SpotFake+: A Multimodal Framework for Fake News Detection via Transfer Learning

Shivangi Singhal, Anubha Kabra*, Mohit Sharma, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru**
(shivangis , rajivratn, tanmoy , pk)@iiitd.ac.in, anubhakabradtu@gmail.com*, mohit.sharma.cs29@gmail.com **

Motivation

Online news platforms are becoming exceedingly popular amongst consumers due to their ease of access and their vast selection of disparate sources.

News obtained from such platforms are generally of full length as compared to the one present on any social media.

Previous studies have solved the problem of multimodal fake news detection on datasets consisting of news in the form of tweets (Boididou and others 2015).

In this paper, we study the problem of detecting fake news on 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.

Dataset

In this paper, we consider the FakeNewsNet repository (Shu et al. 2018) for multimodal fake news detection.

FakeNewsNet consists of full length articles rather than short claims or news in the form of tweets.

The repository 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 and Gossipcop and E! Online, respectively.

Dataset	Politifact	GossipCop
Real	624 (321)	16817 (10529)
Fake	432 (164)	5223 (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

We manually remove logos from the articles, and dropped samples that either lacked images or contained GIFs.

Figure 1: Image samples from FakeNewsNet repository that were discarded after the cleaning process.

Model Architecture

The schematic diagram of the model is shown in Figure 2.

Textual Feature Extractor
This is a sub-module of SpotFake+ that is responsible of extracting the contextual text features from the posts.
It uses pre-trained XLNet that represent words and sentences in a way that best captures underlying semantic and contextual meaning.

Visual Feature Extractor
We employ the pre-trained VGG-19. We extract the output of the second last layer of VGG-19 convolutional network pre-trained on ImageNet dataset.

Multimodal Fusion
The two feature vectors obtained via different modalities are fused together using simple concatenation technique to obtain the desired news representation.
This news representation is then passed through a fully connected neural network for fake news classification.

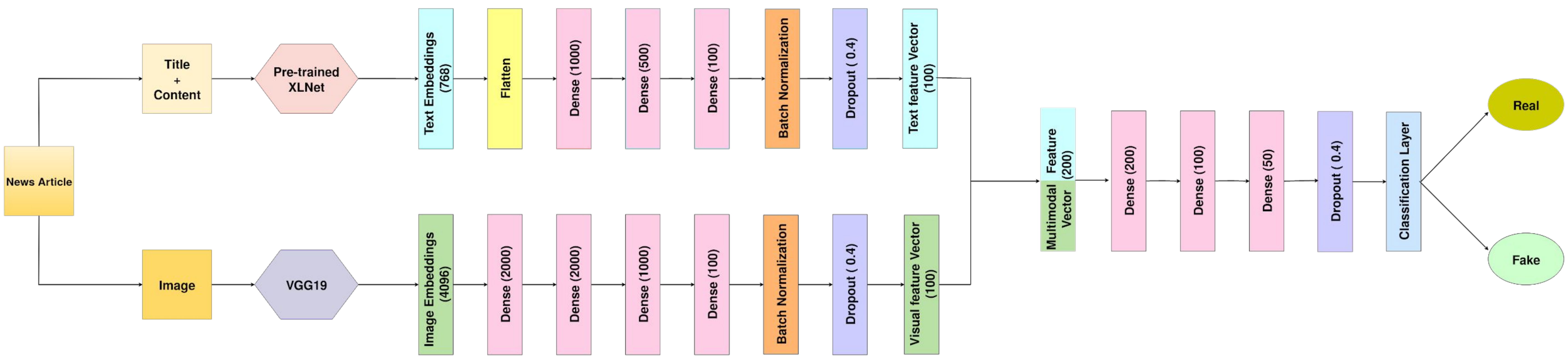

Figure 2: Our proposed SpotFake+ for Fake News Detection

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 3.

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
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.

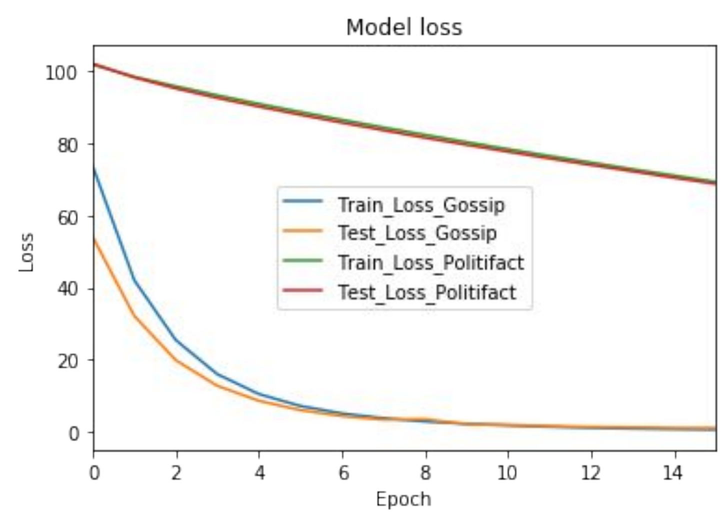

Figure 3: Loss function graphs on GossipCop and Politifact datasets by SpotFake+.

Parameters	Unimodal Text Models (XLNet + dense layer)	Image Models (VGG)
Input feature dim	768	4096
# of dense layers	3 (100,500,100)	3 (2000,1000,100)
Output feature dim	100	100
Dropout	0.4	0.4
Activation	ReLu	ReLu
Optimizer	SGD	SGD
Batch Size	32	32

Table 3: List of Hyperparameters used in SpotFake+.

Samples identified by proposed SpotFake+

Figure 4: Samples of fake data from GossipCop that were incorrectly predicted as real by Unimodal text models but were correctly predicted as fake by SpotFake+.

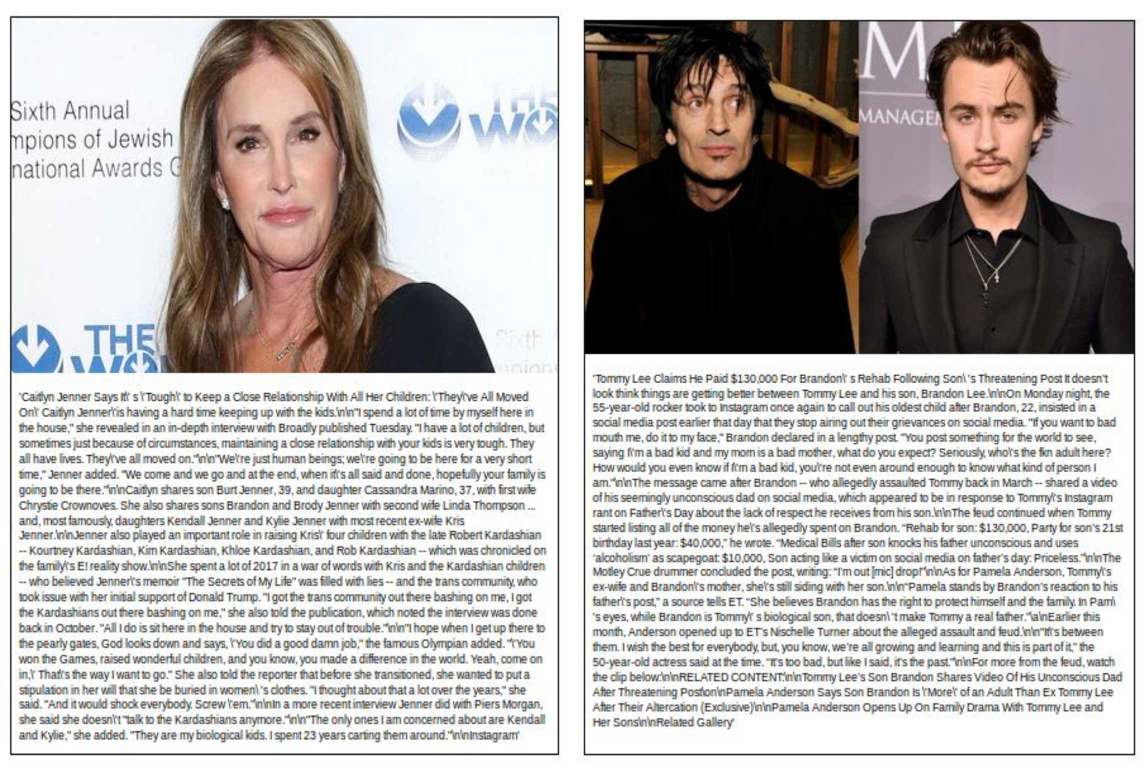

Figure 5: Samples of real data from GossipCop that were incorrectly predicted as fake by Unimodal text models but were correctly predicted as real by SpotFake+.

References

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

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.

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.

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

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

Code and dataset is available at: https://github.com/shiivangii/SpotFakePlus

MIDAS@IIITD
Multimodal Digital Media Analysis Lab