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

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

5 Methodology

Our proposed model structure is very similar to Seq-GAN[4] 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[1, 2, 3] 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	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 comparing by ablation experiments. It compare our proposed model with model which don't use generated comments in order to testify to importance of generating comments by Seq-GAN.

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

References

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

[2] 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.

[3] 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).

[4] 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.

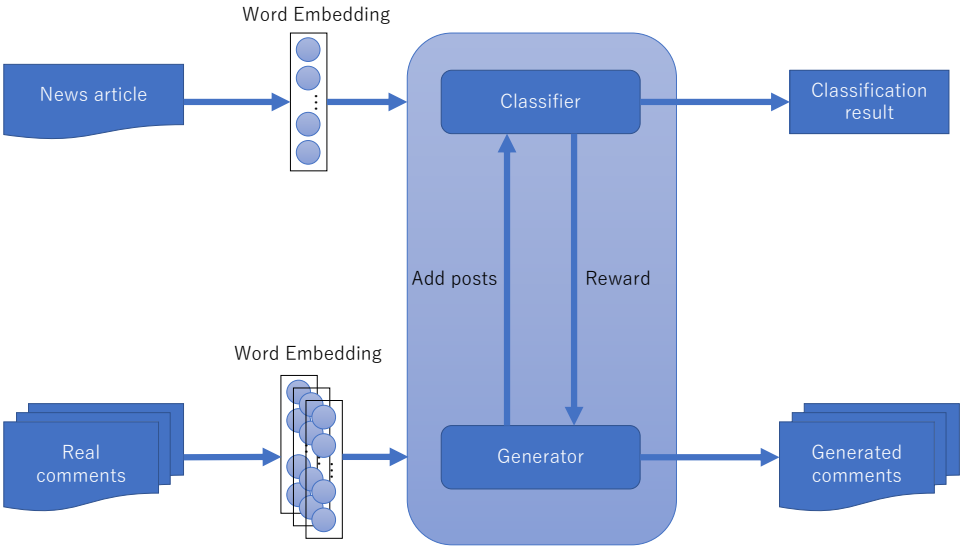

Figure1　The structure of our planning proposed model.