Detecting Online Fake News by Considering Stance of Headline+Body+Text on Online News Article

Vicky Kumar Prasad, Umesh Saraswat, Professor Priyanka Meel

Department of Information Technology
Delhi Technological University(Formerly Delhi College of Engineering), Delhi – 110042 , India

***Abstract*-** Due to the increasing popularity of online social networks, fake news for various commercial and political purposes has been appearing in large numbers and widespread in the online world. It has adversely affected both online social network systems as well as offline communities and conversations. An important goal in improving the trustworthiness of information in online social networks is to identify the fake news regularly. This paper aims at investigating the principles, methodologies and algorithms for detecting fake news articles and title from online social networks as well as news article from daily sources and evaluating the corresponding performance. In order to address this issue we propose a fake news detection of online available news material (texts or title) by building a model that can classify articles into either Fake or Real with high accuracy. A LSTM(Long Short Term Memory) based approach is used to find a latent pattern in the news title as well as in its body(content) so that the text can be classified as the Fake or Real news instance. Finally both title and content can be combined to get a highly accurate news instance detector.

***Index Terms*-** Fake news articles, fake news detection, deep learning , LSTM.

I. INTRODUCTION

"Fake news" is not a new thing. The root of it existed in the society long back, but the damage done by it to mankind made it a serious issue to be solved by the research community. In 1835, New York Sun published a series of articles about "the discovery of life on the moon". Soon the fake stories were printed in newspapers in Europe. Similarly, fake news widely exists in our daily life and is becoming more widespread following the Internet's development. The change to neoteric methods of news consumption in recent times has brought the issue of fake news and misinformation to the forefront of discussion. As thousands of new news articles proliferate social media networks every day, with each having neither a credibility nor a validation check, an ecosystem fuelled by misinformation (the inadvertent sharing of false information) and disinformation (the deliberate creation and sharing of information known to be false) has been established.

The term fake news has become a buzz word these days. However, an agreed definition of the term "fake news is still to be found. It can be defined as a type of yellow journalism or propaganda that consists of deliberate misinformation or hoaxes spread via traditional print and broadcast news media or online social media. These are published usually with the intent to mislead in order to damage a community or person, create chaos, and gain financially or politically. Since people are often unable to spend enough time to cross-check reference and be sure of the credibility of news, automated detection of fake news is indispensable.

Detecting misinformation on social media is an extremely important but also a technically challenging problem. The difficulty comes in part from the fact that even the human eye cannot accurately distinguish true from false news; for example, one study found that when shown a fake news article, respondents found it "'somewhat' or 'very' accurate 75% of the

time”, and another found that 80% of high school students had a hard time determining whether an article was fake . In an attempt to combat the growing misinformation and confusion, several fact-checking websites have been deployed to expose or confirm stories (e.g. snopes.com). These websites play a crucial role in combating fake news, but they require expert analysis which inhibits a timely response. As a response, numerous articles and blogs have been written to raise public awareness and provide tips on differentiating truth from falsehood. While each author provides a different set of signals to look out for, there are several characteristics that are generally agreed upon, relating to the text of an article, the response it receives, and its source. The most natural characteristic is the text of an article. Advice in the media varies from evaluating whether the headline matches the body of the article, to judging the consistency and quality of the language. Attempts to automate the evaluation of text have manifested in sophisticated natural language processing and machine learning techniques that rely on hand-crafted and data-specific textual features to classify a piece of text as true or false. These approaches are limited by the fact that the linguistic characteristics of fake news are still not yet fully understood. Further, the characteristics vary across different types of fake news, topics, and media platforms. A second characteristic is the response that a news article is meant to illicit. Advice columns encourage readers to consider how a story makes them feel – does it rovoke either anger or an emotional response? The advice stems from the observation that fake news often contains opinionated and inflammatory language, crafted as click bait or to incite confusion . For example, the

New York Times cited examples of people profiting from publishing fake stories online; the more provoking, the greater the response, and the larger the profit [26]. Efforts to automate response detection typically model the spread of fake news as an epidemic on a social graph, or use hand-crafted features that are social-network dependent, such as the number of Facebook likes, combined with a traditional classifier.

II. RELATED WORK-IDENTIFY, RESEARCH AND COLLECT IDEA

It's the foremost preliminary step for proceeding with any research work writing. While doing this go through a complete thought process of your Journal subject and research for it's viability by following means:

1) Read already published work in the same field.
2) Goggling on the topic of your research work.
3) Attend conferences, workshops and symposiums on the same fields or on related counterparts.
4) Understand the scientific terms and jargon related to your research work.

III. WRITE DOWN YOUR RESULTS, STUDIES AND FINDINGS

Now it is time to articulate the research work with ideas gathered in above steps by adopting any of below suitable approaches:

A. Bits and Pieces together

In this approach combine all your researched information in form of a journal or research paper. In this researcher can take the reference of already accomplished work as a starting building block of its paper.

Jump Start

This approach works the best in guidance of fellow researchers. In this the authors continuously receives or asks inputs from their fellows. It enriches the information pool of your paper with expert comments or up gradations. And the researcher feels confident about their work and takes a jump to start the paper writing.

B. Use of Simulation software

There are numbers of software available which can mimic the process involved in your research work and can produce the possible result. One of such type of software is Matlab. You can readily find M Files related to your research work on internet or in some cases these can require few modifications. Once these Files are uploaded in software, you can get the simulated results of your paper and it eases the process of paper writing.

As by adopting the above practices all major constructs of a research paper can be written and compiled to form a complete research ready for Peer review.

IV. CONCLUSION

Existing features-based approaches to news verification on news articles and headlines ignore the very important part of news that is related image with article in news. In this paper, we focus on text articles to improve the verification performance. We find that apart from their popularity and great impact on news diffusion, images also have distinctive distribution patterns for the real and fake news visually and statistically.

The concept of short term memory can be included in the fake image as well to improve the performance. As only using LSTM on text approach will be not as good as combination of CNN in Images and LSTM in text. Different other parameters in the Text dataset can too be involved in order to increase the performance of the Text classifier. The most important future can be that we can build the classifier which combinely classify fake images and text so that our news detector can have more use cases.

REFERENCES

[1] Zhiwei Jin, Juan Cao, Yongdong Zhang, Senior Member, IEEE, Jianshe Zhou, and Qi Tian, Fellow, IEEE, "*Novel Visual and Statistical Image Features for Microblogs News Verification*," IEEE Transactions on Multimedia, VOL. 19, No. 3, March 2017.

[2] Jiebo Luo Department of Computer Science University of Rochester, *Multimodal Fusion with Recurrent Neural Networks for Rumor Detection on Microblogs* MM'17, October 23–27, 2017, Mountain View, CA, USA

[3] Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. *MVAE: Multimodal Variational Autoencoder for Fake News Detection.* In Proceedings of the 2019 World Wide Web Conference (WWW '19), May 13–17, 2019, San Francisco, CA, USA. ACM, New York, NY, USA.

[4] Hunt Allcott and Matthew Gentzkow. 2017. *Social Media and Fake News in the 2016 Election*. 31 (05 2017), 211–236

[5] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. *Fake news detection on social media: A data mining perspective*. ACM SIGKDD Explorations Newsletter 19, 1 (2017), 22–36.

[6] J. Z. Pan, S. Pavlova, C. Li, N. Li, Y. Li, and J. Liu, "*Content based fake news detection using knowledge graphs*," in International Semantic Web Conference (1), ser. Lecture Notes in Computer Science, vol. 11136. Springer, 2018, pp. 669–683

AUTHORS

First Author – Author name, qualifications, associated institute (if any) and email address.

Second Author – Author name, qualifications, associated institute (if any) and email address.

Third Author – Author name, qualifications, associated institute (if any) and email address.

Correspondence Author – Author name, email address, alternate email address (if any), contact number.