Fake News Detection with Generated Comments for News Articles

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

Ryohei Orihara
Department of Informatics
the University of Electro-Communications
Tokyo, Japan
orihara@acm.org

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. It suggests that our proposed model will help to spot fake news on social networks.**

***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".

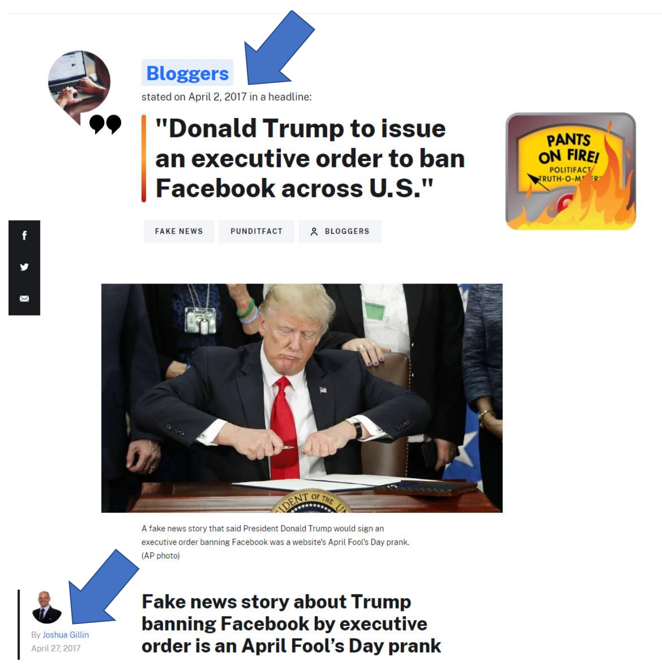

Fig. 1. An example of fact-checking. It confirmed that April Fool's prank. The blue arrows shows the posted date of fake news and fact-checking. There are 25 days between posting fake news and fact-checking.

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 online. Directer of

General of the WHO called this problem “infodemic” and he told that fake news spreads faster and more easily than itself [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]. Nowadays, fact-checking is the most token method to spot fake news. This is a process of evaluating for news by people who has knowledge of news topic. Fig. I is an example of fact-checking [3]. However, this takes so long time and it is hard to spot before spreading. 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 it harder to classify by a 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 [4]. 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 [5]. 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 [6] 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

To detect and classify fake news is not a new topic because it is so similar to detecting spam [7], rumor [8], and illegal advertisement [9]. Following some previous works [10]–[12], we define fake news as news that is intentionally fabricated and can be verified as false.

A. Detecting fake news

There are many works which detect fake news with only news content. In-text features, writing styles [13] and amount of emotions [14] were considered because commonly fake news has original styles and emotions. Besides, using deep neural networks achieved better results in classification on some works [15]–[17].

Many works consider the social context of news content. The Social context feature is generated by user-based [18]–[20], post-based [21]–[23], and network-based [24], [25].

Considering the social context, it must wait for moments from posted because social contexts are made by users who are exposed. Therefore, a Two-Level Convolutional Neural Network with User Response Generator(TCNN-URG) was proposed [5]. This generates comment by hidden variables which are trained by a probable distribution of comment appearance. Generating comments can give additional information to classify posts and get even if the news is just posted. However, this generates only words that have a high probability of appearance and there are no grammar elements.

B. Generating fake news

In generating natural language articles, the Grover model made so natural neural fake news articles [6]. This model is trained by news which separated into news domain, author, posted date, title, and article and evaluate the prediction. The interesting thing is that human beings are more likely to be fooled by generated articles being real ones. We tried to extend this model and generate naturally comments.

III. Methodology

Like II-B, the original Grover model was trained by news which had five parts. Each part is attached start and end token and some of them are dropped to predict. We replaced the other part of the article to three comments and tried to predict one of the comments. We modeled by the joint distribution alongside the original one:

$$p = (\text{article}, \text{comment\_1}, \text{comment\_2}, \text{comment\_3}) \tag{1}$$

This model's diagram is Fig. 2. Basically it was replaced for comments from Grover model's news structure except for article. The purpose of our model was to generate comments more likely written by humans.

The last token of integrated sequences were `[CLS]` and this was used for classification of credibility following the original one. The original one was made for generation of fake news but our proposed model was arranged to generate comments. Fig.3 shows our process of experiment.

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 from the FakeNewsNet dataset [26]. 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 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”(approx. 1.5%).
- “via” was also the top frequency of generated word from both real and fake.

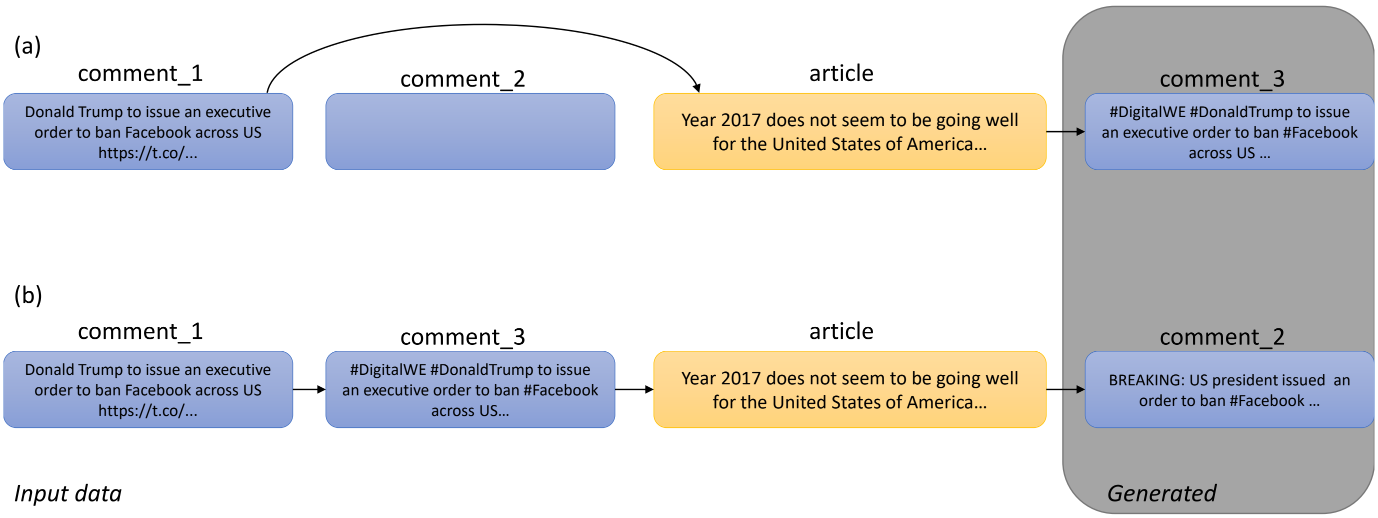

Fig. 2. An ordinary diagram of two generations of our proposed method. (a) shows that one of the comments was generated from partly dropped contexts for comments. (b) shows that another one was generated from contexts that include a generated comment from (a).

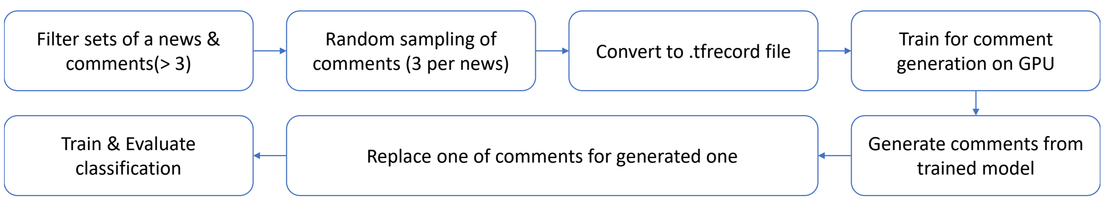

Fig. 3. The process of our experiment.

- The second and third were “trump” and “obama” but both of their percentages were under 1%.
- The generated comments seemed to be not accurate in grammar.

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

- The percent about the 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 approx. 0.9 percentage point.
- “breaking:” was the second of the most percentage point between frequency(fake was more than real). the delta was approx. 0.7 percentage point.

B. *Quality of classify*

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

A. *Generating comments*

According to trends of words in generated comments, our proposed method seemed to be trained by the credibility of news articles. Most of the generated comments referred to topics of politics and this may be caused by the character of the dataset.

The interesting word of generated comment is "breaking:". Our experiment results showed that this word is more generated by fake news than real news. This phenomenon was also confirmed in the research of TCNN-URG [5].

The quality of the grammar was clearly not as good as it should have been by human A/B testing. This is caused by a lack of dataset scale. Grover article used 120 gigabytes of dataset [6]. We need to search or get a more large dataset of sets of articles and tweets.

B. classification

According to TABLE I, our proposed model made the best score of recall but precision was the worst score. This means the proposed model can detect more fake news than another model which doesn't use generated comments even if social contexts are limited. The trend suggests that this model helps for people who search for news which are needed to fact-checking. However, the model also detected more not fake news than another one so this is a point of improvement. We will check if the trend is changed by scale of dataset.

C. Suggestions of improvements

APPENDIX
SETTINGS OF EXPERIMENTS

- Trained on Ubuntu 16.04 on Docker in Linux server with TITAN X (Pascal).
- Our proposed model was extended from grover repository by forking on GitHub.
- Model size was Grover-Base but we reduced vocabulary a little bit in order to fit for extension.

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. staff and 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] J. Gillin, "Politifact - fake news story about trump banning facebook by executive order is an april fool's day prank," Apr 2017. [Online]. Available: https://www.politifact.com/factchecks/2017/apr/27/blog-posting/fake-news-story-about-trump-banning-facebook-execu/

[4] 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

[5] 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

[6] 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

[7] H. Shen, F. Ma, X. Zhang, L. Zong, X. Liu, and W. Liang, "Discovering social spammers from multiple views," *Neurocomputing*, vol. 225, pp. 49–57, 2017.

[8] Z. Jin, J. Cao, Y. Jiang, and Y. Zhang, "News credibility evaluation on microblog with a hierarchical propagation model," in *2014 IEEE International Conference on Data Mining*, 12 2014, pp. 230–239.

[9] H.-H. Huang, Y.-W. Wen, and H.-H. Chen, "Detection of false online advertisements with dcnn," in *Proceedings of the 26th International Conference on World Wide Web Companion*, ser. WWW '17 Companion. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2017, pp. 795–796. [Online]. Available: https://doi.org/10.1145/3041021.3054233

[10] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *SIGKDD Explor. Newsl.*, vol. 19, no. 1, pp. 22–36, Sep. 2017. [Online]. Available: http://doi.acm.org/10.1145/3137597.3137600

[11] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, ser. CIKM '17. New York, NY, USA: ACM, 2017, pp. 797–806. [Online]. Available: http://doi.acm.org/10.1145/3132847.3132877

[12] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, "Eann: Event adversarial neural networks for multi-modal fake news detection," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '18. New York, NY, USA: ACM, 2018, pp. 849–857. [Online]. Available: http://doi.acm.org/10.1145/3219819.3219903

[13] M. Potthast, J. Kiesel, K. Reinartz, J. Bevendorff, and B. Stein, "A stylometric inquiry into hyperpartisan and fake news," *CoRR*, vol. abs/1702.05638, 2017. [Online]. Available: http://arxiv.org/abs/1702.05638

[14] C. Guo, J. Cao, X. Zhang, K. Shu, and M. Yu, "Exploiting emotions for fake news detection on social media," *CoRR*, vol. abs/1903.01728, 2019. [Online]. Available: http://arxiv.org/abs/1903.01728

[15] W. Y. Wang, ""liar, liar pants on fire": A new benchmark dataset for fake news detection," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 422–426. [Online]. Available: https://www.aclweb.org/anthology/P17-2067

[16] H. Karimi and J. Tang, "Learning hierarchical discourse-level structure for fake news detection," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 3432–3442. [Online]. Available: https://www.aclweb.org/anthology/N19-1347

[17] H. Karimi, P. Roy, S. Saba-Sadiya, and J. Tang, "Multi-source multi-class fake news detection," in *Proceedings of the 27th International Conference on Computational Linguistics*. Santa Fe, New Mexico, USA: Association for Computational Linguistics, Aug. 2018, pp. 1546–1557. [Online]. Available: https://www.aclweb.org/anthology/C18-1131

[18] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th International Conference on World Wide Web*, ser. WWW '11. New York, NY, USA: ACM, 2011, pp. 675–684. [Online]. Available: http://doi.acm.org/10.1145/1963405.1963500

[19] K. Shu, S. Wang, and H. Liu, "Understanding user profiles on social media for fake news detection," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, 4 2018, pp. 430–435.

[20] K. Shu, X. Zhou, S. Wang, R. Zafarani, and H. Liu, "The role of user profile for fake news detection," *CoRR*, vol. abs/1904.13355, 2019. [Online]. Available: http://arxiv.org/abs/1904.13355

[21] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu, "Unsupervised fake news detection on social media: A generative approach," in *AAAI*, 2019.

[22] E. Tacchini, G. Ballarin, M. L. D. Vedova, S. Moret, and L. de Alfaro, "Some like it hoax: Automated fake news detection in social networks," *ArXiv*, vol. abs/1704.07506, 2017.

[23] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News verification by exploiting conflicting social viewpoints in microblogs," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, ser. AAAI'16. AAAI Press, 2016, pp. 2972–2978. [Online]. Available: http://dl.acm.org/citation.cfm?id=3016100.3016318

[24] L. Wu and H. Liu, "Tracing fake-news footprints: Characterizing social media messages by how they propagate," in *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, ser.

WSDM '18. New York, NY, USA: ACM, 2018, pp. 637–645. [Online]. Available: http://doi.acm.org/10.1145/3159652.3159677

[25] F. Monti, F. Frasca, D. Eynard, D. Mannion, and M. M. Bronstein, "Fake news detection on social media using geometric deep learning," *CoRR*, vol. abs/1902.06673, 2019. [Online]. Available: http://arxiv.org/abs/1902.06673

[26] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media," *ArXiv*, vol. abs/1809.01286, 2018.