Unsupervised Fake News Detection

Kalit Inani Preetika Verma

I. Introduction

A. Background

Fake news dissemination has been a concerning issue since the invention of the printing press in the 15th century. The spread of fake news and misinformation has brought disastrous consequences many times. Therefore, it is important to develop an algorithm to curb the spread of fake news which can create panic and confusion among people. Fact-checking websites like Politifact and Gossipcop are being extensively used which check the accuracy of statements. Researchers have divided fake news into seven categories - false news, polarised content, satire, misreporting, commentary, persuasive information, and citizen journalism. It has also been found that fake news articles are less factual, less grammatically correct, and have more emotionally charged claims. Analyzing the linguistic features of text has also been proved to help in classification.

B. Motivation

However, manual identification of fake news requires experts in various fields and a huge workload. Therefore, it is crucial to develop algorithms to curb the spread of fake news. Most existing methods of fake news detection are supervised. This requires extensive time and effort to build reliable datasets. Thus, there arises a need to automate this using some tool in an unsupervised manner. Unsupervised approaches seem more practical today and are a quite new and evolving area in research. There are majorly four features that serve to classify a particular article as fake or not - text content, social context, propagation information, and multimedia content.

C. Objective

Our work aims to use these features and use techniques like measuring reconstruction errors generated by autoencoder training, clustering articles in a lower-dimensional space (using latent vector generated by autoencoder), and trying out a semi-supervised approach where we use the latent vector produced by an autoencoder to fine-tune for the task. We want to develop a multimodal variational autoencoder that learns the shared representations of the modalities. It would reconstruct all the modalities for the learned shared representation and thus discover correlations among them.

II. Related Works with Comparison

A. MVAE: Multimodal Variational Autoencoder for Fake News Detection

Current approaches for the detection of fake news is their inability to learn a shared representation of multimodal (textual + visual) information. So, the authors of the paper [1] proposed an end-to-end network that uses a bimodal variational autoencoder coupled with a binary classifier. VAEs can learn probabilistic latent variable models by optimizing abound on the marginal likelihood of the observed data.

The components of the model are presented below:

- Encoder:
  - Input: Text of the post as well as the image and social context features attached to it
  - Output: a shared representation of the features learned from all modalities
  - Textual Encoder: Extract features using RNNs and LSTMs
  - Visual Encoder: Use pre-trained VGG-19 network
- Decoder: to reconstruct data from the sampled multimodal representation
  - Textual Decoder
  - Visual Decoder

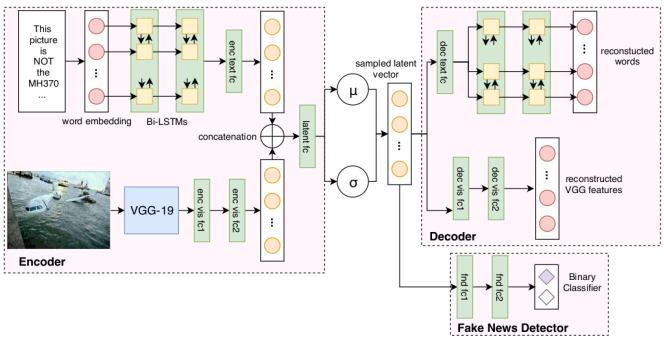

Fig. 1. Network architecture of MVAE having encoder, decoder, and classifier

B. Unsupervised Fake News Detection Based on Autoencoder

The authors of the paper 'Unsupervised Fake News Detection Based on Autoencoder' [2] regard fake news as an anomaly in the collection of articles. They have combined text content, images, propagation, and user information and fed these into an autoencoder model.

The basic idea for the approach is that the article which gives a reconstruction error higher than a specified threshold tends to be a fake article. The main features of the model architecture are presented below:

- Fusion of the features: concatenation
- Error function: RMSE
- Components:
  - Encoder: BI-GRU - Repeat - Dense

- Self-Attention
- Decoder: BI-GRU - Dense - Time Distributed

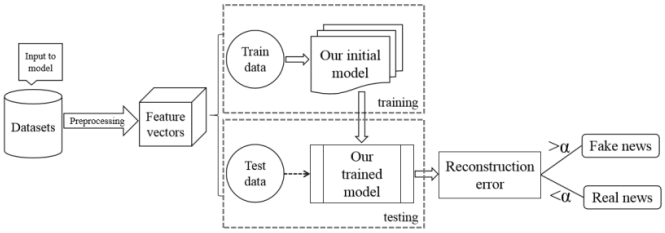

Fig. 2. Using a threshold for reconstruction error

C. Multimodal Multi-image Fake News Detection

The authors of this paper [3] proposed a multimodal multi-image system. They used BERT to better capture the underlying semantic and contextual meaning of the text for textual representation. For the visual representation, they extracted image tags from multiple images that the articles contain using the VGG-16 model. They also used the similarity between the image and the title of the articles. The text-image similarity can be a valuable indicator, especially in the cases that the images are chosen randomly and do not correspond to the article.

III. Proposed Techniques and Algorithms

Surveying all the above research papers, we decided to experiment on the following techniques:

- Our approach is based upon the reconstruction error classification. We plan to combine text and image features as well as social context features. Moreover, we would add the title text - image similarity feature in the input.
- The next idea is to cluster the articles in a lower-dimensional space generated by the latent vector. The real and fake articles could get separated into several separate clusters.
- One more approach is to model the problem in a semi-supervised way. We could use the latent vector generated by the autoencoder and pass it as an input to a neural network classifier. Moreover, we could add the social context features as well as the text-image similarity features which were not being considered by other authors.

IV. Datasets used with description

Since we wanted to model the problem in a multimodal way, we had to find an appropriate dataset with structured multimedia content along with social context features. Here are some of the datasets we went through:

- Twitter Dataset: Released as a part of MediaEval for verifying multimedia use task. The task was aimed at detecting fake multimedia content on social media. The dataset consists of tweets and each tweet has textual content, image, and social context information.
- Weibo Dataset: Consists of data collected from Xinhua News Agency, an authoritative news source of China, and Weibo, a Chinese microblogging website. Collected data is verified by Weibo's official rumor debunking system.
- FakeNewsNet Dataset: This dataset [5] is specifically created for research in the area of fake news detection. They have collected two types of articles - Politifact and Gossipcop. This dataset has multiple dimensions of data like news articles, tweets, retweets, user profile information, user timeline tweets, user followers, and user following.

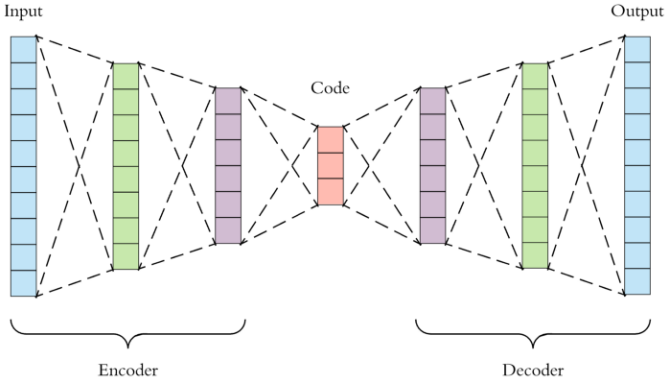

Fig. 3. Latent vector generated by encoder which is fed into the decoder

We decided to try our approach on the FakeNewsNet repository. Statistics for the dataset are shown in the Figure 4. The collection of the dataset was a tedious task and it took about a week to download it completely. The articles for real and fake news are downloaded in separate folders, thereby serving as a label for our dataset.

Flowchart for dataset integration:

- The headlines of the news articles were used as search queries for collecting user engagements using Twitter API.
- URLs are collected from the tweets and are further used as search queries to collect additional tweets.
- User response toward these posts such as replies, likes, and reposts is fetched. For spatial information, locations explicitly provided in user profiles are obtained. For temporal information, timestamps of user engagements are recorded.

V. Experiments and Results

A. Experiments

- We started with creating an encoder-decoder model for the text of the news articles.
- The text was preprocessed and words were converted into numeric representations using word embeddings. Pre-trained GloVe embeddings were used.
- To keep the input text of each data point have a similar structure, different text summarizer pre-trained models have been experimented with Extractive text summarization and Abstractive text summarization

	Type	Features	Politifact		Gossipcop	
			Fake	Real	Fake	Real
News content	Linguistic	#News articles	432	624	5323	16817
		#News articles with text	420	528	4947	16694
	Visual	#News articles with images	336	447	1650	16767
Social context	User	#Users posting tweets	95553	249887	26155	80137
		#Users involved in likes	113473	401363	348852	145078
		#Users involved in retweets	106195	346459	239483	118894
		#Users involved in replies	40585	186675	106325	50799
	Post	#Tweets posting news	164892	399237	519581	876967
	Response	#Tweets with replies	11975	41852	39717	11912
		#Tweets with likes	31692	93839	96906	41889
		#Tweets with retweets	23489	67035	56552	24955

Fig. 4. FakeNewsNet dataset statistics

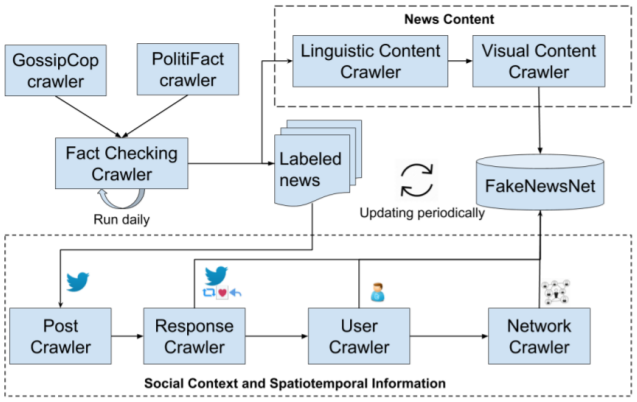

Fig. 5. Web scraping for generating the dataset

- We created a simple neural network to check if there was any problem with preprocessing. We experimented with two encoder-decoder architectures.
- The first architecture has a bi-directional LSTM with max pooling. The encoder has two LSTM layers while the decoder has an LSTM layer followed by pooling and a dense layer. We use mean square error, batch size of 64, Adam optimizer, and train for 20 epochs.
- After calculating the reconstruction error, we searched for the optimal threshold.

B. Results

We found out that there is no clustering as we had expected and the model is unable to reconstruct properly. The accuracy and F1 score were around 0.506.

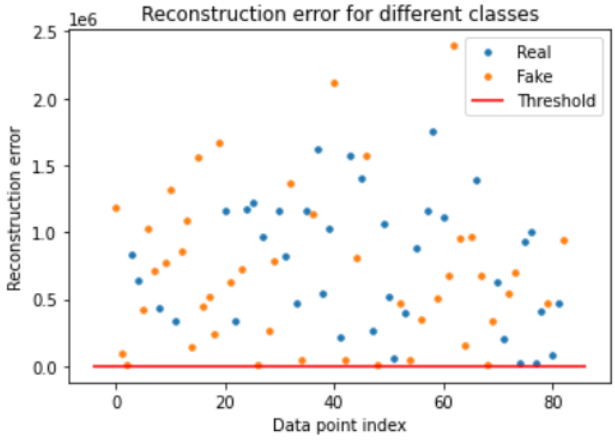

Fig. 6. Results for vanilla autoencoder on politifact

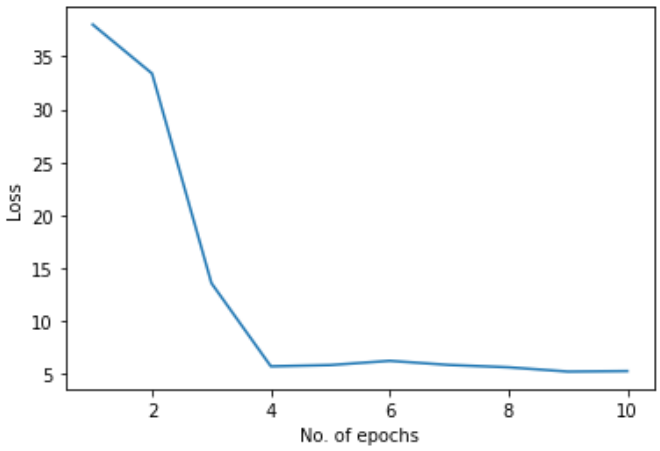

Fig. 7. Training curve for VAE

C. Major challenges

- Completely downloading the FakeNewsNet repository with all the features like news article text, user tweets, retweets, user profile information, images, and many more was a tedious task. Retweets for the Gossipcop are still being downloaded.
- The data provided is not complete in the sense that some articles have news article information missing while others have no tweets or retweets. This makes it difficult to design a model architecture that works for all types of data inputs.
- Considering the images as features, some of the articles contain images that are not relevant. Thus, there arises a need to manually filter all such articles.

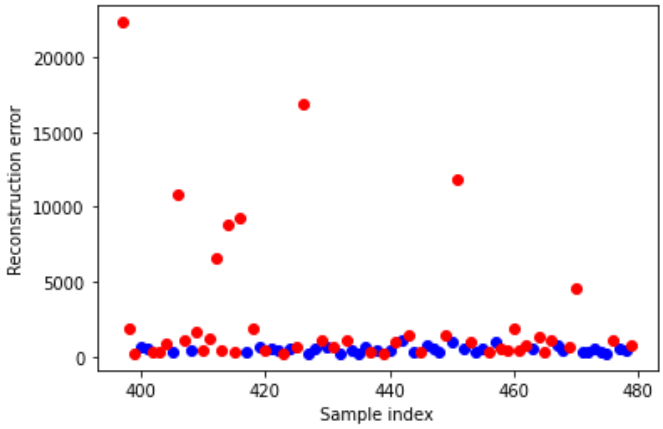

Fig. 8. Reconstruction error plot, where red points: fake articles, and blue points: real articles

- The size of the Gossipcop dataset is huge to work with. For eg: While trying to generate a summarized text for the news article content, we were facing a 'GPU out of memory error'. Thus, there is a need for better processing (GPU) and storage (high RAM) devices.

VI. Conclusion and Future Work

We need improved methods for text reconstruction. The fake and real data are not clustering as we had expected. Reconstruction errors are in a very narrow range for the entire training set. Even using a variational autoencoder after a vanilla autoencoder is not making much difference. We also need to incorporate more features as only textual features are not giving proper results. The images in the dataset have to be manually selected as each article has several images. Text-image similarity can also be incorporated into the feature vector.

VII. Work Update

A. Work Completed

Given below is the timeline of the work completed:

- August: Understanding Natural Language Processing concepts and implementing the FIRE tasks
- September: Understanding transformers, writing paper for FIRE task, and improving results
- October: Surveying recent fake news detection methods
- November: Analyzed several techniques, surveyed datasets, downloaded FakeNewsNet dataset, performed data preprocessing, text summarization.
- December: Experimenting with the autoencoder and variational autoencoders for text reconstruction on Politifact and Gossipcop datasets.

VIII. Work Remaining

We are doing a literature review on text reconstruction using autoencoders. We have to manually select the images from the datasets to add visual features. We also plan to incorporate text-image similarity features and social context and survey different fusion techniques.

References

[1] K. Dhruv, G. J. Singh, G. Manish, and V. Vasudeva, "MVAE: Multimodal variational autoencoder for fake news detection," in Proc. ACM World Wide Web Conf., 2019, pp. 2915–2921.

[2] Li, D., Guo, H., Wang, Z., & Zheng, Z. (2021). Unsupervised Fake News Detection Based on Autoencoder. IEEE Access, 9, 29356-29365.

[3] Giachanou, Anastasia, Guobiao Zhang, and Paolo Rosso. "Multimodal multi-image fake news detection." 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2020.

[4] Singhal, S.; Shah, R.; Chakraborty, T.; and Kumaraguru, P. 2019. Spotfake+: A multimodal framework for fake news detection via transfer learning.

[5] 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.