Assessing Online Media Reliability: Assigning a trust Metric Value and Detecting Fake News

Nicholas Snell, Jeremy Straub, Brandon Stoick, Terry Traylor, William Fleck

Department of Computer Science, North Dakota State University, Fargo, North Dakota, USA

ABSTRACT

Fabricated information is easily distributed throughout social media platforms and the internet. This allows incorrect and embellished information to misinform and manipulate the public in service of an attacker's goals. Falsified information – also commonly known as "fake news" – has been around for centuries. In modern day, it presents a unique challenge because of the difficulty of tracing news items origin, when spread electronically. Fake news can affect voting patterns, political careers, businesses' new product launches, and countless other information consumption processes. This paper proposes a method that uses machine learning to identify "Fake News" stories. The conditional probability that a story is fake is calculated, given the presence of feature predictors inside a news story. A concise summary of the qualitative methods used to study Fake News stories is presented. This is followed by a discussion of computational social science and machine learning methods that can be used to train and tune a classifier to detect fake news. Some of the main linguistic trends, identified in social media platforms, that are associated with fake news are identified. A larger integrated system that can be used to identify and mitigate the impact of falsified content is also proposed.

Key words: fake news, reliability, machine learning, web 2.0 technologies, data processing, internet, social sciences, system integration

INTRODUCTION

So-called 'Fake news' or falsified information, has been around for many years. However, the form of this content hasn't always been the same. In the past, falsified information was distributed principally from person to person and occasionally in mainstream media such as newspapers, radio or television. The problem has existed "since the first protohuman whispered the first malicious gossip"[1]. While the creation of deliberately fake content has been going on for a long time, many have been able to spot when something is fake, thus making it hard for one to spread falsified content through mass distribution methods.

More recently, new ways to distribute and obtain information have been developed which remove some of the content filtering mechanisms. Most young people today do not obtain their news through newspapers, nor do they ritually watch the news on television[2]. Instead, the internet and, in many cases, social media are commonly used to learn about happenings in the world. With these new ways to obtain information, came new systematic flaws that allow for falsified information to easily spread from person to person[3]. Many still do not know whether or not something is really true on the internet.

While many websites distribute news-like content (some of which is accurate, some of which is accidentally inaccurate and some of which is intentionally deceptive), one of the largest areas of deliberately deceptive content distribution is via social media, where falsified content spreads like proverbial wildfire. On social media, one may receive inaccurate or falsified content via his or her network of co-workers, friends and acquaintances. Some individuals may originate inaccurate and even deliberately misleading content. In many other cases, individuals will simply forward or link to content provided to them by others.

This is problematic because it has been shown in past studies that when 'fake news' is spread throughout friends and acquaintances on social media, many will tend to place greater levels of trust in the information and won't question it[4]. Real world consequences can and do occur because of this falsified content that has been spread via social media platforms.

One example of this is the 2016 United States presidential election. In one instance, a resident of Washington read an article that said that "Comet Ping Pong, a pizza restaurant in northwest Washington, was harboring young children as sex slaves as part of a child-abuse ring led by Hillary Clinton"[5]. After reading the article the individual, drove to the pizzeria and fired an AR-15 rifle, in an attempt to resolve the aforementioned situation[5].

This paper reviews the methods that can be used to combat falsified content. It also discusses the key characteristics that can be used to identify and classify fake (i.e., deliberately deceptive or manipulative) and real (i.e., legitimate reporting, potentially with accidental errors) news articles. This paper presents several methods that have been utilized to detect and potentially combat against falsified content. The creation of a neural network to analyze and classify news articles and the development of a data set to support its development, training and testing are also presented.

2. BACKGROUND

'Fake' news is designed to be as indistinguishable as possible from real news content with the intent that it is able to pass as 'real' news and manipulate the public. There are potentially multiple goals for the creation of this illicit content. For example, if someone wanted to create chaos or execute a crime, fake news could be used for this purpose. As in the case of the aforementioned Washington resident[5], one could use a fabricated story to potentially manipulate another individual to perform an act for them.

Since the 2016 United States presidential election, Wang contends that "the world has witnessed a growing epidemic of fake news"[6]. While some individuals have learned how to detect falsified content, at least some of the time, significant amounts of inaccurate news items persist. Fundamentally, the problem (and what is relied on by nefarious individuals) is that individuals have busy lives, and many do not take the time to critically evaluate or research the accuracy of what they have read.

According to information compiled by Allcott and Gentzkow[3], the majority of adults in the united states (62 percent) get news from social media[7] and popular 'fake' social media news stories are shared more via Facebook than "the most popular mainstream news stories"[3,8]. Further, Silverman and Singer-Vine contend that in many cases, readers of these 'fake' articles believe what they read [9].

While public skepticism of information sources appears to be increasing (trust in media networks has gone down since 2016 [10]), there is currently no indication that 'fake news' content creation and the problems that it causes will be disappearing anytime soon. This suggests that a system that acts as a 'fake news' detector and is able to identify falsified content while a user is surfing the web or browsing social media platforms would have significant utility. An effective system with this capability would potentially be an excellent asset in combating falsified content and reducing the efficacy of the use of deceptive content for the manipulation of the public.

3. KEY CHARCTERISITCS

This research explored the key differences between 'real' and 'fake' news articles through a qualitative analysis of manually collected and classified articles. The articles were assembled into a database that was used for this work. In particular, the work focused on how these attributes could be used as inputs to a custom naive Bayesian classifier or a neural network.

The first characteristic that differentiated the two was the articles' titles. Previous work showed that proper noun and verb usage could be indicative of the legitimacy of an article. Additionally, fake news titles often have words in caps. Both of these could be used as inputs for a classifier. Table 1 shows an example of a fake and real news title for thematically similar stories.

Table 1. Example of political bias in titles[11].

Title Topic	Fake News Article Title	Real New Article Title
Google Political Bias	"If Google is not stopped, the rogue search engine will STEAL every election for Democrats from this day forward" Source: News Target	"Google CEO Says He Leads 'Without Political Bias' In Congressional Testimony" Source: NPR

The second characteristic that was indicative of article status was its listed author. Analysis demonstrated that the same authors were associated with multiple fake news articles. By comparing an author name with an existing database that tracks authors and article legitimacy, another prospective input is attained. Figure 1 shows different authors associated with fake articles in the database used for this work and the number of articles they wrote.

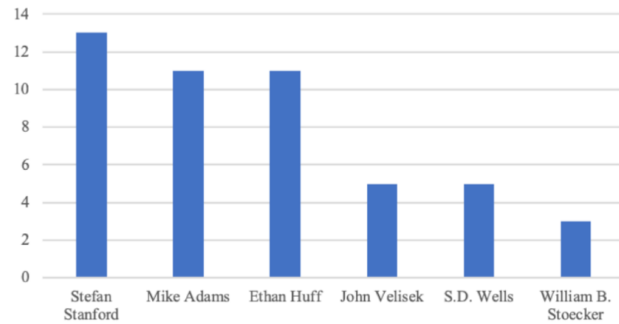

Figure 1. Fake Article Frequency for Certain Authors[11].

The third characteristic found through this qualitative analysis process was the evidence and source used and referenced. While further research is needed to find a way to create an input out of this article attribute autonomously, the links presented in articles and a lexical analysis of the evidence presented in articles could assist in determining their legitimacy. Table 2 provides an example of this.

Table 2. Accuracy of article content example[11].

Topic	**Article Evidence / Source Comparison**
Jerome Corsi's Involvement in Robert Mueller Prove	"Jerome Corsi, Friend of Roger Stone, Is in Plea Talks With Mueller" Real article source: New York Times "Mueller proves his own probe is BOGUS by focusing on Jerome Corsi while ignoring blatant lawbreaking by Hillary Clinton" Fake article source: Deepstate.news <u>Author Legitimacy</u> The real article provided biography links for both reporters that contained their credentials while the fake article provides no author credentials, only links to other articles written by them. These credentials give claims made by authors credibility. <u>Article Sourcing / Reference Legitimacy</u> The real article references The Washington Post while the fake article cited TheGatewayPundit.com and TheNationalSentinel.com as its sources. Politifact's Fake News Almanac classifies TheGatewayPundit.com as having some fake stories and TheNationalSentinel.com has been manually classified as having fake stories. <u>Presentation of Evidence and Claim Legitimacy</u> The real article made claims based on verifiable evidence while the fake article made unsubstantiated claims repeatedly.

The fourth article characteristic identified was its origin or publisher. Two potential inputs can be derived from this attribute. First, the publisher can be cross referenced to an existing database of publishers and their truth quality scores. Second, the article's text can be cross compared with the text of articles on other sites to see if it appears on one or more sites known to publish fake news. Out of 16 sites containing fake news in the database created for this project, 14 had duplicate fake articles on them from the other sites.

The fifth characteristic that was identified is the article's political perspective. Articles that had an alt-right or alt-left perspective tended to be less legitimate than articles with a liberal, centrist, or conservative perspective. Another input to a identification system, thus, could be based on a lexical analysis process that determines the political perspective of an article. Table 3 gives examples of how political perspective influences language presented in articles.

Table 3. Political policy example[11].

Political Topic	Alt-Left / Alt-Right	Conservative, Liberal, or Centrist Language
Green New Deal	"Bottom line: AOC's "Green New Deal" is a systemically corrupt plan to disembowel the entire functioning U.S. infrastructure, debilitate all farming capabilities and initiate a catastrophic food collapse, wipe out the entire middle class, confiscate all guns, private property and private businesses, and put in place a tyrannical police-state government that will never be able to be removed without a full-scale revolution." Source: Natural News	"A draft text of the plan circulating Congress includes a framework aimed at eliminating greenhouse gas emission from manufacturing and agriculture and "dramatically" expanding energy sources to meet 100 percent of power demand through renewable sources. The proposal describes this effort as a "historic opportunity to virtually eliminate poverty in the United States and to make prosperity, wealth and economic security available to everyone participating in the transformation." The proposal also calls for a job-guarantee program offering a "living wage job to every person who wants one," a plan to aid workers affected by climate change, universal health care and basic income programs, among other things." Source: Fox News

Further research is needed to transform some of these characteristics into useable inputs. However, they demonstrate that by using key characteristics, one can differentiate between real and fake articles with a higher probability.

4. NEURAL NETWORK DEVELOPMENT

Using the analysis described in the previous section, the team worked to create a neural network that was able to identify fake or real news articles. This network was trained from manually classified articles and tested via an additional set of manually classified articles. The neural network operated by comparing parts of speech. It analyzed multiple articles and found similarities between fake and real articles. One of the characteristics that separated the articles was that different nouns, verbs, adjectives, and adverbs were being used in fake versus real articles. However, fake articles had similar characteristics to each other and so did real articles. Table 4 shows the parts of speech that were implemented in the neural network.

Table 4. Part of speech symbology[12].

Key (Part-of-Speech)	Description
#	pound sign (currency marker)
$	dollar sign (currency marker)

Key (Part-of-Speech)	Description
JJ	Adjective
JJR	Adjective, comparative

Key (Part-of-Speech)	Description
RP	Particle
SYM	Symbol (Scientific)

''	close quote
''	close quote
(	open parenthesis
)	close parenthesis
,	comma
.	period
:	Colon
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner
EX	Existential there
FW	Foreign word
IN	Preposition/sub. conjunction

JJS	Adjective, superlative
MD	Modal
NN	Noun, singular or masps
NNP	Proper noun, singular
NNPS	Proper noun plural
NNS	Noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative

TO	to
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund/present participle
VBN	Verb, past participle
VBP	Verb, non-3rd ps. sing. present
VBZ	Verb,3rd ps. sing. present
WDT	wh-determiner
WP	wh-pronoun
WP$	Possessive wh-pronoun
WRB	wh-adverb

Using these parts of speech, and two-hundred and seventy-seven of the database's manually classified articles, the neural network was able to achieve an accuracy of 82% in detecting other articles that had been manually classified as real or fake articles[12]. Figure 2 demonstrates the impact of the training data size on neural network performance. It has, thus, been shown that parts of speech and a properly trained neural network are able to achieve the highest level of accuracy, compared to using N-grams. The N-grams approach involves having the neural network identify the most common N-grams throughout the data set. The most common N-grams are then tested on the data set.

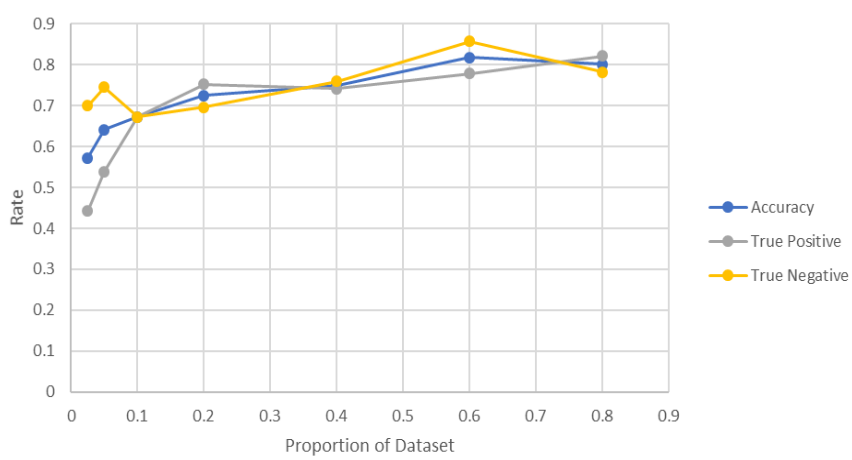

Figure 2. Article accuracy and true positive and negative rates compared to the proportion of the dataset used for training purposes[13].

Comparing the N-grams and the parts of speech approaches shows that, based on the data collected, the parts of speech approach was more accurate than the N-grams approach. The overall accuracy using N-grams was 60%. This would suggest that the parts of speech neural network may be able to reliably detect key differences between real and fake articles.

These results present a number of prospective topics for future work. First, a hybrid technique which combined the N-grams and parts of speech approaches could be evaluated. Additionally, adding other key characteristics that are associated more strongly with either fake or real articles, could increase the accuracy of the neural network. In support of these objectives, a larger data set will need to be created. Ideally, this would include thousands of articles that have been manually classified as fake or real.

5. PROFANITY IDENTIFICATION

Another approach for identifying fake news was also evaluated. This approach used the same parts of speech paradigm, described above. However, the classifier was also configured to identify profanity present in articles, as well. This was tested to see how well profanity detection alone would be able to detect fake news articles. The parts of speech, approach, was also tested by itself, as well. Finally, the parts of speech and profanity approaches were combined together.

While using just the parts of speech approach, the classifier had an overall accuracy rate of 62%. After optimizing the characteristics utilized, the accuracy was increased to 72%. The profanity approach initially had an accuracy of 59%. With optimization, this increased to 65%. The combination of the two approaches resulted in the highest accuracy of 76%.

While these approaches, on their own, did not outperform the neural network, they may be synergistic. The approaches, generally, performed better than the N-gram frequency approach. However, they were still outperformed by the neural network. Future work, in this area, will focus on identifying techniques for combining these techniques together to enhance overall performance.

6. CONCLUSIONS AND FUTURE WORK

This work has shown that several techniques can be used to identify fake versus real news and performed an initial characterization of the efficacy of several of these techniques. The potential efficacy of combining techniques has also been demonstrated; however, this is an area where additional work is required as more potential combinations may exist. Additionally, future work should investigate the potential increase in classifier accuracy that can be attained by adding additional key characteristics.

Problematically, the adversary is not standing still and numerous individuals and groups are continuing to research the best way to fool the public with misinformation. Because of this, fake news is becoming more difficult to detect each day. Fake news articles explicitly try to make the reader believe and spread their content.

The techniques that have been proposed to combat ‘fake news’ (such as neural networks and classifiers) can also be used to create or verify the creation of this illicit content, making it more difficult to detect whether or not the article is fake. Overall, the automated detection of ‘fake news’ is difficult and may not be a long-term solution as adversaries adapt their creation approaches to confound identification techniques. Fundamentally, an approach based on identifying key details of articles and fact checking may be the most prudent, in the longer term.

Further work is needed to be able to fully detect all forms of fake and real news. A larger manually classified database is needed for this and increasing the diversity of items and languages would aid future work, as well. Future research can also focus on identifying new key characteristics that can be incorporated into the classifiers. This will be aided by the collection of more data and furthering the analyzing the characteristics of fake and real articles.

ACKNOWLEDGMENTS

This work was partially supported by the U.S. National Science Foundation (award # 1757659). Thanks is given to the members of the NDSU Fake News Research Group team that helped generate the classified dataset that was used in this work.

CONFLICTS OF INTREST

The author declares that there is no conflict of interest regarding the publication of this article.

REFERENCES

[1] Feature, N., Waldrop, M., Natl, P., Sci, A., Gillmor, D., Entrepreneurship, D. M., Gillmor, D. and Co, N., "Erratum: News Feature: The genuine problem of fake news (Proceedings of the National Academy of Sciences of the United States of America 114 (12631-12634) DOI: 10.1073/pnas.1719005114)," Proc. Natl. Acad. Sci. U. S. A. **114**(52), E11333 (2017).

[2] Marchi, R., "With Facebook, Blogs, and Fake News, Teens Reject Journalistic "Objectivity"," J. Commun. Inq. **36**(3), 246–262 (2012).

[3] Allcott, H. and Gentzkow, M., "Social Media and Fake News in the 2016 Election," J. Econ. Perspect. **31**(2) (2017).

[4] Ordway, D.-M., "What research says about how bad information spreads online," Journalist's Resour. (2018).

[5] Kang, C. and Goldman, A., "In Washington Pizzeria Attack, Fake News Brought Real Guns," New York Times (2016).

[6] Wang, W. Y., "'Liar, Liar Pants on Fire': A New Benchmark Dataset for Fake News Detection," 422–426 (2017).

[7] Gottfried, J. and Shearer, E., "News Use Across Social Media Platforms 2016," Pew Res. Cent. (2016).

[8] Silverman, C., "This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook," BuzzFeed News (2016).

[9] Silverman, C. and Singer-Vine, J., "Most Americans Who See Fake News Believe It, New Survey Says," BuzzFeed News (2016).

[10] Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J. and Zittrain, J. L., "The science of fake news," Science (80-.). **3**(9), 1094–1096 (2018).

[11] Fleck, W., Snell, N., Traylor, T. and Straub, J., "Development of a 'Fake News' machine learning classifier and a dataset for its testing," Disruptive Technol. Inf. Sci. II **11013**, 9, SPIE (2019).

[12] Kalvoda, B., Stoick, B., Snell, N. and Straub, J., "Evaluation of Algorithms for Fake News Identification," Proc. SPIE (2019).

[13] Stoick, B., Snell, N. and Straub, J., "Fake news identification: a comparison of parts-of-speech and N-grams with neural networks," Big Data Learn. Anal. Appl., 12, SPIE (2019).