Hybrid Deep Learning Model for Fake News Detection in Social Networks (Student Abstract)

Bibek Upadhayay, Vahid Behzadan

University of New Haven, West Haven, CT, USA
bupadhayay@newhaven.edu, vbehzadan@newhaven.edu

Abstract

The proliferation of fake news has grown into a global concern with adverse socio-political and economical impact. In recent years, machine learning has emerged as a promising approach to the automation of detecting and tracking fake news at scale. Current state of the art in the identification of fake news is generally focused on semantic analysis of the text, resulting in promising performance in automated detection of fake news. However, fake news campaigns are also evolving in response to such new technologies by mimicking semantic features of genuine news, which can significantly affect the performance of fake news classifiers trained on contextually limited features. In this work, we propose a novel hybrid deep learning model for fake news detection that augments the semantic characteristics of the news with features extracted from the structure of the dissemination network. To this end, we first extend the LIAR dataset by integrating sentiment and affective features to the data, and then use a BERT-based model to obtain a representation of the text. Moreover, we propose a novel approach for fake news detection based on Graph Attention Networks to leverage the user-centric features and graph features of news residing social network in addition to the features extracted in the previous steps. Experimental evaluation of our approach shows classification accuracy of 97% on the Politifact dataset. We also examined the generalizability of our proposed model on the BuzzFeed dataset, resulting in an accuracy 89.50%.

Introduction

Fake news is defined as "fabricated information that mimics news media content in form but not in organizational process or intent" (Lazer et al. 2018). Fake news can also be defined as news that is false based on its authenticity (false or not), intention (bad or not), and whether the information is news or not (Zhou and Zafarani 2018). Previous work by Upadhayay et. al. (Upadhayay and Behzadan 2020) proposed the Sentimental-LIAR dataset which is an extended version of LIAR dataset (Wang 2017), by adding sentiment and emotions. This work is based on the Undeutsch hypothesis (Undeutsch 1967) and the Four-Factor theory (Zuckerman, DePaulo, and Rosenthal 1981), which claim that fake news statements are expressed with different emotions and sentiments than the genuine news. The aforementioned work follows the recent trend in the field in analyzing the written style of fake news using deep architectures, including Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs). However, the creators of fake news can deceive such style-based detection techniques by simply modifying the written style. To address this issue, a promising avenue is to explore the network structure of fakes news propagation in social media. Fake news travel faster with distinct propagation patterns than genuine news, and also differ in their cascade depth and width size (Zhou and Zafarani 2018). Accordingly, we propose a hybrid deep learning approach to fake news detection that exploit the written style of fake news in tandem with the features of their dissemination in social networks.

The evaluation of our approach is performed on the FakeNewsNet benchmark(Shu et al. 2019), which is composed of two datasets, namely Politifact and BuzzFeed. The features included in this benchmark are news text, user-user relationship, news-user relationship, and user profiles. The Politifact dataset consists of 240 news and 23866 user nodes, whereas BuzzFeed consists of 182 news and 15258 user nodes, and both datasets contain an equal numbers of fake and real news.

Hybrid Model

Our proposed method is a hybrid of both style-based detection and network-based detection. This hybrid process flow diagram is shown in Fig1. We first clean the news articles and then add the emotions and sentiments and pass the dataset to the BERT-base model and CNN model (pretrained on Sentimental LIAR dataset(Upadhayay and Behzadan 2020)) to extract the news representation, which is learned from the written style of news. The users' features are generated by passing the user-user relationship to the custom feature extraction method in the model. The user features are derived based on the network spreader pattern given by Zhou et al. (Zhou and Zafarani 2019) and consist of the following features for each user: *frequency of sharing fake news, frequency of sharing real news, number of involvements, frequency of involvements, closeness centrality, degree centrality, in the degree of user node, out-degree of user node*. The news feature dimension from the CNN model is padded to match the dimensionality of user profiles before creating a heterogeneous network. The network

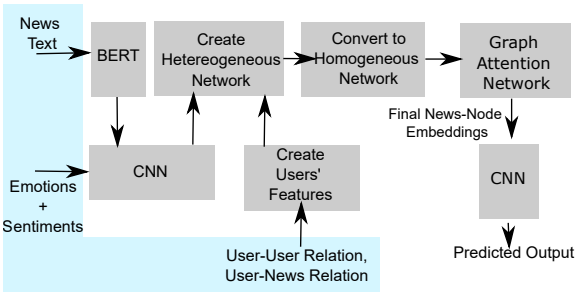

Figure 1: Hybrid Process Flow Diagram

only consists of a single meta-path of news-user-news which motivated the conversion of heterogeneous news networks to the homogeneous news network. The heterogeneous network is then converted to a homogeneous graph with one type of nodes, which is then passed to GAT to obtain node embeddings.

The importance of each news is determined by the users sharing them, hence we calculate the node level attention of news based on its neighboring users by using self-attention mechanism (Vaswani et al. 2017). We only perform masked level attention with first-degree neighbors (including the news node itself). Then we used the weight coefficients for each news-users node pair to determine the final embedding of the particular news node. Additionally, we implemented multi-head attention to stable the learning process. We train the GAT based on cross-entropy loss cost function the cross-entropy loss on the Politifact dataset.

Results and Discussion

By using network based detection in addition to the style based detection we received an accuracy of 97% on Politifact dataset, which is a significant improvement over the style-based benchmark accuracy of 70% reported by Upadhayay et. al. (Upadhayay and Behzadan 2020). This supports the hypothesis that augmenting network-based features improves the performance of style-based fake news classifiers. To examine the generalization of our model, we tested the trained model on the BuzzFeed Dataset, resulting in an accuracy of 89.50%.

The training loss versus validation loss graph is shown in Fig2 for 300 epochs. It is observed that the training loss is fluctuating, the probable reason could be due to training of large parameters with a small dataset of 240 news where each news is connected to different users, resulting in local convergence.

Conclusion and Future Works

Our proposed hybrid approach of exploiting both the written style and the news network properties successfully classifies the news into real or fake. The proposed method not only includes the exaggerated written style of news but will also include the network behavioral pattern of fake news media outlets. The experiment revealed that analyzing the news disseminated network will increase the potential of detecting

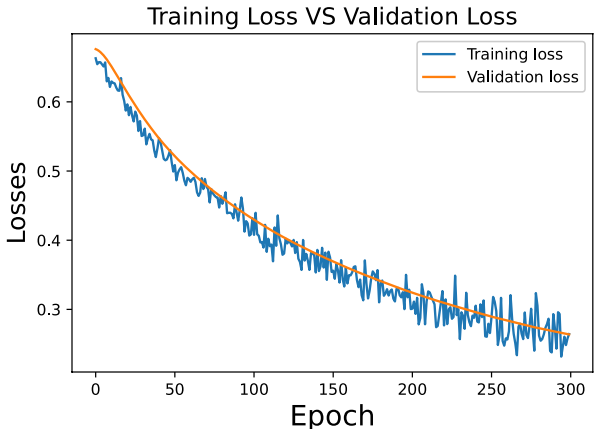

Figure 2: Training VS Validation Loss Plot

fake news. The generality of the model suggests the potential implementation across different network datasets such as Facebook and Reddit.

References

Lazer, D. M.; Baum, M. A.; Benkler, Y.; Berinsky, A. J.; Greenhill, K. M.; Menczer, F.; Metzger, M. J.; Nyhan, B.; Pennycook, G.; Rothschild, D.; et al. 2018. The science of fake news. *Science*, 359(6380): 1094–1096.

Shu, K.; Mahudeswaran, D.; Wang, S.; Lee, D.; and Liu, H. 2019. FakeNewsNet: A Data Repository with News Content, Social Context and Spatialtemporal Information for Studying Fake News on Social Media. arXiv:1809.01286.

Undeutsch, U. 1967. Beurteilung der glaubhaftigkeit von aussagen. *Handbuch der psychologie*, 11: 26–181.

Upadhayay, B.; and Behzadan, V. 2020. Sentimental LIAR: Extended Corpus and Deep Learning Models for Fake Claim Classification. In *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 1–6. IEEE.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762*.

Wang, W. Y. 2017. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. arXiv:1705.00648.

Zhou, X.; and Zafarani, R. 2018. Fake news: A survey of research, detection methods, and opportunities. *arXiv preprint arXiv:1812.00315*, 2.

Zhou, X.; and Zafarani, R. 2019. Network-based fake news detection: A pattern-driven approach. *ACM SIGKDD Explorations Newsletter*, 21(2): 48–60.

Zuckerman, M.; DePaulo, B. M.; and Rosenthal, R. 1981. Verbal and nonverbal communication of deception. In *Advances in experimental social psychology*, volume 14, 1–59. Elsevier.