Masters Thesis Proposal

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

Yuta Yanagi

Department of Informatics, The University of Electro-Communications

Main Advisor	:	Yasuyuki Tahara
Advisor	:	Akihito Ohsuga
Advisor	:	Yuichi Sei

1 Abstract

2 Introduction

In this era, social media is one of important parts of our lives. Social media makes more easy 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[7]. In addition, In the U.S., fake news created some mayhems not only online, but also offline(real incidents) i.e. 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 which reports better results than only considering news text[2]. However, social contexts are not able to get before spreading. This means they can be only used after spread. Hence, there is also a work which generate words of comments from news by Multimodal VAE to detect fake news when they are just posted[3]. Their work tries to generate comments, but generated ones are only have 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[8]. This model train not only news features but also generating comments. In training sequence includes real posted comments but test sequence doesn't 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.

3 Related Works

4 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 gives 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 [8] as comment generation with real comments which are posted in Twitter.

5 Methodology

Our proposed model structure is very similar to Seq-GAN[8] 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.

6 Preliminary Results and Discussion

6.1 Dataset

In order to input our proposed model, we obtained FakeNewsNet[4, 5, 6] 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

| Platform | True | | Fake | |
	News	Comments	News	Comments
PolitiFact	1	1	1	1
GossipCop	1	1	1	1
Overall	1	1	1	1

6.2　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.

6.3 Plan of discussion

6.3.1 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.

6.3.2 EQ2: ablation experiments

We also compare by ablation experiments. It does our proposed model with ones which don't 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.

6.3.3 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.

7 Implications of Research

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, we fear. 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] 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.

[3] Dhruv Khattar et al. "MVAE: Multimodal Variational Autoencoder for Fake News Detection". In: *The World Wide Web Conference.* WWW '19. San Francisco, CA, USA: ACM, 2019, pp. 2915–2921. ISBN: 978-1-4503-6674-8. DOI: 10.1145/3308558.3313552. URL: http://doi.acm.org/10.1145/3308558.3313552.

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

[5] Kai Shu et al. "Fake News Detection on Social Media: A Data Mining Perspective". In: *ACM SIGKDD Explorations Newsletter* 19.1 (2017), pp. 22–36.

[6] 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).

[7] 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/.

[8] Lantao Yu et al. "SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient". In: (2017). URL: https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14344.
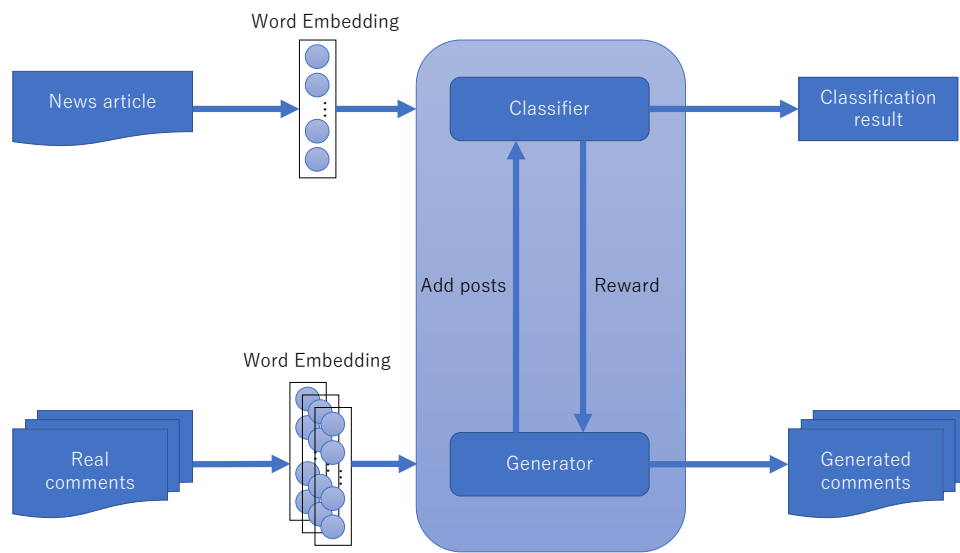
Figure1　The structure of our planning proposed model.