Fake News vs Satire: A Dataset and Analysis

Jennifer Golbeck, Matthew Mauriello, Brooke Auxier, Keval H Bhanushali, Christopher Bonk, Mohamed Amine Bouzaghrane, Cody Buntain, Riya Chanduka, Paul Cheakalos, Jeannine B. Everett, Waleed Falak, Carl Gieringer, Jack Graney, Kelly M. Hoffman, Lindsay Huth, Zhenye Ma, Mayanka Jha, Misbah Khan, Varsha Kori, Elo Lewis, George Mirano, William T. Mohn IV, Sean Mussenden, Tammie M. Nelson, Sean Mcwillie, Akshat Pant, Priya Shetye, Rusha Shrestha, Alexandra Steinheimer, Aditya Subramanian, Gina Visnansky
University of Maryland
jgolbeck@umd.edu

ABSTRACT

Fake news has become a major societal issue and a technical challenge for social media companies to identify. This content is difficult to identify because the term "fake news" covers intentionally false, deceptive stories as well as factual errors, satire, and sometimes, stories that a person just does not like. Addressing the problem requires clear definitions and examples. In this work, we present a dataset of fake news and satire stories that are hand coded, verified, and, in the case of fake news, include rebutting stories. We also include a thematic content analysis of the articles, identifying major themes that include hyperbolic support or condemnation of a figure, conspiracy theories, racist themes, and discrediting of reliable sources. In addition to releasing this dataset for research use, we analyze it and show results based on language that are promising for classification purposes. Overall, our contribution of a dataset and initial analysis are designed to support future work by fake news researchers.

KEYWORDS

fake news, datasets, classification

ACM Reference Format:
Jennifer Golbeck, Matthew Mauriello, Brooke Auxier, Keval H Bhanushali, Christopher Bonk, Mohamed Amine Bouzaghrane, Cody Buntain, Riya Chanduka, Paul Cheakalos, Jeannine B. Everett, Waleed Falak, Carl Gieringer, Jack Graney, Kelly M. Hoffman, Lindsay Huth, Zhenye Ma, Mayanka Jha, Misbah Khan, Varsha Kori, Elo Lewis, George Mirano, William T. Mohn IV, Sean Mussenden, Tammie M. Nelson, Sean Mcwillie, Akshat Pant, Priya Shetye, Rusha Shrestha, Alexandra Steinheimer, Aditya Subramanian, Gina Visnansky. 2018. Fake News vs Satire: A Dataset and Analysis. In *Proceedings of 10th ACM Conference on Web Science (WebSci'18).* ACM, New York, NY, USA, Article 4, 5 pages. https://doi.org/10.1145/3201064.3201100

1 INTRODUCTION

"Fake news" was never a technical term, but in the last year, it has both flared up as an important challenge to social and technical systems and been co-opted as a political weapon against anything (true or false) with which a person might disagree. Identifying fake news can be a challenge because many information items are called "fake news" and share some of its characteristics. Satire, for example, presents stories as news that are factually incorrect, but the intent is not to deceive but rather to call out, ridicule, or expose behavior that is shameful, corrupt, or otherwise "bad". Legitimate news stories may occasionally have factual errors, but these are not fake news because they are not intentionally deceptive. And, of course, the term is now used in some circles as an attack on legitimate, factually correct stories when people in power simply dislike what they have to say.

Figure 1: Fake news.

If actual fake news is to be combatted at web-scale, we must be able to develop mechanisms to automatically classify and differentiate it from satire and legitimate news. To that end, we have built a hand coded dataset of fake news and satirical articles with the full text of 283 fake news stories and 203 satirical stories chosen from a diverse set of sources. Every article focuses on American politics and was posted between January 2016 and October 2017, minimizing the possibility that the topic of the article will influence the classification. Each fake news article is paired with a rebutting article from a reliable source that rebuts the fake source.

We were motivated both by the desire to contribute a useful dataset to the research community and to answer the following research questions: RQ1: Are there differences in the language of fake news and satirical articles on the same topic such that a word-based classification approach can be successful?

WebSci'18, May 27–30, 2018, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5563-6/18/05...$15.00
https://doi.org/10.1145/3201064.3201100

RQ2: Are there substantial thematic differences between fake news and satirical articles on the same topic?

Initial experiments show there is a relatively strong signal here that can be used for classification, with our Naive Bayes-based approach achieving 79.1% with a ROC AUC of 0.880 when differentiating fake news from satire. We also qualitatively analyzed the themes that appeared in these articles. We show that there are both similarities and differences in how these appear in fake news and satire, and we show that we can accurately detect the presence of some themes with a simple word-vector approach.

2 RELATED WORK

We are interested in *truly* fake news in this study - not stories people don't like, stories that have unintentional errors, or satire. We define the term as follows:

> Fake news is information, presented as a news story that is factually incorrect and designed to deceive the consumer into believing it is true.

Our definition builds on the work and analysis of others who have attempted to define this term in recent years, including the following.

Fallis [4] examines the ways people have defined disinformation (as opposed to misinformation). His conclusion is that "disinformation is misleading information that has the function of misleading." More specifically about fake news, researchers in [11] look at the uses of the term. They found six broad meanings of the term "fake news": news satire, news parody, fabrication, manipulation (e.g. photos), advertising (e.g. ads portrayed as legitimate journalism), and propaganda. They identified two common themes: intent and the appropriation of "the look and feel of real news."

In [10], Rubin breaks fake news into three categories: Serious fabrications, large scale hoaxes, and and humorous fakes. They don't explain why they chose these categories instead of some other classification. However, they do go into depth about what each category would contain and how to distinguish them from each other. They also stress the lack of a corpus to do such research, and emphasize 9 guidelines for building such a corpus: "Availability of both truthful and deceptive instances", "Digital textual format accessibility", "Verifiability of ground truth", "Homogeneity in lengths", "Homogeneity in writing matter", "Predefined timeframe", "The manner of news delivery", "Language and culture", and "pragmatic concerns".

The impact of fake news has become increasingly an important issue, due to its potential to impact important events. For example, [1] examined how fake news articles are shared on social media; their analysis suggests that the average American adult saw on the order of one or perhaps several fake news stories in the months around the election and (through a large scale survey) they found that consumers of fake news were more likely to believe stories that favor their preferred candidate or ideology.

In [9], the authors examine the impact of cognitive ability on the durability of opinions based on fake news reports. Four hundred respondents answered an online questionnaire, using a test-control design to see how their impressions and evaluations of an individual (test condition) changed after being told the information they received was incorrect. They found that individuals with lower cognitive ability adjusted their assessments after being told the information they were given was incorrect, but not nearly to the same extent as those with higher cognitive ability. Those with higher cognitive ability, when told they received false information, adjust their assessments in line with those who had never seen the false information to begin with. This was true regardless of other psychographic measures like right-wing authoritarianism and need for closure. This study suggests that for those with lower cognitive capability, the bias created by fake news, while mitigated by learning the initial information was incorrect, still lingers.

Pew Research Center conducted a survey of 1002 U.S. adults to understand attitudes about fake news, its social impact, and individual perception of susceptibility to fake news reports [2]. A majority of Americans believe that fake news is creating confusion about basic facts. This is true across demographic groups, with a correlation between income and the level of concern and across political affiliations. Still, they feel confident that that can tell what is fake when they encounter it, and show some level of discernment between what is patently false versus what is partially false. Seeing fake news more frequently increases the likelihood an individual believes it, creates confusion, and decreases the likelihood that one can tell the difference. Whether this is due to the accuracy of their perception that they can tell the difference, or their predilection to see news as fake is unknown, as the data is self-reported. Twenty three percent acknowledge sharing fake news, with 14% doing so knowingly.

Using the GDELT Global Knowledge graph, which monitors and classifies news stories, the researchers in [12] examined the topics covered by different media groups (such as fake news websites, fact-checking websites, and news media websites) from 2014 to 2016. By tracking the topics discussed across time by these three groups, the researchers were able to determine which groups of media were setting the agenda on different topics. They found that fake news coverage set the agenda for the topic of international relations all three years, and for two years the issues of economy and religion. Overall, fake news was responsive to the agenda set by partisan media on the topics of economy, education, environment, international relations, religion, taxes, and unemployment, indicated an "intricately entwined" relationship between fake news and partisan media. However, in 2016, the data indicates that partisan media became much more responsive to the agendas set by fake news. The authors suggest that future research should look at the flow from fake news to partisan media to all online media.

Researchers in [8] describe several studies investigating the relationships between believing fake news, Cognitive Reflection Test (CRT) scores, tendency to "over claim" (e.g., claim to recognize the name of a fictional historical figure), scores on the authors "bullshit receptivity task" (e.g. rating the profundity of meaningless jargon), and motivated reasoning. They conclude that "people fall for fake news because they fail to think; not because they think in a motivated or identity-protective way."

Our work address several recent calls to action regarding Fake News and the spread of misinformation online (e.g. [6]) by creating a datasets that can be used to (i) analyze and detect fake news and (ii) be used in replication studies.

2.1 Detecting and Classifying Fake News

Looking at how fake news can spread in social media - and what to do about it - [7] describes a potential automated policy for determining when to have a human intervene and check a story being shared (to be used by Facebook/Twitter). They found that automated agents, attempting to pass on only good news and to fact check when appropriate, can actually amplify fake news and lend credibility to it. Their simulations offer insights into when fake news should be addressed and investigated by social media platforms.

In [13], Wang introduced a human-labeled and fact-checked dataset of over 12,000 instances of fake news, in contexts such as political debate, TV ads, Facebook posts, tweets, interview, news release, etc. Each instance was labeled for truthfulness, subject, context/venue, speaker, state, party, and prior history. Additionally, Wang used this new dataset to evaluate three popular learning-based methods for fake news detection, logistic regression, support vector machines, long short-term memory networks (Hochreiter and Schmidhuber, 1997), and a convolutional neural network model (Kim, 2014). Wang goes on to show that a neural network architecture that integrates text and meta-data was more accurate at identifying fake news than the text-only convolutional neural networks baseline.

[3] goes into detail about assessment methods from two approaches: linguistic cues and network analysis. The latter involves information we dont have in our dataset, namely incoming and outgoing links to the article and relevant topics which can be used to create a network. They break the former problem down into data representation and analysis. Their review suggests that the bag of words approach may be useful in tandem with other representations, but not individually. Instead, they suggest a parse tree, as well as using attribute:descriptor pairs to compare with other articles. They also theorize that using a Rhetorical Structure Theory (RST) analytic framework as the distance measure for clustering or other types of algorithms. Finally, they suggest using sentiment as a classifier, as there are often negative emotional undertones in deceptive writing.

The dataset we present in this work contains 283 fake news articles and 203 satirical stories. All articles are focused on American politics, were posted between January 2016 and October 2017, and are in English. The dataset contains the title, a link, and the full text of each article. For fake news stories, a rebutting article is also provided that disproves the premise of the original story.

Below, we describe the process of collecting and labeling stories and the characteristics of the data.

2.2 Collection and Annotation

We established several guidelines at the beginning of this project to guide the collection of fake news and satirical stories:

- A definition of Fake News - "Fake News" has many definitions, but we chose to use "Fake news is information, presented as a news story that is factually incorrect and designed to deceive the consumer into believing it is true." This eliminates legitimate news that may have a factual inaccuracy, satire, and opinion pieces from the scope of our definition.
- American Politics - While fake news is certainly not limited to American politics, we restricted our dataset to that domain to ensure a consistency of topics among all articles. This minimizes the chance that topical differences between fake and satirical stories could affect a classifier.
- Recent articles, posted after January 2016 - The logic here echoes that above; we wanted to ensure that the topics discussed in the articles were similar.
- Diverse sources - There are many fake news and satire websites online and each has hundreds, if not thousands, of articles. It can be tempting to build a large dataset from a few of these sources. However, we wanted to create a highly diverse set with articles from many different sources. Thus, we restricted our dataset to have **no more than five articles from a single website**. Again, this minimizes any chance that a classifier could pick up on the language or style of a certain site when building a model.
- No Borderline Cases - There is a spectrum from fake to satirical news, and this is a fact that we found was exploited by fake news sites. Many fake news websites include disclaimers at the bottom of their pages that they are "satire", but there is nothing satirical about their articles; they simply use this as an "out" from the accusation that they are fake. While working on the borderlines between satire and fake news will be interesting, there is a more pressing challenge to simply differentiate the most obvious cases of each. Thus, we decided our dataset would eliminate any articles that researchers believed fell in a grey area. The fake news stories are all factually incorrect and deceptive. The satirical stores are quite obviously satirical.

Researchers began by identifying fake news and satirical websites. While our goal was not to create a list of sites, this process served our purpose of creating a diverse set of sources. By enumerating websites first, researchers could take responsibility for all the articles taken from an existing site and work would not be duplicated. Each researcher did just that, claiming several fake news or satire sites and providing no more than five articles from each to the dataset. For each article, the researcher provided a text file with the full text and, if the story was a fake news story, they provided a link to a well-researched, factual article that rebutted the fake news story. That may be an article from a fact checking site that specifically debunks a story, or a piece of information that disproves a claim. For example, one fake news story claimed that Twitter banned Donald Trump from the platform. A link to Donald Trump's very active Twitter account proved that this story was false.

When the initial data collection was complete, each article was then reviewed by another researcher. They checked it against all the criteria listed above. Articles that could not be rebutted, that were off topic or out of the time frame, or that were borderline cases were eliminated from the dataset. Inter-rater agreement given by Cohen's kappa was 0.686 with an accuracy of 84.3%.

3 CLASSIFICATION

With a labeled dataset in hand, we could now address RQ1: Are there differences in the language of fake news and satirical articles on the same topic such that a word-based classification approach can be successful?

Table 1: Detailed accuracy measurements for classification of Fake News vs. Satire.

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.811	0.236	0.828	0.811	0.819	0.572	0.880	0.907	Fake
	0.764	0.189	0.742	0.764	0.752	0.572	0.880	0.847	Satire
Weighted Avg.	0.791	0.217	0.792	0.791	0.791	0.572	0.880	0.882	

Table 2: Distribution of theme pairs

Pair	C-H	H-S	C-S	H-R	D-H	C-R	C-D	F-H	C-F	R-S	H-P	F-R	D-S	F-S	D-R	C-P	D-P
Overall	19.50%	6.40%	4.10%	7.00%	2.90%	1.90%	1.20%	1.20%	1.40%	1.20%	1.20%	1.00%	0.40%	0.40%	0.40%	0.40%	0.20%
Satire	9.40%	3.40%	0.00%	7.40%	1.50%	1.50%	0.00%	0.50%	1.50%	1.50%	2.00%	1.50%	0.00%	0.00%	0.50%	1.00%	0.50%
Fake	26.90%	8.50%	7.10%	6.70%	3.90%	2.10%	2.10%	1.80%	1.40%	1.10%	0.70%	0.70%	0.70%	0.70%	0.40%	0.00%	0.00%

Our goal with this research question was not to do a deep linguistic analysis of the types of articles, but rather to understand if the basic word usage patterns differed substantially enough that it would allow for relatively accurate classification. With no additional analysis, we built a model to classify an article based only on the language it used. Each article was represented as a word vector with a class of Fake or Satire. We used Weka [5] to train a model using the Naive Bayes Multinomial algorithm and tested with 10-fold cross validation. We achieved accuracy of 79.1% with a ROC AUC of 0.880. Detailed accuracy measurements are shown in table 1. This high-performing model suggests strong differences in the type of language used between the fake news and satire in our dataset.

4 THEMES OF FAKE NEWS VS. SATIRE

After collecting the data, we explored the themes of our articles more deeply. Unlike [10] which looks at the mechanism for sharing fake news, we look at the types of content that are shared. Using a grounded theory open coding approach, our team developed a code book with major themes that appeared across the dataset. We settled on seven codes:

- H - Hyperbolic position against one person or group (e.g. Trump, Clinton, Obama, Islam, refugees)
  Example headline: "Obama Signs Executive Order Banning The Pledge of Allegiance In Schools Nationwide"
- F - Hyperbolic position in favor of one person or group (e.g. Trump, Clinton, Obama, Islam, refugees)
  BECAUSE TRUMP WON THE PRESIDENCY, FORD SHIFTS TRUCK PRODUCTION FROM MEXICO TO OHIO!
- D - Discredit a normally credible source
  MIT Researchers: Global Warming Data Is Complete Bunk
- S - Sensationalist crimes and violence
  George Zimmerman Found DEAD Just Hours After Bragging About Killing Trayvon Martin
- R - Racist messaging
  Trump Has Fired Muslim Sharia Judge Arrested And Charged
- P - Paranormal theories (e.g. Aliens, Flat Earth)
  Donald Trump Says The Earth Is Flat
- C - Conspiracy theories
  Hillary Clinton Busted in the Middle of Huge Pedophilia Ring Cover Up At State Department

Researchers could also leave an article unlabeled if none of the codes applied.

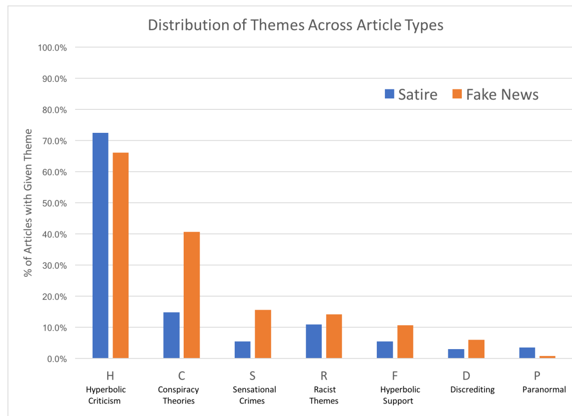

Figure 2: Distribution of themes across article types

Once the code book was finalized and researchers had trained on a subset of articles, they labeled each article with the appropriate codes. Articles could be labeled with multiple codes.

Overall, we found that hyperbolic criticism of a person - usually Trump, Obama, or Clinton - was the most common theme, appearing in more than 2/3rds of articles. Conspiracy theories were also common, appearing in almost 30%. While we were able to identify a common practice of attempting to discredit normally credible sources or the use of paranormal theories (e.g. aliens), these were relatively uncommon, both appearing in less than 5% of articles.

We then compared the distribution of themes within each article type to see if there were major differences in how the themes appeared in fake news vs. satire. Figure 2 shows a side-by-side comparison. The themes followed a generally similar distribution in each article type. However, conspiracy theories were notably more common in fake news stories than in satire. Descriptions of sensationalist crimes were also more common in fake news. Paranormal themes, though uncommon overall, were more apparent in satire than in fake news.

As noted above, articles could have more than one theme, and many did. Overall, 213 articles (43.8%) had multiple themes. This was much more common in fake news ($N = 157, 55.5\%$) than in satire ($N = 56, 27.6\%$). By far the most common pair of themes to appear together were Conspiracy Theories and Hyperbolic Criticism of a person. Examples of these topics include articles about President Obama's birth certificate, the accusation that "illegal aliens" cast 3 million votes in the last election, and that the murder of DNC

Table 3: Accuracy and ROC AUC for Classifying the Themes of Articles

Theme	Accuracy	ROC AUC
H	56.3%	0.583
C	80.1%	0.754
S	89.3%	0.750
R	89.8%	0.669
F	92.4%	0.610
D	96.3%	0.433
P	98.7%	0.672

staffer Seth Rich was orchestrated by George Soros (it was not). This combination appeared on 19.5% of all articles, 26.9% of fake news articles, and 9.4% of satire. It was the most common pairing for both types of article. Among satire articles, the only two other combination that appears on more than 10 articles was Hyperbolic criticism and racist themes. This pair occurred in 7.4% of satirical articles and 6.7% of fake news. Fake news also had popular pairings of Sensationalist Crimes appearing with Hyperbolic Criticism (8.5% of articles) and with Conspiracy Theories (7.1% of articles). Table 2 shows the full data for these theme pair distributions.

4.1 Themes and Classification

4.1.1 Using Themes To Distinguish Fake News from Satire. Our bag of words approach to classification described above was successful, but we wanted to see if including the themes of an article alongside the word vector would improve classification. To do this, we duplicated our word vector dataset and included the themes. We compared the results with this data to those achieved without the themes and found no significant difference in the accuracy or AUC. This suggests that the themes are not providing any real differentiating information that was not already detectable in the word vector itself. We hypothesize if we used these in a classifier that also considered real news, the themes may be more useful.

4.1.2 Detecting Themes from Language. Because some of these themes, which would be uncharacteristic of *real* news, are common in fake news and satire, we investigated if we could automatically determine whether an article contained a particular theme based on the words in the article. Using the same word vector as before, we built a model for each theme separately using only the word vector features.

As shown in table 3, we achieve high accuracy and medium or strong effects for many of the themes. While this is just a preliminary evaluation, it indicates that building useful theme classifiers may be possible and that, in turn, may be useful for understanding and detecting articles that are not "real" news.

5 DISCUSSION

The dataset we have created here, including full text of the articles, labels indicating their type (fake news or satire) and themes, and debunking articles for all fake news is available at https://github.com/jgolbeck/fakenews.git. We hope this dataset will be of use to the community of researchers studying fake news from a variety of perspectives.

Our initial thematic analysis offers insight that may be useful for both automated and qualitative analysis of fake news. Specifically, the fact that Hyperbolic Criticism and Conspiracy Theories are so common in fake news may mean that the presence of these themes may be useful for automatically detecting fake stories. Our preliminary results show that some of these themes can be detected quite accurately, and we believe this is an interesting space for future work.

One concern that arose in our initial discussions and that drove this work is the potential to conflate fake news and satire. Both are untrue stories with differences in intent. And while we found thematic similarities between the two, we also showed that a simple word vector classifier can strongly distinguish between the two. Again, there is much future work to be done here, but the good results we achieved on this dataset suggest that fake news detectors should also be able to tell the difference.

6 CONCLUSIONS

In this paper, we built a dataset of Fake News stories and Satire to serve as a contribution to the Fake News research community. The publicly available dataset includes full text of articles, links to the original stories, rebutting articles for fake news, and thematic codes. We included satirical articles because they, like fake news, are untrue, but vary in their *intention* and we showed preliminary results that indicate it is possible to automatically distinguish between the two types.

We hope this dataset is useful to the research community and that these preliminary results spark future work on understanding the nature of fake news and ways of fighting it.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. *Social media and fake news in the 2016 election.* Technical Report. National Bureau of Economic Research.
[2] Michael Barthel, Amy Mitchell, and Jesse Holcomb. 2016. Many Americans believe fake news is sowing confusion. *Pew Research Center* 15 (2016).
[3] Niall J Conroy, Victoria L Rubin, and Yimin Chen. 2015. Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology* 52, 1 (2015), 1–4.
[4] Don Fallis. 2014. A Functional Analysis of Disinformation. *iConference 2014 Proceedings* (2014).
[5] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. 2009. The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter* 11, 1 (2009), 10–18.
[6] Nicco Mele, David Lazer, Matthew Baum, Nir Grinberg, Lisa Friedland, Kenneth Joseph, Will Hobbs, and Carolina Mattsson. 2017. Combating Fake News: An Agenda for Research and Action. (2017).
[7] Yiangos Papanastasiou. 2017. Fake news propagation and detection: A sequential model. (2017).
[8] Gordon Pennycook and David G Rand. 2017. Who falls for fake news? The roles of analytic thinking, motivated reasoning, political ideology, and bullshit receptivity. (2017).
[9] Arne et al. Roets. 2017. Fake news: Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence* 65 (2017), 107–110.
[10] Victoria L Rubin, Yimin Chen, and Niall J Conroy. 2015. Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology* 52, 1 (2015), 1–4.
[11] Edson C Tandoc Jr, Zheng Wei Lim, and Richard Ling. 2017. Defining Fake News A typology of scholarly definitions. *Digital Journalism* (2017), 1–17.
[12] Chris J Vargo, Lei Guo, and Michelle A Amazeen. 2017. The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *new media & society* (2017), 1461444817712086.
[13] William Yang Wang. 2017. " Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. *arXiv preprint arXiv:1705.00648* (2017).