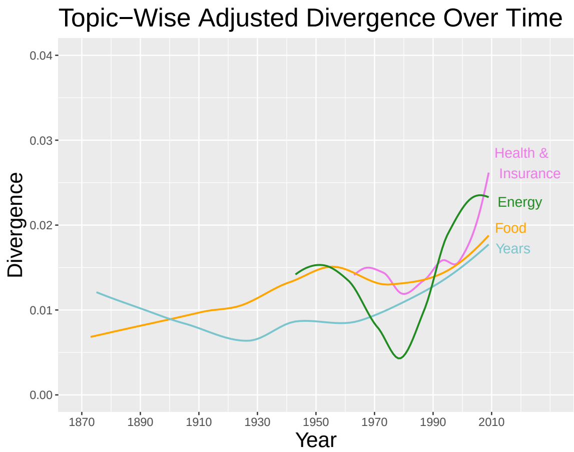


Figure 3: Topic-wise adjusted divergence smoothing splines for the following topics: Health & Insurance, Energy, Food, and Years. The Health & Insurance and Energy divergence trends do not begin in 1873 because early sessions lack sufficient numbers of documents that are primarily about Health & Insurance or Energy.

divergence score from the 1940's onward reflects the modest increase in polarization that followed the post-war period, leading into the cultural upheaval of the 1960's and the economic turbulence of the 1970's.

7.1 Consistency with Prior Research

The polarization series we calculate is also consistent with prior research that has attempted to quantify polarization. We can derive a polarization metric from the DW-Nominate ideology scores computed from spatial models of roll call voting developed by McCarty, Poole and Rosenthal (1997 [6]). We take the difference between average ideology score of Republicans and the same of Democrats for every congressional session between 1873 and 2009, and we call this polarization series $\{Pol_{s,\text{nominate}}\}$. $\{Pol_{s,\text{nominate}}\}$ is also characterized by a u-shape (see appendix E). Additionally, $\{Pol_s\}$ is significantly cointegrated with $\{Pol_{s,\text{nominate}}\}$ according to an Engle–Granger two-step test for cointegration (p-value < 0.01).

In short, we consider our overall polarization series to be valid for two reasons: it is consistent with historical intuitions about the evolution of political polarization, and it shares a common statistical trend with polarization scores derived from previous research. Note that the latter reason is especially convincing and intriguing given that the DW-Nominate ideology scores are derived purely from congressional voting data.

7.2 Polarization: Just a Difference in Subject Emphasis?

We take yet another step to illustrate that the polarization series we computed is indeed measuring polarization. It could be the case that the observed divergence in topic use distributions is due to simple differences in the frequencies with which each party discusses particular subject matters. Disparities in subject discussion frequencies are still an indication of polarization, but, alone, they reflect differences in *what* subjects partisans are talking about, not necessarily *how* they are talking about them. The latter is slightly closer to most people's understanding of polarization.

We provide evidence that $\{Pol_s\}$ is reflecting substantive divergence between partisans, not merely divergence that arises from mentioning subjects with different frequencies. To do this, we repeat or polarization estimation procedure but restrict it to documents which all share the same primary topic, that is the topic assigned the most probability for document i. For example, one of the topics learned by the model is labeled "Health & Insurance", with a coherence score of 0.673. Let H be the set of documents whose primary topic is Health & Insurance. We compute Republicans topic use distribution for Health & Insurance as follows:

$$\hat{p}^{(R)}_{\text{health}} = \frac{1}{|R \cap S \cap H|} \bigoplus_{i \in R \cap S \cap H} p_i$$

and do similarly for Democrats. We then compute a Health & Insurance polarization series by doing the same polarization estimation outlined by equations (3),(4) and (5) but for the distributions $\hat{p}^{(R)}_{\text{health}}$ and $\hat{p}^{(D)}_{\text{health}}$.

The Health & Insurance polarization series's smoothing spline is displayed in figure 3 (in pink), along with smoothing splines for the separate Energy, Food and Years polarization series. The relative ordering in the most recent session of congress is sensible. Of the four subjects represented, Health & Insurance and Energy are the most contentious subjects between the parties, and Years is the least polarized, as it intuitively ought to be.

The sensibility of a sample of the topic-wise adjusted divergences, and their consistency with the overall adjusted divergence series, serves as evidence that the polarization scores are not merely capturing differences in subject choice between the parties, but substantive differences in the ways they talk about common subjects.

8 Model Utility

The model may be used not only to describe documents' content at a high level, but also to investigate the granular differences between two particular congressional speakers on a substantive topic. For example, the following table reports the topic use distributions of Senator Mitch McConnell (R-KY) and Senator Harry Reid (D-NV) during the 2009-2011 session of congress.

McConnell and Reid				
#	McConnell	p_k	Reid	Reid p_k
1	budget	10%	recess	8%
2	political parties	9%	thought	6%
3	Middle East	6%	sports	4%
4	healthcare	6%	healthcare	4%
5	sports	5%	slap	3%
6	nomination	3%	political parties	3%
7	slap	3%	attempt	3%
8	attempt	3%	quietly	2%

The difference in the topics which each speaker uses most frequently is somewhat insightful—while McConnell tends to place the most emphasis on budget issues, Reid devotes more of his emphasis to matters of running the Senate, such as legislative recess. This shows that our model is capturing content relevant to that spoken on the House and Senate floors, though some of these topics seem to be more procedural, a indication that our attempts to avoid identifying procedural speech needs improvement.

We also notice the presence of a few odd topics such as *slap* for both Senators, which may suggest that each speaker is wont to reference underhanded ploys from the opposing party. Consider the following snippets of speech as examples:

> Senator Mitch McConnell (R): *Democrats on Capitol Hill are working behind the scenes on a plan aimed at jamming this massive health spending bill through Congress against the clear wishes of an unsuspecting public.*
>
> Topics: quietly, political parties, unfortunately, attempts, football players

> Senator Harry Reid (D): *No matter what Republicans claim. the government has no intention of choosing any part of your medical plan. Remember, we are talking a public option, a public choice.*
>
> Topics: unfortunately, political parties, intent, attempts, slap

We find it both amusing and insightful that the topic model finds it optimal to reproduce McConnell's words above using the topic *football players*, labeled as such for its proximity to a handful of European soccer athletes. This topic strongly reflects the tactfulness and gamesmanship known to characterize the feud between America's prominent political parties.

Moreover, a noticeable portion of the topics refer to sports, sports lingo, or sports players. It's possible that our model most closely associates the semantics inlaid in this sort of low deliberation with the speech used to narrate athletic sports. Nonetheless, we think that we should carry out additional work to confirm the accuracy of the topic model in order to have greater confidence in the sports-like nature of rivalrous political speech as one of our core findings.

9 Future Work

While there is evidence that our model is capturing lexical differentiation between the two major political parties, there are several aspects of the model that provide opportunities for improvement.

The choice to use a neural approach for topic modeling over more common topic models, such as LDA, is motivated partly by the ability of our model's more flexible form to incorporate additional metadata as input. A model that leverages a speaker's gender and regional information, which may account for some portion of lexical diversity, could help distinguish divergence attributable to these traits from divergence arising from political leanings. The current iteration only nominally incorporates this additional information, relying primarily on a document's text to encode its topic impressions. A more convincing model should account for other types of speaker attributes in a way which allows their influence on a document's topics to be distinguished from one another.

The speech model currently assumes that a span of speech is a strictly additive linear combination of topic impressions. An alternative model of speech could allow for operations more akin to those used to assess embedding's with analogies, such as subtraction.

The topic model architecture has additional opportunities to provide more coherent topics. The number of topics K can be scaled or reduced to meet a desired level of topic granularity. Additional sensitivity analysis should be performed to identify optimal parameter values that produce sufficiently coherent topics and more valuable insights into areas of political contention.

Finally, another key area for improvement is in document curation. Imposing restrictions on speech length proved a limited strategy for increasing the recall of opinion rich spans of text. One possible idea for improving the quality of documents used in training and inference is to create a text classifier that is capable of separating speeches discussing procedural matters from those in which speakers expose their true opinions. Doing this could substantially increase the precision with which opinion-rich documents are identified.

10 Conclusion

In this project, we trained a neural topic model using over a century's worth of congressional speeches delivered on the House and Senate floors. The model learned 100 rather coherent topics, with which we then attempted to measure political polarization between the Democratic and Republican parties over time. We inferred overall and topic-wise polarization trends that are consistent with historical intuition and prior research done to measure polarization using an entirely different form of data (roll call voting). Moreover, we also illustrated how the model can used to infer the granular differences between partisans, and also found that polarized speech explicitly regarding party rivalries can be semantically similar to speech used to narrate athletic sports.

References

[1] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation". In: *J. Mach. Learn. Res.* 3.null (Mar. 2003), pp. 993–1022. ISSN: 1532-4435.

[2] Matthew Corritore, Amir Goldberg, and Sameer B. Srivastava. "Duality in Diversity: How Intrapersonal and Interpersonal Cultural Heterogeneity Relate to Firm Performance". In: *Administrative Science Quarterly* 0.0 (0), p. 0001839219844175. DOI: 10.1177/0001839219844175. eprint: https://doi.org/10.1177/0001839219844175. URL: https://doi.org/10.1177/0001839219844175.

[3] Matthew Gentzkow, Jesse M Shapiro, and Matt Taddy. "Congressional Record for the 43rd-114th Congresses: Parsed Speeches and Phrase Counts". In: *Stanford Social Science Data and Software* (Jan. 2018). URL: https://data.stanford.edu/congress_text.

[4] Matthew Gentzkow, Jesse M Shapiro, and Matt Taddy. "Measuring Group Differences in High-Dimensional Choices: Method and Application to Congressional Speech". In: *Econometric Society* (July 2019). URL: `http://web.stanford.edu/~gentzkow/research/politext.pdf`.

[5] Mohit Iyyer et al. "Feuding Families and Former Friends: Unsupervised Learning for Dynamic Fictional Relationships". In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. San Diego, California: Association for Computational Linguistics, June 2016, pp. 1534–1544. DOI: `10.18653/v1/N16-1180`. URL: `https://www.aclweb.org/anthology/N16-1180`.

[6] Howard Rosenthal Jeffrey B. Lewis Keith Poole. "Congressional Roll-Call Votes Database". In: (2020). URL: `https://voteview.com/`.

[7] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. arXiv: `1412.6980 [cs.LG]`.

[8] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. "GloVe: Global Vectors for Word Representation". In: *Empirical Methods in Natural Language Processing (EMNLP)*. 2014, pp. 1532–1543. URL: `http://www.aclweb.org/anthology/D14-1162`.

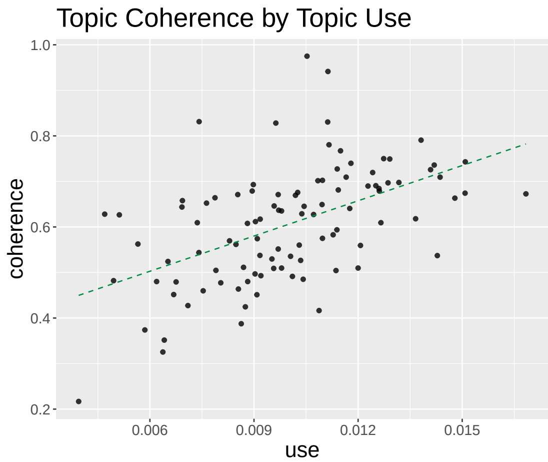
Appendices

A Table of Random Topics

Sample of Random Topics		
coh	Label	Nearest Neighbors
0.22	none	*riddle, stockton, keun, kersee, duluth*
0.23	none	*stamps, reauthorize, enema, appropriation, physiotherapy*
0.30	none	*fabriclive, exceptionalism, adobo, reaganesque, s400*
0.22	none	*belleza, witz, dyfi, qiantang, tropospheric*

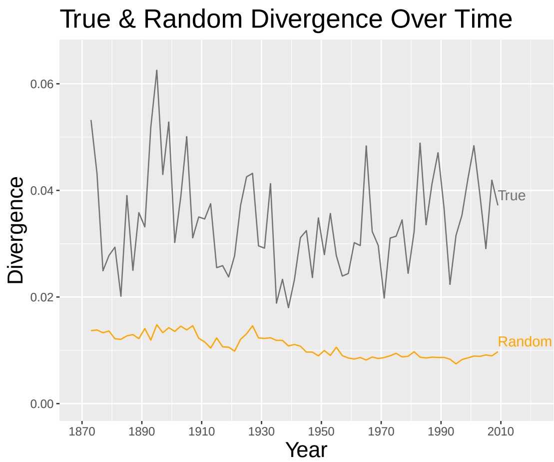
A: Coherence scores and nearest neighbors of a sample of 4 randomly drawn vectors in $\mathbb{R}^{100}$.

B Topic Use by Topic Coherence



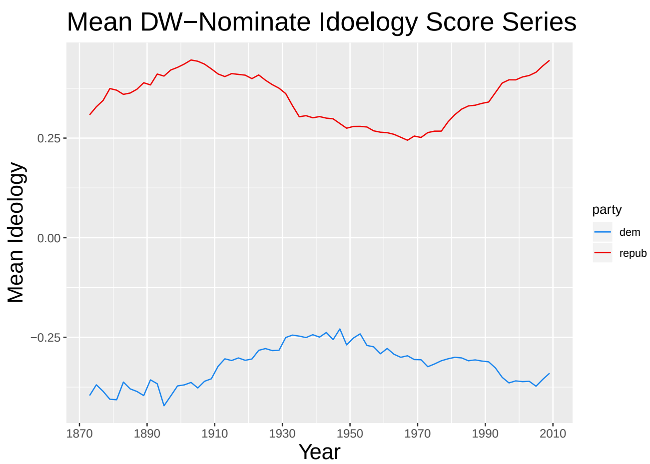
B: Scatter plot of topic use (x-axis) by topic coherence (y-axis) with an overlaid regression line. A linear relationship is highly statistically significant (p-value < 0.001).

C True & Random Divergence



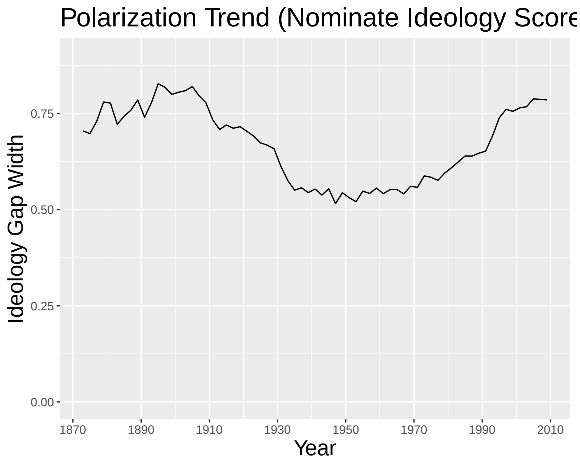
C: True overall divergence and random overall divergence over time.

D Nominate Ideology Score Trends



D: Mean DW-Nominate Ideology Score for Republican and Democratic Parties over time ([6]).

E Nominate Polarization Series



E: Nominate polarization trend, derived by subtracting the mean Republican ideology score (displayed in appendix D) from the mean Democratic ideology score in each session (displayed in appendix D).