

The New Fake News Detection with Generated Comments from News Article

Yuta Yanagi

Department of Informatics
the University of Electro-Communications
Tokyo, Japan
yanagi.yuta@ohsuga.lab.uec.ac.jp

Yasuyuki Tahara

Department of Informatics
the University of Electro-Communications
Tokyo, Japan
tahara@uec.ac.jp

Yuichi Sei

Department of Informatics
the University of Electro-Communications
Tokyo, Japan
sei@uec.ac.jp

Akihiko Ohsuga

Department of Informatics
the University of Electro-Communications
Tokyo, Japan
ohsuga@uec.ac.jp

Abstract—Recently, fake news spread via social networks and make the wrong rumor faster. This problem is serious because the wrong rumor sometimes make social damage via deceived people. Fact-checking is an ordinary solution to measure the credibility of news articles. However this process usually takes a long time and it is hard to make it before spreading the wrong rumor. Automatic detection of fake news is a popular researching topic. It is confirmed that considering not only articles but also social context(i.e. likes, retweets, replies, comments, etc.) supports to spot fake news correctly by them. This type is also hard to detect fake news before spreading the wrong rumor because social context is made with spreading on social networks. We propose a fake news detector with generating part of the social context which is extended from the fake news generator model. This model is trained about generating comments and classification of real/fake by dataset which is combined news and comments. To measure this model's classification quality, we checked classification results from articles with real comments and generated ones by itself. We compared results between classified with attached a generated comment and real comments only and we got results that considering a generated comment help detect more fake news than considering real comments only.

Index Terms—fake news, disinformation, neural network, natural language processing, deep-learning, microblogs

I. INTRODUCTION

In this era, social media is one of the important parts of our lives. Social media makes it easier to get news and share them with friends online. However, at the same moment, there is also information that includes less credibility. Some of them have obvious misinformation that is made by malicious purpose, we call them “fake news”.

Fake news try to make wrong rumors by spreading on social media. This year, there is so much fake news about COVID-19 and sometimes these make wrong rumors. Director of General of the WHO called this problem “infodemic” and he told that fake news spreads faster and more easily than this virus [1]. Besides, fake news created some mayhem not only online, but also offline (real incidents) e.g. in Washington, fake news

about the Pizzagate conspiracy is reported to have motivated the shooting [2]. Spreading fake news also shakes the premise of democracy due to people cannot get accurate information. Therefore, some researches try to spot fake news by machine learning.

The challenging point of this is there are news articles which try to deceive readers and this makes harder to classify by simple rule-based method. To get more information to detection, there are some works which aggregate social context i.e. Retweet, Like, and comments report better results than only considering news text [3]. However, social contexts are not able to get before spreading. Hence, there is also a work that generates words of comments from the news by CVAE to detect fake news when they are just posted [4]. Their work tries to generate comments, but generated ones are only words that have a high probability of appearing.

In this work, we will propose a model that evaluates news credibility by news text and generated comments. This model is modified from generating fake news articles [5] and this model trains not only news features but also generating comments. In training, this sequence includes real posted comments but the test sequence does not use them to simulate operation in real-social media. The skill of generating comments help classification in the test sequence.

We measure the performance of our proposed method by some experiments with a real-posted dataset.

II. RELATED WORKS

- A. *Fact checking*
- B. *Detecting fake news*
- C. *Generating fake news*

III. METHODOLOGY

- A. *Generating comments*
- B. *Classifying news and comments*

IV. RESULTS

A. *Word generation tendency*

First of all, we investigated the difference between generated comments from real and fake news. We generated comments which refer to news articles that are fact-checked by PolitiFact. This dataset contains sets of a news article and tweets (comments) which refer to it. We filtered news which have at least three tweets and sampled three tweets for generating. Both of real and fake labels have 200 sets of an article and comments and we trained to generate comments. We used these indexes: times of used words, percentage of used words, and the gap of a percentage point of used words of generated comments from real and fake. We removed extra elements: stop words by NLTK, url (starts with *http*, *https*), and part of symbols. We didn't remove mentions, colons, and hashtags (i.e. @anyone, analyze: #anything). We found these features of all generated comments:

- The most generated word was “via” (c. 1.5%).
- “via” was also top frequency of generated word from both of real and fake.
- The second and third were “trump” and “obama” but both of their percentages was under 1%.

We also found the difference between generated comments from real and fake news.

- The percent about frequency of “via” in generated comments from fake news article was twice as much as ones from real news.
- “via” was also the most gap of frequency between generated comments from real and fake. The delta was c. 0.9 percentage point.
- “breaking:” was the second of the percentage point between frequency (fake was more than real). the delta was c. 0.7 percentage point.

B. *Quality of classify*

We measured effect of generated comment for classification by comparing classification without a generated comment. We prepared baselines: classify by only news article, with two real-posted comments. We used pair of an article of GossipCop and tweets instead because ones of PolitiFact were too few to make classification accurate. We sampled same rule of IV-A but both of real and fake labels have 2000 sets. The result of classification is Table I. Our proposed method was best of recall score but precision is worse than consider without generated comments.

TABLE I
RESULTS OF CLASSIFICATION

Model name	Precision	Recall	F1 score
Article only	0.647	0.615	0.631
+ Real comment * 2	0.682	0.750	0.714
+ Generated comment	0.590	0.790	0.675

V. DISCUSSION

REFERENCES

- [1] J. Zarocostas, “How to fight an infodemic,” *The Lancet*, vol. 395, no. 10225, p. 676, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S014067362030461X>
- [2] G. s. a. agencies, “Washington gunman motivated by fake news ‘pizzagate’ conspiracy,” 12 2016. [Online]. Available: <https://www.theguardian.com/us-news/2016/dec/05/gunman-detained-at-comet-pizza-restaurant-was-self-investigating-fake-news-reports>
- [3] H. Guo, J. Cao, Y. Zhang, J. Guo, and J. Li, “Rumor detection with hierarchical social attention network,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, ser. CIKM '18. New York, NY, USA: ACM, 2018, pp. 943–951. [Online]. Available: <http://doi.acm.org/10.1145/3269206.3271709>
- [4] F. Qian, C. Gong, K. Sharma, and Y. Liu, “Neural user response generator: Fake news detection with collective user intelligence,” in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 7 2018, pp. 3834–3840. [Online]. Available: <https://doi.org/10.24963/ijcai.2018/533>
- [5] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner, and Y. Choi, “Defending against neural fake news,” in *Advances in Neural Information Processing Systems 32*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 9054–9065. [Online]. Available: <http://papers.nips.cc/paper/9106-defending-against-neural-fake-news.pdf>