Online Social Networks and Media 30 (2022) 100208

Contents lists available at ScienceDirect

Online Social Networks and Media

journal homepage: www.elsevier.com/locate/osnem

A Survey on Political Viewpoints Identification

Tu My Doan *, Jon Atle Gulla

Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway

ARTICLE INFO

Keywords:
Political Viewpoints Identification
Political Ideologies
Political Stance
Political Framing
Bias Detection
Natural Language Processing

ABSTRACT

Political viewpoints identification (PVI) is a task in Natural Language Processing that takes political texts and recognizes the writer's opinions towards a political matter. PVI reduces the ambiguity in texts by identifying the underlying meaning and clarifying the bias margin along the political spectrum (bias leaning). Thus, even non-experts can better understand political texts. For instance, they can identify misinformation, bias, and hidden political agendas. In this paper, we formally define the concept of political viewpoints identification, explain its importance and discuss to what extent current techniques can be used for extracting political views from text. Existing techniques address the problem of PVI inadequately. We outline their deficiencies and present a research agenda to advance PVI.

1. Introduction

Since the advent of the Internet and social media, users are producing a significant amount of data, which have become valuable resources for researchers. This has also resulted in the need to automate the analysis process significantly to overcome the burden of overwhelming amounts of data. Despite a large number of domains out there, in this work, we focus only on political texts. Political texts cover a variety of documents, including media data, floor speeches, political statements, proposed and enacted legislation, committee hearing transcripts, and treaties [1]. Political texts discuss political subjects, such as immigration, taxation, education, or regulations. Therein, speakers or writers express their view or opinion on the subject at hand. A set of such opinions can form an ideology. For instance, Libertarianism[1] values freedom and argues in favor of minimal restrictions for citizens to express themselves. In this work, we define a *political viewpoint* as one of a small number of distinguishable opinions on a political subject. For instance, a speaker could be in favor of raising the minimum wage, building a border wall, or decreasing the corporate tax rate. Some researchers refer to PVI as the problem of political perspective detection [2]. For mining political text, there are a number of techniques to extract political views from published text such as:

- *Content analysis for identifying political viewpoints* to determine to what extend they express political views.
- *Linguistic clues analysis* for the identification of, for example, nouns or phrases commonly used by political parties.
- *Network analysis* for the identification of, for example, the author of the text, the connections between the author and other political parties, or the links among the authors to other members of specific parties.

Let us look at an example. The texts below are exempts from speeches given by former US President Donald Trump (Republican Party) and in response by then Chairperson of the House Nancy Pelosi and then Senate Minority Leader Chuck Schumer (both Democratic Party) about immigration issues[2]:

President Donald Trump: "*... as part of an overall approach to border security, law enforcement professionals have requested $5.7 billion for a physical barrier. At the request of Democrats it will be a steel barrier rather than a concrete wall. This barrier is absolutely critical to border security. It's also what our professionals at the border want and need. This is just common sense. ... Some have suggested a barrier is immoral. Then why do wealthy politicians build walls, fences, and gates around their homes? They don't build walls because they hate the people on the outside but because they love the people on the inside. The only thing that is immoral is the politicians to do nothing and continue to allow more innocent people to be so horribly victimized ...*"

Nancy Pelosi: "*... The fact is, we all agree we need to secure our borders while honoring our values. We can build the infrastructure and roads at our ports of entry. We can install new technology to scan cars and trucks for drugs coming into our nation. We can hire the personnel we need to facilitate trade and immigration at the border. We can fund more innovation to detect unauthorized crossings ...*"

* Corresponding author.
E-mail addresses: tu.m.doan@ntnu.no (T.M. Doan), jon.atle.gulla@ntnu.no (J.A. Gulla).

[1] https://plato.stanford.edu/entries/libertarianism/, accessed on 13 November 2021.
[2] https://www.nytimes.com/2019/01/08/us/politics/trump-speech-transcript.html.

https://doi.org/10.1016/j.osnem.2022.100208
Received 24 September 2021; Received in revised form 27 March 2022; Accepted 15 April 2022
Available online 9 May 2022

Senator Chuck Schumer: "*... We can secure our border without an ineffective, expensive wall. And we can welcome legal immigrants and refugees without compromising safety and security. The symbol of America should be the Statue of Liberty, not a 30 foot wall ...*"

By analyzing the texts, we can see that they are expressing different viewpoints about the same problem — immigration. The Republican party's solution to immigration issues is to build a wall along the southern border to stop illegal immigrants from entering America, hence eliminating issues with drugs, murders, illegal contraband, etc. (*"This barrier is absolutely critical to border security"*) and this is the act of love to their citizens, not a sign of immorality (*"They don't build walls because they hate the people on the outside but because they love the people on the inside"*). On the contrary, Democratic party shows their disagreement on the wall (*"We can secure our border without an ineffective, expensive wall"*) and by having better infrastructure (*"We can build the infrastructure and roads at our ports of entry"*), more personnel (*"We can hire the personnel we need to facilitate trade and immigration at the border"*), issues can be solved. Another thing to note is that both Nancy Pelosi and Senator Chuck Schumer are members of Democratic party (same political network); therefore, convincing us that their views are likely supporting each other whereas the opposite views are true for members from different parties.

As there are various ways for us to extract political viewpoints, we find it more interesting to focus on content analysis as this tells us which parts of the text are used to support the viewpoints of speakers and where they are standing along the political spectrum. Identifying political opinions represents a necessary task in modern political discourse. Social media provides platforms for the exchange of ideas about, among other topics, political issues. Texts on social media sites come with particular features due to technical restrictions. Micro blogging services, such as Twitter, restrict the length of messages. Some users employ the use of emojis to convey more information. The accessibility of tweets has rendered them a welcome source of political texts for many researchers. This survey includes nineteen publications related to social media and political text mining.

The major contributions of this work are threefold: (i) defining political viewpoint identification, (ii) evaluating to what extent existing techniques from related disciplines can be used for political viewpoint identification, and (iii) proposing a research agenda for the task.

The paper structure is inspired by the work of Küçük and Can [3] and is arranged as follows: Section 2 discusses about background information. In Section 3, we focus on different techniques adopted in political viewpoint identification. Information about political annotation guidelines, summary of existing datasets for different tasks, and evaluation metrics are covered in Section 4. Section 5 talks about possible applications of political viewpoint identification before the survey concludes with future work in Section 6.

2. Background

Before the widespread use of social media platforms, users learned to share information with each other through verbal conversations, exchanging emails, personal websites, among other means. Modern digital services also allow users to discuss and share their opinions about movies, books, restaurants, research, and politics. In the domain of politics, understanding and challenging others' opinions matters. Thus, it is crucial to have access to reliable information concerning politics. Unfortunately, the more opinions have been shared online, the more difficult it is for users to filter only useful pieces of information among the vast amount of available information. That is why we need data mining.

Data mining is a broad term with different definitions. Using the metaphor of finding gold nuggets, we define it as searching for useful knowledge in a large, unstructured collection of raw data. Data formats can include multimedia, text, and graphics. Data miners use a multitude of techniques from fields including statistics, information retrieval, machine learning, data visualization, and engineering [4]. Data mining constitutes a vast area of research. In the scope of this survey, we focus exclusively on identifying viewpoints from political texts or *political viewpoint identification* (PVI) for short.

Ideology represents a system of political ideas (see definition[3]). Political parties bring together like-minded individuals who broadly conform with a common ideology. For instance, the Oxford Handbook on Political Ideologies lists Conservatism, Liberalism, Communism, Nationalism, and Populism. These ideologies have some representation in today's Western politics. For instance, the Republican party in the US represents the Conservative ideology. Large collections of individuals, such as parties of large countries, can represent multiple ideologies. For instance, the Communist Party of Vietnam (CPV) represents both Communism and Nationalism. Political parties have to enact laws to govern societal relations. Consequently, parties have to express their views on particular issues. Examples include taxation, sentencing, and regulations. Individuals express their opinions concerning specific issues. Political Viewpoint Identifications aims to determine the opinion or view expressed in text towards some political issue. For instance, a politician states that she wants to increase the value added tax to fund education.

Definition. Given a set of n political viewpoints $P = \{p_1, p_2, \dots, p_n\}$ and a given text T, the task of PVI is to determine to what extent and by which expressed views, the content of T is consistent with the elements of P. Each p_i is described in terms of political text which stands out as the most contrasted piece of information that can help discriminate one viewpoint from another.

The task of identifying a political viewpoint is more challenging than sentiment classification. The latter can use dictionaries assigning sentiment scores to terms. For instance, the review "This vacuum cleaner has worked *quite well* since we bought it last week. We are *very satisfied*" yields two expressions which reveal the sentiment. However, in the context of politics, ideas are more implicitly expressed, for instances as nouns [5]. Besides, political parties use their own terminology to frame information. For instance, Republicans use *death tax*[4] while Democrats use *estate tax*[5] expressing their view on the tax.

For PVI to succeed, systems not only have to find and highlight the viewpoints in political texts. Also, they must explain to users, how they determined the viewpoint. Only, if users can understand the results, the system will ultimately improve the political discourse.

The research field of political viewpoint identification has much space to explore for computer scientists, computational linguists, and political scientists. This survey catalogs existing work related to PVI. Ultimately, PVI should take political texts as input, find the expressed viewpoints, and explain to readers the results. For instance, given T the PVI method determines the sentences $S_v \subset S$ that include viewpoints. Subsequently, the method estimates the match to the given viewpoints P such that the system can show the user pointers and estimates to underlying viewpoints. Previous work concentrates on finding viewpoints. Explaining the model and estimates remains a largely unexplored territory. PVI ought to consider semantic relations and exceed the limitations of statistical models. Consequently, PVI becomes a very unique and challenging task.

We categorize related work into three domains:

- **Political Ideology/leaning/party detection** wherein the system seeks to detect writer's political ideology, leaning, or party affiliation. We think of political ideologies or parties as a set of multiple

[3] https://plato.stanford.edu/entries/law-ideology/, (accessed on 3 December 2021.

[4] See the statement by Senator Thune on 16 November 2021 [6].

[5] See the statement by Congressman John B. Larson on 15 November 2017 [7].

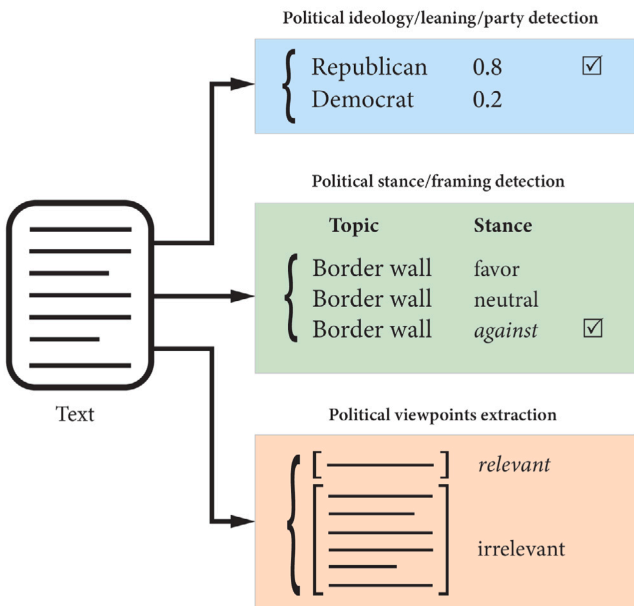

Fig. 1. Illustration of the various domains in the related work.

political viewpoints that share some common themes, $P_{\text{ideology}} \subset P$. Examples of targets for the classification tasks include *liberal vs. conservative*, *left/center/right*, and *Democrat vs. Republican*.

- **Political stance/framing detection** wherein the system aims to recognize the writer's opinion concerning specific political issues or actions related to those. Examples for classification tasks include *in favor/neutral/against*.
- **Political viewpoint extraction**: wherein the system detects parts of texts that relate to political viewpoints. Sentences $\{s_i\}_{i=1}^{M}$ compose the text T. The extraction applies a binary classification such that $T \supset \{s_{\text{relevant}}\} \cup \{s_{\text{irrelevant}}\}$. Subsequently, these parts can be used for either political ideology/leaning/party or stance/framing detection.

Fig. 1 outlines the three ways in which related work processes political texts. In case of detecting the party affiliation, the methods take the text or parts of it and apply a classifier. The classifier estimates the likelihood of the pre-defined classes. In the example, the system assigns an eighty percent chance to a Republican writing the text. In case of detecting the stance towards a political matter, a border wall in the example, the method takes the text or parts thereof and applies a classifier. In the example, the model assigns the highest likelihood to an opponent of a border wall having produced the text. In the case of detecting elements of the text related to political viewpoints, the methods take the full text and filter the sentences related to political viewpoints.

Having discussed the problem of PVI with many media representatives, we define a set of important characteristics for PVI methods. These features need to be fulfilled such that PVI becomes a useful tool for (social) media companies and their readers.

- *Content-driven*: methods should solely rely on content (text, non-stylistic features). In other words, methods should only use the text T. Other information, such as meta-data about the writer, should be ignored.
- *Multi-view*: methods should highlight multiple perspectives such as different political ideologies applying to varying degrees. Concretely, given a political issue, there is a set of discernible opinions $\{p_1, p_2, \ldots, p_k\}$, where k is typically small. The method should take the text and estimate the likelihood that each of the k viewpoints is expressed.
- *Explainability*: methods should show which elements led to the findings and to what extend. Texts comprise building blocks of various sizes: paragraphs, sentences, expressions, words, and characters. In addition to showing that a text likely stems from a particular ideology or stance, the method should visualize what parts of the text were indicative of the prediction. Such highlighting facilitates to better understand the method's abilities and deficiencies and can educate the reader about ways in which opinions are expressed.
- *Comprehensibility*: methods should convey explanations such that users can easily understand them. The success of explanations depends crucially on the recipients' background and understanding. While we expect experts to parse explanations swiftly, laypeople may struggle to see the relations between explanation and text. Hence, systems need to find ways to display hidden political viewpoints more clearly and comprehensibly.

3. Methodologies

In this section, we discuss in detail current techniques and approaches used for political text analysis. We group related work by the addressed problem first, and secondly by the used class of techniques. As illustrated in Fig. 1, we distinguish three methodologies:

- Political ideologies/leaning/party detection
- Political stance/framing detection
- Political viewpoints extraction

With the following sub-categories for each of them:

- *Traditional techniques:* this focuses mainly on using text features such as n-grams, word2vec, or lexicons for classification. Models in this category are Naïve Bayes (NB), Logistic Regression (LR), topic modeling such as Latent Dirichlet Allocation (LDA), Support Vector Machine (SVM), Decision Tree (DT), and Random Forest (RF).
- *Deep learning:* consisting of various deep learning models such as Recurrent Neural Networks (RNN), Long short-term memory (LSTM), Graph Convolutional Networks (GCN), or Transformers.
- *Others:* any techniques that do not belong to two sub-categories above, such as Ruled-based.

Tables 1–3 provide summaries of all the discussed papers in below sub-sections, grouped by techniques.

3.1. Political ideologies/leaning/party detection

The purpose of the detection task is to understand the political leaning of people based on their expressions through different means such as social media platforms, discussion forums, or political speeches. In some early work for ideology detection, the problem was narrowed down to a classification task — which sometimes was binary (Left/Right, Liberal/Conservative). The authors only focus on which ideology text T belongs to in a very general way, mostly depending on the political context of the studied country.

Research for ideologies already started a long time ago and one of them was done by Carbonell [27] back in 1978. He introduced a system called *POLITICS* that was used for reasoning about political ideologies. Later on, a spatial model was developed by Poole and Rosenthal [28] for analyzing the voting in US Senate from 1979 to 1982 by using only one-dimensional space with two targets (*yea/ney*).

By analyzing the text, political party affiliation can also be detected. The early work of this was done by Yu et al. [5] who implemented SVM and NB for political parties affiliation on the Congressional dataset [29]. Their approach took into account the concept of political ideology that one's opinions were highly dependant on their underlying ideologies towards the issues. Moreover, they found that

Table 1
Summary of political ideologies/leaning/party affiliation detection papers.

Technique	Author	Task	Features	Model	Dataset	Year
Traditional techniques	Yu et al. [5]	Party classification	BOW, tf-idf	SVM, NB	2005 Congressional Speeches (House & Senate)	2008
	Høyland et al. [8]	Party detection	bag-of-words, PoS tags, dependency relations	SVM	European Parliament Debates	2014
	Gu et al. [9]	Ideology detection	Twitter links (follow, mention, retweet)	ML-IPM	113 US. Congress (2013–2015)	2017
	Preotiuc-Pietro et al. [10]	Ideology detection	Unigram, LIWC, w2v clusters, emotions, political terms	Logistic Regression, Linear Regression, MTL	Users' survey and Twitter	2017
	Chen et al. [11]	Ideology detection	Entities and relationships, RDF	OKG	US. Congressional floor debate, IBC, Twitter	2017
	Sim et al. [12]	Measuring ideological proportion	Lexicons (cue, bigram, trigram)	CLIP	Ideological books and magazines, political speeches corpus	2013
	Bießmann [13]	Political bias detection	BOW	LR	German parliament speeches dataset	2016
	Wong et al. [14]	Political leaning	Retweet and retweeter information	Optimization framework	Twitter (2012 U.S. presidential election)	2016
	Duthie et al. [15]	Political ethos mining	NER, POS, Domain Specific Rules, Anaphora Resolution, Reported Speech Function, Lexicon	Ethos mining tool	EtHan_Thatcher_3 (UK parliamentary record)	2016
	Prati and Said Hung [16]	Ideological orientation detection	Unigram, tf-idf	NB, SVM, kNN, SVM, RF	Twitter Spanish 24M elections	2017
	Kannangara [17]	Opinion polarity classification, Ideology detection	Topic, sentiment	JEST, JEST-Ideology	Twitter	2018
	Baly et al. [18]	Factuality and Ideological detection	POS, linguistic cues, sentiment scores, complexity, morality, embedding	COR	MBFC dataset	2019
	Balahur et al. [19]	Opinion classification, Party affiliation	Sentiment Lexicon	SVM	Congressional Floor Debates	2009
Deep learning	Iyyer et al. [20]	Ideology detection	Word embedding	RNN	IBC	2014
	Xiao et al. [21]	Ideology detection	Twitter links (follow, mention, retweet)	TIMME	Twitter from politicians	2020
	Li et al. [22]	Ideology detection	n-grams, word embeddings	CB-LSTM	Convote, IBC, Twitter	2017
	Rao and Spasojevic [23]	Party classification	Word embedding	LSTM	Twitter	2016
	Gangula et al. [24]	Political bias detection	Word embedding, attention	Headline Attention Network (LSTM)	Telugu News articles	2019
	Kummervold et al. [25]	Party detection	Tokens sequence, NB-BERT embeddings	Transformer	Norwegian Parliamentary speeches	2021
Others	Djemili et al. [26]	Ideology existence	7 Pre-defined criteria on ideology	Rule-based	Twitter (posted by French Politicians)	2014

classic approaches – which worked well for other sentiment analysis tasks – had failed in the task of political ideology classification due to the lack of emotional keywords and ideas in political context were implicitly expressed in the form of nouns instead of sentimental adjectives.

Sim et al. [12] extended the classic detection by measuring the proportion of ideologies in the text with a new model called "Cue-lag ideological proportions" (CLIP) based on Bayesian Hidden Markov Model (HMM) and tested the model on 2008 and 2012 Presidential elections. The proposed model had two phases: cue extraction and cue-lag ideological measurement. In the first stage, the task was to create "cue lexicons" ($\mathcal{L}$) from ideological corpus (books and magazines). Spares additive generative (SAGE) was adopted to limit the number of cue terms to be classified as ideological cues for "cue lexicon". The latter stage concerned measuring the ideological proportion. Presidential data from the 2008 and 2012 US presidential elections was used to build the corpus. The authors transformed speeches into "cue-lag" representation by matching elements in $\mathcal{L}$ with the text and other elements into numbers by counting the occurrences of non-cue words between two cue-terms ("$\ldots \xrightarrow{3}$ econom_crisi $\xrightarrow{2}$ job_creation $\xrightarrow{5} \ldots$").

In the work by Høyland et al. [8], the authors predict party affiliations from European Parliament (EP) debates using SVM model. The paper is the preliminary work in assessing whether the task can be done using only speeches data. They used data from 5th EP to train model and predict on 6th EP data.

By taking the advantages of multiple link types (follow, mention, and retweet) in Twitter, Gu et al. [9] introduced a Multiple Link Types Ideal Point Estimation Model (ML-IPM) for ideology estimation via multiple link types where different weights were assigned to each link to detect numerical ideology position.

Prati and Said Hung [16] studied ideological orientation by detecting whether they belong into one of three categories: progressive ideology, conservative ideology, or no political orientation. They collected data from Twitter during the Spanish 24M elections and used different machine learning tools (Scikit-learn, SPSS statistical package, cloudtag) to study ideological orientation. Each tweet was converted into numerical vector using Vector Space Model, vector values were unigram weights (using TFIDF) before feeding into machine learning models. Different prediction classifiers were built using Scikit-learn namely Naïve Bayes (NB), Random Forest (RF), K-Nearest Neighbor (KNN), Linear SVM, and Logistic Regression (LR). Models, such as RF, KNN, or SVM can yield some explanations. Still, the type of explanation and users' feedback has not been demonstrated.

Instead of binary classification, Preotiuc-Pietro et al. [10] experimented with political ideology prediction on seven-point scale using a broad range of linguistic features: uni-gram, Linguistic Inquiry and Word Count (LIWC), w2v topic clustering, sentiment & emotions, and political terms. The authors analyzed several data sets obtained from Twitter. The texts had been annotated through surveys extending to a seven point Likert scale unlike other work. To discover the relationships between language and ideological groups per feature set, the authors started by comparing differences among those groups: "Very Conservatives and Very Liberals", "Moderate Conservatives and Moderate Liberals", "Moderates and Extremists", then the differences among political terms used by both parties. For the prediction task, "Cross-Group prediction", "Political Leaning and Engagement prediction" and "Encoding Class Structure" were implemented using Logistic Regression, Linear Regression, and Multi-task learning (MTL) respectively.

The work done by Kannangara [17] not only focused on ideology detection but also on mining a fine-grained political opinion polarity and sarcasm detection on Twitter data. A probabilistic LDA-based model called Joint-Entity-Sentiment-Topic (JEST) was introduced for opinion mining and JEST-Ideology for ideology detection. In the former model, the task was to classify opinion polarity (target, discussed topic, and sentiment). To detect political orientation, JEST-Ideology model was used to emphasize the importance of users' opinion and the topic. The models can also be extended to other domains outside politics.

Bießmann [13] automated the bias detection process using German political dataset of parliament speeches and manifesto statements. He built a set of Logistic Regression (LR) classification models with Bag-of-words (BOW) feature vectors in scikit-learn [30]: party affiliation (using parliament data), government membership (using parliament data), and political views (using manifesto data).

Balahur et al. [19] explore the use of lexica with terms related to emotion, opinion, and attitude to estimate the opinion of speakers in the US parliament towards a set of topics. In addition, they use machine learning in the form of a Support Vector Machine. The work presupposes a simplistic binary case with given topics. Nevertheless, the use of lexica could serve as part of a baseline for PVI. The authors point to an important problem for the work with parliamentary speeches. Responding to a previous speech can induce a high degree of uncertainty regarding the target of positive as well as negative statements. Criticism could be directed both the subject of the debate or comments by previous speakers.

Wong et al. [14] consider the problem of estimating the party-affiliation or likely voting preferences of Twitter users during the 2012 presidential election between Barack Obama and Mitt Romney. The authors observe the retweeting behavior of a set of users and derive a similarity measure. Subsequently, the proposed method regularizes a convex minimization problem finding an approximation of the political preferences. Therein, the authors consider the two option on a real line between -1 and 1. The work shows how difficult estimating political preferences is especially in the absence of sufficient text. Compared to [14], PVI considers individual issues and moves down from the very top-level consideration of a binary choice between two parties.

Duthie et al. [15] introduce an NLP pipeline to identify the support of or attacks on a speaker's ethos. The authors take transcribed debates in the UK parliament and apply a sequence of transformations. These include named-entity recognition, part of speech tagging, domain-based rules, resolving pronouns, and sentiment classification. The proposed system detects the target of comments and visualizes the relationships between speakers as network. The work complements the idea of PVI. While PVI aims to recognize speakers' opinion about political issues, Duthie et al. [15] seek to discern opinions about other speakers.

In the same category for traditional techniques, Baly et al. [18] adopted the Copula Ordinal Regression (COR) model [31] to study the problem of factuality and political ideologies jointly for news media sources with some auxiliary sub-tasks. The task of bias detection is done on 7-pt scale (from extreme-left to extreme-right) and 3-pt scale for factuality (low/mixed/high). For the auxiliary sub-tasks, they mapped the 7-pt scale to various scales such as Bias5-way (5-pt scale for bias detection), Bias3-way (3-pt scale), Bias-extreme (2-pt scale), and Bias-center (2-pt scale). Different features were used for the training including POS tagging, linguistic cues, sentiment scores, document embedding, morality, Twitter meta data, and also web traffic of news media. COR model was trained on MBFC dataset [32].

The work by Chen et al. [11] used both knowledge graph (KG) and probability technique to detect ideologies using the distribution of opinions in the text. The authors utilized entities, their relationships in DBPedia, and Resource Description Framework (RDF) triple to build Opinion-aware Knowledge Graph (OKG) model. Their proposed framework contains three components: opinion estimation, ideology propagation, and ideology detection.

Using Recursive Neural Network (RNN) and word embedding, Rao and Spasojevic [23] considered two different applications: actionable classification (out of scope in this survey) and political leaning which was a binary classification task (Democratic/Republican) on social media text. For the latter application, the authors collected data from Twitter for three months between October 2015 and January 2016 and implemented RNN model using LSTM and word-embedding. The model was able to predict extremely high or low scores which associated with extremely strong or weak political views.

While other models use different features like bag-of-word (BOW), lexicons, or other hand-crafted features, in the work of Iyyer et al. [20], a RNN model was built and tested on the corpus of Thomas et al. [29] for ideology detection. To emphasize the importance of detecting political bias at sentence level, they also created a new bias annotations (sentence and phrase level) dataset based on the Book Corpus (IBC).

In the work by Li et al. [22], the authors used Convolutional and Bi-directional Long Short Term Memory neural network (CB-LSTM) for detecting political ideologies. The model helps capture the target-related context while simultaneously learn semantic representation for text. The target-related representation per sentence is learned through CNN which is concatenated with word embedding layer and fed into Bi-LSTM. Then it goes through a number of hidden layers and finally to softmax layer.

The work done by Xiao et al. [21] used various relation link types (follow/retweet/reply/mention/like, etc.) in Twitter – which was underestimated by other work – for ideologies detection. The authors proposed Multi-task Multi-relational Embedding model (TIMME) model which consisted of two components: (i) the Multi-relation encoder — a Graph Convolutional Networks (GCN) based approach and (ii) a Multi-task decoder containing TIMME decoder and TIMME-hierarchical decoder. Data collection was done on politicians' tweets.

Gangula et al. [24] focused on political bias in news articles based on attentions. Due to the lack of datasets for bias detection in Telugu language (spoken widely in some states of India), the authors created a new dataset with articles taken from Telugu newspapers. They built a Headline Attention Network model using LSTM consisting of different components: headline encoder, article encoder, headline attention layer, and final layer for bias detection. To compare the developed model, they also trained other classifiers such as NB, SVM, and CNN.

Kummervold et al. [25] fine-tuned a classifier for party affiliation detection using Transformers and NB-BERT language model. To train classifier, the authors created a balanced dataset of 6000 speeches from the Norwegian Parliament (Storting) between 1998 and 2016 and kept two major parties (Fremskrittspartiet and Sosialistisk Venstreparti) [33].

Some related work falls neither in the category of traditional machine learning nor deep learning. For the "Other techniques" subgroup, we have the model introduced by Djemili et al. [26] which was the first of its kind that used linguistic rules based on certain criteria to identify whether the text bearing any ideology for tweets posted by French politicians. Since those criteria are just linguistic based, they need to be integrated into a NLP tools. For the first step, the authors used Wikimeta[6] to do POS tagging for the text then checking those POS elements against seven rules. For example, for rule 1, they checked if a tweet contained any temporal elements, or any subjects like I (*je*), you (*vous*), we (*nous*), and so forth. In order to detect a large number of tweets, they developed an Ideological tweet detection system taking a set of tweets as input and output the ideological tweets. Input text went through the analysis step by using Wikimeta API to do POS tagging, then those tagged elements were checked against seven rules. The text would be classified as containing ideology if it satisfied all seven rules.

3.2. Political stance/framing detection

Since stance detection is a wide topic, we only focus on stance detection in political text. Readers can have a look at the work done by Küçük and Can [3] for more detailed review on this topic. Stance detection is about finding the position hold by the speaker towards an idea, a topic, or an object [62]. We can think of stance in the stance detection task as a form of high-level political viewpoint p. Similar to other tasks, stance detection can only give us results by detecting to which category the text belongs. Table 2 provides a summary of related articles about stance detection.

One of the early works for stance detection was done by Thomas et al. [29] who developed a support/oppose classification model for single speech segment and multiple speech segments using SVM. They took into account the relationships of same-speakers constraints and different-speakers agreements.

Focusing on classifying stance for political debates online, Walker et al. [34] combined various features in their classifier — such as count (character, word, sentence), unigrams, bigrams, cue words, repeated punctuation, LIWC features and their dependencies, syntactic and POS dependencies, and opinion words. The authors trained NB, JRIP, and SVM model to (i) classify the posts based on their features and (ii) investigate the effectiveness of classifying speaker-based documents and the use of MinCut algorithm to the final results.

Johnson and Goldwasser [36] built a weakly supervised joint model using Probabilistic Soft Logic (PSL) [63] to analyze politicians framing activities and their patterns through time on Twitter data. The model applied a set of rules – defined using first order logic – to incorporate various information from Twitter (temporal activity patterns, political framing, and their agreement patterns) to improve the accuracy of stance detection. For example, we expect politicians of the same party to have the same stance ($\mathrm{LocalSameStance}_1(P_1, P_2) \rightarrow \mathrm{SameStance}_1(P_1, P_2)$), whereas representatives of two different parties are like to disagree on political matters. If Democrats agree on the issue then the rule will be DEM whereas the disagreed rule for Republicans will be $\neg DEM$ (negation of Democrats).

The work by Bar-Haim et al. [39] focused on the related task of claim stance classification on the first benchmark dataset. By giving the topic and a set claims, the goal is to identify whether each claim supports or contests the topic. The problem was broken down into the subtasks of open-domain target identification and open-domain contrast detection. In the first subtask, the authors implemented the target extraction (logistic regression with L2 regularization) and targeted sentiment analyzer (sentiment matching, sentiment shifters application, sentiment weighting & score computation). For the second subtask, the algorithm was developed to generate the top-K *anchor pair* for complex targets that represented the semantic link between the targets. The authors introduced a new relatedness measure based on the probability of the co-occurrence of the anchor pair with consistent and contrasting cue-phrases for contrast relation task.

In the work published by Skeppstedt et al. [37], the authors detected stance modifiers at seven levels namely: Certainty, Uncertainty, Hypotheticality, Prediction, Recommendation, Concession/Contrast, and Source. The idea is to classify the text at the token-level by detecting useful cue-words rather than classifying at the sentence-level. They also incorporated features from unlabeled data using clustering techniques. Scikit-learn library was used to train SVM and Logistic Regression classifier.

Both work done by Lai et al. [38,41] also focused on stance detection on Twitter data. In Lai et al. [38], a supervised model (Naïve Bayes) was used to detect stance in the US Presidential election in 2016 between Hillary Clinton and Donald Trump. They defined two concepts — "enemies" and "friends" — in order to reflect the relationship of entities in relation to the target. They used (i) four lexica (AFINN, Hu&Liu, LIWC, DAL) as their sentiment features, (ii) structure features (hashtags, mentions, punctuation marks), (iii) contextual features (target of interest mentioned by name, pronoun, target's party, party colleague opposite, target's opposition's party, nobody) and (iv) additional labeled-based features (sentiment, opinion target) as features for NB classifier.

In another work also by Lai et al. [41], the shifting perspective of users through time and network-based features (retweet, quote, and reply) were taken into account. The authors did some network analysis by building graphs (retweet, quote, and reply) that represents the relationship among Twitter users during the debate period. They also introduced new model with three new network-based features — Retweet Communities, Quote Communities, and Reply Communities. SVM classifier was trained for stance detection task.

In the work of Vilares and He [40], the authors introduced a Latent Argument Model (LAM) – a Bayesian approach – for modeling both topics and stance as latent variables. They define switch variable x which can be either *background, topic, argument word*. The model samples it from various distributions based on the word type. Collapsed Gibbs Sampling is adopted for inferring model parameters and latent variables. Part-of-Speech (POS) tags and subjectivity lexicon are used to separate topic and perspective words, resulting into two variants of the model namely *LAM_POS* and *LAM_LEX* respectively. The model is tested on House of Common Debates (HCD) dataset.

Hardalov et al. [54] proposed an end-to-end framework for cross-domain label-adaptive stance detection with respect to unseen labels. They used Mixture-of-Experts with Label Embeddings (MoLE) which was based on input representations from a pre-trained language model with a well-known technique for multi-source domain adaptation Mixture-of-Experts (MoE) and Domain-adversarial neural network (DANN). To solve the challenge with out-of-domain dataset prediction, the authors adopted various methods to learn the probability distribution over the set of test labels such as label embeddings (LEL), hard mapping, soft mapping, and weak mapping. The authors evaluated their model on 16 different datasets (e.g., Arc, Argmin, Emergent, Fnc1, and SemEval2016T6).

In the work done by Vamvas and Sennrich [52], the authors fine-tuned a multilingual BERT model on the newly introduced multilingual dataset — X-Stance (English, Swiss German, French, and Italian). Four baselines were trained to evaluate the task on X-Stance dataset namely (i)*global* majority class baseline, (ii) *target-wise* majority baseline, (iii) bag-of-words baseline, and (iv) multilingual BERT (M-BERT) baseline.

Schick and Schütze [55] introduce a training method that provides soft labels for various natural language tasks. Their work considers a set of tasks among which is stance detection. The training method uses existing language models and a pre-defined pattern to generate

[6] http://www.wikimeta.fr.

Table 2
Summary of Stance/Framing detection papers.

Technique	Author	Task	Features	Model	Dataset	Year
Traditional techniques	Thomas et al. [29]	Stance detection	Unigrams, word-presence vector, weighted links,	SVM	Political debates	2006
	Walker et al. [34]	Stance detection	Count features, unigram, bigram, cue words, punctuations, LIWC features and dependencies, syntactic, POS, opinion words, context	NB, JRIP, SVM	Online political debates	2012
	Salah [35]	Stance detection	BOW, TFIDF, POS, sentiment lexicon	Debate Graph Extraction (DGE)	UK House of Commons parliamentary debate	2014
	Johnson and Goldwasser [36]	Stance detection	Content, frames, temporal activity, logical rules	PSL	Twitter (politicians' tweets)	2016
	Skeppstedt et al. [37]	Stance modifiers detection	Cue words	LR	Political Blogs about Brexit	2017
	Lai et al. [38]	Stance detection	Sentiment, structural, context-based, labeled-based features	NB	Twitter (US 2016 elections)	2017
	Bar-Haim et al. [39]	Claim stance classification	Nouns phrases in claims, sentiment features (lexicon, shifters, weighting, score), contrasting sentiment complex phrases and semantic relations, TF-IDF	Logistic regression classifier,	Claim polarity dataset	2017
	Vilares and He [40]	Stance detection	Topics, perspectives	LAM	UK Parliament House of Commons	2017
	Lai et al. [41]	Stance detection	Diachronic perspective, network-based features (retweet, quote, reply)	SVM	Twitter (Italian tweets)	2018
	Menini et al. [42]	Stance detection	Lexical overlap, Topic position, Similarity with other related/unrelated pairs, Negation, Keyword embeddings, Argument entailment, Argument sentiment	SVM	Nixon and Kennedy - 1960 Presidential campaign	2018
	Lai et al. [43]	Stance detection	Stylistic, Structural, Affective and Contextual features	MultiTACOS	E-USA (English), R-CAT(Spanish–Catalan), E-FRA(French), R-ITA(Italian)	2020
	Tsur et al. [44]	Political framing detection	Topic, n-grams, time series	LDA, Autoregressive Distributed Lag	US. Congress statements	2015
	Naderi and Hirst [45]	Political framing detection	word2vec, syntactic embeddings, POS-tags, dependency relation	SVM	ComArg, argumentative parliamentary statements	2015
	Baumer et al. [46]	Language of framing	Lexicon, grammatical, document, theoretical and dictionaries features	SGD, NB, Perceptron, NN, LR, Passive Aggressive	Framing Annotation Data for News Articles	2015
	Johnson et al. [47]	Political framing detection	n-grams, word2vec, logical rules, phrase indicators	PSL	Congressional Tweets	2017
	Johnson et al. [48]	Political framing detection	n-grams, word2vec, logical rules	Global PSL	Congressional Tweets	2017
	Dahlberg and Sahlgren [49]	Political framing detection	n-grams	Random Indexing	Swedish Blogs	2014
Deep learning	Lehmann and Derczynski [50]	Stance detection	fastText, context-based features (party, politician)	LSTM, MLP	Politicians' quotes in Danish	2019
	Bhavan et al. [51]	Stance detection	Text-based (TF-IDF, LDA-based, NRC Emotion) and graph-based (node2vec)	manModel, govModel	HanDeSeT	2019
	Vamvas and Sennrich [52]	Stance detection	BOW	M-BERT	X-stance	2020
	Sawhney et al. [53]	Stance analysis	Context, embeddings	GPoIS	ParlVote	2020
	Hardalov et al. [54]	Stance detection	Word embeddings, labels mappings	MoLE	16 datasets (e.g., Arc, Argmin, Emergent, Fnc1, SemEval2016T6)	2021
	Schick and Schütze [55]	Stance detection	Text, language models	PET, Transformer	X-stance	2021
	Hardalov et al. [56]	Stance detection	Text, language models	PET, Transformer	15 data sets	2022

Table 3
Summary of political viewpoint extraction papers.

Technique	Author	Task	Features	Model	Dataset	Year
Traditional techniques	Trabelsi and Zaiane [57]	Grouping arguing expressions to viewpoints	Viewpoints	JTV	ObamaCare, Assault Weapons, Gay Marriage 1& 2, Israel–Palestine conflict	2014
	Paul and Girju [58]	Identifying topics and aspects	Topics, aspects	TAM	ACL Anthology, Israeli–Palestinian conflict	2010
	Thonet et al. [59]	Identifying topics and opinions	Topics, viewpoints	VODUM	Israeli–Palestinian conflict	2016
	Lin et al. [60]	Identifying perspectives	Perspectives	NB, LSPM	Israeli–Palestinian conflict	2006
	Menini and Tonelli [61]	Viewpoints comparison	Sentiment, word embeddings, cosine similarity, entailment, lemma overlap, negation	SVM	1960 Elections, 1960 Elections Extended, Debatepedia	2016

soft labels for unlabeled instances. The evaluation focusing on the stance detection uses two patterns to mask input and fine-tune the transformers to provide labels. The stance detection is represented as binary classification: either the author is in favor or against. The method does not allow to detect which passages are related to a specific political matter. As such, the method cannot readily be applied to PVI.

Hardalov et al. [56] extend the work of Schick and Schütze [55] with an label encoder. The encoder allows them to use a varying list of labels that no longer have to be fixed terms. Further, they consider a multi-lingual setting with 15 data sets covering twelve languages. Their method can assign labels to texts for stance detection. Still, the method lacks the functionality to determine passages of a text that convey the target of the opinion. Some adjustments are necessary to apply the method to PVI.

Lai et al. [43] used Stance Detection System (i.e. MultiTACOS) to train different machine learning models (SVM, NB, LR, Neural Networks) on various datasets for Stance detection task (English, Spanish, Catalan, French, and Italian). They investigated the portability of the task across various languages and feature groups of features namely *Stylistic, Structural, Affective* and *Contextual*. The authors also annotated two new dataset for the purpose of evaluation in French and Italian. The evaluation shows that *Structural* and *Stylistic* features works best in supervised context, *Affective* and *Contextual* features are mostly used for dataset with the absence of stance detection target (semi-supervised) and *Contextual* features are useful for stance detecting referendum data. The work also highlighted the importance of different feature groups and traditional methods performed as competitive as the neural networks models (LSTM, biLSTM and CNN).

Sawhney et al. [53] present a combination of neural language model and neural graph attention model to estimate the viewpoint of politicians on particular motions. They use a data set comprising more than 30 000 transcribed speeches in the UK parliament as well as the text of legal motions related to those to refine a BERT language model. Simultaneously, they map speakers, transcribed speeches, and motion documents into a graph. Subsequently, they apply an attention-based approximation to determine suited weights for the edges between the nodes. Their evaluations suggest that the proposed system can classify political stances, as determined by whether politicians voted in favor of the motion, in about four out of five cases. Besides, the authors report that their system facilitates visualizing party-cohesion with the learned attention weights.

One of the rare works that used non-English datasets for stance detection is from Lehmann and Derczynski [50]. Deep learning technique with LSTM and pre-trained fastText word embedding of 300 dimension were used to build two different models for stance detection: Conditional LSTM and Multi-layered perceptron (MLP). Two new trained baselines using Naïve Bayes (NB) and Random Forest (RF) were created to compare the results of the Conditional and MLP model.

Framing is widely used by politicians to put bias into discussions to support their stance [47] — which started originally from social science research [64,65] before it gained more attention in the field of computer science, especially in NLP [44,46–48]. Understanding the language of framing used per person can help us distinguish the characteristics features of their language use; hence, making the speaker stand out among the others.

Dahlberg and Sahlgren [49] investigate the framing of the concept for "outsiders" in Swedish politics both qualitatively and quantitatively. They explore documents published by the two major parties in Sweden and highlight statements with the term. Subsequently, the apply Random Indexing (RI) to a large corpus of 1.5 billion words collected from Swedish blogs. Recently, Transformers have superseded RI when it comes to representing concepts in vector spaces. The work highlights some of the problems with automatic framing detection in political texts.

Tsur et al. [44] presented a statistical framework that used topic modeling and time series analysis to understand the framing strategies. LDA was implemented to identify topics in text in an unsupervised manner and Autoregressive-Distributed-Lag model for time dependency analysis. For time series analysis, they implemented two different time series: weekly and yearly per topic which revealed two seasonal effects.

For models that were implemented using Probabilistic Soft Logic (PSL), there were two related works done by Johnson et al. [47,48]. In Johnson et al. [48], the authors presented a first model for in-depth framing detection using six different weakly supervised models which utilized extracting features from tweets as input data for each global Probabilistic Soft Logic (PSL) model [63]. A set of rules was defined for each model. The latter model inherited from its previous models to improve the final accuracy. For example, in Model 1, the goal was to check whether the tweet contained any unigram that was in the list of unigram keywords for a particular frame ranging from 1 to 17 possible frames, if the result were true then the model would return the frame number ($\mathrm{HasUnigram}_F(T,U) \rightarrow \mathrm{Frame}(T,F)$ where T,U,F were the tweet, unigram and the frame number respectively). Model 2 was the combination for Model 1 plus the condition for Party affiliation (Model 1 + Party: $\mathrm{HasUnigram}_F(T,U) \rightarrow \mathrm{Frame}(T,F) \wedge \mathrm{Party}(T,P) \rightarrow \mathrm{Frame}(T,F)$). As one tweet could have multiple frames making this a multi label classification task.

Another work also by Johnson et al. [47] used ideological phrases from politicians' slogans (from tweets or speeches) for framing detection. Their model consisted two parts: the frame detection and the ideological phrase indicator containing a list of frequent used phrases in each frame. The adopted PSL model was very much similar to the one in [48] which also used n-gram keywords, word similarity, political party as features. But in this model [47], the authors also combined similarity between phrases as the new feature. Similarly, the task was treated as a multi label classification due to the possibilities of one tweet having multiple frames.

Different from previous work, Baumer et al. [46] did not focus on framing detection but on the use of language related to framing. They wanted to focus on where framing happened in the text rather than doing frame detection. A set of different feature subsets was used for classifiers — such as lexicon, grammatical, document, theoretical, and dictionaries features. They performed training and testing on different models: Stochastic Gradient Descent (SGD), Naïve Bayes (NB), Perceptron, Nearest Neighbor (NN), Logistic Regression (LR), and Passive Aggressive.

Naderi and Hirst [45] conducted the framing identification on parliamentary discourse speeches. They also showed how the use of embeddings could improve the results of the task. The authors also introduced a new argumentative parliamentary discourse corpus for the task. The supervised model was trained on the ComArg corpus[7] containing user-postings annotated with seven known frames and test on the argumentative parliamentary statements. The authors compared the semantic textual similarity (STS) between statements and frames. They represented the statements using word2vec, syntactic embeddings, and skip-thought model which was then used to measure the semantic textual similarity (STS) between statements and frames representation. They adopted two similarity metrics to measure this: (i) cosine similarity score and (ii) the concatenation of component-wise product of two vectors and their absolute difference score (P&D). Additional features were also added to the similarity scores to understand more the impact of stance features such as POS-tags, typed dependencies, and distributed representations of statements. The evaluation suggested that the P&D similarity score gave best measure in capturing the similarity and adding stance features to cosine similarity helped improve the accuracy of the model.

Salah [35] investigates the question on how to estimate politicians' stance on particular issues. The analysis takes a set of speeches related to motions in the UK parliament and transforms those into a bag-of-words representation. Subsequently, the author tests sentiment analysis methods to predict the approval or dissent towards the motions. The evaluation finds that the textual representation alone fails to provide meaningful insights. Without additional information, the sentiment prediction fails in two out of five cases. The authors verifies that introducing additional information either by political affiliation or more sophisticated graph representation can boost the binary classification performance.

Bhavan et al. [51] also worked with the debate data from the UK parliament. Their work explores the utility of adding meta-data in the form of social graphs. They map politicians to their parties and use a random walk procedure to arrive at numerical representations. Their evaluation suggests that this information is helpful to determine the political stance of speakers.

Menini et al. [42] take political documents related to the US presidential election in 1960 between Kennedy and Nixon. The work defines a pipeline to automatically process transcribed speeches and published documents to obtain a graph of arguments related to each other. Argument mining relates to political viewpoint identification. Both focus on particular political issues and speakers' or writers' opinion about how to deal with them. The authors identify focus on the relation of arguments to support the work of political scholars and historians. PVI focuses more on the opinion of individuals about how to govern in the future.

3.3. Political viewpoint extraction

In the work by Trabelsi and Zaiane [57], the authors introduced a probabilistic model called Joint Topic Viewpoint (JTV) for mining contentious documents by identifying arguing expressions and group them by viewpoint. For the JTV model, the authors enhanced LDA to model the complex structure of contentious documents. Each term in a document was assigned a pair topic-viewpoint label. They assumed that a document might contain various discussed topics in different proportions and viewpoints were also expressed proportionally for each topic. The JTV model generated a probability distribution for all terms for every topic-viewpoint pair. To group similar topic-viewpoints into clusters, they modified the constrained k-means clustering from Wagstaff et al. [66]. However, their experiments were limited to binary viewpoints provided by the datasets, even though this could be extended to more than two labels. For example, for Gay Marriage 1 (GM1) dataset, two viewpoints "should be illegal" and "should be legal" were used.

The Topic-Aspect (TAM) – LDA-based topic model – was developed by Paul and Girju [58] to capture the topics and aspects (perspectives) from data in unsupervised manner. Beside identifying aspect-neutral word distribution or aspect-dependent distribution of a word, the authors also added to the model a mixture component to group words into background, topic-specific, and aspect-specific tokens.

Topic Model Unifying Viewpoint, Topic and Opinion Discovery (VODUM) by Thonet et al. [59] used probability based on LDA to model viewpoints, topics, and opinions. The authors defined four properties in VODUM: (i) using part-of-speech tagging to identify topical words (nouns) and opinion words (adjectives, verbs, and adverbs), (ii) using sentence-level instead of word-level for topic variables, (iii) defining document-level for viewpoints, and (iv) defining topic distributions to viewpoint-specific because different viewpoints have different dominating topics.

The work done by Lin et al. [60] focuses on identifying viewpoints at both document and sentence level. Beside using the NB model, the authors also developed a Latent Sentence Perspective Model (LSPM) for identifying how strongly the perspective was reflected in the sentence when the annotation was unavailable with three intensity levels: strong, little, and no perspective.

Menini and Tonelli [61] focused on comparing viewpoints between politicians. The authors built a SVM classifier to classify the agreement/disagreement labels from 1960 Elections, 1960 Elections Extended, and Debatepedia dataset. First, a feature vector is learned for each pair of snippets using word2vec embeddings, snippet features (sentiment, semantic, surface features), and subtree features containing keywords for the topics. Then the feature vector is used as input for SVM.

3.4. Limitations in current techniques

The previous sections discussed three problems closely related to political viewpoint identification. Detecting the political ideology, party affiliation, or leaning can indicate the view of someone on particular political matters. We may assume that a member of a party shares the same views on many issues. Still, parties are subject to vivid internal debates where members clash about different opinions on political problems. Related work in this domain typically uses annotated data sets and evaluates classifiers with few classes. More recent research explores the use of deep neural nets. However high an accuracy the sophisticated classifiers report, we may doubt whether they actually reflect a speaker's or writer's true opinion about different issues. Some methods rely on the availability of meta-data. As stated earlier, media representatives as well as operators of large social networks might not have access to such meta-data for the texts they would like to run through the PVI pipeline.

Detecting and extracting parts of text related to political viewpoints represents an essential aspect of PVI. Previous work has established methods to identify sentences as well as smaller text fragments conveying political statements. Still, most of the presented works focus on a rather narrow scope of few topics. A functional PVI system relies crucially on NLP resources suited for its data source. The lack of such resources for languages other than English, Chinese, Spanish, and a few more, makes it difficult to apply the proposed methods readily in other countries. Furthermore, we have to account for differences regarding the political systems and cultures. Thus, extracting political viewpoints

[7] https://takelab.fer.hr/data/comarg/.

Table 4
Summary of contributions and limitations of relevant methodologies in PVI.

Methodology	Contributions	Limitations
Political ideologies/ leaning/party detection	– text preprocessing with NLP – annotated political texts – classifiers for party affiliation/ideology – demonstrations & tools – baseline methods	– simplifying problem as classification with few classes – focus on English and US/UK elections – required additional data (social graph, party affiliation) – lack of work on explainability
Stance/Framing detection	– text preprocessing with NLP – annotated political texts – classifiers for stances/framing – demonstrations & tools – baseline methods	– focus on English and US/UK elections – frequently using other data than text – lack of work on explainability – generalizability remains unclear
Viewpoint extraction	– text preprocessing with NLP – annotated political texts – similarity measures for political texts – demonstrations & tools	– focus on English and US/UK elections – methods tend to rely on external data sources – lack of work on explainability

from political texts remains an open issue with a multitude of urgent research questions to be addressed in the future.

Detecting political stances and framing comes relatively close to PVI. Both problems take text as input and want to identify the viewpoint of the author. Previous work has conducted a variety of experiments to explore methods to discover political views. More recently, deep neural nets have become popular tools to represent language. They allow us to convert text into numerical representation. Mapping political text into such spaces yields some insights into relatedness of concepts or expressions. Still, there are a set of open issues which limit the applicability of proposed methods to the PVI task. Most discussed work represents the task of stance detection as classification problem. The availability of powerful libraries of classifiers allows researchers to quickly come up with measurements in the presence of annotated data. Our requirements towards PVI demand more than a classification. In addition to estimating the likelihood of viewpoints being expressed, we strive for explanations. Explanations help us to better understand what the underlying model reflects. In particular deep neural networks are notoriously hard to explain. We doubt that PVI systems can succeed unless they present credible, comprehensible clues on why the text has been classified as the model suggested. Existing work on either detecting political party affiliation, ideology, leaning, stance, or framing use a variety of machine learning models. Some models can yield explanations. Random Forests produce sequences of rules arriving at the decision. Support Vector Machines summarize the vector space with a minimal number of vectors. Nearest Neighbor methods return a representative data point. Knowledge Graphs can produce paths to link nodes. Still, the literature survey shows limited work on explanations. Other machine learning models remain opaque and hard to interpret. This group includes topic models as well as various kinds of deep learning architectures, such as RNNs, LSTMs, CGNs, and Transformers. These models could support the PVI task. Still, the lack of access to understandable explanations demands more research.

Besides technical limitation, the eco-system around political news, commentary, and official political business raises further limitations. The way in which we consume news today differs markedly from news consumption in the time when some of the data sets emerged, which the related work uses. In addition to news consumption, language has changed. We may ask, how representative of a language debates from the 1960s still are today. Social media presents additional challenges. For once, user contributed texts exhibit less grammatical cohesion and more stylistic deviances than pre-written political speeches. Political texts can be valuable assets for PVI. Still, there is much work needed to establish a sound link towards the everyday language on social media. A considerable set of related work uses data especially from the micro blogging service Twitter. Their micro blogs, or tweets, have limited space disallowing the sender to convey deep political thoughts. The lack of sufficient text impedes the ability of systems to automatically recognize political viewpoints. Finally, a hard problem for detecting political viewpoints comes from language usage. Unlike sentiment classification, political viewpoints avoid obviously positive or negative adjectives. Instead speakers or writers use more opaque expressions. Besides, the usage of sarcasm can obfuscate the true viewpoint of speakers or writers.

Table 4 summarizes the contributions and limitations of current relevant methodologies related to PVI.

4. Annotations, dataset and evaluation metrics

In this section, we go through the annotations guidelines for creating datasets used in the related work. Our suggestions for creating dataset for PVI are also included, followed by some discussions about evaluation metrics for evaluating model performance. A summary of datasets for each task will also be provided in Tables 5, 6, and 7.

4.1. Annotations

Even though most datasets use Twitter posts, people still work on other data sources such as news articles, books, speeches, debates. Beside many English datasets, there are also many datasets in other languages (Italian, Danish) which are discussed in the following sub-section.

4.1.1. Ideologies/leaning/party detection datasets

Iyyer et al. [20] used Crowdflower crowd-sourcing to create a new dataset (IBC) for ideological bias detection at sentence and phrase level instead of author or document level as in other datasets. The authors combined both Convote dataset [29] and the Ideological Book Corpus (IBC) data to build a new one which targeted ideologies at sub-sentential level. The main annotations classes were *liberal* or *conservative*. They hired Crowdflower workers to do the annotations. Those workers were requested to select the path that had ideologies (e.g., contained top-ten partisan unigram or open class constituencies). To ensure quality of the work, the annotators must qualify some requirements before taking the full task such as residing in US, having basic understanding about politics, correctly annotating 6 over 8 gold paths, and keeping the accuracy level at 75%. The result dataset contains 3412 sentences with 13 640 annotated nodes.

Gebhard and Hamborg [67] published POLUSA dataset with 0.9M political articles taken from 18 news outlets. After getting data from those outlets the authors did five different tasks to ensure the quality of the data. They did base selection in the first task by collecting subset of all data while still maintaining the reasonable number of articles for each time frame. In second task, almost-duplicate items were removed using nearest neighbor clustering technique from the base collection which reduced base data by 5%. The next task involving removing all non-English articles and keeping only those discussing about policy topics. Filtering articles was done both manually and also

Table 5
Summary of political datasets.

Author	Dataset	Domain	Annotations	Target	Size	Link
Iyyer et al. [20]	IBC (English)	Ideological books	Annotated nodes with labels (liberal, conservative, neutral)	Political Ideologies	3412 sentences with 13,640 annotated nodes	[a]
Menini and Tonelli [61]	1960 Elections Dataset	Discourses and official declarations	Agreement/disagreement, sentiment, debated topic similarity	Political Ideologies Comparison	350 snippets	[b]
Rao and Spasojevic [23]	Twitter (English)	Twitter	Democrat/Republican	Political leaning	27,130 users	[c]
Gebhard and Hamborg [67]	POLUSA Dataset (English)	News articles	Authors, publication date, URL, outlet name, political leaning, temporal	Political leaning, temporal	0.9M articles	[d]
Høyland et al. [8]	European Parliament Debates	Debates	Annotated with political parties	Political Leaning	689 speeches	NA
Wong et al. [14]	2012 U.S. presidential election	Twitter	Liberal/Conservative/Neutral and political leaning scores per label	Political leaning	100 sources	NA
Kummervold et al. [25]	The Talk of Norway (Norwegian)	Parliamentary speeches	Political parties, speakers' metadata	Political leaning	250,373 speeches	[e]
Solberg and Ortiz [68]	Norwegian Parliamentary Speech Corpus (Norwegian)	Political speeches	Political parties, speakers' meta data	Political leaning	65,000 sentences	[f]
Djemili et al. [26]	Twitter - politicians' tweets (French)	Twitter	NA	Ideologies existence	34,273 tweets	NA[g]
Prati and Said Hung [16]	Twitter - elections (Spanish)	Twitter	Ideologies orientations in tweets	Ideological orientation (progressive ideological trend, conservative ideological trend and no political orientation)	24,900 tweets	NA
Gangula et al. [24]	Telugu Newspapers (Indian)	News articles	5 political parties (BJP, TDP, Congress, TRS, YCP) and None	Political bias detection	1329 articles	NA
Baly et al. [32]	MBFC (English)	News media	Factuality (3-pt scale: high/mixed/low) & bias (7-pt scale: from extreme-left to extreme-right)	Political factuality and bias	1066 news media	[h]
Bießmann [13]	Political text (German)	Parliament discussions and party manifesto	Political party (Parliament speeches), political view (manifesto statements)	Political bias detection	29,451 political statements	[i]
Duthie et al. [15]	EtHan_Thatcher_3 (UK parliamentary record)	Political debates	Source-person, Target-person, Ethos support, Ethos attack	Political ethos mining	60 sessions, 739 segments, 253 speakers	[j,k]

[a] https://people.cs.umass.edu/~miyyer/ibc/index.html.
[b] https://dh.fbk.eu/2016/10/agreement-disagreement-datasets/.
[c] https://github.com/klout/opendata/blob/master/political_leaning/README.md.
[d] https://zenodo.org/record/3946057#.X9YJaS1Q3OQ.
[e] https://www.duo.uio.no/handle/10852/71356.
[f] https://www.nb.no/sprakbanken/en/resource-catalogue/oai-nb-no-sbr-58/.
[g] NA: Not Available.
[h] https://github.com/ramybaly/News-Media-Reliability.
[i] https://github.com/felixbiessmann/fipi.
[j] http://corpora.aifdb.org/Ethan3Train.
[k] http://corpora.aifdb.org/Ethan3Test.

by a trained convolutional neural network (CNN) model. This resulted in 13% data removal from base selection. Next, they assigned political leaning to news outlets in fourth step and then balanced the dataset in the last step to reduce the distortion in temporal distribution and over-representation of some news outlets.

4.1.2. Stance detection datasets

The most popular dataset for stance detection is the SemEval-2016 Stance Dataset created by Mohammad et al. [71]. The authors created a list of three query hashtags categories: *for hashtags, against hashtags, neither hashtags* for stances and collected tweets from Twitter that contained hashtags at the end of tweets that belonged to one of those categories. They used crowdsourcing (CrowdFlower) to annotate tweet–target pair data. A set of instructions was provided to annotators to label the data. To ensure quality of the annotations, the authors manually annotated 5% of data and distribute them among with other pairs for accuracy checking and annotators did not know about. In order to continue the work, they must maintain the accuracy rate at 70%. Beside the stance annotations, the authors also created the sentiments for the tweets (without targets). They asked annotators to label data with *positive, negative, neutral*. There are 4163 items in the Stance Dataset (2914 for training and 1249 for test set).

Table 6
Summary of political datasets (cont.).

Author	Dataset	Domain	Annotations	Target	Size	Link
Abercrombie and Batista-Navarro [69]	Hansard Motion Policies	Parliamentary Debates	Crowd-sourced, Manual	Opinion Topic Classification	592 motions	[a]
Baturo et al. [70]	UNGDC (UN General Debates)	Political Speeches	Speakers, States	Political Text Mining	7300 speeches	[b]
Thomas et al. [29]	Congressional Floor Debates (English)	Political debates	Speaker index, bill index, party, voting, page index, direct mention	Political party/stance	3857 speech segments	[c]
Walker et al. [34]	ConvinceMe.net debates (English)	Online debates	For/Against	Political stance	4731 posts	NA
Menini et al. [42]	Nixon and Kennedy during 1960 Presidential campaign	Political speeches	attack/support/no relation	Political stance	1462 pairs of arguments	[d]
Skeppstedt et al. [37]	Brexit Blogs (English)	Blogs	Certainty, Uncertainty, Hypotheticality, Prediction, Recommendation, Concession/Contrast and Source	Stance modifiers	2095 sentences	NA
Johnson and Goldwasser [36]	Twitter (politicians tweets) (English)	Twitter	Stance (For/Against), party agreement	Politicians' stance and their party agreement	99,161 tweets	NA
Lai et al. [38]	ConRef-Stance-ITA (Italian)	Twitter	For/ Against/None	Political stance	963 triplets (one tweet, one re-tweet, one reply)	[e]
Mohammad et al. [71]	SemEval-2016 Stance (English)	Twitter	For/Against/Neither	Political stance and sentiment	4163 tweet–target pairs	[f]
Lehmann and Derczynski [50]	Political quotes from politicians (Danish)	News articles	For/Against/Neutral	Political stance	898 quotes with stance	[g]
Bhavan et al. [51]	HanDeSeT	Political speeches	Speaker, party, policies	Political stance	1251 samples (607 politicians and their speeches with motions)	[h]
Salah [35]	UK House of Commons parliamentary debate	Political debates	Speakers, Aye/No	Political stance	1251 debates, 341,404 speeches	NA
Sawhney et al. [53]	ParlVote	Debate transcripts	Speaker name, party, Aye/No	Political stance	33,461 debates	NA
Lai et al. [43]	E-USA, R-CAT, E-FRA, R-ITA	Twitter	5400 Catalan, 5400 Spanish, 2000 French, 1000 Italian tweets	Political stance (Multilingual)	Favor/Against/None	[i]
Vamvas and Sennrich [52]	X-stance dataset	Political election comments	150 Questions, 65,000 answers, comments	Political Stance (multilingual)	Favor/Against	[j]
Bar-Haim et al. [39]	Claim polarity (English)	Debates	Pro/Con, claims, sentiment	Claim stance classification	55 topics with 2394 claims	NA

[a] https://data.mendeley.com/datasets/j83yzp7ynz/1.
[b] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/0TJX8Y.
[c] www.cs.cornell.edu/home/llee/data/convote.html.
[d] https://dh.fbk.eu/2017/11/political-argumentation/.
[e] https://github.com/mirkolai/Stance-Evolution-and-Twitter-Interactions/tree/v1.0.0.
[f] https://www.saifmohammad.com/WebPages/StanceDataset.htm.
[g] https://github.com/rasleh/Political-Stance-in-Danish.
[h] https://data.mendeley.com/datasets/j83yzp7ynz/1.
[i] https://github.com/mirkolai/MultilingualStanceDetection.
[j] https://zenodo.org/record/3831317#.YZQOUy8w1pQ.

In the work by Lai et al. [41], authors introduced a new dataset called *CONREF-STANCE-ITA* for stance detection using Twitter posts in Italian. Annotation classes include: *Favor, Against, None*. The authors focused on annotating stances at user level by creating a triplet of tweets: one tweet, one re-tweet and one reply of the same user within a temporal phrase. Two domain experts were asked to annotate those triplets. If there was any disagreement between them for any specific triplets, crowd-sourcing platform CrowdFlower was hired to handle the annotations of those tweets. There were some requirements for annotators from CrowdFlower that they must be Italian and living Italy and pass the evaluation test before given the task. Therefore, it will require from two to five annotators to complete the task. Lai et al. also followed the inter-annotation agreement (IAA) by Mohammad et al. [71] to determine the final label for a triplet. The final dataset contains 963 triplets.

Bar-Haim et al. [39] added Pro/Con annotation to the IBM dataset [73]. The dataset contains 55 topics from International Debate Education Association[8] with 2394 claims taken manually from Wikipedia articles. The annotations were done by five annotators where majority annotation was selected for final label.

Due to the lack of Danish dataset, Lehmann and Derczynski [50] introduced a new stance dataset containing Danish politicians' quotes from news articles. They created a poll online with a shortlist of

[8] http://idebate.org.

Table 7
Summary of political datasets (cont.).

Author	Dataset	Domain	Annotations	Target	Size	Link
Tsur et al. [44]	US. Congress statements (English)	Political Congress statements	Topics/Frames (Health care, economy/budget, corruption, etc.)	Political framing	134,000 statements	[a]
Johnson et al. [48]	Congressional Tweets Dataset (English)	Twitter	17 frames	Political framing	2050 framing labeled	NA
Naderi and Hirst [45]	ComArg, Parliamentary discourse	Users' statements, parliament debates	pro/con,various frames	Political framing	ComArg (198 statements & 7 frames), Parliamentary discourse (121 sentences, 366 paragraphs, 7 frames)	[b]
Baumer et al. [46]	Framing Annotation Data for News Articles (English)	News articles	Annotated words for framing by their positions	Language of framing	74 articles (59,948 annotated words)	[c]
Card et al. [72]	Framing Annotations Across Issues	News articles	frame categories	frame categories	20,037	[d]
Paul and Girju [58]	ACL Anthology, Israeli–Palestinian conflict	Scientific abstract, Online articles	Palestinian/Israeli	Topics and aspects identification	594 P/I editorials	[e]
Trabelsi and Zaiane [57]	ObamaCare (OC), Assault Weapons (AW), Gay Marriage (GM) 1& 2, Israel–Palestine conflict	Survey, online articles	Various viewpoints	Grouping arguing expressions to viewpoints	OC: 942 docs, AW: 394 docs, GM1: 98 docs, GM2: 450 docs	NA

[a]https://votesmart.org.
[b]https://takelab.fer.hr/data/comarg/.
[c]https://ecommons.cornell.edu/handle/1813/39216.
[d]https://github.com/dallascard/media_frames_corpus.
[e]http://www.aclweb.org/anthology-new/;http://www.bitterlemons.org.

possible topics which would be included into the dataset. To avoid the over-representation of any specific politician, they chose all politicians from all political parties in the parliament to include into the dataset. Data labeling process followed the work from Mohammad et al. [74] with three classes: *for, against, neutral*. Since some quotes were difficult to classify on a specific category, the authors created subtopics (*national policy* and *centralization*) for dataset. If a quote belonged to two subtopics, they would create a duplication of it and assign each of them to each subtopic. Detailed annotations guidelines per annotated class were provided for annotators. The final dataset contains 898 quotes in total.

Lai et al. [43] provided two newly annotated datasets in French and Italian for stance detection. Since there already existed benchmark dataset for SemEval2016 in English [75] and IberEval2017 in Spanish and Catalan [76], the authors decided to extend these two into Italian and French. They collected tweets (no retweets were included) in these two languages using the same criteria as in the English, Spanish–Catalan dataset. For the French dataet (E-FRA), they collected tweets about 2017 presidential elections between Emmanuel Macron and Marine Le Pen using these keywords: *macron*, *#presidentielles2017*, *lepen*, and *le pen* and randomly selected 2000 of them. For the Italian dataset (R-ITA) Italian tweets about Referendum in Italy on December 42016 were collected. Debate topics with hashtag *referendumcostituzionale* were collected and 1000 of them were randomly sampled. The authors asked domain expert to annotate 100 gold standard tweets and hired Crowdflower workers who lived in France and Italy to annotate data for French and Italian dataset respectively.

In the work by Vamvas and Sennrich [52], the authors introduced a large dataset for stance detection in multi-languages (English, Swiss German, French, and Italian) and multitude of topics and targets. They extracted responses in candidates elections in Switzerland from the voting advice application *Smartvote*. Dataset contains 150 questions and 65 000 answers given by the candidates running political office in Switzerland. Every question was interpreted as the NLP target and the comments as input to be classified.

Abercrombie and Batista-Navarro [69] take transcriptions from the UK House of Commons and augment them with both labels from a parliamentary monitoring organizations and own labels. Subsequently, they apply a Support Vector Machine classifier and find that the texts contribute very little to the models. Instead, they report that the title and to a lesser degree the meta-information constitute the building blocks for identifying the motions' topics. The paper also shows that we can apply opinion mining to estimate speaker's opinion towards the motion. The goal setting is relatively close to PVI. The texts suffer from a very procedural language and crowd-sourced labels reveal some issues.

Baumer et al. [46] created a new dataset using political news data for detecting framing language. After collecting news articles from 15 online news websites, they hired annotators from Mechanical Turk (MTurk) to highlight framing words or phrases from the articles by their positions in the text. To ensure quality of the work, the authors applied bonus scheme by giving more money for annotators whenever their annotations are matched with at least two different workers and deducted money when there was no match. There were also other rules that annotators must follow and collaboration among annotators was not allowed. The resulted dataset has 74 articles including 507 annotations of 59 948 annotated words.

In the work of [48], two graduate students were asked to annotate 17 frames for 2050 tweets from US House and Senate representative using the Policy Frames Codebook by [77]. However, only 14 over 15 frames from the codebook were used and authors also proposed three new frames for the task. Since it was difficult to assign only one frame to some tweets, annotators were allowed to give one or more frames to a tweet with the first frame was the most relevant one. After one month working on the task, they gathered together to decide one main frame for tweets with two or more annotated frames.

Naderi and Hirst [45] built a corpus based on the debates on same-sex marriage in Canadian Parliament both in 1999 and 2005. They selected two set of data: sentences and paragraphs. For the first set, they asked three annotators to annotate data with stance (pro/con/no) and pre-existing frames from ComArg corpus. For the second set, two annotators were asked to annotate the speeches with frames taken from ComArg without stances. They adopted the Weighted Kappa metric to check for annotation reliability. The final corpus consists of 121 sentences and 366 paragraphs.

4.1.3. Political viewpoint extraction datasets

For political viewpoint extraction, the most well-known dataset is the Palestinians and Israelis dataset — which was first used in the work of Lin et al. [60]. The authors collected a set of 594 articles between 2001–2005 from the *bitterlemons* website focusing on the conflict between Israeli and Palestinian. These articles were annotated by the side as the authors (Palestinian or Israeli).

The dataset introduced by Menini and Tonelli [61] focused on viewpoints comparison. They identify whether two speakers agree or disagree on a specific topic. To build dataset, the authors collected discourses and official declarations data given by Nixon and Kennedy during the 1960 presidential campaign from The American Presidency Project.[9] The authors define 38 topics and extract sentences containing at least one keyword in each topic and form the *text excerpt* by combining also the before and after sentence. To create *snippets*, five pairs of *excerpts* from Nixon and Kennedy are randomly paired. The authors ask two trusted annotators to manually annotate these *snippets* with agreement/disagreement relation, sentiment, and similarity of the solution proposed with respect of the debated topic. This has resulted into a corpus of 350 *snippets* for all topics.

In addition, there are also other datasets containing stances as labels and were previously used in the work of Trabelsi and Zaiane [57] namely the ObamaCare (for/against), Assault Weapons (allow/not), or Gay Marriage (illegal/not or hurt/no).

For the task of PVI, we suggest that annotations should focus on political ideologies/viewpoints not only at a high level – for identifying the overall ideology of the text, but also at a low level – for explaining purposes which can be used to explain what contributes to overall ideologies/viewpoints. We annotate at the paragraph/sub-paragraph level because (i) the document level is too general for expressing viewpoints, especially when the written text might contain various viewpoints, (ii) the sentence level is too specific and lacking context; hence, leaving the paragraph/sub-paragraph level the most comprehensive way to capture viewpoints.

Having annotations about the stance and framing (for understanding speakers' position/language use), or speaker/publisher/author name (for political network link) might also give extra information (optional). Moreover, to be able to annotate political text, annotators must have enough knowledge about this domain; Crowdflower and MTurk are two popular places to look for suitable workers. Even though this is not a simple process and the task is quite new, the guidelines from Iyyer et al. [20] can be used as a good guide for this.

4.2. Dataset

Despite the domination use of Twitter-based datasets [9–11,14,16,17,21–23,26,36,38,48,71], other political-related data such as political news articles [24,32,46,50,67], speeches, statements and debates [5,11,13,20,29,34,44,70] or blogs [37] also serve as a valuable source for political research. Although politics does not get much attention comparing to other research domains, there is also a wide range of data in different languages (French [26], Danish [50,78], Italian [38,41], Spanish [16], German [13,78], Norwegian [33,68], Swedish [78], etc.).

Details of used datasets are listed clearly in Tables 5–7 with information about owners, data sources, annotations used, targets, and the size of data. Most of them are available online for download.

Currently, there is no standard in the available datasets used in related work and most of them are task-oriented. For instance, a dataset annotated for political leaning detection task [14] is limited to this task. Some of them – for example [14,20] – are partly compatible for the first task of PVI (the classification) but not the second component (the explanation) as it was not annotated for political viewpoints detection.

To create dataset for the task, most authors used data from Twitter as the staring point. Due to the fact that tweets are short in length and noisy, they are unsuited for PVI. To choose data for PVI, our focus is on longer text with less noise – such as political news articles or political speeches – which is more helpful for us to target political viewpoints at different levels such as paragraph level or sub-paragraph level. Unfortunately, the available datasets are not annotated this way. For instance, stance datasets are designed for identifying stance of speakers — a high level of viewpoint and framing dataset focuses on annotating data based on a list of frames which does not suit for any components of PVI.

Another challenge when annotating dataset for PVI is that the annotation process is time-consuming and very expensive. The task requires annotators with good domain knowledge. Even though many authors chose to use crowd sourcing, this is not the ideal way to build the dataset because crowd souring workers are sometime unreliable in terms of quality. Thus, raising the need to have a proper annotated dataset for PVI.

Due to the lack of datasets for PVI, users are expected to include additional annotations in current datasets or create new ones following the guidelines above. People can also begin with the Israeli–Palestinian conflict dataset which was commonly used in various work in viewpoint extraction [57–60]. Even though this dataset only focuses on annotation at document level, it still serves as a good starting point.

4.3. Evaluation metrics

A majority of related work models their problem with political texts as classification task. Consequently, they adopt classification evaluation measures. For instance, twenty-nine of the presented works use either *precision*, *recall*, F_1, or *accuracy* for their evaluation [5,8,10,11,13,15,16,20–23,29,34,36–39,41–43,45–48,50–52,54,61]. In addition, some related work uses *mean absolute error* (MAE) [18] or *area under the curve* (AUC) [9,21]. These evaluation criteria reflect the performance of system well in conditions where labels from a gold standard are available. They fail to work with scenarios without these annotations such as unsupervised learning.

Sometimes, due to the special nature of the task, results were validated manually by human annotators [26]. Manual evaluation is task-specific — which is designed based on the evaluation targets, purposes and varied from task to task; hence, there is no standardization and it is very time-consuming. But it is suitable to fulfill complex evaluation that the automatic evaluation cannot achieve, even when there are no gold targets.

As evaluation for PVI is quite limited, none of the current measures was designed to fully satisfy the task. Practically, PVI consists of two components — which means each of them must be evaluated accordingly. Hence, raising the complexity for evaluation task. We require a criteria reflecting the performance in recognizing political viewpoints. Existing classification measures capture this aspect partially. We doubt that there is a one-to-one match between texts and political viewpoints. Thus, our ideal criteria would have to reflect the presence or absence of evidence for a multi-class problem. Imagine, for instance, that we have texts authored by a Libertarian-leaning member of the Republican party and a progressive member of the Democratic party. Both writers could agree that legalizing certain drugs would be an idea they support. Still, the viewpoints of both writers would be different. Thus, a classification evaluation with the labels "support" and "oppose" will fall short for the PVI task. Besides, we need to complement the performance criteria with a second criteria showing the explainability. The explainability criteria should reflect how well the model's predictions can be understood. Our literature review found very limited work on explaining the predictions for viewpoints in political texts. The research on political texts has yet to provide candidates for such metrics, protocols to measure them, and user studies to verify their utility.

[9] https://www.presidency.ucsb.edu.

5. Applications

In this section, we discuss some possible application ideas making PVI more useful in terms of balancing information and encouraging diversity in recommending news.

Political bias-aware tools: modern people are being surrounded by different kinds of information and reading news online has become a frequent habit for them. However, the affects of news on readers might vary depending on the domains. When it comes to political news, not many people can easily identify the political leaning behind a written article and most of the time the task might require help from experts with domain knowledge. Since political bias in news might affect one's political leaning, knowing clearly what they are going to read will minimize bias judgements towards a controversial discussed topic or issue. Therefore, it will be beneficial if readers have access to unbiased articles on news media and *political bias detector* is a promising tool. Moreover, having an *unbiased news recommender system* will increase the balance in recommending political news to readers, giving them a wide variety of articles with neutral views and to minimize the number of bias items.

Political surveillance tool: Another possible application for PVI is a tool that can keep certain media outlets balanced in terms on political views. It would be more interesting for readers to know whether a particular news media is leaning to the left or right, having unbiased views or not. Moreover, when recommending news, diversity is also an important factor because people should be confronted with different point of views rather being pointed to one particular direction for the news media benefits. A political surveillance tool acts as an important component that can prevent filter bubbles, echo chambers in news outlets, and avoid fake and manipulation news as well. As a result, giving readers a more democratic platform with more trustworthy information and free from hidden bias.

In the digital era, having more access to different types of information means having more knowledge, more power, and more chances to succeed. Unfortunately, it is not clear to what extent and how democratic it has to be for the public to have access to the kind of information they need; whether the published news on the social world unbiased enough for readers to have fair views about the subjects is still an open question. Having limited access to information means blocking our views on things, developing bias understanding and making ways for fake news to spread out, intrude and damage our lives. This has raised concerns about the danger of biased news in various domains, especially the political domain as they can do as much damage as fake news. Furthermore, freedom of information is also crucial in building good knowledge base, wise judgemental minds for people.

The improvements of modern techniques in NLP have opened many doors for researchers in various domains including politics. Unfortunately, misconducting research might lead people to a wrong direction as in the case of Cambridge Analytica data scandal.[10] The company was known to take advantages from people's political profiles building from illegal-obtained data and misused them in different political activities including the 2016 US presidential campaigns to affect people's voting behaviors.

Political viewpoint identifying can become a very powerful technique with high capabilities in accomplishing things, but what happened with Cambridge Analytica has also rang a bell to all of us about the possible damages caused by the misuse of technologies. At the same time, pointing out that doing ethical evaluations before conducting any research work is very important. If PVI is applied in a right way, it can bring out a greater good, especially for the media sector — such as increasing the level of transparency, higher diversity and more balance in political news articles.

[10] https://en.wikipedia.org/wiki/Cambridge_Analytica.

6. Conclusion and future work

This survey has reviewed related work in the domain of political text mining. The domain of political text mining includes problem such as classifying ideologies, leaning, or party affiliation, identifying framing, stance, and viewpoints. We have focused on techniques, datasets, evaluation criteria, and possible applications. As it stands, the current state of the art falls somewhat short of our goal with automatic political viewpoint identification in political texts. A majority of related work concentrates on the English language and elections in the United States of America or the United Kingdom of Britain. We doubt that reported results translate into other languages, cultures, and political systems. Technically, related work contributes a set of useful resources such as methods, datasets, and contextual insights. Still, in particular the aspect of explainability has not been explored. We hypothesize that understandable explanations will be crucial to gain the trust of users and ultimately become widely accepted and used.

Future research can address the limitation of the state of the art. Using *multi-lingual* data sets could help to verify how well proposed methods generalize to other cultures and political systems. Related work has started looking into other languages, especially with the use of Twitter — see examples for Spanish [16], French [26], or Italian [41]. Social media texts come with their challenges, such as neglecting conventional language standards, using emojis and other non-standard characters, and brevity due to technical restrictions. Ultimately, research should leverage both resources from citizens as well as politicians. Furthermore, political text mining is a prime subject to bring together social scientists and computer scientists for *cross-domain* collaboration. Computer scientist can lend their technical expertise to process the data and define new algorithms and models. Social scientists can contribute by helping annotate and defining concepts and goals. The explainability of PVI models represents a foremost candidate for such collaboration. Determining how explanations affect users requires knowledge from various disciplines including human computer interaction, psychology, and political science.

In the short-term, we propose a research agenda:

- Create datasets with sentence/paragraph level annotations concerning political viewpoints.
- Apply existing stance detection models to obtain predictions for political viewpoints.
- Conceptualize explanatory features.
- Implement a system capable of generating explanations for PVI predictions.
- Conduct user studies to obtain feedback to explanations.

Having completed the proposed tasks, the research community can evaluate the potential benefit of PVI systems as well as their best configuration with respect to methods, data, and optimization criteria. At this stage, an international committee should ponder the pros and cons of such technology. On the one hand, PVI technology promises to help educate the public and combat misinformation. On the other hand, the technology could become a tool for more advance propaganda. Suppose, the PVI system gains the trust of the broader public. Propagandists could mimic the user experience and launch systems to further their agenda. We currently observe these struggles in the domain of news personalization, where deliberate misinformation has become an apparent problem for societies to reconcile.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is done as part of the Trondheim Analytica project and funded under Digital Transformation program at Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway.

References

[1] Justin Grimmer, Brandon M. Stewart, Text as data: The promise and pitfalls of automatic content analysis methods for political texts, Polit. Anal. 21 (3) (2013) 267–297.

[2] Chang Li, Dan Goldwasser, Using social and linguistic information to adapt pretrained representations for political perspective identification, in: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, pp. 4569–4579.

[3] Dilek Küçük, Fazli Can, Stance detection: A survey, ACM Comput. Surv. 53 (2020) 1–37, http://dx.doi.org/10.1145/3369026.

[4] Han Jiawei, Pei Jian, Kamber Micheline, Data Mining: Concepts and Techniques, in: The Morgan Kaufmann Series in Data Management Systems, vol. 3, Morgan Kaufmann, ISBN: 9780123814791, 2011, URL http://search.ebscohost.com/login.aspx?direct=true&db=e000xww&AN=377411&site=ehost-live.

[5] Bei Yu, Stefan Kaufmann, Daniel Diermeier, Classifying party affiliation from political speech, J. Inf. Technol. Polit. 5 (2008) 33–48, http://dx.doi.org/10.1080/19331680802149608.

[6] 167 Congressional record S1534, 2021, Accessed on 13 November 2021.

[7] 163 Congressional record H9365, 2017, Accessed on 24 March 2022.

[8] Bjørn Høyland, Jean-François Godbout, Emanuele Lapponi, Erik Velldal, Predicting party affiliations from European parliament debates, in: Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science, 2014, pp. 56–60.

[9] Yupeng Gu, Ting Chen, Yizhou Sun, Bingyu Wang, Ideology detection for Twitter users via link analysis, in: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, ISBN: 978-3-319-60239-4, 2017, pp. 262–268, http://dx.doi.org/10.1007/978-3-319-60240-0_32.

[10] Daniel Preotiuc-Pietro, Ye Liu, Daniel Hopkins, Lyle Ungar, Beyond binary labels: Political ideology prediction of Twitter users, in: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017, pp. 729–740, http://dx.doi.org/10.18653/v1/P17-1068.

[11] Wei Chen, Xiao Zhang, Tengjiao Wang, Bishan Yang, Yi Li, Opinion-aware knowledge graph for political ideology detection, in: IJCAI, 2017, pp. 3647–3653.

[12] Yanchuan Sim, B.D.L. Acree, Justin Gross, N.A. Smith, Measuring ideological proportions in political speeches, EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference (2013) 91–101.

[13] Felix Bießmann, Automating political bias prediction, 2016, CoRR, abs/1608.02195. URL http://arxiv.org/abs/1608.02195.

[14] Felix Ming Fai Wong, Chee Wei Tan, Soumya Sen, Mung Chiang, Quantifying political leaning from tweets, retweets, and retweeters, IEEE Trans. Knowl. Data Eng. 28 (8) (2016) 2158–2172.

[15] Rory Duthie, Katarzyna Budzynska, Chris Reed, Mining ethos in political debate, in: Computational Models of Argument: Proceedings from the Sixth International Conference on Computational Models of Argument (COMMA), IOS Press, 2016, pp. 299–310.

[16] Ronaldo Prati, Elias Said Hung, Predicting the ideological orientation during the Spanish 24M elections in Twitter using machine learning, AI Soc. (2017) http://dx.doi.org/10.1007/s00146-017-0761-0.

[17] Sandeepa Kannangara, Mining Twitter for fine-grained political opinion polarity classification, ideology detection and sarcasm detection, in: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, in: WSDM '18, Association for Computing Machinery, New York, NY, USA, ISBN: 9781450355810, 2018, pp. 751–752, http://dx.doi.org/10.1145/3159652.3170461.

[18] Ramy Baly, Georgi Karadzhov, Abdelrhman Saleh, James Glass, Preslav Nakov, Multi-task ordinal regression for jointly predicting the trustworthiness and the leading political ideology of news media, in: Proceedings of NAACL-HLT, 2019, http://dx.doi.org/10.18653/v1/N19-1216.

[19] Alexandra Balahur, Zornitsa Kozareva, Andrés Montoyo, Determining the polarity and source of opinions expressed in political debates, in: International Conference on Intelligent Text Processing and Computational Linguistics, Springer, 2009, pp. 468–480.

[20] Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, Ps Resnik, Political ideology detection using recursive neural networks, in: 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference, Vol. 1, 2014, pp. 1113–1122, http://dx.doi.org/10.3115/v1/P14-1105.

[21] Zhiping Xiao, Weiping Song, Haoyan Xu, Zhicheng Ren, Yizhou Sun, TIMME: Twitter ideology-detection via multi-task multi-relational embedding, in: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020, pp. 2258–2268, http://dx.doi.org/10.1145/3394486.3403275.

[22] Xilian Li, Wei Chen, Tengjiao Wang, Weijing Huang, Target-specific convolutional bi-directional LSTM neural network for political ideology analysis, in: Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint Conference on Web and Big Data, Springer, 2017, pp. 64–72.

[23] Adithya Rao, Nemanja Spasojevic, Actionable and political text classification using word embeddings and LSTM, 2016, arXiv preprint arXiv:1607.02501.

[24] Rama Rohit Gangula, Suma Duggenpudi, Radhika Mamidi, Detecting political bias in news articles using headline attention, in: Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, 2019, pp. 77–84, http://dx.doi.org/10.18653/v1/W19-4809.

[25] Per E. Kummervold, Javier De la Rosa, Freddy Wetjen, Svein Arne Brygfjeld, Operationalizing a national digital library: The case for a Norwegian transformer model, in: Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa), 2021, pp. 20–29, URL https://aclanthology.org/2021.nodalida-main.3.

[26] Sarah Djemili, Julien Longhi, Claudia Marinica, Dimitris Kotzinos, Georges-Elia Sarfati, What does Twitter have to say about ideology? in: NLP 4 CMC: Natural Language Processing for Computer-Mediated Communication/Social Media-Pre-Conference Workshop at Konvens 2014, Vol. 1, Universitätsverlag Hildesheim, 2014.

[27] Jaime G. Carbonell Jr., POLITICS: Automated Ideological Reasoning*, Cogn. Sci. 2 (1) (1978) 27–51, http://dx.doi.org/10.1207/s15516709cog0201_3, URL https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog0201_3.

[28] Keith Poole, Howard Rosenthal, A spatial model for legislative roll call vote analysis, Am. J. Polit. Sci. 29 (1983) http://dx.doi.org/10.2307/2111172.

[29] Matt Thomas, Bo Pang, Lillian Lee, Get out the vote: Determining support or opposition from congressional floor-debate transcripts, 2006, arXiv Preprint cs/0607062.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[31] Robert Walecki, Ognjen Rudovic, Vladimir Pavlovic, Maja Pantic, Copula ordinal regression for joint estimation of facial action unit intensity, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4902–4910, http://dx.doi.org/10.1109/CVPR.2016.530.

[32] Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, Preslav Nakov, Predicting factuality of reporting and bias of news media sources, 2018, http://dx.doi.org/10.18653/v1/D18-1389, arXiv preprint arXiv:1810.01765.

[33] Emanuele Lapponi, Martin G. Søyland, Erik Velldal, Stephan Oepen, The talk of Norway: a richly annotated corpus of the Norwegian parliament, 1998–2016, Lang. Resour. Eval. 52 (3) (2018) 873–893.

[34] Marilyn Walker, Pranav Anand, Rob Abbott, Jean Fox Tree, Craig Martell, Joseph Eichenbaum, That is your evidence?: Classifying stance in online political debate, Decis. Support Syst. 53 (2012) 719–729, http://dx.doi.org/10.1016/j.dss.2012.05.032.

[35] Zaher Salah, Machine Learning and Sentiment Analysis Approaches for the Analysis of Parliamentary Debates (Ph.D. thesis), University of Liverpool, 2014.

[36] Kristen Johnson, Dan Goldwasser, Identifying stance by analyzing political discourse on Twitter, in: Proceedings of the First Workshop on NLP and Computational Social Science, 2016, pp. 66–75, http://dx.doi.org/10.18653/v1/W16-5609.

[37] Maria Skeppstedt, Vasiliki Simaki, Carita Paradis, Andreas Kerren, Detection of stance and sentiment modifiers in political blogs, in: International Conference on Speech and Computer, ISBN: 978-3-319-66428-6, 2017, http://dx.doi.org/10.1007/978-3-319-66429-3_29.

[38] Mirko Lai, Delia Hernandez Farias, Viviana Patti, Paolo Rosso, Friends and enemies of clinton and trump: Using context for detecting stance in political tweets, in: Mexican International Conference on Artificial Intelligence, ISBN: 978-3-319-62433-4, 2017, pp. 155–168, http://dx.doi.org/10.1007/978-3-319-62434-1_13.

[39] Roy Bar-Haim, Indrajit Bhattacharya, Francesco Dinuzzo, Amrita Saha, Noam Slonim, Stance classification of context-dependent claims, in: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, Association for Computational Linguistics, Valencia, Spain, 2017, pp. 251–261, URL https://aclanthology.org/E17-1024.

[40] David Vilares, Yulan He, Detecting perspectives in political debates, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 1573–1582.

[41] Mirko Lai, Viviana Patti, Giancarlo Ruffo, Paolo Rosso, Stance evolution and Twitter interactions in an Italian political debate, in: International Conference on Applications of Natural Language to Information Systems, 2018, http://dx.doi.org/10.1007/978-3-319-91947-8_2.

[42] Stefano Menini, Elena Cabrio, Sara Tonelli, Serena Villata, Never retreat, never retract: Argumentation analysis for political speeches, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, 2018.

[43] Mirko Lai, Alessandra Teresa Cignarella, Delia Irazú Hernández Farías, Cristina Bosco, Viviana Patti, Paolo Rosso, Multilingual stance detection in social media political debates, Comput. Speech Lang. (ISSN: 0885-2308) 63 (2020) 101075, http://dx.doi.org/10.1016/j.csl.2020.101075, URL https://www.sciencedirect.com/science/article/pii/S0885230820300085.

[44] Oren Tsur, Dan Calacci, David Lazer, A frame of mind: Using statistical models for detection of framing and agenda setting campaigns, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015, pp. 1629–1638, http://dx.doi.org/10.3115/v1/P15-1157.

[45] Nona Naderi, Graeme Hirst, Argumentation mining in parliamentary discourse, in: Principles and Practice of Multi-Agent Systems, Springer, 2015, pp. 16–25.

[46] Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, Geri Gay, Testing and comparing computational approaches for identifying the language of framing in political news, in: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015, pp. 1472–1482.

[47] Kristen Johnson, I.-Ta Lee, Dan Goldwasser, Ideological phrase indicators for classification of political discourse framing on Twitter, in: Proceedings of the Second Workshop on NLP and Computational Social Science, 2017, pp. 90–99, http://dx.doi.org/10.18653/v1/W17-2913.

[48] Kristen Marie Johnson, Di Jin, Dan Goldwasser, Modeling of political discourse framing on Twitter, in: ICWSM, 2017.

[49] Stefan Dahlberg, Magnus Sahlgren, Issue framing and language use in the Swedish blogosphere, in: A.R. Kaal, I. Maks, A.M.E. van Elfrinkhof (Eds.), DAPSAC, John Benjamins, ISBN: 9789027206466, 2014, http://dx.doi.org/10.1075/dapsac.55.

[50] Rasmus Lehmann, Leon Derczynski, Political stance in Danish, in: Proceedings of the 22nd Nordic Conference on Computational Linguistics, Linköping University Electronic Press, Turku, Finland, 2019, pp. 197–207, URL https://www.aclweb.org/anthology/W19-6121.

[51] Anjali Bhavan, Rohan Mishra, Pradyumna Prakhar Sinha, Ramit Sawhney, Rajiv Shah, Investigating political herd mentality: A community sentiment based approach, in: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, 2019, pp. 281–287.

[52] Jannis Vamvas, Rico Sennrich, X-Stance: A multilingual multi-target dataset for stance detection, 2020, CoRR abs/2003.08385. URL https://arxiv.org/abs/2003.08385.

[53] Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, Rajiv Shah, GPolS: A contextual graph-based language model for analyzing parliamentary debates and political cohesion, in: Proceedings of the 28th International Conference on Computational Linguistics, 2020, pp. 4847–4859.

[54] Momchil Hardalov, Arnav Arora, Preslav Nakov, Isabelle Augenstein, Cross-domain label-adaptive stance detection, 2021, ArXiv abs/2104.07467.

[55] Timo Schick, Hinrich Schütze, Exploiting cloze questions for few-shot text classification and natural language inference, Comput. Res. Repos. (2020) arXiv:2001.07676. URL http://arxiv.org/abs/2001.07676.

[56] Momchil Hardalov, Arnav Arora, Preslav Nakov, Isabelle Augenstein, Few-shot cross-lingual stance detection with sentiment-based pre-training, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 36, 2022.

[57] Amine Trabelsi, Osmar R. Zaiane, Mining contentious documents using an unsupervised topic model based approach, in: 2014 IEEE International Conference on Data Mining, 2014, pp. 550–559, http://dx.doi.org/10.1109/ICDM.2014.120.

[58] Michael Paul, Roxana Girju, A two-dimensional topic-aspect model for discovering multi-faceted topics, in: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, in: AAAI'10, AAAI Press, 2010, pp. 545–550.

[59] Thibaut Thonet, Guillaume Cabanac, Mohand Boughanem, Karen Pinel-Sauvagnat, VODUM: a topic model unifying viewpoint, topic and opinion discovery, in: 38th European Conference on Information Retrieval (ECIR 2016), Vol. 9626, Springer, Padua, IT, 2016, pp. 533–545, http://dx.doi.org/10.1007/978-3-319-30671-1_39, URL https://oatao.univ-toulouse.fr/16879/. Thanks to Springer editor. This papers appears in Volume 9626 of Lecture Notes in Computer Science ISSN : 0302-9743 The original PDF is available at: https://link.springer.com/chapter/10.1007/978-3-319-30671-1_39.

[60] Wei-Hao Lin, Theresa Wilson, Janyce Wiebe, Alexander Hauptmann, Which side are you on? Identifying perspectives at the document and sentence levels, in: Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X), Association for Computational Linguistics, New York City, 2006, pp. 109–116, URL https://aclanthology.org/W06-2915.

[61] Stefano Menini, Sara Tonelli, Agreement and disagreement: Comparison of points of view in the political domain, in: COLING 2016, the 26th International Conference on Computational Linguistics, 2016, pp. 2461–2470.

[62] Swapna Somasundaran, Janyce Wiebe, Recognizing stances in online debates, in: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2009, http://dx.doi.org/10.3115/1687878.1687912.

[63] Stephen H. Bach, Matthias Broecheler, Bert Huang, Lise Getoor, Hinge-loss Markov random fields and probabilistic soft logic, 2017, arXiv preprint arXiv:1505.04406, 18(1):3846–3912. ISSN 1532-4435.

[64] Dennis Chong, James Druckman, Framing theory, Annu. Rev. Political Sci. 10 (2007) http://dx.doi.org/10.1146/annurev.polisci.10.072805.103054.

[65] Robert Entman, Framing: Toward clarification of a fractured paradigm, J. Commun. 43 (1993) 51–58, http://dx.doi.org/10.1111/j.1460-2466.1993.tb01304.x.

[66] Kiri Wagstaff, Claire Cardie, Seth Rogers, Stefan Schrödl, Constrained K-means clustering with background knowledge, in: ICML, 2001.

[67] Lukas Gebhard, Felix Hamborg, The POLUSA dataset: 0.9M political news articles balanced by time and outlet popularity, in: Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020, in: JCDL '20, Association for Computing Machinery, New York, NY, USA, ISBN: 9781450375856, 2020, pp. 467–468, http://dx.doi.org/10.1145/3383583.3398567.

[68] Per Erik Solberg, Pablo Ortiz, The Norwegian parliamentary speech corpus, 2022, arXiv preprint arXiv:2201.10881.

[69] Gavin Abercrombie, Riza Theresa Batista-Navarro, Identifying opinion-topics and polarity of parliamentary debate motions, in: Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Association for Computational Linguistics, 2018.

[70] Alexander Baturo, Niheer Dasandi, Slava J. Mikhaylov, Understanding state preferences with text as data: Introducing the UN general debate corpus, Res. Polit. 4 (2) (2017) 2053168017712821.

[71] Saif M. Mohammad, Parinaz Sobhani, Svetlana Kiritchenko, Stance and sentiment in tweets, ACM Trans. Internet Technol. (ISSN: 1533-5399) 17 (3) (2017) http://dx.doi.org/10.1145/3003433.

[72] Dallas Card, Amber Boydstun, Justin H. Gross, Philip Resnik, Noah A. Smith, The media frames corpus: Annotations of frames across issues, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2015, pp. 438–444.

[73] Ehud Aharoni, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, Noam Slonim, A benchmark dataset for automatic detection of claims and evidence in the context of controversial topics, in: Proceedings of the First Workshop on Argumentation Mining, Association for Computational Linguistics, Baltimore, Maryland, 2014, pp. 64–68, http://dx.doi.org/10.3115/v1/W14-2109, URL https://aclanthology.org/W14-2109.

[74] Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, Colin Cherry, A dataset for detecting stance in tweets, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), European Language Resources Association (ELRA), Portorož, Slovenia, 2016, pp. 3945–3952, URL https://www.aclweb.org/anthology/L16-1623.

[75] Mariona Taulé, M. Antonia Martí, Francisco M. Rangel, Paolo Rosso, Cristina Bosco, Viviana Patti, et al., Overview of the task on stance and gender detection in tweets on Catalan independence at IberEval 2017, in: 2nd Workshop on Evaluation of Human Language Technologies for Iberian Languages, IberEval 2017, Vol. 1881, CEUR-WS, 2017, pp. 157–177.

[76] Steven Bethard, Marine Carpuat, Daniel Cer, David Jurgens, Preslav Nakov, Torsten Zesch, Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). 2016.

[77] Amber E. Boydstun, Dallas Card, Justin Gross, Paul Resnick, Noah A. Smith, Tracking the Development of Media Frames within and across Policy Issues, Technical report, Carnegie Mellon University, 2018, http://dx.doi.org/10.1184/R1/6473780.v1.

[78] Christian Rauh, Jan Schwalbach, The ParlSpeech V2 Data Set: Full-Text Corpora of 6.3 Million Parliamentary Speeches in the Key Legislative Chambers of Nine Representative Democracies, Harvard Dataverse, 2020, http://dx.doi.org/10.7910/DVN/L4OAKN.