Identifying Framing Bias in Online News

FRED MORSTATTER, USC Information Sciences Institute, USA
LIANG WU, Arizona State University, USA
URAZ YAVANOGLU, Gazi University, Turkey
STEPHEN R. CORMAN and HUAN LIU, Arizona State University, USA

It has been observed that different media outlets exert bias in the way they report the news, which seamlessly influences the way that readers' knowledge is built through filtering what we read. Therefore, understanding bias in news media is fundamental for obtaining a holistic view of a news story. Traditional work has focused on biases in terms of "agenda setting," where more attention is allocated to stories that fit their biased narrative. The corresponding method is straightforward, since the bias can be detected through counting the occurrences of different stories/themes within the documents. However, these methods are not applicable to biases which are implicit in wording, namely, "framing" bias. According to framing theory, biased communicators will select and emphasize certain facts and interpretations over others when telling their story. By focusing on facts and interpretations that conform to their bias, they can tell the story in a way that suits their narrative. Automatic detection of framing bias is challenging since nuances in the wording can change the interpretation of the story. In this work, we aim to investigate how the subtle pattern hidden in language use of a news agency can be discovered and further leveraged to detect frames. In particular, we aim to identify the type and polarity of frame in a sentence. Extensive experiments are conducted on real-world data from different countries. A case study is further provided to reveal possible applications of the proposed method.

CCS Concepts: • **Human-centered computing** → **Collaborative and social computing**; • **Information systems** → *Web mining*;

Additional Key Words and Phrases: Bias, framing, machine learning, natural language processing

1 INTRODUCTION

There are many levels at which text can be biased. One of the ways in which text can be biased is at the "framing" level. This is a more subtle level of bias that is introduced when specific aspects of a story are reinforced or emphasized over others. This can take many forms, from emphasis of

information to the selective presentation of information within a text. Since this level of bias is more subtle, it can be difficult for readers to detect it in a story. Due to their subtlety, they can be used to manipulate the opinion of the reader [6, 51]. Because of their ability to sway the opinions of their readers, it is important that framing information is presented to readers so that they can make an informed decision about the information they consume as well as the opinions they form based upon that information. Due to the large volume of information published online, the burden falls to automated approaches to identify these frames within text. In this work, we address the question of how to identify frames automatically within text. This is done in order to better inform both users and recommendation systems about the underlying bias in news documents.

News can be biased at different levels [6, 51], of which at least two formal levels have been established. At the first level, news agencies can choose to cover only stories which reinforce their views. It is intuitive that a news outlet cannot cover everything due to time or space restrictions; however, this type of bias goes beyond mere space restrictions. It is a systematic bias in the way that news is covered. It has been established that news agencies often select a subset of stories that pertain to a narrative [39]. This type of media bias is called "agenda setting" [17], and it is relatively easily detected by measuring the attention given to different stories [33, 39], or counting which sources are referenced in their published texts [18, 56]. The second, more subtle level of bias is when news agencies reinforce specific pieces of information. That is, without consideration to the amount of space dedicated to the story itself, news agencies can bias their stories by selecting different pieces of information from within the same story and focusing on them [22, 26]. This is a case of the observation made by Luntz that "it is not what you say, but how you say it," [51]. News agencies are framing their articles by selectively introducing information into articles that supports a predetermined hypothesis. This is usually done through the wording of the article, where journalists pay special attention to making certain pieces of evidence and arguments seem salient. In doing this, the news producers can affect the way the story is consumed [55]. This type of bias is called "second-order agenda setting," or more commonly, "framing." This type of bias is much more difficult to detect as the facts that are mentioned are often the same, but the way in which they are presented differs. The difficulty comes from going beyond mere counts of news occurrences to understanding the underlying bias from the text of the news stories. Being able to identify framed text is important as it can help us to better understand the bias underlying a news source from a computational perspective. This understanding can then be applied to recommender systems or to help users to understand the underlying bias in their data.

In this work, we take a machine learning approach to identifying frames in text. The underlying approach is guided by the intuition that frames are attempts to sway the reader by exposing them to the same idea repeatedly. This repetition should introduce redundancy into the data which can be learned by a classifier. We test this hypothesis at three levels of granularity. First, we test the ability of a machine learning binary classifier to identify a sentence as "framed" or "not framed." We then move on to test the ability of the classifier to identify certain *types* of frames, and finally their polarity. Finally, we see how well we can predict frames as a function of time and show how this information can be applied to other applications of text analysis.

The main contributions of this article are as follows:

(1) We investigate how framed text can be detected at different levels of granularity. We first start with a binary classification task ("is the sentence framed or not?") and continue to identify the frames, and finally their polarity.
(2) We show that the frame detection task is resilient to time. We research if we can learn a classifier that can identify framed sentences in future articles, making it appropriate for use in real-world scenarios.

(3) We present a case study detailing how the implications of framing from communications theory can be used to improve the detection of other types of opinion bias.

2 RELATED WORK

We organize the related work into two distinct parts. First, we discuss the psychological implications of framing. This is important as it shows both the importance of framing and guides how they are implemented in the real world. Next, we discuss attempts to automatically identify media bias in text. Finally, we conclude the related work by discussing how our work builds on both aspects in order to automatically identify frames in text data.

2.1 Psychological Effects of Framing

Ultimately, framing is an attempt to persuade the opinion of the reader. It is a subtle persuasion carried out covertly. Entman defines framing as "select[ing] some aspects of a perceived reality and make them more salient in a communicating text" [22]. Other work has refined this definition to better fit text analysis [24]. The way text is framed can greatly affect the way that readers adopt that idea. The way people understand certain concepts depends on the way those concepts are framed. Tversky and Kahneman found that by presenting two choices in different ways, readers would largely prefer one choice over the other depending on the way they framed the choices, even though the choices were exactly the same [55]. They were able to convince their participants that the different choices were ideal by focusing on different facts of the same policy, thus making them more salient. This is the underlying principle behind framing effects. Framing effects have been observed in news media [22], where news agencies frame stories to match their opinion by selecting particular facts, concepts, and quotes that match their individual bias. This is done in an effort to convince the reader that the choice or policy they are promoting is favorable [51]. Framing detection [19] is to process contextual terms across large volumes of text.

Given that we know that frames can have real effects on the readers such as directly affecting the policy they identify as best [55], it is important that we are able to identify them in order to better understand the motivations of the underlying news agencies. The way many have done this is through content analysis [36]. Next, we will discuss the body of work that seeks to identify media bias in text, and discuss how we can apply the principles from these studies to our own.

2.2 Detecting Media Bias

Herein we discuss work that aims to identify media bias in text. We separate the approaches into two sets: those that use text alone for classification, and those that use external knowledge and resources in order to identify frames.

2.2.1 Media Bias with External Knowledge. Many of the approaches in this category come from social sciences literature. In general, media bias is a well-studied problem in communications and social sciences literature. [41] investigate online biased content by looking at the reader-base of particular articles. Similarly, in the economics literature, [29] assessed media bias by measuring which media outlets cited which "think tanks" in their text. [20] studied how media bias can affect the opinions of voters, similar to the seminal work by [39]. Researchers use sentiment analysis for detecting not only biased news but also comments made by news readers [45]. The problem inherent to sentiment analysis is that labeled data is not readily available to train classifiers [13] for many tasks. Individuals are affected by different kinds of biased content [25]. Using quotes within the text, researchers have also mapped the bias of different news agencies [18]; however, this is done with respect to a "liberal" or "conservative" paradigm.

Specific to framing, some work has been done in this area. [11, 15] proposed a coding scheme for content analysis to identify issue-specific frames. This is aimed at building a corpus across issues. This corpus has been applied to study how frames develop across time [10]. The labeling strategy that we use is different because it is focused on building a corpus of frames within a single specific issue, which is different than previous work.

2.3 Media Bias in Computer Science Literature

Media bias has been studied previously in computer science. [44] discussed news production process and aspect level classification to avoid it. [50] proposed an unsupervised method to characterize as gatekeeping, coverage, and statement biases. The results of framing were used by [1] to find events in political disputes. By mapping the frequency with which framed texts occur over time, the authors were able to find events in their data. While the authors of this work use frames in their analysis, we go beyond this by predicting the presence of frames in the text. One way to accomplish this goal is to understand how individuals identify framing and propose techniques for identifying frame-invoking language [7]. Language analysis related to identifying common linguistic cues over framing bias was performed by [48].

Outside of framing, [42] looks at the quotations that different news agencies choose when reporting on a story, a type of agenda setting. [28] used the results of framing from the perspective of how the speaker constructs his message, looking for specific constructions in the text. The authors show that the way the writers present their ideas can be used to improve the task of sentiment analysis. Furthermore, computationally discovering an actor's "stand," or "opinion" has been studied. For example, using the text of political speeches, researchers have mapped the position of different political candidates [52]. Public sources contain controversial texts. [3] used co-following relationships inferred by Twitter to find ideological maps of news sources. Twitter-like media sources [2] also increase political diversity of news. Others [40, 47] proposed semi-supervised and unsupervised sentiment analysis approaches to detect polarity of words. [14] discussed semantic inferences require opinion mining instead of standard natural language processing (NLP) techniques. [54] built a model to understand language dynamics in politics using communication theory and probabilistic topic models. [4] study the extent to which diverse opinions spread throughout the Facebook network.

Computational approaches have been applied to framing. [7] found that "entailment," "implicature," and "subjectivity" help in identifying framing bias. Similarly, [28] were able to show a connection between lexical semantics and readers' perception. This is important as they proposed a computational approach to identify this relationship based upon sentiment.

The approach we take in this work is different from the related work. First, we discuss our efforts to curate a labeled dataset on framing. This is different from previous work as it focuses on the many different frames which can be used within one topic instead of coding them across many topics [11]. Additionally, we are using a different feature space than previous work [7]. This guides the hypotheses we can make from the data. Furthermore, we do not stop at identifying the frame *types* [48], but we go beyond this to also identify their polarity.

3 FRAMING IN NEWS MEDIA

In order to study framing, we first develop a dataset with known examples of framing. The dataset used consists of news articles pertaining to the topic of the construction of a ballistic missile defense system in Europe. Hereafter, we refer to this topic as "BMD." We focus on this topic for two reasons: (1) this is a sensitive topic that will elicit framed text from the different countries participating in the discussion, and (2) by focusing on a specific topic we can alleviate any indications of first-order agenda setting bias in the data. The dataset consists of 823 news articles and 31,121

Table 1. Statistics of Data Used in this Work

Country	Language	Articles	Sentences	Gov.	Frame Type (+/−)									
					CS	DS	DB	GT	PT	PE	RP	RR	ST	TR
Czech Republic	English	82	3,164	52.91%	85/4	3/0	16/3	65/0	47/22	176/52	83/19	75/4	61/6	42/24
France	English	121	4,207	57.62%	48/11	39/17	3/1	52/0	79/16	104/45	60/6	96/2	95/24	60/39
Germany	English	100	2,846	21.29%	31/12	3/2	8/1	18/2	95/17	100/66	89/6	156/3	76/13	74/48
Poland	Polish	97	6,085	65.88%	32/8	13/1	44/16	11/1	22/2	139/54	40/15	70/4	31/4	29/13
Russia	Russian	68	3,313	46.73%	25/5	2/5	4/1	21/6	32/2	111/22	97/12	113/8	31/17	91/40
Spain	English	152	3,284	54.72%	51/13	6/1	13/3	25/4	68/9	116/18	60/4	81/3	46/11	61/22
United Kingdom	English	65	2,715	53.55%	43/18	9/1	3/1	35/7	80/15	100/49	46/8	100/10	62/11	28/31
United States	English	138	5,507	52.79%	121/18	10/0	25/2	79/2	120/33	364/58	130/3	114/12	178/17	70/77

Since we will focus on sentence-level classification for frames, we report counts at the sentence level for the "frame type" column group. The acronyms for each "frame type" corresponds to the acronyms introduced in Section 3. The "Gov" column indicates the percentage of sentences generated by government outlets.

sentences crawled from the internet, aggregated by country. Statistics of the dataset are shown in Table 1. The acronyms introduced in the "Frame Type" columns correspond to those introduced in the subsequent subsection. "Articles" refers to the total number of news articles originating within that country, "Sentences" refers to the number of sentences, and "Gov." refers to the fraction of articles which are written by a government entity. The mechanism for determining government texts will be described in a subsequent section.

Four communications scholars were hired as coders to hand code each sentence in each of the articles. Native Polish and Russian speakers were hired to code the Polish and Russian text. In the case of Czech, French, German, and Spanish, the articles were machine translated into English because native speakers could not be found to code for these languages. While there is a possibility that this translation approach will introduce some inaccuracy into the system, we are confident that the high translation accuracy within Western languages [5] will not introduce a major limitation to this work. Coders were trained on randomly sampled texts from the corpus. Training involved iterations of coding documents, calculating reliability, discussing disagreements, and adjusting category definitions and rules until acceptable reliability was achieved. Texts were treated as the unit of observation and a coder's results were represented as a vector of frame counts for each text coded. Reliability was assessed by computing Krippendorff alpha (interval) on all coders' vectors for a given text. In the final training round, reliabilities for 10 texts were in the range $0.725 \leq \alpha \leq 0.967$ (mean $\alpha = 0.873$). In production, two coders coded each text, and where necessary discussed disagreements to arrive at a final code for a disputed sentence. The frames themselves were identified *a priori* by the same experts using a held-out sample of documents. The frames are as follows:

(1) **General Threat (GT):** BMD is either a response to the threat of a nuclear attack or to any other unspecified threat. This can include any non-actor, such as a geographic area such as "the middle east."

(2) **Specific Threat (ST):** BMD is a necessary response to an external threat. This can be from a named agent or actor. This can be specific mentions of named entities such as people or states, as well as terms like "terrorist groups."

(3) **Collective Security (CS):** BMD is a goal of NATO member states and proof of a cooperative effort with the rest of NATO or other allies.

(4) **Deterrence System (DS):** BMD is a useful addition to ongoing deterrent/security and nuclear deterrent efforts, but it does not replace those efforts.

(5) **Domestic Benefit (DB):** BMD provides some external benefit to the actor other than the obvious benefit of increased security. Examples of these benefits include greater scientific resources, increased/upgraded technology, and closer partnerships, among others.

(6) **Progress/Effectiveness (PE):** This frames BMD with a focus on its development. It also highlights progress or success in the program. This specifically includes successful missile tests, missile installations, and summits. It constructs BMD as a successful program, highlighting cost effectiveness of the system.

(7) **Political Tensions (PT):** This frame focuses on the political nature of BMD. It highlights any contentions and complicated matters of countries' relationships with regard to BMD. This frame makes contrasting political opinions salient, and includes internal debate and tensions. The key to this frame is that the text tagged with it concerns only political relations. Security issues are not considered.

(8) **Threat to Russia (TR):** This frame focuses on the ways in which BMD poses a threat to the interests and/or security of Russia. Text that contains this frame is either from the Russian perspective, or about the Russian perspective.

(9) **Russian Roadblocks (RR):** This frame highlights Russia's opposition to BMD. It includes political threats and posturing made by the Russian side. This frame centers on Russia's response to BMD from a third-party perspective. It focuses on actions made by Russia to discourage, delay, or prevent the BMD system.

(10) **Partnership with Russia (PR):** This constructs the relationship between NATO and Russia as a "partnership." Language in this frame stresses Russia as a partner with NATO and highlights the position that NATO and Russia should collaborate to achieve shared goals.

In addition to these framing types, each sentence is also tagged with a polarity, meaning each frame is either positive or negative. This means in support of or opposing the frame. This is *not* a "gain/loss" frame such as those conceptualized in [55]. For example, if a sentence discusses any success in moving forward in the BMD program, then it will be labeled with "Progress/Effectiveness+." If a sentence discusses how the program has failed to succeed, then it will be tagged with "Progress/Effectiveness−." Example sentences with their labels (including polarity) are shown in Table 2.

Furthermore, beyond framing, we view these articles as coming from one of two "types" of outlets: government, or media. Government outlets are those news agencies that are officially sponsored by the government, and media outlets are those that are not. We do not differentiate between news outlets that are perceived as "pro-government," or "anti-government," which would be ultimately subjective. The categorization is done based upon the publicly acknowledged affiliation of the news agency.

4 IDENTIFYING FRAMES WITHIN TEXT

In this section, we investigate the ability to classify sentences based upon their frames. It is important to distinguish frames from no frames as this process can shed light on what types of linguistic patterns are useful for identifying sentences containing the loaded facts and pieces of evidence discussed in the Introduction. Framing is nuanced [51], and this can mean that a machine learning algorithm may have difficulty in identifying frames within sentences. Here we measure how well we can capture the way countries *frame* the articles they publish. Following the definition of framing in the Introduction, we attempt to find *sentences* that frame an argument. This is more useful as most of a document will be unframed, so it is more useful to identify specific sentences

Table 2. Example Sentences by Frame Type

Frame	Sentence
PE+	"We would get multiple, coordinated opportunities to stop an incoming missile, rather than individual nations going it alone."
PE−	"The anti-missile system remains unproved after tens of billions of dollars of development work."
TR+	"The proposed placement of about 10 interceptor missiles in Poland and a radar tracking station in the Czech Republic would pose a 'clear threat' to Russia."
TR−	"The missile system is limited and not meant to counter Russia, with its huge nuclear arsenal, but to protect against a 'rogue state.'"
No frame	"According to pool reports, Putin did not meet with Gates and Rice until 40 minutes after their arrival."

This is intended to give the reader a sense of the difference between the positive and negative polarity of frame labels. "No frame" is a sentence that was *not* identified by the coders as containing a frame.

which exert framing bias. In other words, we leverage the labeled data to build a classifier that can detect sentences that contain frames.

In this work, we focus on the sentence as the basic unit for classification. This is because the sentence is the basic unit of written text [30]. Sub-sentence units such as clauses may not provide enough information to make an accurate classification. Coarser views such as the paragraph level may include additional noisy sentences such as irrelevant sentences and sentences containing other frame types. The choice of sentence level also is useful for potential applications of this tool, such as the use of the tool in identifying biased text [48]. While these concerns led this work to focus on the sentence level, expanding the approaches proposed here to other granularities remains a point for future work.

4.1 Identifying the Presence of Frames

In order to find framed sentences, we build a classifier that uses only the language of an article in order to classify its sentences as "framed" or "not framed." This is to test the classification ability of a binary classifier at identifying framed sentences. Here we define a framed sentence as one of those that was identified during the coding process as a sentence containing one of the frames defined in Section 3.

In this experiment, we follow a standard supervised machine learning setup where instances are sentences, and features are those described in the subsequent paragraphs. The features consist of *n*-gram features, as well as structural features extracted from the sentence. Since the articles in this study consist of multiple languages, we are unable to extract all features for all countries as mature NLP tools do not exist for many non-English languages. The feature groups used in this study are as follows. Feature groups with the "(ML)" tag next to their name are multi-lingual feature groups that can be automatically extracted in any language:

(1) *Unigram (ML)*. The raw frequency counts of the *word-level* unigrams. This counts single word sequences. For example, in the sequence "missile defense program," "missile," "defense," and "program" are all unigrams.

(2) *Bigram (ML).* The raw frequency counts of the *word-level* bigrams. This counts two-word sequences. For example, in the sequence "missile defense program," "missile defense" and "defense program" are all bigrams.

(3) *Part-of-Speech Unigrams and Bigrams.* Using the part-of-speech tagger provided by Python's NLTK [8], we extract part-of-speech tags for each word in the corpus. Using these tags, we extract the word-level part-of-speech unigrams and bigrams.

(4) *Has Quote (ML).* A single binary feature that indicates whether the sentence contains a quote. This is a special "unigram" that arises as a side effect of how we tokenize the sentence. We hypothesize that news agencies may use framing to make their narrative seem more salient, and include this feature to see if it aids the classifier in finding framed sentences.

(5) *Parse Tree Height.* Using Python's NLTK, we extract the syntactic structure of the sentence, also called the parse tree [34]. We extract a single integer which indicates the height of the parse tree. A higher number indicates a taller parse tree, which means that the sentence is more complex. Since framing is nuanced, we hypothesize that more complicated sentences are likely to elicit framing.

(6) *Sentence Complexity.* A floating point number designed as another measure of sentence complexity. In the same vein as parse tree height, we try to approximate the complexity of the sentence. This feature is calculated as the number of *unique* part-of-speech tags divided by the number of words in the sentence.

(7) *Has Named Entity.* We include a single binary feature that indicates whether the sentence contains a named entity. The named entity extraction tool we use is provided by Python's NLTK [8]. Motivated by the definitions of frames such as "Specific Threat," we notice that the presence of named entities may be important for frame detection. Similar to the "Has Quote" feature, we hypothesize that framed sentences may be more likely to allude to specific people or other entities to enforce their slant.

The definition of framing given in [22] is that framing is to make aspects of a story more salient. These aspects consist of pieces of information (i.e., evidence), and arguments. Thus, we need to find the information that a news agency is trying to make more salient. To help our classifier detect this, we apply TF-IDF weighting to the features. This gives us a focus on not only the importance of the term to the document (TF), but also the term's importance with respect to the corpus (IDF). By taking both the term frequency and inverse document frequency into consideration, we may be able to find sentences that are focusing on aspects not emphasized by others.

Using these features and this normalization technique, we examine several classifiers from different backgrounds for their ability to separate framed from not framed sentences. Because of the difficulty of the problem, we do not know which type of classifier will perform the best. Thus, we compare the results of several classifiers that employ a wide range of strategies. We investigate the following classifiers.

Trivial Classifiers

C1 *Majority Classifier*: This classifier always predicts the majority class label based upon the distribution of the labels present in the training data.

Text-Based Classifiers

C2 *Topic Classifier*: We use a topic modeling algorithm to create a feature vector from the sentence. Topic modeling algorithms try to find the underlying "topics," or themes, in the text. Simultaneously, these models also measure how strongly each document is associated with each topic. Topic associations used as features have shown to be useful in some classification tasks. Using Latent Dirichlet Allocation [9], we treat sentences as

documents, and measure their association with each of the K latent topics discovered by the topic modeling algorithm. The features in this model are the probability of the sentence belonging to each of the K topics (yielding K total features). We then train a support vector machine (SVM) model using these features. We choose SVM as it gave the best performance in a pilot test we conducted with several other candidate models. The underlying assumption in this model is that framed sentences will draw from a different set of topics than non-framed sentences. Empirically, we set $K = 100$, as it gave the best performance in this task.

C3 *LSTM*: Long Short-Term Memory [31] is a model that takes into account the order in which inputs are given to the system. We hypothesize that this method may be useful as the order in which words occur *can* contribute to framing. Furthermore, LSTM classifiers have shown superior performance in text classification [27]. Thus, we test this classifier for its ability to recognize framed sentences. We feed the words in the sentence in order as this classifier does not follow a bag-of-words assumption.

Generative Classifiers

C4 *Multinomial Naïve Bayes Classifier*: We train a Multinomial Naïve Bayes Classifier using the features described above. This classifier has proven to be a strong contender in many text classification problems [38], so we evaluate it for this task.

Discriminative Classifiers

C5 *Decision Tree*: We train a decision tree [46] using the described features. We choose this model as it allows an interpretation of the feature importance.

C6 *Linear SVM*: We train an SVM using the features described above. We use a linear kernel to separate the instances [23].

C7 *Logistic Regression*: We train a logistic regression classifier using the features described above. By using the feature weights, we will be able to obtain a ranking of the feature importance with regard to each class.

C8 *Logistic Regression without English-Specific Features*: We train a logistic classifier with a subset of the features described above: only the multi-lingual features. Note that for non-English languages, this is identical to C6.

Ensemble Classifiers

C9 *Random Forest Classifier*: We train a random forest classifier [12] using the features described. This classifier is unique as it is a type of ensemble classification scheme. We choose this as a representative ensemble classifier.

To test these classifiers, we use a standard 10-fold cross-validation experimental setup. This means that we randomly split the data into 10 equally sized chunks, and use 9 of those chunks to train the classifier. The remaining chunk is then used to test the performance of the model [35]. The average of all 10 folds for all classifiers is shown in Figure 1. We do not balance the datasets, instead using the results of "C1" to show the result of a random guess. The results of this experiment show a ubiquitous winner in the logistic regression without English-specific features ("C8") classifier. These results are significant in all cases except Poland and Russia. This shows that frames are best detected with the language of the text, and that bigrams are strong enough to pick up the nuances of the text. Furthermore, part-of-speech tags and sentence complexity do not contribute to detecting whether a sentence is framed or not, but instead add noise to the classifier. As a result, we use the "C8" classifier in subsequent experiments.

4.2 Differentiating Frame Types

In the previous section, we used a Boolean classification scheme: sentences are either framed or they are not. While this is useful to see the ability of the classifier to find frames, we must go
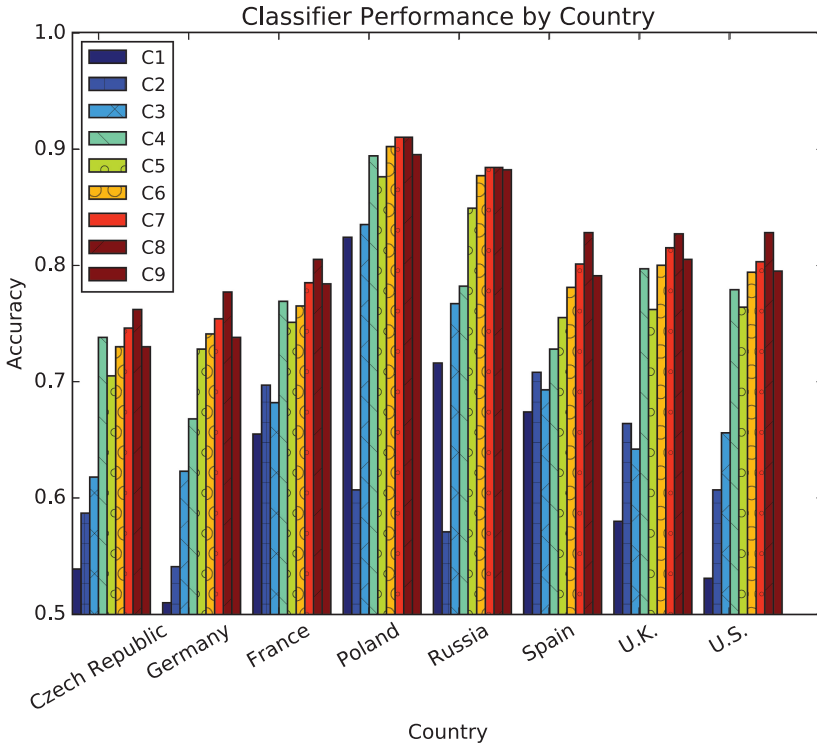
Fig. 1. Performance of classifiers in terms of accuracy (*y*-axis) by country (*x*-axis). Note that the range of the *y*-axis spans from 0.5 to 1.0 as 0.5 is the worst possible performance in any binary classification. The logistic regression model without English-specific features "C8" attains the best performance across all countries.

beyond this to truly detect frames in the dataset. The dataset contains not just Boolean labels as to whether a sentence is framed, but also the *type* of frame that is exhibited in the sentence. In our earlier section where we introduced the data, we discussed the 10 types of frames that appear in the dataset. In this experiment, we try to differentiate the frames themselves, and we ignore the possibility of a "not framed" class, which is what we differentiated in the previous experiment. We use a 1-vs.-all approach to do the classification: in each experiment we treat the class under study as the positive case and all others as the negative case. We do this because this class was investigated in the previous experiment. Additionally, the results would be skewed: since the "not framed" class is prominent in the text, our results would be skewed by a preference toward this class. We want to measure how well our classifier can distinguish between the types of frames.

The results of this experiment, shown in the "10-Class Accuracy" column of Table 3 show that we are able to detect the correct class the majority of the time. We see that when we are only differentiating frame types, our classifier still does well, beating the baseline classifier by at least 10%, with an average improvement of 18% across the countries. This shows that our classifier is able to pick up on the differences between these features.

4.3 Detecting Polarity of Frames

The last step to fully representing the frame labels is to predict not just the frame type, but also the polarity. Frames are not merely discussions of topics, but are an attempt to skew the perception

Table 3. Results of Predicting the Class Type

Country	Majority Class	10-Class Accuracy
Czech Republic	0.283	0.441
France	0.179	0.401
Germany	0.201	0.469
Poland	0.354	0.459
Russia	0.199	0.320
Spain	0.211	0.382
United Kingdom	0.223	0.441
United States	0.297	0.478

The 10-class accuracy consists only of the different frame types. The "Majority Class" column is the performance of a baseline which always guesses the most prominent class.

Table 4. Results of Predicting the Class Type with Polarity

Country	Majority Class	20-Class Accuracy
Czech Republic	0.222	0.382
France	0.108	0.349
Germany	0.179	0.430
Poland	0.258	0.339
Russia	0.179	0.278
Spain	0.180	0.361
United Kingdom	0.133	0.363
United States	0.258	0.434

The "20-Class Accuracy" column consists only of the 10 different frame types which can be slanted in two ways. The "Majority Class" column is the performance of the majority-class baseline, which always guesses the most prominent class.

of the topic in a certain way. Thus, it is important that we can also predict which direction the sentence is biased toward. Here, we test the ability of our classifier to detect the polarity of frames. This task is harder for two reasons. First, the number of predicted classes double, which can further confuse a classifier. Second, the labels in this case are even more nuanced. We have "zoomed in" to not only consider the frame, but also the polarity, which can be determined by only very small nuances to the language.

Extending our experiment in the previous section, we double the number of framed classes. That is, we have 10 framed classes, which can be of 2 possible polarities (positive, or negative; there is no neutral polarity). Following the previous experiment, we do not have a "not framed" class, as we are interested in seeing how well our classifier differentiates among the frame classes. The results of this experiment are shown in Table 4. The results are quite positive: despite doubling the number of classes, the performance is only reduced by 1.0% on average, a minor impact.

5 TEMPORAL STABILITY OF FRAME PREDICTION

So far we have employed a standard cross-validation experimental setup where the model is left to predict a completely random 10% of the sentences in the corpus. While it is useful to understand how much signal we can extract from our features, as well as the feasibility of the task as a whole, this setup is not a valid representation of how the classifier will be used in the real world. A

Table 5. Number of Articles Per Year Per Country

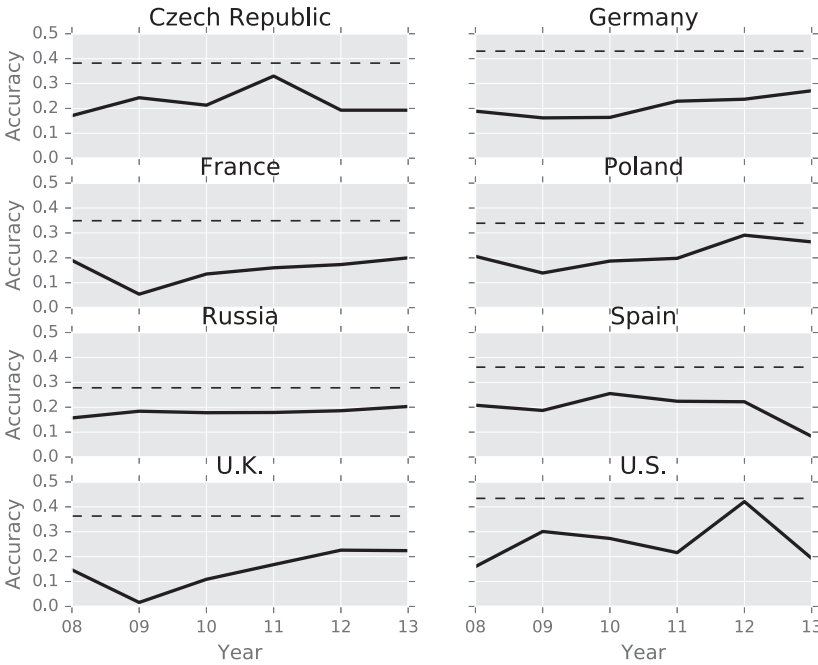| Year	Czech Republic	Germany	France	Poland	Russia	Spain	U.K.	U.S.
2007	21	10	20	7	11	20	6	20
2008	20	11	20	10	11	20	9	18
2009	16	14	21	12	3	32	5	20
2010	14	16	21	14	9	20	14	20
2011	13	13	18	14	12	20	8	20
2012	4	20	16	20	10	20	14	20
2013	8	16	5	20	12	20	9	20

Fig. 2. Performance of framing classifiers by year, trained only on the preceding year. The x-axis indicates the year of the test data, and the y-axis is the accuracy of the classifier. The solid line is the performance of the by-year classifier, and the dashed line is the results of the earlier cross-validation experiment. This is provided to give a sense of the relative performance of the model.

typical scenario would be that the model would be deployed and then make predictions on *future sentences*, that is, sentences that are generated strictly after the model is built. In order to simulate this environment, we test how well our model performs when trained on past data and tested on future data. We perform this experiment by building a classifier that is trained on year $i - 1$ and testing it with data from year i. This experiment is carried out for $2007 < i \leq 2012$. In this experiment, we are testing how good our classifier is at predicting framed sentences generated in the future. We use the 20-class setup where we predict the frame types as well as their polarity. To give an idea of the volume of information generated by year, we present the number of articles by year in Table 5.

The results of these experiments are shown in Figure 2. The solid line represents the results of our experiment by year, and the dashed line represents the results of the "20-Class Accuracy"

experiment. The year in the x-axis is the year of the *test* data, with the training data in all cases being only the previous year.

The first thing we notice is that there is a performance hit involved with this experimental setup. In only one instance (the United States when tested on 2012) do we achieve similar performance. In fact, on average there is a 16.8% loss in accuracy in this experimental setup. On average, Russia encumbers the lowest performance hit of 9.7%, and the U.K. takes the largest performance hit with 22.1%. Secondly, there is a small trend toward better performance as time goes on, with an average increase in accuracy of 2.5% from 2008 to 2013. There is an average Spearman's ρ correlation of 0.126 between the country and the year across all countries. This means that there is a slight trend toward predictions getting better with time.

6 CASE STUDY: APPLYING FRAMES TO DIFFERENTIATE MEDIA AND GOVERNMENT NEWS

So far, we have tried to predict the frames present in the sentences. Framing is an exertion of an opinion in a nuanced way. It is reasonable to assume that public media outlets may frame their news in a different way from government outlets, as they may have a differing set of opinions on the topic. One example of this phenomenon came in the push to The Iraq War. The government used framing in an attempt to encourage other countries to support their goal, while many media outlets opposed the effort [21]. Because of their difference in opinion, we should be able to differentiate media texts from government texts by using the results of framing.

In this section, we examine the feasibility of using the results of framing in order to detect another type of opinion bias between media and government outlets. The ability to do this can say as much about the classifier as the level of freedom under which the outlet operates.

The hypothesis in this experiment is that in countries with more press freedom, the public media will be allowed more control over how it discusses different aspects of a story. In other words, the public media will be allowed to disagree with, or express a different opinion from the government-generated news. We hypothesize that the "media" or "government" label will be *easier* to predict in countries with *more* press freedom as media outlets are able to express different opinions than those expressed by the government. The contraposition is that this prediction task should perform worse in countries with less press freedom as the media in those countries are unable to express different opinions from the government, making their text more similar between the two types of sources. To test this hypothesis, we build a simple classifier to predict the origin of different sentences by country. Here we define "origin" as the type of outlet that the article originates, either "media" or "government." The government texts were from official government websites. Government sources were identified as those that appeared on NATO's website. Nato provides links to relevant government agency websites for each country. For example, anything from BBC, VOA, RT, Sputnik, and so forth was included in the media category.

The classifier we build is a logistic regression classifier, inspired by the results of the previous framing experiments. The features used consist of the unigrams and bigrams of the text in the sentence, as well as the predicted frame label using a modified version of the fine-grained "polarity" classifier introduced earlier. This modified classifier also takes into account the possibility that the sentence is not framed. This is a binary classification task where we are predicting either a "media" or "government" label. Since most countries have a roughly even distribution of government/media sentences (as shown in the "% Gov." column of Table 1), we do not balance the data. We use a traditional 10-fold cross-validation scheme. A separate model is built for each country.

The results of this experiment are shown in the "Predicted" column of Table 6. The results indicate that this classification task is tractable in all countries, with every country achieving ≥79% accuracy in the prediction task. To give a sense of how these results are impacted by using the

Table 6. Results of the "Media" or "Government"
Classification Task by Country

Country	Predicted	Hand Coded	PFI
Czech Republic	86.2 ± 1.5	86.6 ± 1.5	10.17
France	83.0 ± 1.6	83.0 ± 2.2	21.60
Germany	87.3 ± 1.4	88.5 ± 2.0	10.24
Poland	84.8 ± 1.3	85.0 ± 1.6	13.11
Russia	80.8 ± 2.1	80.0 ± 2.4	43.42
Spain	79.0 ± 2.5	80.2 ± 2.9	20.50
United Kingdom	89.6 ± 1.3	89.8 ± 1.3	18.22
United States	89.2 ± 1.3	89.7 ± 1.9	16.89
Corr. with PFI (ρ)	−0.55	−0.65	—

The "Predicted" column contains the results of using the predicted
frame labels, and the "Hand Coded" column contains the prediction us-
ing the ground truth labels. "PFI" is the results the Press Freedom Index
(PFI) scores each country based upon press freedom, with *higher* scores
indicating countries with *less* press freedom.

framing labels predicted by our classifier, we rerun the experiment by using the labels present in
the ground truth. The results of this experiment are shown in the "Hand Coded" column of Table 6.
We see that using predicted labels only introduces a minor hindrance to performance. This shows
that the classifier is effective at separating these documents, and more importantly that the results
of our framing classifier developed in the previous experiments are useful at detecting other forms
of bias.

While we were able to achieve reasonable classification results, we have not confirmed our
hypothesis regarding separability as a function of the country's freedom. To test this, we compare
our classification with the Press Freedom Index (PFI),[1] an annual scoring of countries made by the
organization Reporters Without Borders. The scoring is manually done by reporters and is based
upon the freedom that news agencies enjoy within their respective country. More details about
how the scoring is performed can be found in the PFI's annual report [49]. When we compare
our classifier's performance for each country with the PFI score for that country in Table 6, we
see that there is a correlation between the two values. The Pearson's ρ between the performance
metrics and the PFI can be seen at the bottom of Table 6. We see a moderate negative correlation
between each performance measure and the PFI. While the correlation between the labels and PFI
is not significant ($p = 0.13$ for "Predicted," and $p = 0.08$ for "Hand Coded"), it represents a clear
statistical trend between the classifiers performance and the PFI. This tells us that when the PFI is
higher (*less* press freedom), that the performance will be lower (the media and government exhibit
more similar linguistic patterns, and are harder to separate).

6.1 Feature Evaluation

In the previous experiments, we saw that we are able to differentiate between media and govern-
ment documents. While high performance is useful, we want to know *why* the classifier makes its
predictions. To do this, we analyze the features that can achieve this high performance. Following
previous examples [32], we inspect the magnitude of the feature weights learned by the logistic re-
gression classifier for each country. Since we have a Boolean classifier, and the features are scaled,

---

[1]https://index.rsf.org/.

Table 7. Top Weighted Individual Features in the "Media," and "Government" Labels for Each Country

Country	Label	Top 5 Features
Czech Republic	Government	**frame:DomesticBenefit-**, **frame:CollectiveSecurity+**, **frame:GeneralThreat+**, uni:think, uni:thank
Czech Republic	Media	uni:", uni:moscow, **frame:RussiaPartnership+**, uni:shield, uni:americans
France	Government	uni:mr., **frame:ProgressEffectiveness-**, uni:france, **frame:GeneralThreat+**, uni:BMD
France	Media	uni:Moscow, uni:Washington, **frame:RussianRoadblocks+**, uni:Russian, uni:shield
Germany	Government	**frame:RussiaPartnership+**, **frame:NOFRAME**, uni:security, uni:policy, uni:example
Germany	Media	uni:shield, **frame:ProgressEffectiveness-**, uni:missiles, uni:putin, **frame:RussianRoadblocks+**
Poland	Government	**frame:NOFRAME**, uni:dziękuję ("thank you"), uni:proszę ("please"), uni:poseł ("envoy"), uni:przeciwrakietowej ("missile")
Poland	Media	uni:tarczy ("shield"), uni:rosji ("russia"), uni:tarcza ("shield"), uni:", uni:rzeczpospolita ("republic")
Russia	Government	uni:Спасибо ("thank you"), **frame:NOFRAME**, uni:К. Тоцкий ("K. Trotskiy") uni:PC, uni:O.A
Russia	Media	uni:", uni:президент ("the president"), uni:Буш ("Bush"), uni:системы ("system"), uni:союзников ("allies")
Spain	Government	bi:", not", uni:initiative, **frame:PoliticalTensions+**, **frame:ProgressEffectiveness-**,uni:mr
Spain	Media	uni:U.S., uni:Brussels, uni:Berlin, uni:EFE, uni:year
United Kingdom	Government	**frame:NOFRAME**, **frame:GeneralThreat+**, uni:forces, uni:early, bi:missile defence
United Kingdom	Media	uni:", **frame:ProgressEffectiveness-**, uni:shield, **frame:RussianRoadblocks+**, uni:?
United States	Government	uni:gates, **frame:CollectiveSecurity+**, uni:block, **frame:NOFRAME**, **frame:ProgressEffectiveness+**
United States	Media	uni:", uni:shield, **frame:PoliticalTensions+**, **frame:ProgressEffectiveness-**, **frame:RussianRoadblocks+**

These are the features that are given the highest weight by the logistic regression classifier. Words in parentheses are English translations of non-English words. Features starting with "uni" are unigram features, "bi" are bigram features, and "frame" are framing features. We also include the results of punctuation unigrams if they appear in the top 5; e.g., "uni:?" indicates the question mark unigram.

the weights can be used to give an indication of the most important features with respect to each class. The results of this experiment are shown in Table 7. The results show a clear preference for the presence of frames as the top features, with 25 of the 80 features being frame features.[2] Additionally, we notice that there are no frame features used in the Russia classifier. Because the media and government are highly correlated in terms of their opinion, frames are not useful to differentiate the class labels. Additionally, we see that the unigram and bigram features largely point to actors and locations related to the topic of ballistic missile defense. This shows that the classifier is using this information to make its differentiation between media and government text.

7 CONCLUSION AND FUTURE WORK

Framing is very important as it is a subtle bias introduced by authors to skew their readers into preferring certain outcomes. In this work, we study the degree to which framed text can be detected automatically. We study the ability of a machine learning classifier to detect frames in sentences of a news corpus pertaining to ballistic missile defense across eight different countries from 2007 to 2013. We predict frames at different granularities, from a binary classifier simply looking for the existence of any frame, to a classifier which predicts not only the frame type, but also the polarity. We also find that a classifier *can* be deployed to detect frames in future articles, albeit with a drop in performance. We find that simple linguistic features perform best in this classification task, and that *n*-grams can give reasonable prediction in finding frames in text. The results of our experiments show great promise in terms of frame detection, and pave the way for future researchers to propose sophisticated classifiers to take on this task.

Additionally, we find that frames are useful for predicting other types of opinion bias. We used frames as features to separate "media"-generated and "government"-generated documents. We find

---

[2]To compare the significance of these results, we also build a decision tree classifier for each country and inspect its top features. The decision tree also contains many framing features in its top features list, meaning that the results generalize to other classifiers.

that a classifier built to do this task most heavily relies on framing features to differentiate these two types of documents.

The results of this work have implications for improving existing web technology. For example, if recommender systems are better able to understand the bias of news articles through their framing, they will be able to better suit their users' taste [16]. On the other hand, the results can be used by news aggregators who wish to provide a less-biased view of news stories to their audience [26, 37, 53], and to overcome "filter bubbles" [43].

This article opens the doors for several areas of future work. First, designing classifiers that can predict future frames is important in order to make this task more fit for the real world. Also, researchers can use the results of this article to improve the performance of this classification task. Additionally, in this study we used a labeled dataset provided by experts in framing literature. This dataset is immensely useful, but also incredibly rare. To get around this labeling bottleneck, one area of future work is to study framing from an unsupervised perspective, eliminating the need for labeled data entirely. Finally, in this article the frames themselves were predetermined by experts. This is another time-consuming task that requires expert human effort in order to carry out. Future work is to identify the frame types automatically, without the need for human intervention.

REFERENCES

[1] Saud Alashri, Sultan Alzahrani, Lenka Bustikova, David Siroky, and Hasan Davulcu. 2015. What animates political debates? Analyzing ideological perspectives in online debates between opposing parties. In *Proceedings of SocialCom*.

[2] Jisun An, Meeyoung Cha, Krishna Gummadi, and Jon Crowcroft. 2011. Media landscape in Twitter: A world of new conventions and political diversity. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*.

[3] Jisun An, Meeyoung Cha, Krishna P. Gummadi, Jon Crowcroft, and Daniele Quercia. 2012. Visualizing media bias through twitter. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*.

[4] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.

[5] Ethan M. Balk, Mei Chung, Nira Hadar, Kamal Patel, W. Yu Winifred, Thomas A. Trikalinos, and Lina Kong Win Chang. 2012. Accuracy of data extraction of non-English language trials with Google Translate.

[6] David P. Baron. 2006. Persistent media bias. *Journal of Public Economics* 90, 1 (2006), 1–36.

[7] Eric Baumer, Elisha Elovic, Ying Qin, Francesca Polletta, and Geri Gay. 2015. Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of HLT-NAACL*. The Association for Computational Linguistics, 1472–1482.

[8] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*. O'Reilly Media, Inc.

[9] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent Dirichlet allocation. *The Journal of Machine Learning Research* 3 (2003), 993–1022.

[10] Amber E. Boydstun, Dallas Card, Justin Gross, Paul Resnick, and Noah A. Smith. 2014. Tracking the development of media frames within and across policy issues.

[11] Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2013. Identifying media frames and frame dynamics within and across policy issues. In *Proceedings of New Directions in Analyzing Text as Data Workshop*.

[12] Leo Breiman. 2001. Random forests. *Machine Learning* 45, 1 (2001), 5–32.

[13] Pedro Henrique Calais Guerra, Adriano Veloso, Wagner Meira, Jr., and Virgílio Almeida. 2011. From bias to opinion: A transfer-learning approach to real-time sentiment analysis. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'11)*. ACM, New York, 150–158. DOI:http://dx.doi.org/10.1145/2020408.2020438

[14] Erik Cambria, Bjorn Schuller, Yunqing Xia, and Catherine Havasi. 2013. New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems* 2 (2013), 15–21.

[15] Dallas Card, Amber E. Boydstun, Justin H. Gross, Philip Resnik, and Noah A. Smith. 2015. The media frames corpus: Annotations of frames across issues. In *Proceedings of ACL (2)*. 438–444.

[16] Abhinandan S. Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. 2007. Google news personalization: Scalable online collaborative filtering. In *Proceedings of the 16th International Conference on World Wide Web*. ACM, 271–280.

[17] James W. Dearing and Everett Rogers. 1996. *Agenda-Setting*. Vol. 6. Sage Publications.

[18] Stefano DellaVigna and Ethan Kaplan. 2006. *The Fox News Effect: Media Bias and Voting*. Technical Report. National Bureau of Economic Research.

[19] Nicholas Diakopoulos, Amy X. Zhang, Dag Elgesem, and Andrew Salway. 2014. Identifying and analyzing moral evaluation frames in climate change blog discourse. In *Proceedings of ICWSM*.

[20] James N. Druckman and Michael Parkin. 2005. The impact of media bias: How editorial slant affects voters. *Journal of Politics* 67, 4 (2005), 1030–1049.

[21] Jill A. Edy and Patrick C. Meirick. 2007. Wanted, dead or alive: Media frames, frame adoption, and support for the war in Afghanistan. *Journal of Communication* 57, 1 (2007), 119–141.

[22] Robert M. Entman. 1993. Framing: Toward clarification of a fractured paradigm. *Journal of Communication* 43, 4 (1993), 51–58.

[23] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. 2008. LIBLINEAR: A library for large linear classification. *The Journal of Machine Learning Research* 9 (2008), 1871–1874.

[24] William A. Gamson. 1989. News as framing: Comments on Graber. *American Behavioral Scientist* 33, 2 (1989), 157–161.

[25] R. Kelly Garrett and Brian E. Weeks. 2013. The promise and peril of real-time corrections to political misperceptions. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. ACM, 1047–1058.

[26] Matthew Gentzkow and Jesse M. Shapiro. 2011. Ideological segregation online and offline. *The Quarterly Journal of Economics* 126, 4 (2011), 1799–1839.

[27] Alex Graves and Jürgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks* 18, 5 (2005), 602–610.

[28] Stephan Greene and Philip Resnik. 2009. More than words: Syntactic packaging and implicit sentiment. In *Proceedings of NAACL*. Association for Computational Linguistics, 503–511.

[29] Tim Groseclose and Jeffrey Milyo. 2005. A measure of media bias. *The Quarterly Journal of Economics* (2005), 1191–1237.

[30] Michael Halliday, Christian M. I. M. Matthiessen, and Christian Matthiessen. 2014. *An Introduction to Functional Grammar*. Routledge.

[31] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[32] David W. Hosmer Jr. and Stanley Lemeshow. 2004. *Applied Logistic Regression*. John Wiley & Sons.

[33] Shanto Iyengar and Adam Simon. 1993. News coverage of the Gulf crisis and public opinion: A study of agenda-setting, priming, and framing. *Communication Research* 20, 3 (1993), 365–383.

[34] Dan Jurafsky and James H. Martin. 2014. *Speech and Language Processing*. Pearson.

[35] Ron Kohavi and others. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of IJCAI*, Vol. 14. 1137–1145.

[36] Klaus Krippendorff. 2004. *Content Analysis: An Introduction to its Methodology*. Sage.

[37] Juhi Kulshrestha, Muhammad Bilal Zafar, Lisette Espin Noboa, Krishna P. Gummadi, and Saptarshi Ghosh. 2015. Characterizing information diets of social media users. In *Proceedings of the 9th International AAAI Conference on Web and Social Media*.

[38] Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze, and others. 2008. *Introduction to Information Retrieval*. Vol. 1. Cambridge University Press.

[39] Maxwell E. McCombs and Donald L. Shaw. 1972. The agenda-setting function of mass media. *Public Opinion Quarterly* (1972), 176–187.

[40] Arturo Montejo-Raez, Eugenio Martínez-Cámara, M. Teresa Martin-Valdivia, and L. Alfonso Urena-Lopez. 2012. Random walk weighting over sentiwordnet for sentiment polarity detection on Twitter. In *Proceedings of the 3rd Workshop in Computational Approaches to Subjectivity and Sentiment Analysis*. Association for Computational Linguistics, 3–10.

[41] Sean A. Munson and Paul Resnick. 2010. Presenting diverse political opinions: How and how much. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1457–1466.

[42] Vlad Niculae, Caroline Suen, Justine Zhang, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2015. QUOTUS: The structure of political media coverage as revealed by quoting patterns. In *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 798–808.

[43] Eli Pariser. 2011. *The Filter Bubble: What the Internet is Hiding from You*. Penguin UK.

[44] Souneil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. 2009. NewsCube: Delivering multiple aspects of news to mitigate media bias. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 443–452.

[45] Souneil Park, Minsam Ko, Jungwoo Kim, Ying Liu, and Junehwa Song. 2011. The politics of comments: Predicting political orientation of news stories with commenters' sentiment patterns. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*. ACM, 113–122.

[46] J. Ross Quinlan. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

[47] Delip Rao and Deepak Ravichandran. 2009. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 675–682.

[48] Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of ACL (1)*. 1650–1659.

[49] Reporters Without Borders. 2015. 2015 World Press Freedom Index: How We Compiled the Index. Retrieved February 20, 2017 from http://fr.rsf.org/IMG/pdf/methodology-en.pdf.

[50] Diego Saez-Trumper, Carlos Castillo, and Mounia Lalmas. 2013. Social media news communities: Gatekeeping, coverage, and statement bias. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*. ACM, 1679–1684.

[51] Dietram A. Scheufele and David Tewksbury. 2007. Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication* 57, 1 (2007), 9–20.

[52] Yanchuan Sim, Brice D. L. Acree, Justin H. Gross, and Noah A. Smith. 2013. Measuring ideological proportions in political speeches. In *Proceedings of EMNLP*.

[53] Mitja Trampuš, Flavio Fuart, Daniele Pighin, Tadej Štajner, Jan Berčič, Blaz Novak, Delia Rusu, Luka Stopar, and Marko Grobelnik. 2015. DiversiNews: Surfacing diversity in online news. *AI Magazine* 36, 4 (2015).

[54] Oren Tsur, Dan Calacci, and David Lazer. 2015. A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of ACL*.

[55] Amos Tversky and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *Science* 211, 4481 (1981), 453–458.

[56] Shaomei Wu, Jake M. Hofman, Winter A. Mason, and Duncan J. Watts. 2011. Who says what to whom on Twitter. In *Proceedings of the 20th International Conference on World Wide Web*. ACM, 705–714.