Fake News Detection Using Knowledge Verification and Natural Language Processing

Muhtasim Mahmud
Dept. of CSE
BRAC University
muhtasim.mahmud@g.bracu.ac.bd

Prabal Kumar Chowdhury
Dept. of CSE
BRAC University
prabal.kumar.chowdhury@g.bracu.ac.bd

Mohammad Rahat Khan
Dept. of CSE
BRAC University
md.rahat.khan@g.bracu.ac.bd

Md. Farhadul Islam
Dept. of CSE
BRAC University
md.farhadul.islam@g.bracu.ac.bd

Md Sabbir Hossain
Dept. of CSE
BRAC University
md.sabbir.hossain1@g.bracu.ac.bd

Annajiat Alim Rasel
Dept. of CSE
BRAC University
annajiat@gmail.com

***Abstract*—As a direct result of the campaign for the presidential election in the United States in 2016, the term "fake news" rose to prominence on a global scale. It is related to the practice of spreading false information and/or information that is misleading in an effort to influence the opinion of the general public. The process in question is referred to as disinformation. One of the primary weapons employed in information warfare, which is recognized as a developing risk to cybersecurity, is this tool. In this paper, we investigate the use of "fake news" as a method of disinformation. In this article, we provide a review of previous attempts to define and automate the process of detecting "fake news." We propose a new, more malleable definition of "fake news" based on the degree of relative bias and the accuracy of reported facts. In this paper, we propose a novel framework for the detection of fake news that makes use of a machine learning model and is based on the definition that we have proposed.**

***Index Terms*—NLP methods, Machine Learning, topic modeling, Natural Language Processing, text segmentation, word embedding**

Introduction

What exactly is the "false news"? In the context of information warfare, one of the goals of this study is to define what exactly constitutes fake news and then to offer an automated system for detecting fake news based on this definition. This is one of the targets of this work. The information warfare danger was identified as an emerging cybersecurity concern in the Symantec 2019 Internet Threat Report. In the context of information warfare, the fabrication of assertions for the purpose of manipulation and their subsequent dissemination are both common practices.

Warfare has been around for as long as information warfare has, so the two are not mutually exclusive. In his book titled "The Art of War," which was written in the 5th century BC, Sun Tzu stated that "All warfare is based on deception." The ties that bind politics and warfare are extremely strong. After Russia's defeat in the elections in 2016, the former President of Bulgaria, Rosen Plevenliev, issued a warning that the Russian government is attempting to influence the outcome of the elections in Bulgaria. In an article published in 2017 by the Wall Street Journal, Parkinson and Kantchev made the assertion that a Bulgarian security agency had, according to their claims, obtained a document from a Russian spy that outlined a Russian campaign to interfere in the elections that were being held in Bulgaria. "The document offered advice on how to burnish the candidate's image by planting stories with Moscow-friendly news outlets," which was the advice that was offered in the document. In order to generate the greatest possible impact and ultimately serve as election talking points for the party, the stories were to be carefully coordinated and initially published in alternative blogs before being distributed widely across mainstream media outlets. There was neither confirmation nor denial made by the Bulgarian security agency regarding the detection of Russian interference. Russia has denied all allegations and suggested that those allegations are fabricated news.

If a piece of information does not have any concrete evidence to back it up, then the factual accuracy of that information cannot be established. What some people consider to be "breaking news" may be considered fake news by others. Therefore, in order to develop an appropriate computational solution, it is necessary to first develop a universal definition. In Section 2 of this paper, we examine a variety of definitions of fake news and propose a new definition based on absolute factual accuracy and relative reliability of the source. This definition is based on absolute factual accuracy and relative reliability of the source. In Section 3, we discuss the work that has been done previously to automate the process of detecting fake news. In Section 4, we present an original framework for the detection of fake news. This framework makes use of both manual and automated knowledge verification as well as stylistic characteristics. In Section 5, we will go over the results of our study. In Section 6, we will discuss the work that will be done in the future, and in Section 7, we will draw conclusions.

Problem Statement

The difficulty lies in defining "fake news." People have a propensity to label as "fake" any piece of news that does not

align with their views or the agendas they wish to pursue. Edson and colleagues presented a classification system for various types of fake news. They analyzed 34 distinct papers on fake news that were published between 2003 and 2017, and based on the definitions provided in those papers, they developed a framework for the various categories of fake news. In the context of information warfare, each of these various types, which include propaganda and advertising/public relations, can be utilized to influence public opinion regarding a specific subject. In this article, our primary focus is on the use of "fake news" as a method of disinformation. According to Lazar et al., "fabricated information that mimics news media content in form but not in organizational process or intent" is the definition of "fake news." The findings of Horne and Adali's research, on the other hand, suggest that there are discernible variations in format, particularly when it comes to the titles of fake news articles. According to the definition offered by Lazar et al., the article published in the Wall Street Journal that depicts Russian interference in Bulgarian elections is likely to be considered "fake news." This is due to the fact that the primary source that can verify the factual accuracy of the claims made in the article, namely the Bulgarian secret service, is refusing to do so. This gives the impression that the story was made up.

How likely is it then that this account is totally made up? It is equally likely that it is false and accurate, given that its veracity cannot be established either way. Since the Wall Street Journal enjoys a solid reputation, this possibility increases, albeit probably not by much. After all, reputable media outlets have been known to publish unreliable reporting in the past.

In 2002, an Associated Press reporter was let go after it was discovered that he had lied about his sources and facts for at least two years. Even if a piece of information is published by a reputable news agency with stringent organizational processes in place, that does not make it accurate. However, if the contrary cannot be demonstrated, information from a generally trustworthy source is more likely to be accurate. The boost is proportional to the source's general credibility.

What other elements might increase the possibility? Is there a report that provides even more evidence? Regarding the Wall Street Journal's report on Russian meddling in overseas elections, yes. As early as 2008, a Czech secret 224 M. D. Ibrishimova and K. F. Li service agency pointed to evidence that contradicted Russian claims that they did not meddle in the politics of other countries. Unambiguously, "operations of intelligence services of the Russian Federation... are by far the most active ones in our territory," as the Czech Security Information Service put it. Moreover, the Czech Security Information Service issued a warning that Russia has revived and repurposed "active measures" tactics from the Soviet era. It is clear that this evidence does not conclusively prove the claim made in the Wall Street Journal article about Russian interference in the Bulgarian election. Given the similarities between Bulgaria and the Czech Republic (which used to be part of Czechoslovakia), the likelihood that this statement is based on fact does increase slightly.

The line between a fabricated report and a report that is factually accurate becomes blurry in situations in which it is difficult to establish the factual accuracy of a piece of information. This is the reason why the website that checks facts was created. When evaluating political claims, Politifact employs a tool called the Truth-o-Meter. Within the scope of this paper, we present "the fake news spectrum." It takes into account reports that cannot be factually verified or disputed, even by professional fact checkers, and even if they come from reliable sources. This is the case regardless of whether or not the reports originate from credible sources.

Specifically, if a report or claim's veracity can't be confirmed or denied, there's a 50% chance that it's made up. This percentage drops slightly in relation to the overall reliability of the source if it is known to be credible or if a similar claim appears in a credible source (s). Consideration can also be given to whether or not the same or similar claims can be found in other sources that adhere to the same or similar standards of organization. There is also the option of considering whether the claim seems to be grounded in fact or opinion. Based on these findings and our prior work in incident classification, we present a machine learning model in Section 4. The next part of this article will focus on the current techniques used to identify fake news.

Research Objectives

Fake news is an unwelcome phenomenon that can be found in today's society. It's possible that this will lead to confusion and misunderstanding regarding significant social and political issues. It's possible that reading fake news could be bad for your health. The dissemination of false information makes it more difficult for people to recognize the truth. Fake news stories create the appearance of legitimate news sites by utilizing technology and social media. You could be the target of organizations and political groups using advertisements designed to look like news stories. While hackers create multiple social media accounts with the help of bots, which are pieces of software, they then use those accounts to spread false information. Simply because it appears to have been shared by a large number of people, this can give the impression that a made-up story is true.

Because of this, our primary objective is to determine, through the application of NLP techniques, whether the news is real or fake. The people will benefit from being aware of the honest news.

Literature Review

[1] Establishing the factual accuracy of a claim is crucial in determining whether it is "fake news" by most definitions of "fake news". Several manually generated tools for identifying the factual accuracy of a given claim exist. However, such approaches rely on humans who may or may not be objective. Additionally, evidence to support the factual accuracy of the claim might not be available as in the case of the Wall Street Journal article on Russian interference in Bulgarian elections.

Various automated fact- checking methods have also been proposed.

[2] Thorne and Vlachos provide a comprehensive survey of existing automated fact-checking methods. One method uses Recognizing Textual Entailment (RTE) where "RTE-based models assume that the textual evidence to fact check a claim is given" as part of the claim. Another method relies on checking a claim against a knowledge database of proven facts. Yet another method attempts to verify claims by profiling their source and implementing "credit history" of individual sources. Thorne and Vlachos identify issues with all of these methods. Namely, RTE-based methods fail when there is no evidence to support the claim, the "database of proven facts" methods fail when presented with novel claims, and the

[3] "profiling the source" methods fail when the source is new. There is an even greater issue associated with fact-checking political news. As Coleman suggests, "Political truth is never neutral, objective or absolute" [4]. Even computational giants such as Google could not tackle this issue and had to shut down their fact-checking tool out of concerns over inaccuracy [22]. Although individually these methods all have weakness, it is worth studying different combinations of them.

[4] In addition to the factual accuracy of a claim, researchers also studied extensively whether its stylistic form can reveal if it is fabricated. Oshikawa et al. provide a comprehensive survey on methods using Natural Language Processing (NLP). Another survey by Groendahl and Asokan asserts that "while certain linguistic features have been indicative of deception in certain corpora, they fail to generalize across divergent semantic domains". However, they do admit that "some results have been replicated in multiple studies".

[5] Groendahl and Asokan focus primarily on fake news detection methods at the document level as opposed to at the level of the news title. Their survey does not include the work of Horne and Adali who show that the title of a news article is often sufficient to detect if it is fake news. An ordinary news title is typically written in a way to entice the reader to read the entire article. Political disinformation campaigns' main purpose is to spread their narratives to as many people as possible, including to people who do not like to read much. A fake news title as a tool for political disin- formation is typically a summary of the entire article.

[**?**] Horne and Adali explore a wide range of syntactic, psychologic, and stylistic features for machine learning models using several different datasets of political news and come to the conclusion that fake news titles are generally longer, have "significantly fewer stop-words and nouns, while using significantly more proper nouns and verb phrases". Of the many different features they test, Horne and Adali identify the top 4 features for classifying fake news' titles: "the percent of stopwords, number of nouns, average word length, and FKE readability." They achieve an accuracy of about 70%.

[7] Researchers have also studied hybrid frameworks that employ natural language features and verification features. Conroy et al. survey the various different fake news detection technologies and outline a hybrid framework, which uses content cues (natural language processing tools to detect deceptive language) as well as information about the network, and source verification. However, as Tschiatschek et al. point out, "it is difficult to design methods based on estimating source reliability and network structure as the number of users who act as sources is diverse and gigantic (e.g., over one billion users on Facebook); and the sources of fake news could be normal users who unintentionally share a news story without realizing that the news is fake".

[8] Tschiatschek et al. propose a system that relies on crowd-sourcing fake news detection using trusted users to flag potentially deceptive content, which is then forwarded to professional fact-checkers for further investigation. However, there are ethical implications to be considered when profiling users. Zhang and Ghourbani provide a comprehensive survey on fake news detection, "an exhaustive set of hand- crafted features, and the existing datasets for training supervised models" and also acknowledge the importance of having a clear definition of fake news [30].

METHODOLOGY

A hybrid framework has been proposed by us for the detection of fake news. This framework repurposes the machine learning model for incident classification that was described in the previous paragraph. Our model for classifying incidents is comprised of five NLP features in addition to three knowledge verification features, which take the form of questions concerning the extent, the breadth, and the dependability of the source of the incident. The ternary responses to these questions are obtained from the user after they have submitted a textual incident report, and the system is able to independently verify these answers.

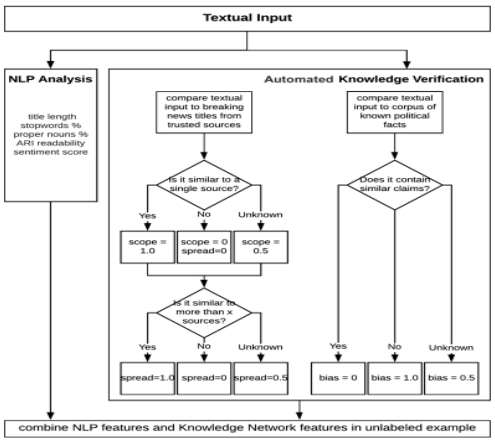

Fig. 1. Automated knowledge verification

For the detection of fake news titles, we propose using the same general model where Stopword density, proper noun to common noun ratio, title length, ARI readability, and Google NLP API-based overall text sentiment are the five Natural Language Processing (NLP) features.

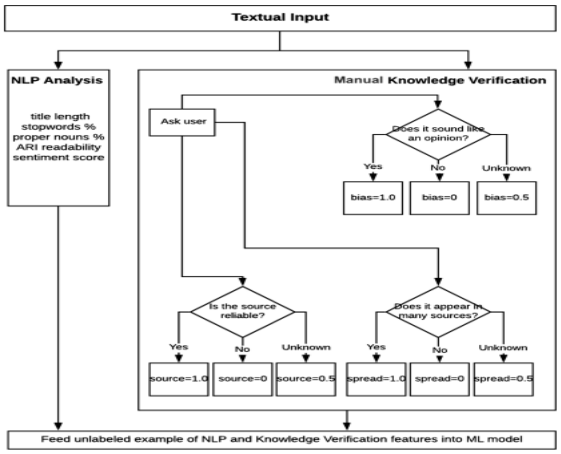

Fig. 2. Manual knowledge verification

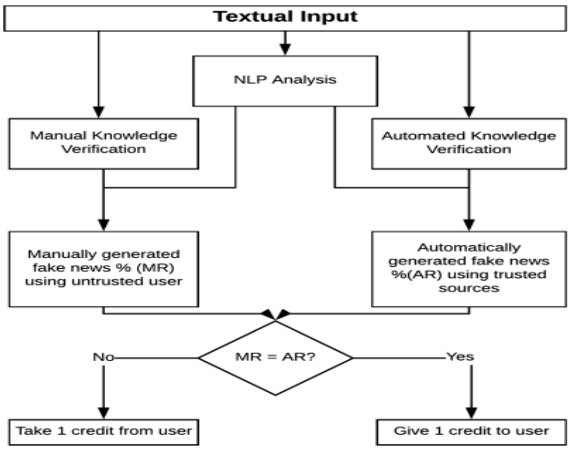

Fig. 4. Verifying user's ability to verify knowledge stated in news title

The three knowledge verification features are ternary answers to the questions of whether or not the title is comparable to a recent title in a reliable source, whether or not similar titles appear in more than x sources, and whether or not the title appears to originate from on facts or opinions.

Unlike Horne and Adali's approach, ours makes advantage

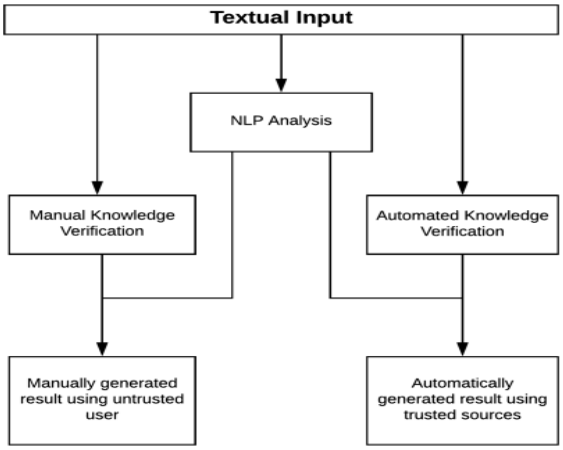

Fig. 3. Simultaneous manual and automated knowledge verification

of knowledge verification characteristics in addition to natural language processing ones. Additionally, we employ the automated readability index (ARI), a somewhat alternative readability measure for English texts, and include an additional element linked to sentiment analysis into our model. Since Horne and Adali found that false news tends to increase its use of proper nouns while decreasing its usage of generic nouns, we opted to employ the ratio of proper nouns (entities) to generic nouns instead.

Our method is a hybrid structure, like the one proposed by Conroy et al. Our suggested verification characteristics, on the other hand, are distinct in that they compare the claim to comparable claims made by a reliable source in the recent past. The claim's credibility is evaluated by looking at how widely it has been shared on reputable websites. At last, we can determine whether or not the allegation is founded on truth or opinion.

Multiple applications exist for our system. With the use of reliable sources and datasets containing established political facts, such as FEVER: a large-scale dataset for Fact Extraction and Verification, proposed by Thorne et al. [33] or ClaimBuster, developed by Hassan et al. [34], the verification process may be automated as shown in Fig. 1.

Figure 2 demonstrates a manual method of generating a verification process by asking the user questions about the information's credibility, its dissemination, and the user's own perceptions of any potential bias. We recommend including users in the investigation by having them respond to the verification questions shown in Fig. 2. In this approach, the proportion of false news results (MR) may be calculated by hand. Behind the scenes, the same trained model is also used to automatically provide a percentage result (AR) for bogus news, as seen in Fig. 1. Following this, the system does a comparison of MR and AR, as illustrated in Fig. 3. In cases where the two sets of findings strongly disagree, or if both sets of findings point to the claim being false, the claim is sent to an expert fact checker for additional study. As an alternative, the method may be used to evaluate the skill of social network members in spotting bogus news. One such use case is shown in Figure 4. This user-verification method is distinct from DETECTIVE, which was described by Tschiatschek et al. [29] since it does not depend on people to manually identify news they believe to be false. Instead, it relies on the user to verify the information while also doing its own automatic check. It generates two labeled examples to be used in training a machine learning model, and then evaluates the model's outputs for the two sets of inputs.

Result Analysis

We used the Google NLP API [40] for the natural language processing analysis and the News API [41] for the fact-checking aspects of the system shown in Fig. 1. As our

machine learning model for false news identification, we used the logistic regression model reported in our study on event classification [31, 32]. Kaggle's Fake News Dataset [42] was used for both training and testing. To ensure the accuracy of our approach, we created a second test dataset consisting of genuine news headlines published by many credible news sites that adhere to strict journalistic principles. As shown in Table 1, our technology has evaluated the likelihood that the headlines in question are fictitious.

Table 1. The probabilities that real news titles are fake as determined by our system

Real news	Probability of being fake news
US expected to allow lawsuits against foreign companies doing business in Cuba	10%
Centre-right opposition wins Estonia election as far-right populists make inroads	30%
Vatican to open secret archives of wartime Pope Pius XII	12%
Netherlands summons ambassador to Iran amid diplomatic spat	39%
Residents evacuate as flash floods hit southern Afghanistan	39%
Rep. Dingell on what Cohen's testimony means for future investigations of Trump	45%
OxyContin drug maker mulls bankruptcy due to myriad lawsuits	37%
Fox News confirms exit of Eboni K. Williams	26%
AG William Barr not recusing himself from Russia probe, official says	5%

To test our algorithm, we created a fresh dataset of fake news by removing key phrases from real news headlines and running each sample through it. Our system's determination of the likelihood that each of these false news headlines is phony is shown in Table 2.

Table 2. The probability that fake news titles are determined as fake

Fake news	Probability of being fake news
US expected to disallow lawsuits against foreign companies doing business in Cuba	80%
Centre-right opposition doesn't win Estonia election as far-right populists make	95%
Vatican won't open secret archives of wartime Pope Pius XII	71%
Netherlands doesn't summon ambassador to Iran amid diplomatic spat	68%
Residents don't evacuate as flash floods hit southern Afghanistan	80%
Rep. Dingell on what Cohen's testimony doesn't mean for future investigations of Trump	84%
OxyContin drug maker doesn't mull bankruptcy due to myriad lawsuits	71%
Fox News denies exit of Eboni K. Williams	74%
AG William Barr recusing himself from Russia probe, official says	40%

Future Work

The study proposes an automated knowledge verification system that mainly utilizes the concept of similarity. Semantic similarity is very important. There have been a number of cutting-edge algorithms developed specifically for this purpose over the last 20 years, including latent semantic analysis, latent relational analysis, explicit semantic analysis, temporal semantic analysis, and distributed semantic analysis. The next step in implementing our suggested framework for automated false news identification would be to conduct an evaluation of available algorithms and choose the one that proves most effective.

Conclusion

In this paper, we provide a working definition of "fake news" within the larger framework of "information warfare." In this article, we examine the history of attempts to automate the process of detecting fake news and briefly discuss the societal and political effects of this phenomenon. Based on our prior work in automating incident classification, we conclude that a framework that combines source and fact verification with NLP analysis is the most promising approach for detecting fake news.

References

[1] Symantec Internet Security Threat Report 2019. https://www.symantec.com/en/sg/security- center/threat-report. Accessed 07 Mar 2019

[2] Tzu, S.: The Art of War. China. 5th Century BC

[3] Kaiser, D.: Politics and War: European Conflict from Philip I1 to Hitler, pp. 149 and 172. Harvard University Press, Cambridge (1990)

[4] Coleman, S.: The elusiveness of political truth: from the conceit of objectivity to intersubjective judgement. Eur. J. Commun. 33(2), 157–171 (2018). https://doi.org/10.1177/ 0267323118760319

[5] Fried, D., Polyakova, A.: Democratic defence against disinformation.

[6] Correra, G.: Bulgaria warns of Russian attempts to divide Europe. BBC (2016). https://www. bbc.com/news/world-europe-37867591. Accessed 19 Feb 2019

[7] Parkinson, J., Kantchev, G.: Wall Street J. (2017). https://www.wsj.com/articles/how-does- russia-meddle-in-elections-look-at-bulgaria-1490282352. Accessed 19 Feb 2019

[8] Vassilev, I.: Russia's elections manual for BSP and the role played by Bulgarian sociological agencies. Bulgaria Analytica (2017). http://bulgariaanalytica.org/en/2017/03/31/

[9] Sputnik News: While West Claims Moscow is Meddling in Bulgarian Elections, Sofia Blames Turkey (2017). https://sputniknews.com/europe/201703251051961027-bulgarian-election-meddling-accusations/, Retrieved 04.04.2019

[10] Edson, C., Tandoc, Jr., Lim, Z.W., Ling, R.: Defining fake news. Digit. J. 6(2), 137–153 (2018). https://doi.org/10.1080/21670811.2017.1360143

[11] Horne, B.D., Adali, S.: This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. arXiv preprint arXiv:1703.09398 (2017)

[12] Wang, P., Angarita, R., Renna, I.: Is this the era of misinformation yet: combining social bots and fake news to deceive the masses. In: Companion Proceedings of the Web Conference 2018 (WWW 2018). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, pp. 1557–1561 (2018)

[13] Factcheck. https://www.factcheck.org/about/our-mission/. Accessed 02 Apr 2019

[14] Snopes. https://www.snopes.com/about-snopes/. Accessed 02 Apr 2019

[15] PolitiFact. https://www.politifact.com/truth-o-meter/article/2018/feb/12/principles-truth-o- meter-politifacts-methodology-i/. Accessed 02 Apr 2019

[16] Thorne, J., Vlachos, A.: Automated fact checking: task formulations, methods and future directions. In: Proceedings of the 27th International Conference on Computational Linguistics (COLING 2018) (2018)

[17] O'Brien, N., Latessa, S., Evangelopoulos, G., Boix, X.: The language of fake news: opening the black-box of deep learning based detectors. In: Workshop on AI for Social Good, NIPS 2018 (2018). http://hdl.handle.net/1721.1/120056

[18] Rashkin, H., Choi, E., Jang, J.Y., Volkova, S., Choi, Y.: Truth of varying shades: analyzing language in fake news and political fact-checking. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2931–2937 (2017)

[19] Oshikawa, R., Qian, J., Wang, W.Y.: A Survey on Natural Language Processing for Fake News Detection. Computation and Language (2018). arXiv:1811.00770v1 [cs.CL]

[20] Gröndahl, T., Asokan, N.: Text Analysis in Adversarial Settings: Does Deception Leave a Stylistic Trace? Computation and Language (2019). arXiv:1902.08939v2 [cs.CL]

[21] Conroy, N.J., Chen, Y., Rubin, V.L.: Automatic deception detection: methods for finding fake news. In: The Proceedings of the Association for Information Science and Technology Annual Meeting (ASIST 2015), 6–10 November, St. Louis (2015)