Fake News in the News: An Analysis of Partisan Coverage of the Fake News Phenomenon

Xunru Che
Northwestern University
Evanston, IL 60201, USA
xunruche2019@u.northwestern.edu

Danaë Metaxa-Kakavouli
Stanford University
Stanford, CA 94305, USA
metaxa@cs.stanford.edu

Jeffrey T. Hancock
Stanford University
Stanford, CA 94305, USA
jeff.hancock@stanford.edu

Abstract

Since the 2016 U.S. election cycle, "fake news" (a term describing verifiably false and misleading news articles) has garnered increasing public attention. This work sheds insight onto this phenomenon by examining the way 10 popular partisan media sites discuss "fake news". We use linguistic analysis techniques including Linguistic Inquiry and Word Count (LIWC), word embedding models, and supervised learning classifiers to analyze news stories containing the phrase "fake news" from left- and right-leaning news sites. Our results yield several insights, including that article text can be used to classify political affiliation with high accuracy, and that left-leaning sites focus on specific fake news stories and individuals involved, while right-leaning sites shift the focus to a narrative of mainstream media dishonesty more broadly.

Author Keywords

Fake news; news media; political partisanship; text analysis

Introduction

Following the 2016 U.S. election cycle, the issue of "fake news" online has come into sharp public focus. As defined by Allocott and Gentzkow (2017), "fake news" refers to news articles that are verifiably false, and intentionally misleading [1]. In the months leading up to the 2016 presidential elections, U.S. social media users were exposed to

several widely-shared fake news stories, which some have suggested increased the public attention given to and belief in the initial fake stories [1, 4, 6, 7].

Relevant to different contexts surrounding the discussion of fake news is the highly partisan nature of the U.S. political climate. A study on congressional speech showed that partisanship began to increase dramatically in the 1990s [3]. U.S. media organizations are similarly polarized. A 2017 report from the Berkman-Klein Center reports that some popular media outlets on the political Right are highly partisan while those on the Left are more objective in their coverage, resulting in an asymmetrical media environment [2].

Given the highly polarized nature of U.S. news media as well as the evolving and nebulous nature of fake news, this research is a step towards comparing and contrasting the ways left- and right-wing news organizations treat the concept of fake news. We first identify five popular left- and right-leaning news sites, and then collect the text of news articles discussing fake news from those sites. We analyze the total of over 200,000 news stories using simple machine learning classifiers, word embedding models, and Linguistic Inquiry and Word Count (LIWC, a tool for finding social and psychological insights from text) [8]. We find that news sources on each side frequently refer to the other, providing further evidence of political polarization in discussions of this topic. We also find some key differences—while left-leaning sources discuss specific examples of fake news, the narrative in right-leaning sources focuses on mainstream media as a whole. These insights and others serve as a first step in understanding the impact of political polarization on the discussion of fake news.

Left-leaning news top informative features	Right-leaning news top informative features
donald	leftwing
rightwing	narrative
barack	caller
hed	antitrump
news	breitbart
guardian	stated
chute	democrats
farright	liberal
value	the
child	clintons

Table 1: Most informative words used by the logistic regression in predicting political slant

Data Set
Using the results of a 2017 study, we chose five left-leaning and five right-leaning new sites which were highly polarized and active in online social media: left-leaning: Huffington Post, PoliticusUSA, Daily Kos, the Guardian, and Raw Story; right-leaning: Breitbart, FOX News, The Washington Examiner, The Daily Caller, and The Right Scoop [2]. We scraped a dataset consisting of 201,101 total news articles using Python's Scrapy library; 5,289 articles containing the term "fake news" in the title or body, and 195,812 articles not containing it.

Methods and Results
Based on initial inspections of the data, we trained a classifier to classify the source's political slant, and used word embeddings and LIWC analysis to explore partisan patterns in the contexts in which fake news was discussed.

Classification
We trained a logistic regression classifier to predict the political slant (left- or right-leaning) of news articles with the phrase "fake news". We used unigrams from the body text of the 5,289 articles related to fake news from as input, performed a 10-fold cross-validation to test the accuracy, and recorded the top 10 most informative unigrams.

Classification Success and Informative Classifier Features
Our classifier achieved a moderately high mean cross-validation score, $\mu = 0.74$, well above chance ($\mu < 0.5$ for binary classification). We examined the 10 most informative features to gain insights into the characteristics of articles that best distinguished left- from right-leaning sources (see Table 1). Most notably, each side frequently referred the the opposite side of the political spectrum, as well as to notable figures in the political climate.

Left-leaning news corpus	
Feature	Score
misinformation	0.76
mainstream media	0.74
russian propaganda	0.72
conspiracy theories	0.71
alternative facts	0.69
disinformation	0.68
pizzagate	0.67
alt-right	0.66
clickbait	0.66
falsehoods	0.66

Right-leaning news corpus	
Feature	Score
mainstream media	0.67
misinformation	0.65
biased	0.63
establishment media	0.62
dishonest	0.62
hysteria	0.62
disinformation	0.62
falsehoods	0.62
media	0.62
liberal media	0.61

Table 2: Phrases that were most similar to the query term "fake news" in each corpus

Word Embeddings

To further explore the contexts in which fake news is discussed, we trained word embedding models using Gensim's Word2Vec [5], training two word embedding models using unigrams and bigrams from news article's body text as inputs.

Word Embedding Outputs

We queried each word embedding model for the top ten features with the highest similarly to the term "fake news" to gain insights into the ways in which the term is leveraged in these different media environments. The output terms and similarity scores for each term are reported in Table 2.

Linguistic Inquiry and Word Count (LIWC)

In order to examine more subtle insights in our data set, we used LIWC, a research tool that identifies linguistic traits in a text data set [8]. We ran LIWC on the two corpora of left- and right-leaning news articles related to fake news, and identified categories in which the two groups significantly differ (see Figure 1). We find that left-leaning content contains more positive emotion words, $F(1,1) = 110.151$, $p < 0.001, d = 0.326$; fewer negative emotion words, $F(1,1) = 22.188, p < 0.001, d = -0.175$; more future-focus words, $F(1,1) = 11.023, p = 0.001, d = 0.062$; fewer past-focus words, $F(1,1) = 68.650, p < 0.001, d = -0.444$; and fewer female-related words, $F(1,1) = 9.430$, $p = 0.002, d = -0.067$.

Discussion

We find that left- and right-slanted media sources use discernibly different language when discussing fake news underscoring a fundamental difference in definitions and portrayals of the issue, and that the two sides frequently focus on each other presumably in the spirit of blaming the other for this problem.

Classification

A logistic regression classifier is relatively well-able to predict political slant based on page text suggesting that the language used on the Left and Right to discuss fake news is substantively different. Examining the most informative features of our classifier (Table 1), we see that sources on both ends of the political spectrum frequently refer to the other in articles about fake news, underscoring the lack of agreement in the current political climate about the cause of this issue; each side blames the other for the problem.

Word Embeddings

Querying our word embedding models for the term "fake news" shows that sources on both sides agree on some synonyms and related words, including "misinformation," "disinformation," and "falsehoods" (see Table 1). The phrase "mainstream media" also appears in the results for both models, indicating that the role of traditional news media is central to the concept of fake news. We also find evidence that partisans on each side blame the other: left-leaning news sites tended to mention related concepts associated with the Right, like "pizzagate," "alternative facts," and "conspiracy theories," suggesting that left-leaning sources attempted to call out this harm in order to defend the reputation of the party. In contrast, right-leaning sites appear more concerned with media dishonesty in general, using terms such as "establishment media," "dishonest," and "hysteria."

LIWC

We find that fake news-related news in left-leaning sites tend to have more moderately positive emotion words and slightly fewer negative-emotion words than right-leaning sites; this is slightly surprising given the left-leaning news coverages' focus on specific instances of fake news, but could be reflective of the more negative view right-leaning sources have about the broader media environment. We

also see that right-leaning sites use more past-focus and female-related words; the Right may be more engaged in discussing the losing female Democratic Party candidate and their perception of her role in this issue.

Limitations and Future Work
This work is a first step towards identifying descriptive differences between conservative and liberal news sites in their discussions of fake news. However, our analysis only compares ten most highly partisan left-leaning and right-leaning sources. Further analyses should consider a greater number of sources in order to verify these results, and could focus on other political affiliations besides the Right and Left, as well as citizens' perceptions of of this coverage.

Conclusion
The work aims to better understand fake news in the context of a politically polarized mainstream media by investigating linguistic differences in the discussion of the topic by left- and right-leaning news sources. We apply three approaches towards that understanding: text classification, word embeddings, and LIWC. We find that both sides of the political spectrum refer to the other in the context of fake news, and that left-leaning sources discuss specific fake news-related events and individuals while right-leaning sources more commonly describe the topic in the context of mainstream media dishonesty. These results underscore the politically polarized nature of partisan news sources, as well as the two parallel conversations—one about conspiracy theory-like examples and the other about media trustworthiness—which surround the topic of fake news.

REFERENCES

1. Hunt Allcott and Matthew Gentzkow. 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31, 2 (May 2017), 211–36. DOI:http://dx.doi.org/10.1257/jep.31.2.211

2. Yochai Benkler, Robert Faris, Hal Roberts, and Ethan Zuckerman. 2017. Study: Breitbart-led right-wing media ecosystem altered broader media agenda. *Columbia Journalism Review* (2017).

3. Matthew Gentzkow, Jesse M. Shapiro, and Matt Taddy. 2016. *Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech*. Working Paper 22423. National Bureau of Economic Research. DOI:http://dx.doi.org/10.3386/w22423

4. Alice Marwick and Rebecca Lewis. 2015. *Media Manipulation and Disinformation Online*. Technical Report. Data & Society Research Institute.

5. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and Their Compositionality. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'13)*. Curran Associates Inc., USA, 3111–3119.

6. Craig Silverman. 2016. This Analysis Shows how Fake Election News Stories Outperformed Real News on Facebook. *Buzzfeed News* (2016).

7. Craig Silverman and Jeremy Singer-Vine. 2016. Most Americans Who See Fake News Believe It, New Survey Says. *BuzzFeed News* (2016).

8. Yla R. Tausczik and James W. Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29, 1 (2010), 24–54.
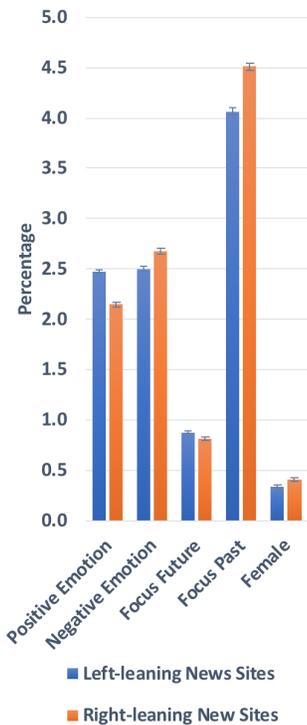
Figure 1: Results from LIWC analysis