

A Study on Impact of Fake News and Methods for Identifying It

Shashank H.

MCA Scholar, School of CS & IT, Dept of MCA, Jain (Deemed-to-be) University, Bangalore, India.

ABSTRACT

The study focuses on the impact of fake news to the people and the effects which are caused by the circulation of fake news using social media network, internet and also by word of mouth. The study mainly focuses on finding the origin from where the fake news is spread to a large group of people and also focuses on the ways which are used currently to detect the genuineness of the content before it is put into circulation. The study also covers the part about the machine learning approach which is used to find if the news is fake or real, with the help of machine learning algorithms we can speed up the process of detecting if the content is real or fake and it can also be automated in social media networks which helps in removal of fake content prior to publish. It also focuses on how fake news is being a threat to modern day journalism as well as journalists whose job is to deliver the happenings of the world to the public. The fake news mostly contains made up stories upon a certain topic or targeted at a certain individual in order to harm their reputation. Checking the facts in the news finding out the source of the news becomes very important while dealing with the random news. The news is one of the important things which helps us to understand the happenings around the world and also lets us the current affairs in the country, if a news which is not real or verified by an organization is board then people will end up believing that it is real. It also has an impact on healthcare domain as well as political domain it is said to have impacted the US election 2016 and also about spreading the fake news during the time of COVID-19 pandemic which has caused distress and panic among people.

1. INTRODUCTION

The study focuses on how the fake news has impact on public and also the measures which are evolving in order to take preventive measures to control the spread of fake news, also discuss about the methods which are currently deployed by several organizations as well as social media networks which are preventing the fake news being spread from their network. The important part of controlling fake news circulation is detecting if the news data is uploaded is real or fake, we use several machine learning techniques such as natural language processing (NLP) which is used to teach the computer about processing natural language in which the news data is present. We also study about the newly evolving techniques for detecting fake news and the algorithms used by new models and the effectiveness of then also the prediction accuracy which is produced by the machine learning model.

2. LITERATURE SURVEY

[1] In this paper, the author discusses about using several machine learning techniques such as neural, statistical and external features to compute the efficient solution for the identifying fake news. The author uses the approach of neural embedding with deep recurrent model and statistical features with n-gram bag-of-words. Also, the author discusses about how social networking sites does not use any automation techniques in order to identify the news is fake before posting it into the network. The result of prediction which is found after using engineering heuristics of neural and statistical method is 85.68 percentage.

[2] In this paper, the author discusses about features and deep representation learning for stance detection in order to check for fake news. First the author discusses about how the fake news is becoming a threat or a pressing concern to modern day high-quality journalism and how most of the times this fake news is used to achieve secondary gains. The steps which are defined by author to prevent the fake news is to check what other news agencies such as verified news agencies are displaying at the time and check if both the contents are same this is called stance detection. The result achieved by this method is 83.38 percentage.

[3] In this paper, the author discusses about a tough-to-beat baseline for detecting fake news using stance detection. The technique of automatic stance evaluation is used by the author in order to process fact checking and detection of fake news and the algorithm also had claimed third place in first stage of fake news detection challenge. The author combines mechanism of term frequency (TF) and term frequency-inverse document frequency (TF-IDF) representations and features. The result of the following approach is calculated to be FCN-1 score of 81.75 percentage.

[4] In this paper, the author discusses about automatic stance detection with the help of end-to-end memory networks, the mechanism works at paragraph level and also integrates the mechanism of convolutional and recurrent neural networks. Further discusses about how the internet aims of raging from affecting people's beliefs and their decisions. The approach uses the mechanism of checking for relevant paragraphs in order to predict the news genuineness.

[5] In this paper, the author discusses on the heuristic-driven ensemble framework for detecting COVID-19 fake news, the author stresses the point how the fake news is causing the panic among people during the raging pandemic COVID-19. The model here has achieved 8th place on the leader board for detection of fake news, the methods which are used here are text processing, tokenization, ensemble and backbone model architectures. The result has initially improved while used with soft-voting with combinations of various models.

[6] In this paper, the author discusses about the performance of a broad set of modern-language model as well as the technique of fine tuning in order to interpret the modern-language model. In this model the author takes three digits of tweet id as a main factor because it contains the information on the date in order to check if the news is genuine or not. As an example, the author discusses about the COVID-19 pandemic misinformation which can cost lives of people. The benchmark twitter dataset used for the model contains around 100 million tweets and the model is found to be effective under handling large datasets.

[7] In this paper, the author makes a comparison between present deep learning techniques by using vector space and mathematical operations with vector spaces. The various vector space representations used are Word2Vec, which is used to combine words into a single paragraph which later can be given to the model. One-hot encoding, this is more of a straightforward approach to vector space which will have a length of 5000 words in order to cover all words. Doc2Vec, it is an extension for Word2Vec which is capable of preserving the order of words. The models of deep learning which are used are artificial neural network (ANN), convolutional neural network (CNN) and recurrent neural network (RNN). The final model is found to be very efficient because of using multiple models and multiple vector space representations.

[8] In this paper, the author discusses about early detection of COVID lies with the help of content, prior knowledge and source information. The author first discusses about how social media is vulnerable to fake news dissemination which will have negative consequence and panic among healthcare domain this can cause even lives of people where public is fed with a hoax news which is not true. The proposed model uses BERT language model and external sources such as Wikipedia for reliability. At the end of model author presents us with a neural network model which integrates both textual and external content for detection of fake COVID news, the future work of this model focuses on using structured data to encode more source knowledge.

[9] In this paper, the author discusses about automating the fake news detection in social networks, the first discussion about how social networks like facebook allow its users to freely post content on their accounts and the mechanism used here will allow us to detect the content is fake or not by considering the number of likes and also based on the users who have liked the model which will help us to know if there is any like-bot used in order to like the fake post. The model uses logistic regression approach and an adoption of Boolean crowdsourcing algorithms, the facebook dataset which is used for the model contains around 15500 posts and 909236 user accounts. The author also uses cross-validation analysis of logistic regression combined with harmonic algorithm for BLC. The accuracy found by the model is 99.4 percentage while used with harmonic algorithm.

[10] In this paper, the author discusses about a hybrid deep model for fake news detection, the first thing about fake news articles the author discusses is about the article itself and then talks about the response it receives and also about the source users which are promoting the news the model mainly focuses on finding source parties or users who post fake content into the internet. The CSI model proposed by the author mainly has three modules called Capture, Score and Integrate, it uses recurrent neural network which is used to find the user activity on the article which is posted, the second module reads about the behaviour of the users and the third model classifies if the news is real or fake.

3. SCOPE

The day-to-day circulation of hoax news with social networks as a medium for spreading the fake news it becomes a major task to automate the task of detecting fake news from real news. The social networks should be able to automate the tasks such as detecting the fake news before posting it in their social network. The machine learning and automation plays a very important role in helping social networks achieving the fake news detection.

4. CONCLUSION

The fake news detection with the help of machine learning with multiple algorithms and automating the process of fake news detection helps us to detect the fake news faster than compared to traditional ways or manual detection. Also, the accuracy which is achieved by the machine learning models is increasing rapidly. Hopefully we will be able to get a machine learning model is fast, accurate and automated which detects the fake news before posting it to the internet.

5. REFERENCES

- [1] A. S. S. S. A. N. Gaurav Bhatt, "On the Benefit of Combining Neural, Statistical and External Features for Fake News Identification," *arXiv*, vol. 1712, no. 03935v1, 2017.
- [2] B. M. P. C. LUÍS BORGES, "Combining Similarity Features and Deep Representation Learning for Stance Detection in the Context of Checking Fake News," *arXiv*, vol. 1811, no. 00706v1, 2018.
- [3] I. A. G. P. S. S. R. Benjamin Riedel, "A simple but tough-to-beat baseline for the Fake News Challenge stance detection task," *arXiv*, vol. 1707, no. 03264v2, 2018.
- [4] R. B. J. G. J. G. L. M. Mitra Mohtarami, "Automatic Stance Detection Using End-to-End Memory Networks," *arXiv*, vol. 1804, no. 07581v1, 2018.
- [5] A. B. S. D. Sourya Dipta Das, "A Heuristic-driven Ensemble Framework for COVID-19 Fake News Detection," *arXiv*, vol. 2101, no. 03545v1, 2021.
- [6] J. D. R. R. Kellin Pelrine, "The Surprising Performance of Simple Baselines for Misinformation Detection," *arXiv*, vol. 2104, no. 06952v1, 2021.
- [7] L. Singh, "Fake News Detection: a comparison between available Deep Learning techniques in vector space," pp. 52-56, 2019.
- [8] Z. B. Ipek Baris, "ECOL: Early Detection of COVID Lies Using Content, Prior Knowledge and Source Information," *arXiv*, vol. 2101, no. 05499v1, 2021.
- [9] G. B. M. L. D. V. S. M. L. d. A. Eugenio Tacchini, "Automated Fake News Detection in Social Networks," *arXiv*, vol. 1704, no. 07506v1, 2017.
- [10] S. S. Y. L. Natali Ruchansky, "CSI: A Hybrid Deep Model for Fake News Detection," *arXiv*, vol. 1703, no. 06959v4, 2017.