Detecting Polarized Topics Using Partisanship-aware Contextualized Topic Embeddings

Zihao He[1,4], **Negar Mokhberian**[1,4], **António Câmara**[2], **Andrés Abeliuk**[3], **Kristina Lerman**[4]

[1]Department of Computer Science, University of Southern California
[2]Department of Computer Science, Columbia University
[3]Department of Computer Science, University of Chile
[4]Information Sciences Institute, University of Southern California

{zihaoh,nmokhber}@usc.edu   a.camara@columbia.edu
aabeliuk@dcc.uchile.cl   lerman@isi.edu

Abstract

Growing polarization of the news media has been blamed for fanning disagreement, controversy and even violence. Early identification of polarized topics is thus an urgent matter that can help mitigate conflict. However, accurate measurement of topic-wise polarization is still an open research challenge. To address this gap, we propose **P**artisanship-**a**ware **C**ontextualized **T**opic **E**mbeddings (**PaCTE**), a method to automatically detect polarized topics from partisan news sources. Specifically, utilizing a language model that has been fine-tuned on recognizing partisanship of the news articles, we represent the ideology of a news corpus on a topic by *corpus-contextualized topic embedding* and measure the polarization using cosine distance. We apply our method to a dataset of news articles about the COVID-19 pandemic. Extensive experiments on different news sources and topics demonstrate the efficacy of our method to capture topical polarization, as indicated by its effectiveness of retrieving the most polarized topics.[1]

1   Introduction

The media environment has grown increasingly polarized in recent years, creating social, cultural and political divisions (Prior, 2013; Fiorina and Abrams, 2008). Although a diversity of opinions is healthy, and even necessary for democratic discourse, unchecked polarization can paralyze society by suppressing consensus required for effective governance (Tworzecki, 2019). In more extreme cases, polarization leads to disagreement, conflict and even violence. The COVID-19 pandemic has exposed many of our vulnerabilities to the pernicious effects of polarization. Public opinions about COVID-19 (Jiang et al., 2020), as well as messaging by political elites (Green et al., 2020; Bhanot

and Hopkins, 2020), are sharply divided along partisan lines. According to a Pew Report (Jurkowitz et al., 2020), partisanship significantly explains attitudes about the costs and benefits of various mitigation strategies, including non-pharmaceutical interventions and lockdowns, and even explains regional differences in the pandemic's toll in the US (Gollwitzer et al., 2020).

In mass media a variety of topics is discussed every day, and polarization can form on different topics. Therefore, identifying nascent disagreements and growing controversies of different topics in news media and public discourse would help journalists craft more balanced news coverage (Lorenz-Spreen et al., 2020; Chen et al., 2020). Different from previous works that study polarization from a more coarse-grained perspective, Demszky et al. (2019) were the first to study polarized topics using tweets about 21 mass shootings to show that some topics were more polarized than others. However, their approach to represent semantic information with word frequencies is less expressive than modern methods allow.

To better capture the topical polarization among partisan (liberal vs. conservative) media sources, we propose **P**artisanship-**a**ware **C**ontextualized **T**opic **E**mbeddings (**PaCTE**). Specifically, given a text corpus containing news articles from both sides, we first extract a set of topics utilizing LDA topic modeling (Blei et al., 2003). Next, we fine-tune a pretrained language model (Devlin et al., 2018) to recognize the partisanship of the news articles so as to render it *partisanship-aware*. Then for each article, we represent its ideology on a topic by a vector, called document-contextualized[2] (DC) topic embedding, by aggregating language model representations of the topic keywords contextualized by the article. Such a representation sheds light primarily on the tokens that appear in the topic keywords and thus concentrates on the topic-

---

[2]We use "article" and "document" interchangeably.

oriented local semantics in the context of the article, instead of the global semantics from the article that might contain irrelevant and noisy information. We further represent the ideology of the news corpus on the topic, what we call corpus-contextualized (CC) topic embedding, by aggregating the DC topic embeddings. As a result, the ideology of the news corpus on a topic is represented by a single vector. Finally, we measure the polarization between two news sources on the topic using the cosine distance between such vectors.

For evaluation, we create ground truth by annotating the polarization of pairs of partisan news sources on a variety of topics. We evaluate the topic polarization scores produced by PaCTE against the ground truth on the task of polarized topics retrieval. Experiments on nine pairs of partisan news sources demonstrate that compared to baselines, PaCTE is more effective in capturing topic polarization and retrieving polarized topics. We argue that public media watchdogs and social media platforms can utilize such a simple-yet-effective tool to flag discussions that have grown divisive so that action could be taken to reduce partisan divisions and improve civil discourse.

2   Related Work

The partisan polarization in the US media is a widely studied topic (Hollander, 2008; Stroud, 2011). During the onset of the COVID-19 pandemic, the polarization among the political elites and the news media causes a lot of confusion. For example, Hart et al. (2020) show that COVID-19 media coverage is politicized and polarized. Other works have been studying the polarization in media from different perspectives. Focusing on the differences in the languages of liberals and conservatives, KhudaBukhsh et al. (2020) analyze political polarization on YouTube using machine translation tools. To analyze how the news outlets frame the events differently, Fan et al. (2019) have collected and labeled 100 triplets of news articles each discussing the same event from three news sources bearing different political ideologies.

In addition to qualitatively analyzing polarization, different approaches to quantifying polarization have also been proposed. Gentzkow et al. (2019) propose two different ways, namely the leave-out estimator and the multinomial regression, to measure the trends of partisanship in congressional speech. Green et al. (2020) define the po-

larization as one's ability to identify the partisanship of a tweet's author based on the contents of tweets and investigate the polarization regarding COVID-19 among political elites on Twitter. Demszky et al. (2019) first measure topic-wise polarization using the leave-out estimator proposed by (Gentzkow et al., 2019); however, they use a token frequency vector to represent an article, which is less expressive and fails to make use of the rich semantics in the context and the pre-knowledge in pretrained language models (Devlin et al., 2018; Liu et al., 2019) or pretrained word embeddings (Mikolov et al., 2013; Pennington et al., 2014); furthermore, they represent the topic using the token frequency vector of the entire document, thus incurring noisy information that might smooth over the target semantics in the locality of topic keywords. In contrast, our method represents the topic embedding in the context of a document, thus generating topic representations with more attention to the target topic keywords as well as making use of the contextualized semantics from the document, as captured by the contextualized embeddings.

Some works have proposed contextualized embeddings to enhance the quality of neural topic models (Bianchi et al., 2020; Chaudhary et al., 2020). However, the scope of this work is to generate better contextualize topic embeddings for articles to capture topic polarization, with a given topic model; the exploration of other topic modeling techniques is beyond the scope of this work.

3   Methodology

The proposed PaCTE framework consists of four components: 1) LDA Topic Modeling, 2) Partisanship Learning, 3) Partisanship-aware Contextualized Topic Embedding Generation, and 4) Measuring Polarization and Ranking Topics. The overall framework is illustrated in Figure 1. In this section we elaborate on each component in detail.

3.1   Problem Definition

The input is a liberal news corpus $D^L = \{d_i^L\}_{i=1}^{|D^L|}$ and a conservative news corpus $D^R = \{d_i^R\}_{i=1}^{|D^R|}$ (L denotes "Left" and R denotes "Right"), where d_i^L is an article from D^L and d_i^R is an article from D^R. A news article is represented as a sequence of tokens: $d_k = (w_i^k)_{i=1}^{|d_k|}$. Given a topic model trained on the combined corpus $D^C = D^L \cup D^R$ with a set of modeled topics $T = \{t_i\}_{i=1}^K$ where t_i represents a topic, we aim to learn a model f
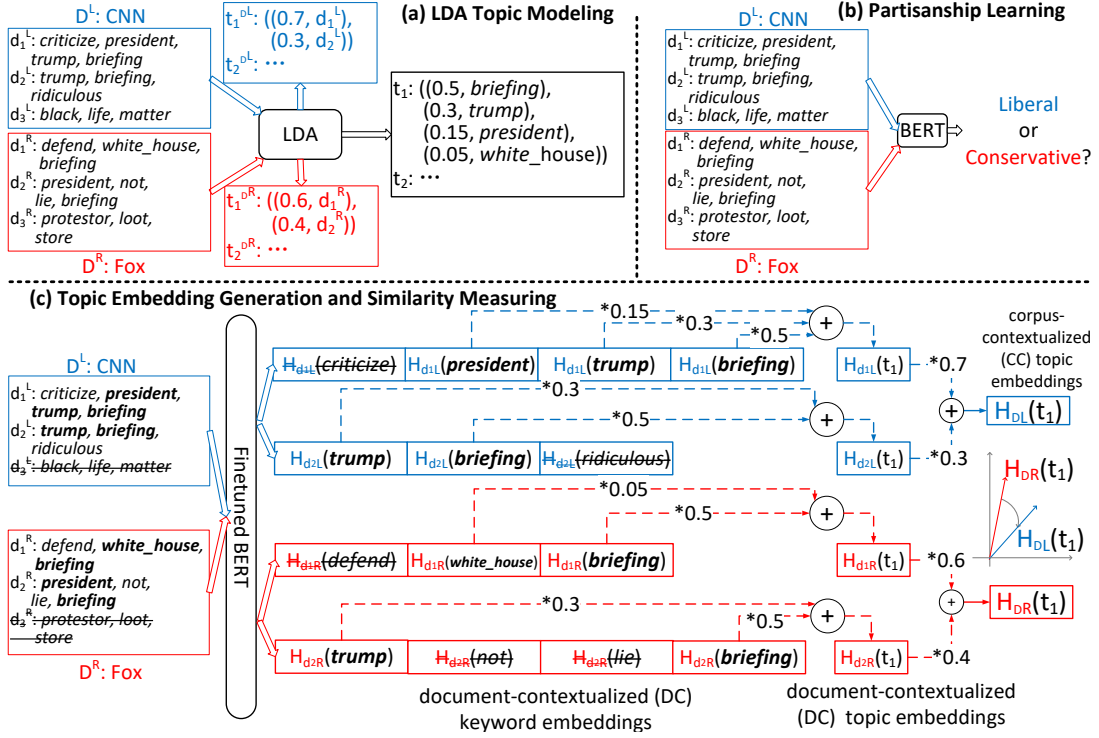
Figure 1: Overview of our PaCTE framework to detect polarized topics in media, illustrated by a toy example on CNN vs. Fox, both consisting 3 documents. **(a)** *LDA topic modeling.* We train an LDA model on the combined corpus and extract 2 topics. Top-4 keywords on topic t_1 are "*briefing*", "*trump*", "*president*" and "*white_house*". Top-2 most relevant documents on topic t_1 are d_1^L and d_2^L for CNN and d_1^R and d_2^R for Fox. d_3^L and d_3^R are not among the most relevant documents of this topic and are excluded in the embedding generation step. Note that we set $K = 2$ (No. of topics), $m = 4$ (No. of keywords), and $n = 2$ (No. of documents), just for clear demonstration. **(b)** *Partisanship learning.* We finetune a pretrained language model to classify the partisanship (liberal vs. conservative) of input documents. **(c)** *Topic embedding generation and similarity measuring.* We provide a step by step illustration of DC keyword embedding → DC topic embedding → CC topic embedding on topic t_1. In the two input corpora, the tokens that are among the top-4 keywords of topic t_1 are highlighted in bold. Take document d_1^L from CNN as an example. The weighted average of the DC keyword embeddings ($H_{d_1^L}(president)$, $H_{d_1^L}(trump)$, and $H_{d_1^L}(briefing)$) is defined as the DC topic embedding $H_{d_1^L}(t_1)$ with keyword coefficients given by Eq. 2; note that $H_{d_1^L}(criticize)$ is excluded because "*criticize*" is not among the top-4 keywords of topic t_1. Similarly we can obtain the DC topic embeddings for d_2^L, d_1^R and d_2^R. The DC topic embeddings are further aggregated into CC topic embeddings $H_{D^L}(t_1)$ and $H_{D^R}(t_1)$ (document coefficients are from Eq. 3) and the cosine distance between them is used as a measure of polarization of the two corpora on topic t_1.

that is able to detect the topic polarization between D^L and D^R on topics in T and output a ranking of topics based on polarization, such that

$$f(D^L, D^R, T) = (t_k)_{k=1}^K,$$
$$i > j \Leftrightarrow \beta(t_i, D^L, D^R) < \beta(t_j, D^L, D^R), \quad (1)$$

where $\beta(t, D^L, D^R)$ represents the polarization score of topic t between D^L and D^R.

3.2 LDA Topic Modeling

We train an LDA topic model using the the combined corpus $D^C = D^L \cup D^R$ and extract K topics $T = \{t_i\}_{i=1}^K$, where t_i is a topic. The modeled topics T apply to both D^L and D^R. An example is given in Figure 1(a).

Representing a topic by keywords. A topic t_i is represented as a distribution of keywords from the global vocabulary of D^C and we only keep the top-m keywords:

$$t_i = ((p_{ij}, w_j))_{j=1}^m, p_{ij} > p_{ik} \Leftrightarrow j < k, \quad (2)$$

where p_{ij} is the probability of observing keyword w_j given topic t_i.

Representing a topic by documents. A document $d \in D^C$ is represented as a distribution over the K topics. Accordingly, we renormalize the probabilities and represent each topic t_i as an (inverse) distribution of documents in D^C and only

keep the top-n most relevant documents, such that

$$t_i^{D^C} = ((q_{ij}^{D^C}, d_j))_{j=1}^n, q_{ij}^{D^C} > q_{ik}^{D^C} \Leftrightarrow j < k, \quad (3)$$

where q_{ij}^D is the probability of observing document $d_j \in D^C$ given topic t_i. Because our goal is to study the polarization between D^L and D^R, instead of using the global documents in D^C, we represent a topic by the top-n documents in D^L and D^R separately and thus obtain $t_i^{D^L}$ and $t_i^{D^R}$ accordingly.

3.3 Learning Partisanship

As we will see in Sections 3.4 and 3.5, the contextualized topic embeddings are generated from a pretrained language model (Devlin et al., 2018) and cosine distance between the topic embeddings from two corpora are used as a measure of topic polarization. The idea is inspired by static word embedding models like GloVe (Pennington et al., 2014), where the authors measure the similarity between words by the cosine similarity between the word embeddings.

However, to apply this measure of similarity, the model should be fitted on the target corpus. To fit the pretrained language model on the news corpora, we can use one of the two training tasks: masked language modeling or partisanship recognition. We decide on the second task because 1) it is more time efficient; 2) it informs the language model of the partisan divisions between different news sources, enhancing the language model's ability to encode the polarization arising from partisan differences in its output. This idea is similar to (Webson et al., 2020) where the authors call the embedding space of the language model after finetuning as "connotation space". As a result, given a document $d \in D^C$, the model is optimized to classify whether it is from D^L or D^R by a binary cross-entropy loss, where the [CLS] embedding is used to represent the document, as shown in Figure 1(b).

3.4 Partisanship-aware Contextualized Topic Embedding Generation

Denote the ideology embedding of A on B as $H_A(B)$, where A represents a news corpus or a document and B represents a topic or a topic keyword. We then represent the ideology of a corpus D on a topic t as corpus-contextualized (CC) topic embedding $H_D(t)$, the ideology of a document d on a topic t as document-contextualized (DC) topic embedding $H_d(t)$, and the ideology of a document d on a topic keyword w as DC keyword embedding $H_d(w)$. We will elaborate on how the CC topic embedding is obtained from a top-down perspective.

According to Equation 3, in order to compute the CC topic embedding $H_D(t_i)$, we can rewrite it as

$$H_D(t_i) = \sum_{j=1}^n q_{ij}^D H_{d_j}(t_i). \quad (4)$$

Hence, we decompose a CC topic embedding into DC topic embeddings from the top-n most relevant documents.

To obtain the DC topic embedding, Demszky et al. (2019) use word frequency vectors; Grootendorst (2020) takes the [CLS] embedding of a pretrained language model that gives a holistic document embedding without encoding the context of a topic. However, while word frequency vectors encode statistical features of words in the document, they neglect their context. In addition, a document is likely to be associated with multiple topics according to the LDA topic model, and therefore using the holistic document embedding as the topic embedding regardless of the specific topic results in identical embeddings for different topics on the same document; moreover, even if a document is only associated with one topic, it might contain information not relevant to that topic and thus the holistic document embedding will encode noisy information. Therefore, we argue that the DC topic embedding should be both contextualized and topic-specific. In this regard, according to Equation 2, we rewrite the DC topic embedding as the weighted sum of DC keyword embeddings where only top-m topic keywords are used instead of all the words in the document, as

$$H_{d_j}(t_i) = \sum_{k=1}^m p_{ik} H_{d_j}(w_k). \quad (5)$$

Finally, in terms of the DC keyword embedding $H_{d_j}(w_k)$, as can be told from its name, it is precisely what a pretrained language model (Devlin et al., 2018) is designed for. Therefore, we take the corresponding final-layer token embedding of w_k when the input to the language model is d_j. Due to the self-attention mechanism (Vaswani et al., 2017) in the pretrained language model, $H_{d_j}(t_i)$ encodes the global context of the document, but since it only takes the sum of topic keyword embeddings, the encoded information is more oriented towards this specific topic t_i, which elegantly resonates with its

name "document-contextualized topic embedding". The step-by-step illustration of the generation of $H_d(w)$, $H_d(t)$ and $H_D(t)$ is shown in Figure 1 (c).

Because the language model used to generate the embeddings is finetuned to encode partisanship, the generated $H_D(t_i)$ also contains this information and is more precisely called partisanship-aware corpus-contextualized topic embedding. For brevity we call it corpus-contextualized (CC) topic embedding.

3.5 Measuring Polarization and Ranking Topics

After obtaining the CC topic embeddings $H_{D^L}(t_i)$ and $H_{D^R}(t_i)$ of the two corpora D^L and D^R on topic t_i, using two different sets of top-n most relevant documents from D^L and D^R respectively, we measure the ideology similarity (and then polarization) based on the cosine similarity between them, such that

$$
\begin{aligned}
c &= \cos\_sim(H_{D^L}(t_i), H_{D^R}(t_i)), \\
\beta(D^L, D^R, t_i) &= 0.5 * (1 - c) \in [0, 1].
\end{aligned}
\tag{6}
$$

A higher value of β indicates more polarization. Therefore, the polarization-based ranked topic list $f(D^L, D^R, T)$ is computed based on the corresponding polarization scores $(\beta(D^L, D^R, t_i))_{i=1}^K$.

4 Experiments and Results

4.1 Dataset

We use the AYLIEN COVID-19 dataset[3] consisting of ~1.5M news articles related to the pandemic spanning from Nov 2019 to July 2020 that are from ~440 global sources. To discover the polarization between politically divided news media, we select six well-known US publishers evenly split between partisan leanings: CNN, Huffington Post (Huff), New York Times (NYT) as liberal sources vs. Fox, Breitbart (Breit) and New York Post (NYP) as conservative sources. After filtering the publishers and remove duplicate articles, 66,368 articles are left spanning from Jan 2020 to July 2020. The statistics of news articles are shown in Appendix A.

4.2 Experimental Setup

Data Preprocessing. We build a global vocabulary containing unigrams and bigrams from the six news sources. We perform lemmatization via

---

[3]https://aylien.com/blog/free-coronavirus-news-dataset

SPACY and remove stopwords via NLTK, where we enrich the stopwords set with "cnn", "fox", "huffington", and "breitbart" since they can bias the language model's predictions during finetuning. We desire the partisanship classification of the language model to be based on the understanding of partisanship, rather than the occurrences of news source names in the news text.

LDA Topic Modeling. We train the topic model using articles from all six sources to create a global topic set. The number of topics K is selected from a grid search in $[10, 50]$ and the model with $K = 39$ produces the best coherence value (Röder et al., 2015). From the 39 topics we remove 9 of them regarding advertisements, sport events, gossip news and recipes, and 30 topics are left; the removed topics are more factual and contain less ideologies from the news media, which is less worth studying. Different from (Demszky et al., 2019) that assigns only one topic with the highest probability to a document, we allow a document to be assigned multiple topics with different probabilities. We represent each topic with its top-10 keywords because given a topic t_i we empirically find that $\sum_{j=1}^{10} p_{ij} > 0.95$; and we keep the top-10 most relevant documents to represent a topic because on some topics, the documents beyond the top-10 list are obviously irrelevant and will bias the polarization study regarding the topic. In Table 1 we show the top-10 keywords of topics that are discussed in this paper. For a complete list of topics please refer to Appendix B.

Learning Partisanship. We finetune the pre-trained bert-base-uncased model from huggingface Transformers (Wolf et al., 2020) to classify the news articles according to their political leanings, or partisanship. To smooth over the differences in style and writing between the sources and render the model primarily sensitive to political divisions, we aggregate CNN, Huff, and NYP to create a holistic Liberal corpus, and similarly aggregate Fox, Breit and NYP to create a holistic Conservative corpus and optimize the model to classify whether an article is from Liberal or Conservative. In fact, finetuning a BERT model to recognize differences only between CNN vs. Fox is likely to make it end up capturing the writing style differences and ignoring political differences, since the former is an easier task. For more details about the training process please refer to Appendix C.

Idx	Top-10 keywords (and two defined stances)
1	*keywords*: police, officer, man, black, protest, people, arrest, kill, protester, matter *stances*: **protests are for social justice** vs. **protests are riots**
2	*keywords*: coronavirus, pandemic, federal, supply, government, make, effort, ventilator, response, agency *stances*: **healthcare supplies are in good condition** vs. **shortage of supplies**
6	*keywords*: case, report, number, death, health, coronavirus, confirm, official, accord, covid
8	*keywords*: state, order, reopen, county, california, governor, business, open, jersey, guideline *stances*: **pro-lockdown** vs. **anti-lockdown**
9	*keywords*: post, twitter, video, facebook, tweet, social_media, share, write, call, make *stances*: **fact-checking is helpful** vs. **fact-checking is misleading**
10	*keywords*: trump, president, white_house, donald, administration, fauci, coronavirus, vice, briefing, task_force *stances*: **critical of white house covid briefings** vs. **defending them**
11	*keywords*: covid, dr, coronavirus, health, disease, drug, expert, risk, treatment, director *stances*: **drugs promoted by Trump are risky** vs. **they are helpful**
12	*keywords*: mr, biden, campaign, election, party, democratic, voter, joe_biden, republican, primary *stances*: **endorsing Biden in Democratic primaries** vs. **endorsing Sanders**
27	*keywords*: year, company, market, stock, price, drop, month, business, global, sale *stances*: **oil/stock prices are falling** vs. **the prices are going up**
28	*keywords*: state, coronavirus, cuomo, florida, texas, york, governor, tuesday, week, monday
29	*keywords*: house, coronavirus, republican, member, bill, senate, democrat, wednesday, washington, thursday
30	*keywords*: country, lockdown, government, coronavirus, measure, people, italy, restriction, travel, border *stances*: **closing borders in Europe** vs. **opening borders**
31	*keywords*: claim, court, judge, law, federal, district, rule, chicago, legal, decision
33	*keywords*: hospital, care, health, patient, medical, covid, center, facility, home, doctor *stances*: **overwhelmed hospitals** vs. **hospitals not overwhelmed**

Table 1: The keywords of topics discussed in the paper and two political stances of 10 labeled topics. The indices of labeled topics are highlighted in bold.

4.3 Annotating Topic Polarization

As ground truth for the evaluation of PaCTE, we annotate the topic polarization scores on a subset of the 30 modeled topics.

We asked three annotators to select 10 topics and define two polarized political stances on each selected topic, and they reached an agreement on $T^{\text{labeled}} = \{t_1, t_2, t_8, t_9, t_{10}, t_{11}, t_{12}, t_{27}, t_{30}, t_{33}\}$, as shown in Table 1. Then on each topic in T^{labeled}, we selected 60 relevant documents (10 from each of the six sources), and asked three annotators to decide which stance they belong to (label it as 0/1). If the document does not have a clear stance, it was labeled as -1. On each document, the majority label from the annotations was used as the final annotation. Please refer to Appendix D for more details about the annotation process.

Denoting the number of negative labels (0) and positive labels (1) in corpus D on topic t as $N_D^t(0)$ and $N_D^t(1)$ respectively, the leaning of the corpus on the topic is quantified as

$$\text{le}(D, t) = (N_D^t(1) - N_D^t(0))/|D| \in [-1, 1]. \quad (7)$$

Intuitively, $\text{le}(D, t)$ reflects how much the corpus is aligned with the stance labeled as 1. Notably, the documents labeled with -1 are not counted because they do not display a clear political standing. Accordingly, the ground-truth polarization score between a liberal corpus D^L and a conservative

corpus D^R on topic t is computed as the difference between the leanings of the two corpora, such that

$$\alpha(D^L, D^R, t) = |\text{le}(D^L, t) - \text{le}(D^R, t)|/2 \in [0, 1]. \quad (8)$$

A higher value of α signifies more polarization. As a result, the ground-truth polarization-based topic ranked list $l_{\text{gt}}(D^L, D^R, T^{\text{labeled}})$ between a liberal corpus D^L and a conservative corpus D^R is computed based on the corresponding ground-truth polarization scores $(\alpha(D^L, D^R, t)|t \in T^{\text{labeled}})$.

4.4 Baselines

We compare PaCTE to the following three baselines.

Leave-out estimator (LOE). For a pair of news corpora D^L and D^R and a given topic t, we take the top-10 most relevant documents from each corpus and feed the token frequency vectors of the documents into the leave-out estimator (Demszky et al., 2019), from which we use estimated partisanship as the polarization score ($\in [0, 1]$) of topic t between D^L and D^R, following the idea of measuring within-topic polarization in their paper. Note that different from their method that extracts topic using embedding-based topic assignment, we use the same LDA topic model in PaCTE to extract topics, so as to ensure fair comparison between PaCTE and LOE.

PaCTE¬FT. A variant of PaCTE without fine-tuning the language model. We compare to it to show the effect of finetuning the language model.

PaCTE-PLS. A variant of PaCTE where the language model is finetuned on news articles with partisanship labels shuffled and thus is confused about the partisanship. We compare to it in order to show the effect of rendering the language model partisanship-aware.

4.5 Quantitative Evaluation with Labeled Topics

To quantitatively evaluate the effectiveness of PaCTE and the baselines in capturing topic polarization, we use the 10 manually labeled topics to create a ground truth ranking of polarized topics and score models on their ability to retrieve the most polarized topics on this ranked list.

Evaluation protocol. Given a liberal news corpus D^L, a conservative news corpus D^R, and a list of 10 topics ranked by ground-truth polarization scores, $l_{gt}(D^L, D^R, T^{labeled})$, as described in Section 4.3, we define the top-3 topics in the list as the target polarized topics that deserve more attention and that should be addressed when trying to prevent polarization from escalating. The target polarized topics between different pairs of news sources are shown in Table 2. Then, given a ranked list of topics $f_{pred}(D^L, D^R, T^{labeled})$ predicted by a model, we evaluate how effectively the 3 target polarized topics are retrieved in this model predicted list using recall@3. In other words, we check how much the overlap is between the top-3 topics in the ground-truth ranking and the top-3 topics in the predicted ranking, of the 10 labeled topics. We call this task *polarized topics retrieval*.

	Fox	Breit	NYP
CNN	1,9,10	9,1,11	9,10,2
Huff	10,1,8	1,11,9	10,12,30
NYT	10,33,1	11,1,33	11,9,10

Table 2: The target polarized topics between different pairs of news sources from human annotations.

Analysis of results. The results of polarized topics retrieval using different methods in nine news corpus pairs are shown in Table 3. The average recall@3 over the nine news source pairs is 0.26, 0.04, 0.26, and 0.52 on LOE, PaCTE¬FT, PaCTE-PLS, and PaCTE respectively, where PaCTE outperforms all other baselines.

Comparing the results of LOE and PaCTE, we see that in most pairs PaCTE outperforms or ties with LOE. We argue that the inferior performance of LOE stems from its inability to capture document semantics due to the use of word frequency vectors. For example, in Huff vs. NYP, topic 12 is one of the target polarized topics, where documents from both stances spend the bulk of the content on the fact about the primaries and then use a few words to explicitly or implicitly endorse Biden or Sanders. Based on the use of words it is difficult to differentiate documents from the two stances, leading to the failure of LOE. In contrast, PaCTE is able to capture the contextual semantics in addition to the statistics of word usages. Therefore, even when word usages are statistically similar, PaCTE manages to discern the semantic difference and capture polarization. However, in Huff vs. Breit, compared to LOE, PaCTE fails to retrieve topic 1 regarding "black lives matter", which is in the target polarized topics. On topic 1 Huff stresses "justice" where the news articles suggest "police knelt on a black man", while Breit stresses "riot" where the articles suggest "the protesters loot stores and attack police". As a result, the word usages of the articles from two stances are significantly different, which is trivial for LOE to capture, and thus LOE ranks topic 1 in a high place in the output list. Despite the difference in word usages, articles from both sources mention "protests" and "violence" a lot and their "negative" semantics is captured by PaCTE, leading to the perceived less polarization by PaCTE.

The worst-performing method is PaCTE¬FT where the language model is not finetuned. On all topics and in all partisan news source pairs, the polarization scores given by PaCTE¬FT are below 0.1 (the full range is $[0, 1]$) which indicates significant alignment. However, this is contradictory to the well-known polarization in news media. Such a phenomenon demonstrates the necessity of fitting a language model on the target corpus before apply cosine similarity between learned embeddings as a measure of word and topic similarities.

In PaCTE-PLS the language model is finetuned on shuffled partisan labels that do not represent real partisanship. Compared to PaCTE¬FT where the model is not finetuned at all, the performance of PaCTE-PLS improves significantly, achieving the performance on a par with LOE. However, neither PaCTE¬FT nor LOE makes use of information about news partisanship, and compared to PaCTE

	Fox				Breit				NYP			
	LOE	PaCTE¬FT	PaCTE-PLS	PaCTE	LOE	PaCTE¬FT	PaCTE-PLS	PaCTE	LOE	PaCTE¬FT	PaCTE-PLS	PaCTE
CNN	1/3	0	0	1/3	1/3	0	1/3	1/3	0	1/3	1/3	2/3
Huff	1/3	1/3	1/3	2/3	2/3	0	1/3	1/3	0	0	1/3	2/3
NYT	1/3	0	1/3	1	1/3	0	1/3	1/3	0	1/3	0	1/3

Table 3: Recall@3 on polarized topics retrieval in nine partisan news source pairs using different methods, where we use the polarization-based topic ranked list from a model predictions $f_{\text{pred}}(D^L, D^R, T^{\text{labeled}})$ to retrieve the top-3 topics from the ground-truth ranked list $l_{\text{gt}}(D^L, D^R, T^{\text{labeled}})$. The row represents the liberal source and the column represents the conservative source in the news source pair.

where partisanship information is leveraged, they are still outperformed.

Insights into partisanship learning. We observe that PaCTE, which is finetuned on partisanship labels, outperforms PaCTE¬FT and PaCTE-PLS. We hypothesize that during the finetuning process of PaCTE, whereas the direct objective is to separate documents based on partisanship labels, the model implicitly learns the two political stances on each topic in an automatic manner; just like in human annotating, the annotators were given two groups of documents from two partisan lines, and the annotators were able to discover the two political stances after reading the documents. Therefore, after finetuning, while the model differentiates document embeddings based on partisan divisions, it separates DC topic embeddings according to the implicitly and automatically learned political stances, bearing resemblance to human annotators' defining two political stances for topics. As a result, we can use the partisanship-aware model to capture topic polarization arising from the partisan divisions.

4.6 Qualitative Analysis with All Topics

In Section 4.5 we quantitatively demonstrate the effectiveness of PaCTE in retrieving polarized topics when evaluating with the 10 labeled topics. We believe that such success generalizes to the case where the input to the model is the complete topic list T containing 30 topics. In this section, we conduct a case study and retrieve the top-3 most polarized topics from T in CNN vs. Fox, Huff vs. Breit and NYT vs. NYP, by PaCTE. Since we do not have the ground-truth target polarized topics from T, for the retrieved topics, we conduct manual inspections on relevant documents and give explanations about the polarization. For the topics in T^{labeled}, the polarization is formed due to the two political stances. Therefore in this section we only focus on the retrieved topics not in T^{labeled}.

CNN vs. Fox. The retrieved top-3 topics are topic 28, 6, 10, where topic 10 is in T^{labeled}. The first retrieved topic is topic 28, where CNN suggests the surge of new COVID cases every day but Fox suggests that the state should reopen. On topic 6 CNN reports the serious situation of coronavirus in the US, including the high number of cases and collapse of quarantine hotels, but Fox focuses more on worldwide coronavirus situation and suggests the high number of cases in Michigan is misleading.

Huff vs. Breit. The retrieved top-3 topics are topic 29, 9, 31, where topic 9 is in T^{labeled}. On topic 29, Huff advocates Pelosi's coronavirus bills while Breit criticizes them. On topic 31, the articles talk about different court cases; however, no clear polarization is discerned between the pair of news sources by manual inspections. We regard it as a failure case of PaCTE. Although the relevant articles are regarding the same topic, they have different subjects or events, and thus misleading PaCTE to perceive polarization between them.

NYT vs. NYP. The retrieved top-3 topics are topic 28, 12, 10, where topic 12 and 10 are in T^{labeled}. On topic 28, just as in CNN vs. Fox, NYT takes the pandemic more seriously and NYP suggests reopening.

As a result, despite a minor error, PaCTE manages to retrieve polarized topics from T on the three pairs of news sources. Although we are not able to verify if the retrieved topics are indeed the ground-truth top-3 most polarized topics, we argue that if given the ground-truth ranking on T, PaCTE will retain its satisfactory quantitative performance in retrieving polarized topics.

4.7 Ablation Study: Document Embedding vs. DC Topic Embedding

In Section 3.4 we propose to use the DC topic embedding to represent the ideology of a document on a topic, instead of using the holistic document embedding. In this section we study the difference between them. We denote the variant of PaCTE that uses document embeddings ([CLS] token em-

beddings) as PaCTE-DE. First, we show the results of polarized topics retrieval using PaCTE-DE and PaCTE in three partisan news source pairs in Table 4.

Method	CNN vs. Fox	Huff vs. Breit	NYT vs. NYP
PaCTE-DE	0	0	0
PaCTE	1/3	1/3	1/3

Table 4: Recall@3 on polarized topics retrieval using PaCTE-DE and PaCTE in three partisan news source pairs.

We observe that PaCTE-DE fails to retrieve any polarized topics in all three pairs of news sources, significantly outperformed by PaCTE. We provide more explanations on the advantages of DC topic embeddings over document embeddings from another perspective, in addition to the capability of DC topic embedding to focus more on the topic-specific semantics in a document. We observe that the polarization scores given by PaCTE-DE in three source pairs on all topics are above 0.98 (the range is [0,1]), suggesting that all topics are highly polarized. Therefore, as the polarization scores cluster within the interval of [0.98,1], the gaps between different scores are barely discernible, in which case the output ranked list is more susceptible to random noise during the language model finetuning and is thus more unstable and erratic. However, the output polarization scores from PaCTE are more evenly distributed in [0,1], and thus are more robust to perturbations during partisanship learning; a small perturbation on a polarization score does not affect the output ranking. As a result, PaCTE enjoys a better chance to outperform PaCTE-DE.

As a matter of fact, the large polarization scores from PaCTE-DE on all topics are expected, because the language model is finetuned to directly separate the document embeddings according to partisan line divisions, resulting in low cosine similarities between document embeddings on every topic, as shown in Figure 2(Left). However, despite the prominent separation of document embeddings, the corresponding DC topic embeddings that are used in PaCTE display more alignment, as shown in Figure 2(Right), where we see on some topics the DC topic embeddings are separated while on other topics the embeddings are more close. Thus, we argue that during the finetuning process, on a given topic, DC topic embeddings retain their similarity if the two partisan news articles agree on this topic, because in these articles the topic-related se-

mantics does not contribute to the forming of the partisanship and thus maintains its position during partisanship learning, while the non-topical semantics (not captured by DC topic embeddings but captured by document embeddings) that contribute to the document partisanship keeps moving apart in the embedding space.
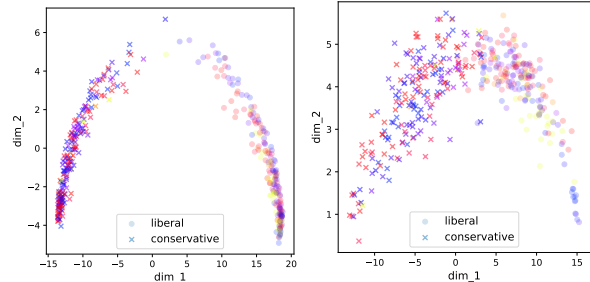

Figure 2: Document embeddings (Left) and DC topic embeddings (Right) on 10 labeled topics in Liberal vs. Conservative. Different colors represent documents categorized to different topics. The original 768-d embeddings are projected into the 2-d space via PCA.

5 Conclusions and Future Work

In this paper, we propose a method to automatically discover topic-level polarization between partisan news sources by contextualized topic embeddings. For evaluation, we create annotations on topic polarization scores in different partisan news source pairs on a variety of topics. Compared to the leave-out estimator (Demszky et al., 2019) that is purely based on statistical features, our method can more precisely and meaningfully capture topical polarization as indicated by the performance on polarized topics retrieval. We hope that more NLP and researchers and practitioners can contribute to this research area that is promising but receiving insufficient attention.

Because detecting polarized topics between partisan news sources is a less established task in the research community, we articulate the data annotation and the model evaluation in great detail and make the method seemingly "complicate". However, we believe that for public media watchdogs and social media platforms to flag the highly polarized topics, our method is simple to implement, because each of the five steps described in Section 3 is based on robust methods in NLP.

For future work, we plan to perform our method on more datasets, such as the tweets with noisy texts (Demszky et al., 2019). In addition, we will study how to finetune the language model when when partisanship labels are not available.

References

Syon Bhanot and Daniel J Hopkins. 2020. Partisan polarization and resistance to elite messages: Results from a survey experiment on social distancing. *Available at SSRN 3593450*.

Federico Bianchi, Silvia Terragni, and Dirk Hovy. 2020. Pre-training is a hot topic: Contextualized document embeddings improve topic coherence. *arXiv preprint arXiv:2004.03974*.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Yatin Chaudhary, Pankaj Gupta, Khushbu Saxena, Vivek Kulkarni, Thomas Runkler, and Hinrich Schütze. 2020. Topicbert for energy efficient document classification. *arXiv preprint arXiv:2010.16407*.

Wei-Fan Chen, Khalid Al-Khatib, Henning Wachsmuth, and Benno Stein. 2020. Analyzing political bias and unfairness in news articles at different levels of granularity. *arXiv preprint arXiv:2010.10652*.

Dorottya Demszky, Nikhil Garg, Rob Voigt, James Zou, Matthew Gentzkow, Jesse Shapiro, and Dan Jurafsky. 2019. Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. *arXiv preprint arXiv:1904.01596*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Lisa Fan, Marshall White, Eva Sharma, Ruisi Su, Prafulla Kumar Choubey, Ruihong Huang, and Lu Wang. 2019. In plain sight: Media bias through the lens of factual reporting. *arXiv preprint arXiv:1909.02670*.

Morris P Fiorina and Samuel J Abrams. 2008. Political polarization in the american public. *Annu. Rev. Polit. Sci.*, 11:563–588.

Matthew Gentzkow, Jesse M Shapiro, and Matt Taddy. 2019. Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica*, 87(4):1307–1340.

Anton Gollwitzer, Cameron Martel, William J Brady, Philip Pärnamets, Isaac G Freedman, Eric D Knowles, and Jay J Van Bavel. 2020. Partisan differences in physical distancing are linked to health outcomes during the covid-19 pandemic. *Nature human behaviour*, 4(11):1186–1197.

Jon Green, Jared Edgerton, Daniel Naftel, Kelsey Shoub, and Skyler J Cranmer. 2020. Elusive consensus: Polarization in elite communication on the covid-19 pandemic. *Science Advances*, 6(28):eabc2717.

Maarten Grootendorst. 2020. Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics.

P Sol Hart, Sedona Chinn, and Stuart Soroka. 2020. Politicization and polarization in covid-19 news coverage. *Science Communication*, 42(5):679–697.

Barry A Hollander. 2008. Tuning out or tuning elsewhere? partisanship, polarization, and media migration from 1998 to 2006. *Journalism & Mass Communication Quarterly*, 85(1):23–40.

Julie Jiang, Emily Chen, Shen Yan, Kristina Lerman, and Emilio Ferrara. 2020. Political polarization drives online conversations about covid-19 in the united states. *Human Behavior and Emerging Technologies*, 2(3):200–211.

Mark Jurkowitz, Amy Mitchell, Elisa Shearer, and Mason Walker. 2020. U.s. media polarization and the 2020 election: A nation divided.

Ashiqur R KhudaBukhsh, Rupak Sarkar, Mark S Kamlet, and Tom M Mitchell. 2020. We don't speak the same language: Interpreting polarization through machine translation. *arXiv preprint arXiv:2010.02339*.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Philipp Lorenz-Spreen, Stephan Lewandowsky, Cass R Sunstein, and Ralph Hertwig. 2020. How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nature human behaviour*, 4(11):1102–1109.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Markus Prior. 2013. Media and political polarization. *Annual Review of Political Science*, 16:101–127.

Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408.

Natalie Jomini Stroud. 2011. *Niche news: The politics of news choice*. Oxford University Press on Demand.

Hubert Tworzecki. 2019. Poland: a case of top-down polarization. *The ANNALS of the American Academy of Political and Social Science*, 681(1):97–119.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Albert Webson, Zhizhong Chen, Carsten Eickhoff, and Ellie Pavlick. 2020. Are" undocumented workers" the same as" illegal aliens"? disentangling denotation and connotation in vector spaces. *arXiv preprint arXiv:2010.02976*.

Thomas Wolf, Julien Chaumond, Lysandre Debut, Victor Sanh, Clement Delangue, Anthony Moi, Pierric Cistac, Morgan Funtowicz, Joe Davison, Sam Shleifer, et al. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45.

A   Data Preprocessing

The statistics of the dataset is in Table 5. We use the summary of each news article to perform the textual analysis, because the summary contains sufficient information to understand the political stance of the article and the whole text is lengthy for the pretrained language model to handle. For a complete list of all documents, please check our public repository[4].

B   LDA Topic Modeling

We use MALLET[5] topic modeling. The top-10 keywords of all 39 topics are shown in Table 6. Among them topic 0, 3, 4, 14, 16, 26, 35, 36, 37 are not used in further analysis because after reading relevant articles we find that they are more about advertisements, sport events, gossip news and recipes and etc., which are more factual and convey limited media ideologies. 30 topics are left after removing the 9 topics. Table 6 lists the top-10 keywords of the 30 topics.

B.1   News Article Examples

On topic 10, we show six examples of news articles, one from each news source. For a complete list of news articles, please refer to our public repository.

CNN: *There has been a concerted effort among aides and allies to get President Donald Trump to stop conducting the daily coronavirus briefings, multiple sources tell CNN. The briefing came a day after Trump had given a lengthy briefing to the media, at one point suggesting it might be possible to treat coronavirus by injecting people with sunlight or disinfectants. Trump asked White House coronavirus task force coordinator Dr. Deborah Birx during Thursday's briefing. A source close to the coronavirus task force said Trump was upset about the "flack" he was taking after those comments and that appears to be part of the reason why the President cut Friday's briefing short. During the earlier questioning from reporters on Friday, Trump said he was being "sarcastic" with his suggestion that people inject themselves with disinfectant, even though he was clearly being serious during Thursday's briefing.*

Fox: *White House press secretary Kayleigh McEnany, during her first official briefing,*

promised that she 'will never lie' to the press in her new role. White House press secretary Kayleigh McEnany, during her first official briefing, promised that she "will never lie" to the press in her new role. McEnany took the podium for the first time Friday, after being tapped as White House press secretary from her post as national spokeswoman for President Trump's re-election campaign earlier this month. TRUMP NAMES KAYLEIGH MCENANY AS NEW WHITE HOUSE PRESS SECRETARY "I will never lie to you," McEnany told reporters. McEnany seemed to signal that the White House would scale back on their daily coronavirus task force briefings, which were regularly led by the president himself, and Vice President Pence, with appearances from Dr. Deborah Birx and Dr. Anthony Fauci to provide public health information.

Huffington Post: *President Donald Trump on Sunday tore into former President Barack Obama, calling him "an incompetent president" after Obama appeared to criticize his response to the coronavirus crisis during two commencement speeches a day earlier. Asked about Obama's remarks, Trump told reporters on the White House lawn that he "didn't hear it" before proceeding to bash his predecessor as "grossly incompetent." President Trump: "[President Obama] was an incompetent president. But earlier this month, Obama reportedly bashed the Trump administration's response to the pandemic as "an absolute chaotic disaster" during a phone call with some of his former White House aides. When a Washington Post reporter last week asked Trump to explain "Obamagate," the president refused.*

Breibart: *New York magazine Washington correspondent Olivia Nuzzi responded angrily to criticism from former White House press secretary Ari Fleischer on Monday evening, tweeting at him: "Oh shut the f*ck up." Fleisher, who served under President George W. Bush, criticized Nuzzi after a Rose Garden press briefing on the coronavirus pandemic in which she asked President Donald Trump: "If an American president loses more Americans over the course of six weeks than died in the entirety of the Vietnam War, does he deserve to be re-elected?" One example is a "fake news" viral photograph of President Lyndon B. Johnson, which was presented by many Trump critics as if Johnson had been expressing grief over the deaths in Vietnam. President Trump is said to be reconsidering*

	Jan	Feb	Mar	Apr	May	Jun	Jul	#sum
CNN	232	589	2958	3293	2444	1789	2120	*13425*
Fox	156	504	3938	5148	3616	2367	2585	*18314*
Huff	17	74	1237	1450	1185	701	828	*5492*
Breit	93	240	1918	2164	1353	777	924	*7469*
NYT	94	369	2117	2177	1730	849	1097	*8423*
NYP	144	405	3063	3470	2278	1733	2152	*13245*
#sum	*736*	*2171*	*15231*	*17702*	*12606*	*8216*	*736*	*66368*

Table 5: The number of news articles from all sources in all months.

Idx	Top-10 Topic Keywords
1	police, officer, man, black, protest, people, arrest, kill, protester, matter
2	coronavirus, pandemic, federal, supply, government, make, effort, ventilator, response, agency
5	coronavirus, virus, test, covid, people, tested_positive, testing, positive, symptom, spread
6	case, report, number, death, health, coronavirus, confirm, official, accord, covid
7	event, plan, june, announce, due, july, hold, cancel, pandemic, date
8	state, order, reopen, county, california, governor, business, open, jersey, guideline
9	post, twitter, video, facebook, tweet, social_media, share, write, call, make
10	trump, president, white_house, donald, administration, fauci, coronavirus, vice, briefing, task_force
11	covid, dr, coronavirus, health, disease, drug, expert, risk, treatment, director
12	mr, biden, campaign, election, party, democratic, voter, joe_biden, republican, primary
13	school, child, student, university, parent, high, kid, year, family, class
15	american, pandemic, crisis, america, nation, make, policy, job, people, economy
17	time, world, space, launch, turn, center, long, life, leave, moment
18	coronavirus, report, outbreak, accord, ship, official, quarantine, military, force, iran
19	city, york, de_blasio, mayor, resident, area, yorker, coronavirus, people, tuesday
20	mask, people, wear, face, service, social_distance, church, sunday, coronavirus, stay
21	people, time, thing, good, work, make, lot, add, give, feel
22	department, official, national, security, fire, investigation, report, threat, call, director
23	employee, worker, company, restaurant, food, store, work, customer, business, amazon
24	china, chinese, world, outbreak, virus, wuhan, organization, coronavirus, global, government
25	time, series, film, show, year, make, movie, live, race, set
27	year, company, market, stock, price, drop, month, business, global, sale
28	state, coronavirus, cuomo, florida, texas, york, governor, tuesday, week, monday
29	house, coronavirus, republican, member, bill, senate, democrat, wednesday, washington, thursday
30	country, lockdown, government, coronavirus, measure, people, italy, restriction, travel, border
31	claim, court, judge, law, federal, district, rule, chicago, legal, decision
32	health, public, people, work, community, include, protect, provide, group, pandemic
33	hospital, care, health, patient, medical, covid, center, facility, home, doctor
34	program, pay, money, fund, economic, job, business, relief, federal, receive
38	coronavirus, office, letter, pandemic, call, send, statement, issue, write, act

Table 6: List of all the 39 topics with corresponding top keywords.

his daily press briefings because journalists use them to grandstand and to score political points, rather than to pursue information. The contrast with press briefings for governors and mayors is stark: there, journalists tend to be more deferential and to ask questions aimed at eliciting information rather than assigning political fault.

New York Times: *WASHINGTON — After several days spent weathering attacks from White House officials, Dr. Anthony S. Fauci hit back on Wednesday, calling recent efforts to discredit him "bizarre" and a hindrance to the government's ability to communicate information about the coronavirus pandemic. On Wednesday, Peter Navarro, Mr. Trump's top trade adviser, published a brazen*

op-ed article in USA Today describing Dr. Fauci as "wrong about everything." All the while, White House officials — including the president and the press secretary — assert in the face of the evidence that there is no concerted effort to attack Dr. Fauci. He has not briefed Mr. Trump in weeks, preferring to work with Dr. Deborah L. Birx, who helps coordinate the administration's coronavirus response, or to send his messages through Vice President Mike Pence. In the piece, Mr. Navarro presented what White House officials have been saying privately about Dr. Fauci, and what Mr. Trump has said publicly: They like Dr. Fauci personally, but he has made mistakes.

New York Post: *President Trump said Wednes-*

day he has a "very good relationship" with White House coronavirus task force member Anthony Fauci, despite an op-ed by one of his top advisers that trashed the immunologist. Trump distanced himself from trade adviser Peter Navarro's op-ed that said Fauci "has been wrong about everything." "I get along very well with Dr. Fauci," Trump told reporters in the Oval Office. On Monday, White House Press Secretary Kayleigh McEnany denied a Washington Post report that said reporters were given "opposition research" to discredit Fauci, including his past remarks early on in the pandemic that the public didn't need to wear masks. "We were asked a very specific question by the Washington Post, and that question was President Trump noted that Dr. Fauci had made some mistakes, and we provided a direct answer to what was a direct question."

B.2 Top-10 Most Relevant Documents on All Topics

We show the topic-10 most relevant document indices on all 30 topics on each source. On some topics there are less than 10 relevant documents on some sources. Note that such topics are not in the 10 labeled topics and are only used for qualitative analysis; in other words, for quantitative analysis, we ensure that on all the 10 labeled topics, there are 10 relevant documents on each source.

Topic 1. CNN: 22873, 62724, 62635, 63979, 60318, 64323, 39686, 23087, 66007, 64346. **Fox**: 64455, 21485, 26889, 21509, 22055, 62291, 21458, 22866, 21404, 65938. **Huff**: 22937, 63328, 40375, 66026, 64381, 61909, 64511, 63335, 64173, 64573. **Breit**: 64348, 52383, 21375, 32945, 64143, 58746, 65060, 37841, 21378, 40036. **NYT**: 22742, 65966, 21562, 58360, 21357, 65875, 65146, 21330, 52138, 21501. **NYP**: 21316, 62547, 64164, 21638, 64765, 7055, 21590, 60740, 21720, 40400.

Topic 2. CNN: 2291, 9180, 42882, 6582, 1629, 1476, 11612, 850, 42172, 2006. **Fox**: 9182, 31226, 2682, 48284, 10207, 46558, 11722, 10312, 3984, 11122. **Huff**: 889, 35891, 48378, 46288, 5361, 708, 2408, 1319, 1938, 8962. **Breit**: 3817, 13126, 6136, 55780, 13045, 3958, 17769, 9004, 48430, 18855. **NYT**: 3723, 44895, 12794, 54586, 8385, 61651, 33770, 51084, 33735, 11857. **NYP**: 46244, 46842, 8039, 62990, 1877, 17378, 66227, 786, 39335, 809.

Topic 5. CNN: 52276, 22960, 63870, 63867, 9240, 22872, 56452, 28994, 13558, 52074. **Fox**: 4863, 14561, 13454, 43376, 14939, 8762, 12185, 8924, 14800, 6561. **Huff**: 12973, 29357, 5156, 12938, 2941, 22574, 14862, 43495, 29785, 29580. **Breit**: 13167, 64051, 8622, 12110, 14544, 62240, 51507, 15519, 18759, 47388. **NYT**: 29543, 19969, 2964, 30581, 49542, 24286, 61430, 35059, 4465, 66042. **NYP**: 15340, 13404, 54124, 64067, 35789, 9325, 22692, 3575, 8217, 15513.

Topic 6. CNN: 53265, 56003, 16985, 15115, 26006, 17063, 19812, 50074, 18014, 21857. **Fox**: 58537, 63076, 53236, 22482, 50794, 42788, 60553, 59539, 63026, 54784. **Huff**: 54194, 37337, 64330, 35448, 10906, 6077, 48061, 48797, 20643, 25306. **Breit**: 36565, 61077, 44474, 49577, 17068, 48689, 14186, 16924, 18078, 29693. **NYT**: 62593, 50117, 28572, 39978, 7679, 56428, 5277, 23283, 18404, 20709. **NYP**: 62856, 66347, 48582, 60094, 6419, 54128, 21187, 59736, 60922, 59286.

Topic 7. CNN: 8968, 47625, 14093, 8928, 10954, 17954, 12589, 3657, 29776, 54903. **Fox**: 14933, 64745, 50666, 8112, 48177, 63686, 8539, 16974, 10530, 62301. **Huff**: 43539, 45395, 41053, 41201, 4411, 14668, 62437, 25455, 17322, 33304. **Breit**: 14659, 16032, 15822, 13584, 4032, 8559, 16522, 6377, 48135, 45140. **NYT**: 36774, 1098, 41903, 6893, 22100, 42130, 40933, 65242, 13346, 12585. **NYP**: 50435, 61426, 4653, 48992, 41460, 41187, 49057, 17932, 59701, 59289.

Topic 8. CNN: 52753, 35955, 53068, 52474, 8618, 36939, 58294, 53262, 54031, 39671. **Fox**: 52787, 36759, 28989, 41859, 38546, 35214, 10486, 51718, 32366, 38501. **Huff**: 52782, 13488, 33939, 26489, 21984, 35994, 58843, 24991, 22420, 54017. **Breit**: 36756, 25982, 26581, 56836, 31828, 31625, 24778, 54023, 29688, 22190. **NYT**: 26442, 10769, 28108, 54705, 12483, 8496, 33668, 29239, 34964, 25184. **NYP**: 54049, 32336, 60474, 10483, 25023, 10508, 58551, 62855, 22591, 6424.

Topic 9. CNN: 24337, 29633, 49300, 18255, 21561, 63481, 65460, 12850, 22029, 5379. **Fox**: 58858, 22378, 24954, 10605, 49169, 28970, 42046, 22859, 17031, 22101. **Huff**: 24163, 22956, 28908, 20469, 62999, 28039, 22394, 64592, 22044, 49612. **Breit**: 58881, 63178, 22588, 51608, 11734, 23200, 63795, 65340, 36749, 49565. **NYT**: 25897, 22208, 65768, 63558, 57615, 61546, 51728, 56320, 16812, 26196. **NYP**: 29025, 53038, 11502, 17723, 49458, 37686, 53745, 24250, 54854, 24424.

Topic 10. CNN: 36977, 9900, 60315, 47434, 52901, 61685, 15586, 33393, 16909, 28711. **Fox**: 33421, 52296, 33431, 33349, 5564, 15986, 46620,

9827, 24355, 53249. **Huff**: 27192, 23019, 37029, 46903, 8395, 30295, 38470, 8936, 32619, 33415. **Breit**: 35799, 54043, 54046, 43323, 47227, 47486, 61645, 18345, 13315, 31312. **NYT**: 53227, 26542, 43823, 33269, 52807, 50983, 10166, 33072, 52406, 14937. **NYP**: 43755, 53340, 43996, 51619, 58930, 16594, 32531, 48513, 60137, 37131.

Topic 11. CNN: 43064, 24419, 43853, 37234, 24293, 62395, 60856, 37716, 65640, 11503. **Fox**: 6852, 44149, 42764, 26980, 37265, 34316, 65050, 28248, 42912, 22639. **Huff**: 10973, 24477, 33168, 30443, 65264, 62143, 32735, 23337, 48333, 38400. **Breit**: 4390, 62298, 46481, 10710, 21938, 44850, 11701, 378, 5428, 8059. **NYT**: 44316, 24463, 45811, 37043, 24655, 39125, 36073, 338, 26193, 33556. **NYP**: 46561, 62407, 65900, 7878, 10542, 52025, 61086, 34882, 43578, 19746.

Topic 12. CNN: 16723, 17545, 13563, 12103, 14653, 42843, 25659, 35193, 15965, 14021. **Fox**: 34526, 25672, 37152, 16593, 12336, 11898, 27791, 15679, 64568, 43431. **Huff**: 12071, 36383, 63407, 12041, 39262, 15806, 43507, 53503, 35194, 16040. **Breit**: 21477, 45924, 15829, 15787, 60055, 46496, 14115, 34548, 57445, 15883. **NYT**: 18071, 18061, 18087, 18091, 18574, 18109, 18044, 2961, 18097, 44621. **NYP**: 13572, 15798, 63085, 6355, 46392, 44686, 16507, 32856, 13753, 59949.

Topic 13. CNN: 52042, 53647, 56269, 48533, 50908, 32977, 52157, 325, 54723, 19008. **Fox**: 55527, 55611, 52290, 50854, 55905, 55553, 9799, 50424, 52626, 55145. **Huff**: 57978, 10357, 12659, 57989, 28854, 15152, 48231, 39256, 41416, 50280. **Breit**: 35932, 53281, 51224, 52932, 52933, 6724, 61389, 54136, 1500, 4253. **NYT**: 51392, 56173, 37773, 17033, 53487, 14635, 37881, 14693, 28559, 29795. **NYP**: 12922, 57494, 62649, 33391, 33069, 48882, 35976, 33935, 66118, 49222.

Topic 15. CNN: 35585, 58100, 63056, 60887, 66162, 23165, 33687. **Fox**: 64255, 57723, 61319, 47541, 57594, 60724, 46614, 65841, 32082, 51435. **Huff**: 50283, 50286, 29216, 64189, 37388, 51475, 43624, 23605, 64125, 38169. **Breit**: 1602, 47078, 51743, 41817, 38117, 8476, 32283, 47210, 21692, 14994. **NYT**: 26295, 26104, 35795, 9649, 169, 53045, 26517, 49375, 10809, 64661. **NYP**: 41409, 23782.

Topic 17. CNN: 21394, 46193, 21796, 50681, 2467, 52824, 23060, 42194, 33416, 33996. **Fox**: 22895, 21313, 21429, 24489, 25071, 24318, 25695, 56398, 37404, 21419. **Huff**: 62406, 22703, 48727, 46191, 51732, 8868, 56959, 44157, 13008, 36388.

Breit: 22979, 48822, 21421, 60683, 39031, 24114, 43545. **NYT**: 21423, 21486, 41181, 38254, 48811, 56904, 53586, 52196, 54239, 51030. **NYP**: 27696, 41257, 33369, 12983, 41190, 35086, 33866, 4693, 42508, 60983.

Topic 18. CNN: 39244, 42833, 687, 47269, 46101, 1227, 46316, 43181, 299, 37315. **Fox**: 61023, 37051, 47122, 34309, 48330, 20285, 20295, 42259, 7247, 9568. **Huff**: 3282, 44890, 37119, 45584, 47564, 154, 29728, 5455, 45893, 19422. **Breit**: 124, 36851, 60988, 2273, 553, 979, 64492, 5952, 7323, 13383. **NYT**: 173, 36987, 41045, 48109, 47310, 1555, 12502, 42449, 56665, 38772. **NYP**: 1007, 48021, 47721, 47477, 3308, 1597, 105, 37866, 34656, 35694.

Topic 19. CNN: 65216, 65919, 47703, 15243, 64710, 35077, 61871, 12372, 7614, 4141. **Fox**: 1690, 12309, 65586, 29089, 3600, 29899, 55457, 22667, 62019, 55717. **Huff**: 21497, 22538, 34750, 19739, 47734, 13671, 9080, 44803, 9790, 6579. **Breit**: 26804, 6509, 30461, 11703, 13818, 35813, 34832, 36035, 34461, 40208. **NYT**: 56435, 65514, 11949, 35123, 60194, 56984, 7009, 24659, 15179, 32155. **NYP**: 31276, 64202, 33547, 36193, 61943, 6662, 50296, 41943, 37147, 12254.

Topic 20. CNN: 29359, 41880, 44391, 23929, 23891, 58610, 16, 7791, 43279, 43686. **Fox**: 41486, 4380, 25130, 44564, 7153, 4102, 62478, 45326, 48008, 44450. **Huff**: 52707, 34296, 43615, 42366, 8483, 61865, 53883, 58217, 58209, 52080. **Breit**: 40577, 50156, 44531, 1704, 224, 45886, 4078, 30607, 36544, 21342. **NYT**: 43680, 43132, 3466, 46829, 16385, 44529, 44498, 13911, 18247, 44490. **NYP**: 58586, 13713, 14492, 22920, 14037, 44445, 36107, 213, 4948, 62594.

Topic 21. CNN: 30694, 38631, 10489, 64595, 46011, 58309, 26764, 39667, 30189, 62513. **Fox**: 54100, 5720, 37070, 53083, 36870, 4434, 51052, 40273, 44540, 58646. **Huff**: 53895, 5143, 64645, 53896, 60453, 50905, 63982, 27568, 26936, 34053. **Breit**: 24115, 26582, 38790, 9481. **NYT**: 2723, 27086, 52363, 10737, 19923, 44353, 6979, 56929, 37268, 8272. **NYP**: 57065, 3175, 39506, 58484, 50308, 53214.

Topic 22. CNN: 23167, 48096, 26205, 22047, 47742, 34234, 33189, 22970, 49527, 39375. **Fox**: 31847, 29072, 30409, 30494, 30000, 30115, 39926, 33583, 29766, 59646. **Huff**: 23126, 47403, 28642, 14644, 7126, 66321, 32087, 60313, 46938, 46902. **Breit**: 35477, 46795, 27951, 59637, 38856, 46490, 33413, 13190, 6100, 28231. **NYT**: 44318, 37139,

44335, 9489, 10618, 24316, 28351, 23162, 25276, 22511. **NYP**: 29730, 27592, 60671, 7102, 47325, 29276, 26228, 14283, 60403, 50133.

Topic 23. **CNN**: 5326, 44622, 44231, 11176, 6496, 52668, 14017, 43584, 12395, 46745. **Fox**: 4017, 31428, 44209, 1964, 37142, 14336, 45345, 46060, 44181, 9230. **Huff**: 12371, 3245, 49665, 49671, 6603, 10079, 52967, 59703, 1622, 5552. **Breit**: 43555, 30686, 29565, 56249, 45016, 953, 9378, 51834, 1867, 37187. **NYT**: 11277, 11450, 61190, 11747, 35565, 62501, 45602, 10102, 55676, 24533. **NYP**: 3088, 43904, 42403, 40986, 12588, 31823, 12860, 43962, 45447, 60394.

Topic 24. **CNN**: 11322, 12476, 51400, 32280, 26578, 7698, 22944, 53884, 22879, 27018. **Fox**: 41909, 8147, 52862, 9992, 11884, 61000, 2816, 42300, 64408, 37554. **Huff**: 20522, 21145, 7456, 20030, 21166, 6756, 62148, 51402, 66041, 9049. **Breit**: 45020, 25401, 36033, 19866, 23527, 16583, 63436, 46336, 20068, 52811. **NYT**: 15450, 30846, 19962, 5814, 46898, 33715, 2936, 17488, 18875, 32108. **NYP**: 52593, 25600, 15066, 25704, 27623, 32851, 20066, 33527, 10198, 31849.

Topic 25. **CNN**: 22679, 52783, 36356, 50190, 50458, 47910, 44222, 40460, 27182, 2412. **Fox**: 63929, 56052, 27995, 63404, 64738, 13451, 2886, 28144, 11928, 45087. **Huff**: 25128, 32938, 53468, 22477, 58245, 13502, 16615, 52208, 51719, 34708. **Breit**: 43320, 58828, 3398, 14846, 51825, 13013, 19562, 49059, 56808, 49844. **NYT**: 46, 62894, 28598, 9397, 56747, 49004, 52510, 22834, 18371, 28077. **NYP**: 27689, 48044, 60205, 51648, 27053, 65173, 62536, 36022, 49470, 55655.

Topic 27. **CNN**: 48318, 39788, 40268, 30754, 35012, 3553, 42657, 42793, 58409, 43689. **Fox**: 10658, 16722, 34052, 16603, 16563, 38831, 12629, 3598, 16775, 3114. **Huff**: 9897, 16530, 55262, 1735, 47491, 17807, 47369, 11247, 15041, 2377. **Breit**: 7705, 16758, 15418, 13101, 19347, 10946, 59668, 17223, 18750, 16521. **NYT**: 44301, 27331, 3481, 16605, 33742, 17125, 16868, 16467, 22856, 38888. **NYP**: 58424, 62470, 58079, 66201, 33378, 34479, 19782, 16541, 41876, 16748.

Topic 28. **CNN**: 31044, 44809, 60338, 8200, 52516, 22992, 1983, 48445, 58510, 55807. **Fox**: 46514, 39209, 39519, 29768, 33664, 39207, 59484, 58943, 35452, 27794. **Huff**: 41088, 2294, 1791, 2527, 16944, 35820, 58940, 36351, 29250, 33344. **Breit**: 59388, 10467, 32176, 21937, 47074, 59991, 40333, 45801, 54859, 10944. **NYT**: 3442, 2528, 36292, 57818, 26841, 79, 5783, 44555, 26664,

13881. **NYP**: 37264, 60701, 63209, 23180, 10178, 61152, 45302, 45076, 21315, 37572.

Topic 29. **CNN**: 5370, 5943, 16279, 41875, 23014, 2499, 6987, 6015, 8575, 5551. **Fox**: 22625, 22575, 8567, 3697, 22756, 38698, 21966, 5688, 9058, 49027. **Huff**: 5705, 3418, 5282, 39977, 878, 38365, 41605, 27619, 33204, 5578. **Breit**: 2357, 34682, 5694, 27927, 8582, 26267, 32384, 15207, 7962, 51745. **NYT**: 38387, 39511, 5632, 30272, 11624, 11802, 5849, 42372, 15295, 1519. **NYP**: 38590, 46541, 43984, 39203, 22612, 8518, 3227, 5616, 35491, 6413.

Topic 30. **CNN**: 12834, 12820, 56912, 2409, 57696, 29817, 50038, 27123, 65625, 56735. **Fox**: 15271, 16197, 15482, 49437, 16316, 11890, 4167, 12690, 63882, 16313. **Huff**: 65420, 14194, 58055, 35495, 52096, 46084, 18917, 43566, 64277, 53090. **Breit**: 7890, 15098, 27515, 57927, 16468, 29475, 23506, 16687, 18931, 54634. **NYT**: 64811, 14150, 11027, 9950, 30034, 62830, 49472, 57934, 12405, 8467. **NYP**: 63157, 16513, 58020, 25525, 27500, 35402, 26423, 44011, 13333, 58093.

Topic 31. **CNN**: 40883, 60073, 45442, 47780, 66073, 47018, 57685, 1530, 1803, 27870. **Fox**: 38325, 5939, 39474, 54708, 44688, 34237, 54184, 1421, 47715, 35102. **Huff**: 6195, 53859, 65214, 65114, 53135, 55111, 47030, 43457, 47202, 12078. **Breit**: 58702, 57208, 51227, 25817, 27911, 40737, 55540, 18717, 42631, 62134. **NYT**: 65623, 13394, 57574, 59446, 56216, 55016, 52453, 32816, 49563, 36481. **NYP**: 47461, 55354, 53865, 57144, 55160, 57763, 51672, 42502, 58347, 65727.

Topic 32. **CNN**: 5228, 18432, 17048, 60294, 31452, 17064, 1250, 43011, 31358, 31484. **Fox**: 24593, 37494, 38517, 41439, 54372, 23690, 45754, 61263, 40552, 1334. **Huff**: 5774, 34939, 2302, 8251, 50885, 15592, 696, 50917, 5142, 50876. **Breit**: 44194, 38015, 7918, 43423, 9199, 27009, 31618, 6376, 38280, 9201. **NYT**: 54223, 28551, 43858, 2159, 26747, 30253, 42923, 27523, 21090, 27586. **NYP**: 37943, 58555, 26131, 59816, 42493, 37261, 33388, 31703, 6127, 40455.

Topic 33. **CNN**: 4440, 46162, 47741, 41021, 38965, 22754, 12074, 39663, 50663, 7134. **Fox**: 32866, 55982, 47482, 45951, 53253, 43766, 42071, 32533, 46577, 6228. **Huff**: 64120, 57480, 48002, 44465, 31905, 6234, 1683, 49307, 16780, 6000. **Breit**: 2210, 36748, 36102, 27893, 51132, 9938, 6838, 4996, 24471, 704. **NYT**: 15462, 51790, 1032, 9831, 36054, 29577, 44683, 31507, 43101, 43299. **NYP**: 4888, 498, 21705, 6785, 26875,

2439, 35833, 54726, 41755, 24068.

Topic 34. CNN: 6266, 65009, 66144, 43941, 7262, 32916, 39176, 32957, 7827, 65108. **Fox**: 54014, 37279, 25070, 25979, 57486, 2458, 55106, 53826, 38417, 42652. **Huff**: 63169, 712, 33532, 42077, 30986, 35798, 56157, 4025, 51157, 54739. **Breit**: 7502, 12430, 1267, 13007, 8824, 24116, 7093, 5218, 25354, 56127. **NYT**: 47350, 46129, 28442, 46107, 58116, 37725, 2418, 33019, 43138, 30807. **NYP**: 38718, 1997, 54988, 41888, 42854, 45033, 35790, 806, 8543, 63421.

Topic 38. CNN: 26089, 9780, 11160, 10135, 38444, 10528, 15396, 21793, 11899, 30939. **Fox**: 9904, 61954, 38592, 1912, 9742, 48158, 47330, 11298, 2145, 1109. **Huff**: 1053, 10636, 5194, 56248, 59653, 10001, 16251, 26935. **Breit**: 9820, 37006, 9757, 10047, 43153, 47998, 15249, 10466, 4255, 6684. **NYT**: 28253, 10549, 21983, 10078, 38217, 14959, 9733, 60099, 9898. **NYP**: 10117, 42971, 54551, 10258, 42732, 37803, 28002, 9832, 31722, 10433.

C Learning Partisanship

The news articles are split into the training set comprising topical documents and validation set comprising non-topical documents. Non-topical documents have small probabilities (<0.15) categorized to all topics. We do such a split because all documents are assigned a partisanship label, but not all of them are topical. For the topical documents from which we will generate contextualized topic embeddings, we use them as the training data to finetune the language model during the training phase. As a result, train set has 30,571 documents and the validation set has 35,797 documents. The model is trained for 30 epochs and we pick the one with the best performance on validation set for the subsequent topic embedding generation. We train the model using Adam optimizer, with learning rate 1e-5 and weight decay 5e-4. We use a batch size of 64 and train the model on 4 RTX 2080 GPUs. Each epoch takes about 10 minutes. The best validation F1 score on classifying partisanship is 91.3.

D Annotating Topic Polarization

We recruit 3 annotators that work as academic researchers in the areas of NLP and social science. For each one of the 30 topics, the annotators are provided with the top-10 topic keywords and the summaries of top-10 most relevant documents from each news corpus (as a total of 60 documents). First, the annotators select 15 topics on which they feel it is straightforward to find two polarized political stances by reading the relevant documents. For example, on topic 12 about Democratic primaries, it is intuitive to perceive the two political stances are "endorsing Biden" and "endorsing Sanders" after reading relevant articles, and then this topic is likely to be selected. We take the overlap of the 15 selected topics from 3 annotators and obtain 10 topics: $T^{\text{labeled}} = \{t_1, t_2, t_8, t_9, t_{10}, t_{11}, t_{12}, t_{27}, t_{30}, t_{33}\}$ with defined polarized political stances. In other words, the annotators reach an agreement that it is more clear on these 10 topics that there are two political stances. We find that on each of these 10 topics, the two stances defined by 3 annotators reach a complete agreement.

We do not annotate all topics because 1) it is difficult for humans to discern the two political stances on some topics, especially when such two stances do not exist at all; 2) we use the vanilla LDA topic modeling which is not the state-of-the-art, so the modeled topics will change using different topic models, in which case the annotating step should be repeated. Nevertheless, we argue that annotating 10 topics is sufficient to quantitatively evaluate the effectiveness of PaCTE.

Given a topic t from T^{labeled}, the defined two stances, and its 60 most relevant documents (10 from each of the six news sources), for each document, we ask the annotators to label which stance it belongs to and label it as 0 or 1; if the annotator is not able to perceive a clear political stance, then the annotator will label it as -1. For each document, the majority vote of the three labels with be used as the final annotation. If no majority vote is achieved, in other words, the three annotators give three different labels to a document, then a fourth annotator will read the document again and decide the final label. For a complete list of all document labels on the 10 selected topics, please refer to our public repository.