SpotFake+: A Multimodal Framework for Fake News Detection via Transfer Learning (Student Abstract)

Shivangi Singhal[1], **Anubha Kabra**[2*], **Mohit Sharma**[3*]
Rajiv Ratn Shah[4], **Tanmoy Chakraborty**[5], **Ponnurangam Kumaraguru**[6]

[1,3,4,5,6]IIIT-Delhi, India, [2]DTU, India

(shivangis[1], rajivratn[4], tanmoy[5], pk[6])@iiitd.ac.in, anubhakabradtu@gmail.com[2], mohit.sharma.cs29@gmail.com[3]

Abstract

In recent years, there has been a substantial rise in the consumption of news via online platforms. The ease of publication and lack of editorial rigour in some of these platforms have further led to the proliferation of fake news. In this paper, we study the problem of detecting fake news on the FakeNewsNet repository, a collection of full length articles along with associated images. We present SpotFake+, a multimodal approach that leverages transfer learning to capture semantic and contextual information from the news articles and its associated images and achieves the better accuracy for fake news detection. To the best of our knowledge, this is the first work that performs a multimodal approach for fake news detection on a dataset that consists of full length articles. It outperforms the performance shown by both single modality and multiple-modality models. We also release the pretrained model for the benefit of the community.

Introduction

Online news platforms are becoming exceedingly popular amongst consumers due to their ease of access and their vast selection of disparate sources. These platforms are further democratizing news distribution by making it exceedingly simple to publish. The flip side of this is that the lack of proper editorial rigour, fact checking and the presence of bad actors have promulgated fake news to an equal extent.

In this paper, we consider the FakeNewsNet repository (Shu et al. 2018) for multimodal fake news detection. In contrast to existing datasets in this space, FakeNewsNet consists of full length articles rather than short claims (Wang 2017) or news in the form of tweets (Boididou and others 2015). FakeNewsNet also contains images associated with each article. Thus we believe that it is more representative of a news article.

Previous studies on FakeNewsNet have used various machine learning techniques (SVM, Naive Bayes, Logistic Regression) and deep learning models (CNN, LSTM, Attention) to perform the fake news detection but they fail to perform well due to following reasons: (1) they lacked the contextual information present in the text and (2) they do not capture the features from the image modality that may seek to emphasize certain facts.

To overcome the above mentioned challenges, we propose SpotFake+, an advanced version of existing multimodal fake news detection system, named SpotFake (Singhal et al. 2019). The proposed architecture leverages pretrained language transformers and pre-trained ImageNet models for feature extraction. These feature vectors are fed into fully connected layers for classification.

Methodology

Our methodology primarily consists of the following: (i) details of the dataset used in experiments, (ii) pre-processing of the dataset, and (iii) details of the text and image sub-module used in SpotFake+ for fake news detection.

Dataset

The FakeNewsNet repository (Shu et al. 2018) consists of two datasets from two different domains- politics and entertainment. Each news article has text and an image associated with it. The ground-truth labels for the political and entertainment domain were collected from Politifact[1] and Gossipcop[2] and E! Online[3], respectively. The number of samples present in the dataset is given in Table1.

Dataset	Politifact	Gossipcop
Real	624 (321)	16817 (10259)
Fake	432 (164)	5323 (2581)

Table 1: The number of samples in the FakeNewsNet repository. The values in the brackets indicate samples fit to use after data pre-processing.

Dataset Pre-processing

Before starting with the experiments, we manually removed logos from the articles, and dropped samples that either lacked images or contained GIFs. This resulted in around five hundred usable samples from Politifact and thirteen

---

[1]https://www.politifact.com/

[2]https://www.gossipcop.com/

[3]https://www.eonline.com/ap
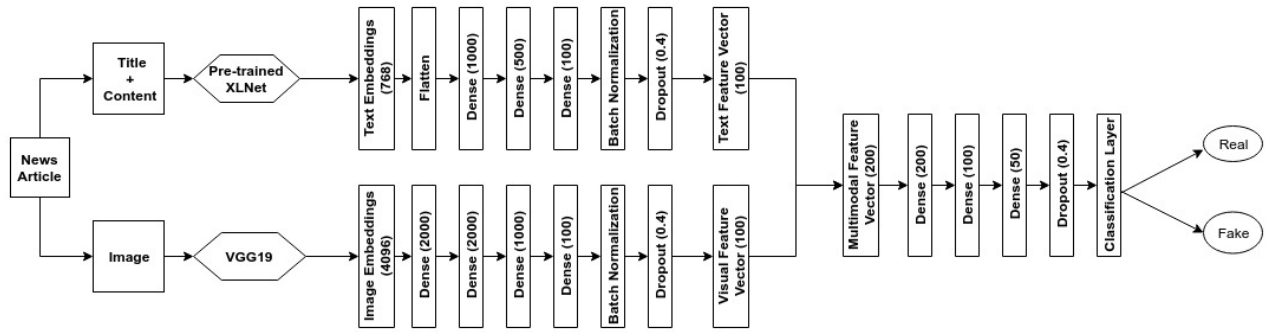
Figure 1: Our proposed SpotFake+ for Fake News Detection.

Modality	Models	Politifact	Gossipcop
Text	SVM	0.58	0.497
	Logistic Regression	0.642	0.648
	Naive Bayes	0.617	0.624
	CNN	0.629	0.723
	SAF (Social Article Fusion)	0.691	0.689
	XLNet + dense layer	**0.74**	**0.836**
	XLNet+ CNN	0.721	0.84
	XLNet + LSTM	0.721	0.807
Image	VGG19	0.654	0.80
Multimodal (Text+Image)	EANN (Wang et al. 2018)	0.74	0.86
	MVAE (Khattar et al. 2019)	0.673	0.775
	SpotFake (Singhal et al. 2019)	0.721	0.807
	SpotFake+ (XLNet + dense + VGG19)	**0.846**	**0.856**

Table 2: Comparison of accuracy on FakeNewsNet dataset. SpotFake+ is our proposed multimodal approach.

thousand samples from Gossipcop. Refer Table 1 for the distribution of real and fake news articles.

Proposed Model Architecture

The proposed SpotFake+ is a multimodal approach that successfully detects fake news in full-length articles. The schematic diagram of the model is shown in Figure 1. The proposed model has two sub-modules- a textual feature extractor and a visual feature extractor.

Results

SpotFake+ is compared against current state of the art text and other multiple-modality models (Wang et al. 2018; Khattar et al. 2019; Singhal et al. 2019) on the same dataset.

The detailed analysis of the results are shown in Table 2. The loss function graphs are also plotted in Figure 2.

Conclusion and Future Works

In this paper, we present SpotFake+, an advanced version of SpotFake. Given a news article, SpotFake+ can classify it into two categories: real or fake.The proposed architecture uses transfer learning to capture the textual and visual features within an article. The experiments performed in this paper further reveal the potential of multimodal features for the problem of fake news detection. In the future, experiments can be performed to study the contribution of each modality towards solving the problem. The work can further be expanded to incorporate meta level feature modalities.
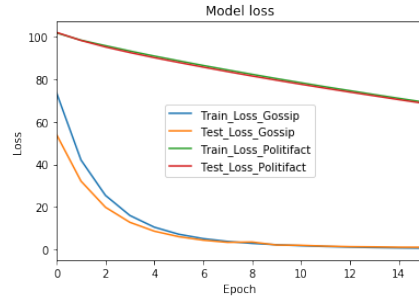

Figure 2: Loss function graphs on GossipCop and Politifact datasets by SpotFake+.

We also release the pretrained model[4] for the benefit of the community. We believe this model would be useful in fake news detection tasks for full length articles.

References

Boididou, C., et al. 2015. Verifying multimedia use at mediaeval 2015. In *MediaEval 2015 Workshop*.

Khattar, D.; Goud, J. S.; Gupta, M.; and Varma, V. 2019. Mvae: Multimodal variational autoencoder for fake news detection. In *WWW*.

Shu, K.; Mahudeswaran, D.; Wang, S.; Lee, D.; and Liu, H. 2018. Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. *CoRR* abs/1809.01286.

Singhal, S.; Shah, R.; Chakraborty, T.; Kumaraguru, P.; and Satoh, S. 2019. Spotfake: A multimodal framework for fake news detection. In *IEEE BigMM*.

Wang, Y.; Ma, F.; Jin, Z.; Yuan, Y.; Xun, G.; Jha, K.; Su, L.; and Gao, J. 2018. Eann: Event adversarial neural networks for multi-modal fake news detection. In *KDD*.

Wang, W. Y. 2017. "liar, liar pants on fire": A new benchmark dataset for fake news detection. *CoRR* abs/1705.00648.

---

[4]https://github.com/shiivangii/SpotFakePlus