Masters Thesis Proposal
The New Fake News Classification with Comment Generation by Seq-GAN

Yuta Yanagi, Yasuyuki Tahara, Akihito Ohsuga, Yuichi Sei

July 30, 2019

Abstract

There are already positive results about classifying between true news and fake news. In addition, some works reported considering comments on social media made results better. However, comments are not appeared when news are just posted and it doesn't suitable for early detection. Therefore, there is a previous work which proposed the model which does not only classifying true news or fake news, but also real posted or generated. This tried to probably distribution of comment appearance and this achieved generate high quality of classifying in test. Our proposed model plan to evolve this approach by generating comments much similar to real ones by Seq-GAN. We plan to experiment this model to measure performance of classifying news, importance of generating comments, and quality of generated comments. If results will be positive, it shows that generating comments like real posted is helpful to detect fake news.

Index terms— Fake News; Seq-GAN; Natural Language Programming; Deep Neural Network; Social Network

1 Introduction

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

Fake news try to make wrong rumors on social media by spreading on social media. Last year, a picture was spread on Russian social media which shows someone with an Estonian flag on their sleeve beating the protester. After this, Estonian volunteers

identified this picture as a fake news[21]. In addition, 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[1]. Spreading fake news also shakes premise of democracy due to people cannot get accurate information. Therefore, there are some researches which try to spot fake news by machine learning.

The challenging point of this is there are news article 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[4]. However, social contexts are not able to get before spreading. Hence, there is also a work which generate words of comments from news by CVAE to detect fake news when they are just posted[12]. Their work tries to generate comments, but generated ones are only words which have high probability of appearing.

In this work, we will propose a model which evaluate news credibility by news text and generated comments by Seq-GAN[26]. This model train not only news features but also generating comments. In training sequence includes real posted comments but test sequence does not use them in order to simulate operation in real-social media. The skill of generating comments help classification in test sequence.

We plan to measure performance of our proposed method by some experiments with real-posted dataset and some state-of-the-art fake news detection algorithms.

2 Related Works

To detect and classify fake news is not a new topic because it is so similar to detecting spam[14], rumor[6], and illegal advertisement[5]. Following some previous works[17, 13, 23], we define fake news as news which is intentionally fabricated and can be verified as false.

There are many works which detect fake news with only news content. In text feature, writing styles[11] and amount of emotions[3] were considered because commonly fake news has original styles & emotions. In addition, using deep neural network achieved better results in classification on some works[22, 8, 9]. There are also many works which consider social context of news content. Social context feature was generated by user-based[2, 15, 19], post-based[25, 20, 7], and network-based[24, 10].

Considering social context, it must wait moments from posted because social contexts are made by users which are exposed. Therefore, Two-Level Convolutional Neural Network with User Response Generator(TCNN-URG) are proposed[12]. This

generates comment by hidden variables which are trained by probably distribution of comment appearance. Generating comments can give additional information to classify posts and get even if news is just posted. However, this generates only words which have high probably of appearance and there are no grammar elements.

In generating natural language sentences, Seq-GAN is proposed[26]. This is arranged from GAN in order to apply natural language processing and this has also generator and classifier(discriminator). Classifier of Seq-GAN try to classify text which is real or generated. Generator create text from classification results and text features from classifier. In this work, we will arrange it for generating comments which will be post for news article.

3 Thesis

The main objective of this research is developing new fake news classification with comment generation and investigate how proposed method is better in operation on social media. To suppress spread of fake news, we have to spot it early enough. Specifically, it is required classifying before spread of fake news if classifier operate in social media.

In classification of fake news, social contexts give strong information. Among social contexts, comments give more information as natural language than retweets and likes. However, it is impossible to get social contexts from news which is just posted on social media. Therefore, we train model not only classifier but also comment generator for fake news detection. This use Seq-GAN [26] as comment generation with real comments which are posted in Twitter.

4 Methodology

Our proposed model structure is very similar to Seq-GAN[26] and it has classifier and comment generator. Fig.1 shows structure of our model. On the one hand, generator create comments from post. On the other hand, classifier evaluates two values to binary classifications with real or generated comments from generator: post's credibility and reality of comments.

Generator is trained by post feature which is leaked from classifier and classifier is trained by label of posts(true, fake) and comments(real, generated). In the test term, classifier only use posts with generated comments in order to simulate operation on social media.

5 Preliminary Results and Discussion

5.1 Dataset

In order to input our proposed model, we obtained FakeNewsNet[16, 17, 18] dataset. This includes tweets which has URL of news. Every news and tweets are labeled true/fake by fact-check result on PolitiFact or GossipCop. We use tweet text as comments and the other information (user, retweet, like, etc.) are not used. Table.1 is statistics of dataset.

Table1 Statistics of FakeNewsNet by fact-checking platforms

	True		Fake	
Platform	News	Comments	News	Comments
PolitiFact	624	370669	432	195901
GossipCop	16817	843933	5323	539491
Overall	17441	1214602	5755	735392

5.2 Plan of experiments

We will make answer of following evaluation questions:

EQ1 Can our proposed model detect fake news more accurate than any other state-

of-the-art fake news detection algorithms?

EQ2 Is generating comments important in fake news detection by Seq-GAN?

EQ3 Are generated comments similar to real comments?

We are planning to answer them by comparing our model with any other state-of-the-art fake news detection algorithms, ablation experiments, and subjective evaluation by human beings.

5.3 Plan of discussion

5.3.1 EQ1: comparing

We will get results of not only our proposed model but also other algorithms which are proposed by related works. All of them use both of news text and comments for equal comparing.

5.3.2 EQ2: ablation experiments

We also compare by ablation experiments. It does our proposed model with ones which do not use generated comments in order to testify to importance of generating comments by Seq-GAN. If proposed model is better than ablated one, generating comments is important part to find fake news.

5.3.3 EQ3: subjective evaluation

When training is over, generated comments will be so similar to real comments. We can measure how far from real comments to generated comments are by subjective evaluation.

6 Research Goal

This research will show how generating comment is important to identify fake news on social media. Fake news constitutes a grave menace to the safety of our world. Therefore, it is important to detect news credibility before spreading on social media. We aim to detect fake news more precisely by generating comments from articles that are only available after they have spread.

In computer science, computer security is the protection of computer systems. Likewise, countering to fake news prevent people from exposing to malicious information on social media. This is similar to computer security for the purpose of protecting the safety of society.

References

[1] Guardian staff and agencies. *Washington gunman motivated by fake news 'Pizzagate' conspiracy*. Dec. 2016. URL: https://www.theguardian.com/us-news/2016/dec/05/gunman-detained-at-comet-pizza-restaurant-was-self-investigating-fake-news-reports.

[2] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. "Information Credibility on Twitter". In: *Proceedings of the 20th International Conference on World Wide Web*. WWW '11. Hyderabad, India: ACM, 2011, pp. 675–684. ISBN: 978-1-4503-0632-4. DOI: 10.1145/1963405.1963500. URL: http://doi.acm.org/10.1145/1963405.1963500.

[3] Chuan Guo et al. "Exploiting Emotions for Fake News Detection on Social Media". In: *CoRR* abs/1903.01728 (2019). arXiv: 1903.01728. URL: http://arxiv.org/abs/1903.01728.

[4] Han Guo et al. "Rumor Detection with Hierarchical Social Attention Network". In: *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. CIKM '18. Torino, Italy: ACM, 2018, pp. 943–951. ISBN: 978-1-4503-6014-2. DOI: 10.1145/3269206.3271709. URL: http://doi.acm.org/10.1145/3269206.3271709.

[5] Hen-Hsen Huang, Yu-Wei Wen, and Hsin-Hsi Chen. "Detection of False Online Advertisements with DCNN". In: *Proceedings of the 26th International Conference on World Wide Web Companion*. WWW '17 Companion. Perth, Australia: International World Wide Web Conferences Steering Committee, 2017, pp. 795–796. ISBN: 978-1-4503-4914-7. DOI: 10.1145/3041021.3054233. URL: https://doi.org/10.1145/3041021.3054233.

[6] Z. Jin et al. "News Credibility Evaluation on Microblog with a Hierarchical Propagation Model". In: *2014 IEEE International Conference on Data Mining*. Dec. 2014, pp. 230–239. DOI: 10.1109/ICDM.2014.91.

[7] Zhiwei Jin et al. "News Verification by Exploiting Conflicting Social Viewpoints in Microblogs". In: *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*. AAAI'16. Phoenix, Arizona: AAAI Press, 2016, pp. 2972–2978. URL: http://dl.acm.org/citation.cfm?id=3016100.3016318.

[8] Hamid Karimi and Jiliang 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: Hu-*

man Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, June 2019, pp. 3432–3442. URL: https://www.aclweb.org/anthology/N19-1347.

[9] Hamid Karimi et al. "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. URL: https://www.aclweb.org/anthology/C18-1131.

[10] Federico Monti et al. "Fake News Detection on Social Media using Geometric Deep Learning". In: *CoRR* abs/1902.06673 (2019). arXiv: 1902.06673. URL: http://arxiv.org/abs/1902.06673.

[11] Martin Potthast et al. "A Stylometric Inquiry into Hyperpartisan and Fake News". In: *CoRR* abs/1702.05638 (2017). arXiv: 1702.05638. URL: http://arxiv.org/abs/1702.05638.

[12] Feng Qian et al. "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, July 2018, pp. 3834–3840. DOI: 10.24963/ijcai.2018/533. URL: https://doi.org/10.24963/ijcai.2018/533.

[13] Natali Ruchansky, Sungyong Seo, and Yan Liu. "CSI: A Hybrid Deep Model for Fake News Detection". In: *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. CIKM '17. Singapore, Singapore: ACM, 2017, pp. 797–806. ISBN: 978-1-4503-4918-5. DOI: 10.1145/3132847.3132877. URL: http://doi.acm.org/10.1145/3132847.3132877.

[14] Hua Shen et al. "Discovering social spammers from multiple views". In: *Neurocomputing* 225 (2017), pp. 49–57.

[15] 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)*. Apr. 2018, pp. 430–435. DOI: 10.1109/MIPR.2018.00092.

[16] Kai Shu, Suhang Wang, and Huan Liu. "Exploiting Tri-Relationship for Fake News Detection". In: *arXiv preprint arXiv:1712.07709* (2017).

[17] Kai Shu et al. "Fake News Detection on Social Media: A Data Mining Perspective". In: *SIGKDD Explor. Newsl.* 19.1 (Sept. 2017), pp. 22–36. ISSN: 1931-0145. DOI: 10.1145/3137597.3137600. URL: http://doi.acm.org/10.1145/3137597.3137600.

[18] Kai Shu et al. “FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media”. In: *arXiv preprint arXiv:1809.01286* (2018).

[19] Kai Shu et al. “The Role of User Profile for Fake News Detection”. In: *CoRR* abs/1904.13355 (2019). arXiv: 1904.13355. URL: http://arxiv.org/abs/1904.13355.

[20] Eugenio Tacchini et al. “Some Like it Hoax: Automated Fake News Detection in Social Networks”. In: *ArXiv* abs/1704.07506 (2017).

[21] Madis Vaikmaa. *Manipulation And Disinformation In Social Media: The Case Of Estonia And #ESTexitEU*. Apr. 2019. URL: https://euvsdisinfo.eu/manipulation-and-disinformation-in-social-media-the-case-of-estonia-and-estexiteu/.

[22] William Yang 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, July 2017, pp. 422–426. DOI: 10.18653/v1/P17-2067. URL: https://www.aclweb.org/anthology/P17-2067.

[23] Yaqing Wang et al. “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*. KDD ’18. London, United Kingdom: ACM, 2018, pp. 849–857. ISBN: 978-1-4503-5552-0. DOI: 10.1145/3219819.3219903. URL: http://doi.acm.org/10.1145/3219819.3219903.

[24] Liang Wu and Huan 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*. WSDM ’18. Marina Del Rey, CA, USA: ACM, 2018, pp. 637–645. ISBN: 978-1-4503-5581-0. DOI: 10.1145/3159652.3159677. URL: http://doi.acm.org/10.1145/3159652.3159677.

[25] Shuo Yang et al. “Unsupervised Fake News Detection on Social Media: A Generative Approach”. In: *AAAI*. 2019.

[26] Lantao Yu et al. “SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient”. In: *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. AAAI’17. San Francisco, California, USA: AAAI Press, 2017,

pp. 2852–2858. URL: http://dl.acm.org/citation.cfm?id=3298483.3298649.

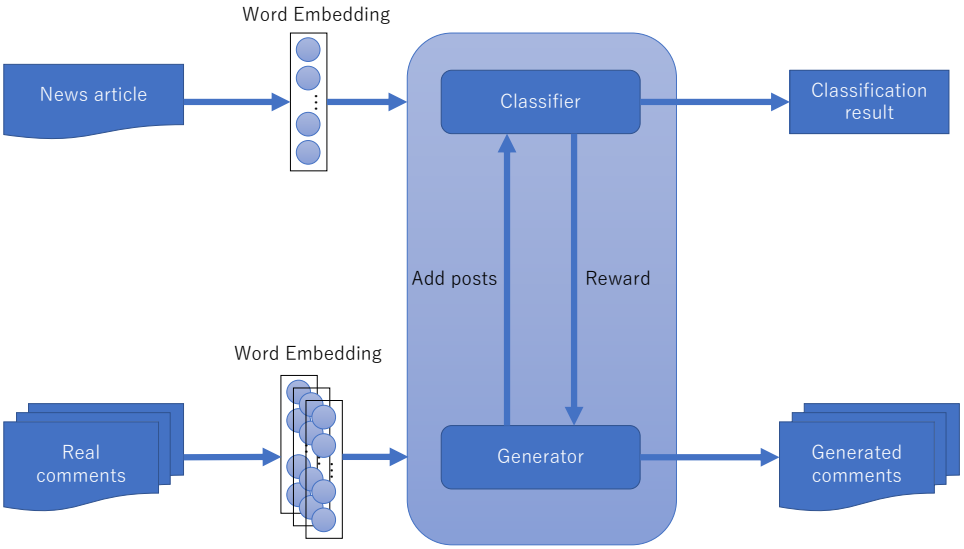

Figure1　The structure of our planning proposed model.