GANによるフェイクニュースの分類と生成

Can we learn more about fake news by generating it directly

Miao XIAO†, Rakesh AGRAWAL††, Kenki NAKAMURA†, Tianwei CHEN†, and Qiang MA†

† School of Informatics, Kyoto University
Yoshida-honmachi, Sakyo-ku, Kyoto, 606–8232, Japan
†† Data Insights Laboratories, University of Wisconsin-Madison
1210 W.Dayton St. Madison, WI 53706–1613, USA
E-mail: †{mxiao,nakamura-kenki,chentianwei}@db.soc.i.kyoto-u.ac.jp, ††rakesha.prof@gmail.com,
†††qiang@i.kyoto-u.ac.jp

Abstract Nowadays, more and more people tend to consume news from social media. However, fake news has been easily and widely propagated by social media due to there is no significant third-party fact-checking the contents of news. It is a challenge for us to classify fake news articles into different types and notify users reliability of news. Related work primarily utilizes machine learning methods to do classification which is based on a large amount of fake news data. However, since it is so difficult to recognize fake news among all the news that we cannot collect fake news easily by some heuristic methods like web crawler, in which case vast manual efforts are required. On the other hand, news patterns are always constantly updated with new events. If we blindly extract fixed features from an existing dataset, our model will be over-fitting. Our proposed method tries to solve these problems by training a pair of adversarial neuron networks simultaneously: Classifier judges whether input fake news data is from Generator or not and categorize real-fake-news, then leak information to guide Generator to write fake-fake-news. Generator captures the distribution of real-fake-news dataset and fools the Classifier. We call this method as Fake News GAN which is short for fake news generative adversarial networks.

Key words Fake News, generative adeversarial networks, deep learning, natural language processing.

1 Introduction

Nowadays, social media plays a more and more important role in human ' s daily life. We almost use social platforms every day such as Twitter, Line or otherwise to check the latest information or just send messages to friends. Social platforms do really provide convenient and seamless access to information.

As a result, more and more people tend to consume news from social media: about 62 percent of Americans get news from social media [1] and more than one-third people from the UK and US will get news from social media firstly when using a smartphone [2]. The reason is that it is often more timely and less expensive to comsume news by social media compared with traditional news organizations, such as newspapers or televisions; and it is more convinient to share, add comments on, or discuss the news with friends directly.

Aside from all the merits, on the other hand, there are also some fatal drawbacks due to the structure of social media which is dramatically different from the previous traditional media technologies: there are no third-parties to filter, check the fact, or to add editorial judgments to these contents relayed by social media. Furthermore, sometimes, a user without reputation can reach as many readers as NHK, or the New York Times. Once he/she relays fake news among his/her followers, false information will be propagated by social media rapidly and widely.

Fake news is news, stories or hoaxes created to deliberately misinform or deceive readers(注1). Fake news has raised more and more concerns since the 2016 US presidential election. The most discussed fake news stories tended to favor Donald Trump over Hillary Clinton, Such as Hillary indictment imminent(注2). So a number of commentators have suggested that Donald Trump would not have been elected president without fake news.

Fake news can affect our society proufoundly. Two typical ways that aimed to interdicting the propagation of fake news

(注1)：https://www.webwise.ie/teachers/what-is-fake-news/
(注2)：https://www.politifact.com/punditfact/statements/2017/oct/31/freedomjunkshuncom/fake-news-site-claims-hillary-clinton-flies-ukrain

has been verified to be effective: fake news identification and classification.

Identify fake news directly requires a large number of manual efforts, furthermore, this method cannot remind the public of the reliability of news from social media timely. So the major challenge for us is to construct a model which is able to classify the fake news articles into different types and remind readers timely. On the other hand, for deep learning based text-classify models, we need a large amount of training data, but fake news dataset is rare and it is too difficult to identify fake news from the real without manual efforts. And the most important feature of fake news is that its' patterns change rapidly, which means that as time goes on, the way to write fake news articles can be changed frequently. So, concentrating too much on the current fake news will make our model be overfitting.

Our main contributions are as follows:

- We proposed a LSTM based neuron network, which converts the words of news articles into dense vectors then utilizes LSTM [3] to classify the fake news articles. The whole network is trained and tested by the real-world fake news dataset from Kaggle(注3). The experimental results verified that our method achieves better performance compared to the recent studies.
- We proposed a generative adversarial network based on the preceding LSTM neuron network, we called Fake News GAN, can generate and classify fake news simultaneously which avoids over-fitting and covers the randomness of fake news patterns effectively.

2 Related Work

Fake news raises major concerns since the 2016 US election. There exist several related work to identify or classify fake news articles recently. In this part, we will introduce several related methods and some deep learning technologies used in our proposed methods.

2 1 Online Fact-checking

Shao et al. [7] proposed an online fact-checking platform for fake news named Hoaxy. This application collects credible scores of news articles from several online fact-checking websites, such as Snope.com, PolitiFact.com, and FactCheck.org. This platform does really work well due to the efficient scores of these websites. On the other hand, information can be propagated by social media rapidly, both evaluating the fake news and collecting the scores spend too much time to avoid breaking fake news to confuse people. For example, Barack Obama being injured by an explosion at the White House. Thus, we should identify fake news as soon as they come out from social media rather than waiting for the score by third-party fact-checking.

2 2 Tensor Decomposition in Ensembles

Hosseinimotlagh et al. [8] proposed a method by modeling the fake news corpus as a 3-mode (article, term, term) tensor which captures spatial terms relation and article-term relations, and then use CP/PARAFAC decomposition to identify latent groups of articles. The highlight of this paper is that they introduced an ensemble method which leverages multiple decompositions of the tensor to further refine the latent groups and produce a clean categorization of articles, which attends higher accuracy than the state of the art. The intuition behind the method is that news articles that tend to frequently appear surrounded each other among different rank configurations of tensor decomposition.

2 3 SeqGAN

GAN [9] is short for Generate Adversarial Network and originally designed for generating images, due to the stochastic gradient descent algorithm [16] and consecutiveness of pixels. But in NLP tasks, the training data are usually word vectors with on consecutiveness, so we must leverage the power of reinforcement learning rather than BP gradient descent to guide the generation process.

There are several related works for GAN in NLP: MaskGAN [19], SeqGAN [10], and LeakGAN [11] which have been verified to perform very well in generating semantic sentences. SeqGAN utilizes Policy Gradient algorithm [17] as the optimization method. The general settings are: regard generator as an agent, word token generated by it at every time stamp as action, and all the tokens generated before as state and then use a reward function formed by the real possibilities of complete sentences judged from discriminator. For incomplete sentences generated at a certain time, Monte Carlo Tree Search [11] is used for completing them.

3 Deep Learning Based Fake News Classification

From above related work, we find it more efficient to focus on the fake news text itself. By classifying fake news articles into different types, we can remind readers of the reliability of news timely. In this section, we proposed two progressive deep learning based methods which classify fake news with higher accuracy than the state of the art.

3 1 LSTM Based Method

In this part, We construct multiple neuron networks to classify fake news text. As shown in the figure1, we get fake news text as the input of word embedding layer and then use 3 kinds of word embedding methods(one-hot, GloVe, and FastText) for comparisons. After that, input the whole or the part of text matrices formed by words vectors to the LSTM.

(注3) : Megan Risdal. Fake news dataset. https://www.kaggle.com /mrisdal/fake-news

Finally, a softmax layer is utilized to classify the feature vectors from LTSM. The following parts will introduce the word embedding and LSTM layer in detail.

3 1. 1 Word Embedding

Recently, distributed representation of words has been widely utilized as the pre-training of natural language processing tasks. Unlike the tensor decomposition, news text is represented as a words co-occurrence matrix.

In this model, we use two typical word embedding methods, GloVe [12] and FastText [13], and one-hot as the comparison, to pre-train our neuron networks. GloVe is an unsupervised learning algorithm. Unlike the Word2Vec [14] model, the training process of GloVe is performed on aggregated global word-word co-occurrence statistics from a corpus, and the results indicate the linear substructures of the word vector space. The following formulation indicates the loss function of GloVe model:

$$J = \sum_{i,j=1}^{V} f(X_{ij})(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - logX_{ij})^2 \qquad (1)$$

where $w_i^T \tilde{w}_j$ are the word vectors we want to get, b_i and $\tilde{b}_j$ are bias variables, $logX_{ij}$ is the weight function and X_{ij} is the entry from word co-occurrence matrix.

FastText is based on Word2Vec model but every single word is not the smallest part within the model. It assumes that n-grams of characters form a single word, and n could range from 1 to the length of the word. For instance, learn is composed of [lea, lear, learn] etc.

3 1. 2 LSTM Layer

LSTM is short for Long Short-Term Memory Network and widely used in some NLP tasks, such as Machine Translation, Text Classification, QA and other fields. LSTM is based on RNN but adds memory cells and three control gates to effectively control historical information especially in long text tasks. Input gate controls the extent of information from the input vector at current moment; forget gate controls how much the historical information influence the information at the current moment; output gate controls the extent to which the value in the cell is used to compute the output. Unlike

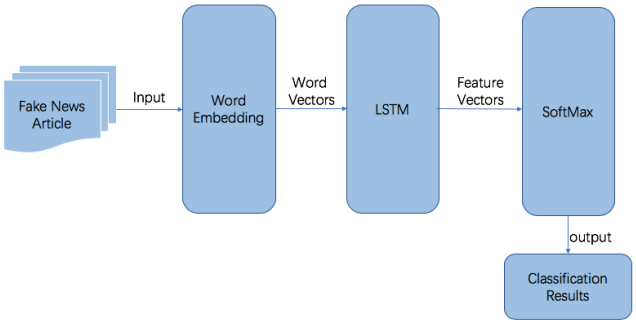

Figure 1 LSTM Based Network

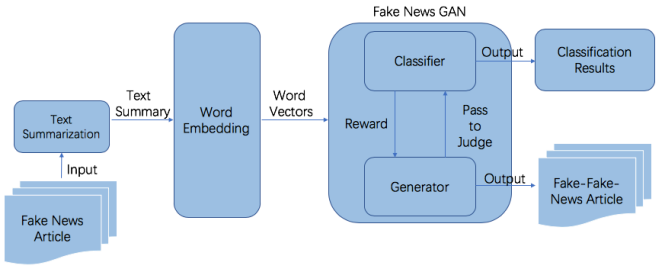

Figure 2 Fake News GAN

RNN, hidden state at the previous moment is not completely washed away, thus this kind of structure enhances its ability to process long text sequences and solves the vanishing gradient problem.

3 2 Fake News GAN

Based on the results of the experiments of the preceding LSTM based neuron network, we find it more efficient to classify the fake news articles by deep learning than our baseline tensor decomposition. On the other hand, deep learning based methods require plenty of training data to modify a large number of parameters. There are only rare datasets existing since it is so difficult to recognize fake news from the all the news that we cannot collect fake news easily by some heuristic methods like web crawler, in which case vast manual efforts are required.

Since we use deep learning, we must pay attention to avoid over-fitting. The most worthy information of one news article describes events from who, what, when, where, how, and why. And if we change any part among them, we can see it as fake news. So due to the special patterns and randomness of news articles, even if we collect all the current fake news as training data, it will be changed quickly in the near future. If we blindly extract fixed features from a current existing dataset, our model will be over-fitting.

In this part, as shown in figure 2, we proposed a GAN based neuron network to improve the accuracy, avoid over-fitting, and get more training data by generating fake-fake-news directly due to the randomness of the generator.

3 2. 1 Classifier

GAN trais two neural networks Generator and Discriminator simultaneously. General discriminator tries to judge whether the input image data is from the generator or not, and generator tries to capture the distribution of real dataset then fools discriminator mistakes its inputs as real. The whole process can be formalized as a min-max game with the following loss function between G and D:

$$\max_G \min_D V(G,D) = E_{X\sim p_d(x)} logD(X) + E_{Z\sim p_z(z)} log[1-D(G(z))] \qquad (2)$$

where $p_d(x)$ is the data distribution from training data,$p_z(z)$ is a prior on input noise variables. In our network, discrim-

inator will be modified as a classifier which categorizes fake news articles into several types and judge the real or fake at the same time. Since words vectors of news articles has no consecutiveness, we must leverage the power of reinforcement learning at every turn of the training process: the classifier will leak the extracted features information from the real-fake-news as a reward signal to guide the generator to imitate the real.

3.2.2 Generator

Generator tries to capture the distribution of the real-fake-news dataset to make generated sentences be judged as true by classifier. In our network, generator also receives the reward signal from classifier to guide the generation process. But due to the instability of existing reinforcement learning applied to GAN, as the length increases, the generated sentence will be more and more non-semantic. In leakGAN and seqGAN, the limitation length of generated sentences is about 30 words, but fake news articles has more than 300 words in general.

Our idea is to mine the abstract of the whole text before inputing them into the network, for example, by using textrank [20], which means that classifier categorize news articles by only using abstracts. This idea has been verified to be effective from LSTM based neuron network. And the generator also only need to generate the fake news abstract which contains less than 30 words.

4 Experiments

In this part, we conduct a series of experiments on our proposed LSTM based network by using various forms of word vectors as the input of the LSTM network. As for experiments on GAN in this part will be held on classifier and generator respectively.

4.1 Dataset

We evaluate the accuracy of all the proposed networks on the same real-world dataset as Tensor Decomposition [8] from Kaggle for comparison. This dataset contains more than 12,000 fake news articles which are labeled by BS detectors(注4). Table 1 demonstrates various types with the description of fake news.

4.2 LSTM based network

We only use the news text and labels of the dataset for training and validation. After deleting the stop words, as we have mentioned above, 3 kinds of word embedding methods will be performed for experiments: one-hot word vectors in 1,000 dimensions, GloVe in 300 dimensions, and Fasttext in 300 dimensions. Both GloVe and Fasttext are advanced word embedding methods, and the reason why we still use naive one-hot method is that we want to prove the quality of word vectors impact the results a lot.

(注4) : B.S.Detector. Fake news classifications. http://bsdetector.tech, 2017.

Table 1 Fake News Labels from BS Detector

Category Names	Descriptions
Bias	Sources that traffic in political propaganda and gross distortions of fact.
Conspiracy	Sources that are well-known promoters of kooky conspiracy theories.
Satire	Sources that provide humorous commentary on current events in the form of fake news.
Hate	Sources that actively promote racism, misogyny, homophobia, and other forms of discrimination.
Junksci	Sources that promote pseudoscience, metaphysics, naturalistic fallacies, and other scientifically dubious claims.
State	Sources in repressive states operating under government sanction.

Furthermore, based on word vectors, in order to make our neuron network understand the text in more detail, we use Stanford Parser [15] to mine the sentence structures in grammar, for instance, find the subjects and objects within one sentence.

From figure 3 and figure 4, as expected, we can find that GloVe and Fasttext performed much better than one-hot, but grammar structures impact little on results. Moreover, in satire and junksci, we find apparent downtrend as the number of top news increase. We think this is because the models are overfitted or the instability of BS Detectors.

4.3 CNN based network

CNN based network will be utilized as the classifier of our Fake news GAN, and we think it can extract features of text easily than LSTM with lower computational cost. In this experiment, we used the text-rank algorithm to extract the abstract of every piece of fake news corpus then delete items longer than 30, for which we can ensure this pre-processed dataset can also be applied to the generator. The CNN network is quite light and only use the one-hot word embedding method with several convolutional and pooling layers. We

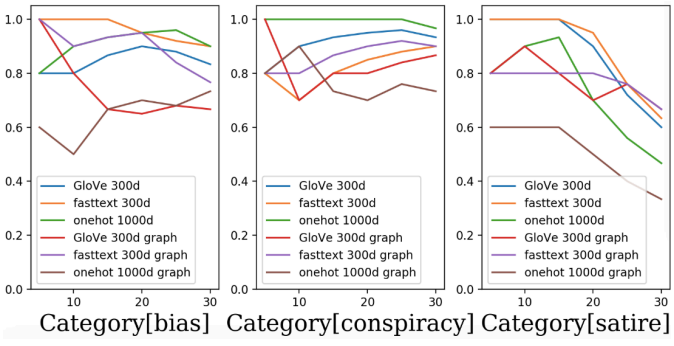

Figure 3 Results from top 30 news of bias, conspiracy, and satire respectively.

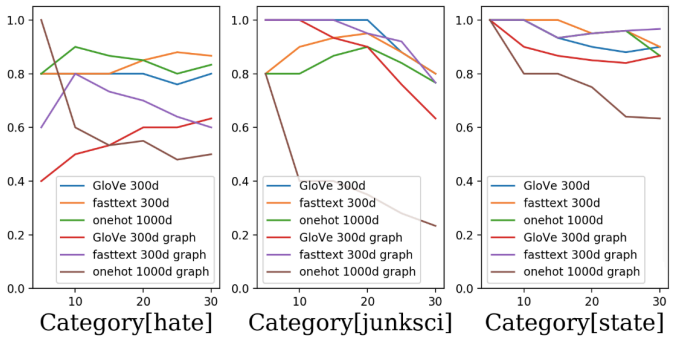

Figure 4 Results from top 30 news of bias, junksci and state respectively.

apply 10,195 pieces of news abstract as input and the whole network are trained for 3 epochs in 1950 steps. Finally, it achieved a quite well result as figure 5 illustrates.

4 4 Fake Fake-News Generating

Here, we modified the seqGAN model to be fit for our dataset and purpose then get some generations seems like to be semantic. And here are some examples of our results in table 2.

5 Conclusion and Future Work

In this paper, we proposed two neuron networks for fake news classification. LSTM based network has been verified to be more efficient than the baseline Tensor Decomposition by our experiments. As for GAN based network, we just construct the generator and classifier respectively, and from the results, we can see that CNN based classifier achieves better accuracy than LSTM. But on the other hand, the results of the generator is not well as expected.

In the future, we need to combine our classifier and generator together to get a better result and still need to conduct more experiments to prove our idea. Besides, in order to label the fake fake-news, we consider setting one particular

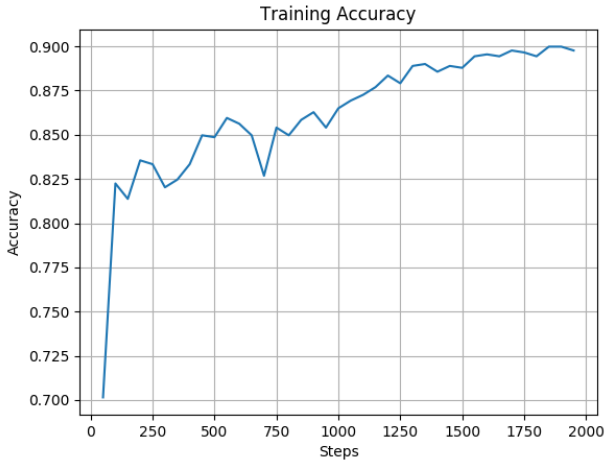

Figure 5 CNN based news abstract classification result

Table 2 Fake Fake-news Examples

1	a few people had a finding of a trump supporters to the presidential election and promises his connection to be the fact of the servers on the accused of dating.
2	Hillary Clinton ' s campaign chair massing the cities and for emails nor will stanss work in a collect her child from action here to get a supporters who support Hillary Clinton s campaign chairman.
3	Investigation of the remaining millions of the report issued by the white house with republicans in the united states is said sensation of the republicans would stand for healthy ones on Wednesday through the life.
4	The video described in the republicans is a fact to he stop near georgial community and promised to the will be out that they are the treated is a regular.
5	The emails start as hillary clinton email scandal that he is not enforcing the law some it s nom that the role the donald trump and healthy ones of the fbi and investigation into the presidential election of his committed to grab his pack and industry.

GAN for each kind of fake-news. And at present, our work is just concentrated on how to classify the fake-news due to the dataset, in the future we may find the datasets labeled by fake and real to train the Neuron networks to identify the fake news. And recently, some deeply contextualized word representation methods, such as BERT [21], have been verified to improve the accuracy of basic NLP tasks a lot. We want to apply them to our models in the future.

6 Acknowledgements

This work is partly supported by JSPS KAKENHI(16K12532) and The Kyoto University Foundation.

References

[1] Allcott, Hunt, and Matthew Gentzkow. "Social media and fake news in the 2016 election." Journal of Economic Perspectives 31.2 (2017): 211-36.

[2] Gottfried, Jeffrey, and Elisa Shearer. News Use Across Social Medial Platforms 2016. Pew Research Center, 2016.

[3] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[4] Zafarani, Reza, Mohammad Ali Abbasi, and Huan Liu. Social media mining: an introduction. Cambridge University Press, 2014.

[5] Centola, Damon. "The spread of behavior in an online social network experiment." science 329.5996 (2010): 1194-1197.

[6] Kempe, David, Jon Kleinberg, and Éva Tardos. "Maximizing the spread of influence through a social network." Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003.

[7] Shao, Chengcheng, et al. "Hoaxy: A platform for tracking online misinformation." Proceedings of the 25th international conference companion on world wide web. International World Wide Web Conferences Steering Committee, 2016.

[8] Hosseinimotlagh, Seyedmehdi, and Evangelos E. Papalexakis. "Unsupervised Content-Based Identification of Fake News Articles with Tensor Decomposition Ensembles." (2018).

[9] Goodfellow, Ian, et al. "Generative adversarial nets." Ad-

vances in neural information processing systems. 2014.

[10] Yu, Lantao, et al. "SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient." AAAI. 2017.

[11] Guo, Jiaxian, et al. "Long text generation via adversarial training with leaked information." arXiv preprint arXiv:1709.08624 (2017).

[12] Pennington, Jeffrey, Richard Socher, and Christopher Manning. "Glove: Global vectors for word representation." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014.

[13] Bojanowski, Piotr, et al. "Enriching word vectors with subword information." Transactions of the Association for Computational Linguistics 5 (2017): 135-146.

[14] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.

[15] Danqi Chen and Christopher D Manning. 2014. A Fast and Accurate Dependency Parser using Neural Networks. Proceedings of EMNLP 2014.

[16] Robbins, Herbert, and Sutton Monro. "A stochastic approximation method." Herbert Robbins Selected Papers. Springer, New York, NY, 1985. 102-109.

[17] Sutton, Richard S., et al. "Policy gradient methods for reinforcement learning with function approximation." Advances in neural information processing systems. 2000.

[18] Browne, Cameron B., et al. "A survey of monte carlo tree search methods." IEEE Transactions on Computational Intelligence and AI in games 4.1 (2012): 1-43.

[19] Fedus, William, Ian Goodfellow, and Andrew M. Dai. "Maskgan: Better text generation via filling in the _." arXiv preprint arXiv:1801.07736 (2018).

[20] Mihalcea, Rada, and Paul Tarau. "Textrank: Bringing order into text." Proceedings of the 2004 conference on empirical methods in natural language processing. 2004.

[21] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).