COVFake: A Word Embedding Coupled with LSTM Approach for COVID Related Fake News Detection

Muhammad Usama Islam
School of Computing and Informatics
University of Louisiana
at Lafayette

Md. Mobarak Hossain
Department of CSE
Dhaka University of
Engineering and Technology

Mohammod Abul Kashem
Department of CSE
Dhaka University of
Engineering and Technology

ABSTRACT

Coronavirus (COVID) took a substantial toll on human life with its unprecedented arrival in human sphere. An unforeseen circumstance which lead to various types of guidelines of procedures directed from the monitoring bodies including face-mask guideline, hand-wash guidelines and so forth. However, with the advent of this disease, misinformation became a causal factor to this scenario albeit claiming millions of life in the process. A threatening disease coupled with misinformation has created a disastrous scenario in human life. Our approach, exploits the power of natural language processing, specifically word embedding and Long short term memory (LSTM) to detect the COVID related fake news. Our model performs with a promising accuracy of 96% which concludes our effort of contribution to this massive outbreak from a linguistic standpoint.

General Terms

Coronavirus, Fake news

Keywords

Coronavirus news, Fake news, Natural language processing, text analysis, Long short term memory, Word embedding, Fake news detection

1. INTRODUCTION

Fake news as a term gained immense popularity with the advent of social media where alternative media companies started twisting the truth towards a certain narrative to feed false narratives or promote conspiracy theories [23]. While one of its major blows was at elections [3], the major effect was human becoming polarized more than ever in taking certain decisions [26].Fake news not only has changed how a government shall respond to an issue but also changed the perspective of how its population would react to the policies. To a larger extent, misinformation and fake news became intertwined to form own policy guidelines that started dictating the lives of the people [5]. Computational linguists having understood the harsh rhetoric and effect of fake news, has successfully started working towards natural language processing based models to detect the fake news [19]

Coronavirus, a virus of SARS family, affected the lifespan of this entire planet starting from late December 2019 by claiming mil-

lions of lives ultimately transforming to a pandemic [18].With new guidelines being issued from the regulating authorities such as World Health Organization (WHO), Center for Disease Control(CDC) and so forth, it was anticipated that the polarized community coupled with misinformation would be exposed to fake news in COVID era [16, 4].

In our investigation, we tried to accumulate these two factors in to demonstrate the effect of fake news on COVID followed by a detection model inspired by LSTM [9, 24] and word embedding [10, 8] to classify and detect the COVID related fake news.

The rest of the paper is organized as follows. Section 2 discusses the related works carried out on COVID related fake news followed by the impending motivation to carry out the research on this field. Section 3 contains our proposed approach that includes the dataset description, dataset preparation, and language processing model.Afterwards, Section 4 discusses about implementation and evaluation of our system followed by an analysis of results and lastly Section 5 contains the conclusion and future scope of this work.

2. LITERATURE REVIEW

Various research works has been carried out in natural language processing and understanding with the aid of machine learning and deep learning algorithms [12, 30].Ahmed and colleagues [2] has investigated fake news and built a classifier by utilizing unigarm features and support vector machine (SVM) classifer of linear nature to classify the fake news. Similar works on SVM as well as logistic regression is performed in [27].Granik.et.al [11] took a different approach and used naive bayes classifier to detect the fake news. An interesting co-similarity of all the research work above is the utilization of machine learning algorithms for fake news detection approach which later paved the way for word embedding [10, 8], LSTM [9, 24] and deep learning [22] to take over. LSTMs as well as word embedding has been well utilized for text based analysis for the case natural language processing, sentiment analysis and opinion mining [10, 8]. Although, use of LSTM has seen a boom since the data explosion phenomenon since 2013 albeit being developed in late 90's but word embedding has been utilized widely for text based representation and understanding. Liang Wu and Huan Liu [28] has investigated various embedding including graph embedding to understand and perceive the footprint of fake-news delivery through social network.Similar research work on fake news detection has been carried out in [7]. While substantial work has

been carried out on fake-news detection, but little is investigated into COVID related fake news.The importance pertaining to the matter is that, there are various guidelines, facts in place by monitoring bodies which are being twisted to the extent of false narrative [25].A modeling based predictor has been proposed in a study carried out in Nigeria [4].Ensemble method has been used in [1] followed by Xue and colleagues approach to detect fake news in twitter [29].Machine learning approaches has been investigated to understand the learning patterns in [13] that explores the mental health and effect mechanism of COVID on human-life. The insinuation of fake news extremely is rapid because of early suppression of COVID news by various governments in several parts of the world [17].The importance of research carried out in fake news detection not only proliferates in the domain of misinformation but to fight injustice and racism that has been created through the propaganda as discussed in [21]. All this factors has contributed to the importance of carrying out research work for fake news detection in COVID news articles which became the penultimate aim of this paper.

3. PROPOSED METHODOLOGY

3.1 Dataset Description

The dataset that we have used for our experiment is taken from COVID dataset developed by Sumit et.al [6] that contains a total of 10201 text article headlines which are labelled into 2 classes of classification of being fake and not-fake. In our proposed methodology We did not directly use the dataset rather performed some pre-processing stages before feeding into our model architecture. A description of dataset based on values related to classes is depicted in Figure 1 which shows that the dataset is skewed towards Class 0 more than that of Class 1.
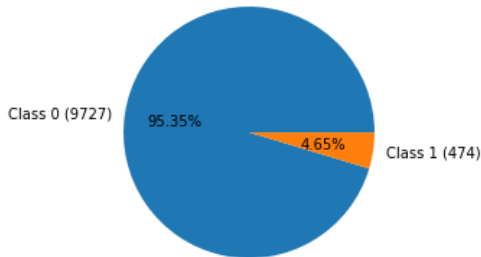

Fig. 1. Percentage of Class 0 and Class 1 values

3.2 Dataset Preparation and Pre-processing

The dataset is cleaned to drop any null values that it may contain. The sample examples of labelled dataset can be visualized in Figure 2. We can observe that, there are texts of COVID related news followed by a labelled outcome or target.Before dividing into train-test split dataset, we vectorize our dataset after having gone through the procedure of eliminating the stop-words often categorized as unnecessary words and stemming the text. For this purpose, we have used porter stemmer [20] and tokenized our words for the purpose of word-embedding followed by the inclusion of LSTM. The theory and process of word-embedding and LSTM is discussed in next section.

3.3 Word Embedding

Word embedding is a type of dense vector space representation of texts that succeeds the vector space model [15]. The theory suggests that, for a series of words, $T_1....T_n$, if there is a text T_i then, the context of T_i would be provided by its temporal locality.The concept is visualized in Figure 3. The probability of the concept
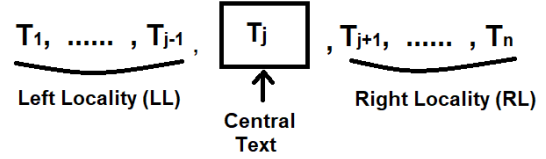

Fig. 3. Embedding Locality Concept

is mathematically implemented with the vectorized term of the T value (can be word, byte, text) and represented through softmax activation function shown in equation 1

$$P(T_0|T_i) \quad = \frac{e^{db}(V_{ti} \times V_{to}^T)}{\sum_{c=1}^{c} e^{dc}(V_{ti} \times V_{tj}^T)} \quad \text{for} \quad b = 1, 2, ...c \quad (1)$$

The softmax activation function itself can be visualized below from where the vectorized form of probability for word embedding in equation 1 has been derived is showcased below:

$$s(d)_b \quad = \frac{e^{db}}{\sum_{c=1}^{c} e^{dc}} \quad \text{for} \quad b = 1, 2, ...c$$

To predict the context of the central word, T_j as discussed in Figure 3 we characterize the objective function as follows in equation 2:

$$OF_T = \frac{1}{N} \sum_{i=1}^{N} \sum_{j \epsilon i}^{N} log(P(T_j|T_i)) \quad (2)$$

Following the prediction model of contexts, the loss function, is characterized by the following equation characterized as equation 3.

$$LF_T = -\frac{1}{N} \sum_{i=1}^{N} \sum_{j \epsilon i}^{N} log(P(T_j|T_i)) \quad (3)$$

In case of binary classification problem which is the case in our investigation, the concept of negative sampling comes into play as its idea is to basically transform a multi-class classification to a binary classification where the model learns to differentiate between the true and negative pairs. This idea leads to the change of our objective function to the following equation where instead of softmax, sigmoid function has been used thus changing our equation from equation 1 to equation 4. The equation 4 describes the exploitation of negative sampling method that leads to the following equation.

$$P(T_0|T_i)(log(\sigma(T_{wo} \times T_{wi}^T))) + \sum_{j=1}^{k} log(\sigma((-T_{wj} \times T_{wi}^T)) \quad (4)$$

```
                                        headlines  outcome
2947  If These 10 States Beat Covid, India Can Win, ...      1
                                        headlines  outcome
2946  The Spanish Police acts to sanction citizens o...      0
                                        headlines  outcome
2945  A video (already removed from YouTube) in whic...      0
                                        headlines  outcome
2948  Google launches new global Covid-19 map for jo...      1
                                        headlines  outcome
2944  The Head of Government of the City of Buenos A...      0
                                        headlines  outcome
2949  How to make COVID vaccine at home' was the sec...      1
```

Fig. 2.   COVID Fake news dataset sample.

3.4 Long Short Term Memory (LSTM)

Before we go into the details of LSTM [9, 24], we need to contextualize by understanding the problem of Recurrent Neural Netowrks (RNN) and its vanishing gradient problem [14]. It can be defined as the in-ability to remember a long sequence of memory which leads to forgetting and leaving out important information. To address this problem, we have used LSTM in our method where, cell state and gates acts as a form of highway that transfers information through the sequence chain including the relevant information from earlier times hence solving the vanishing gradient problem. The figure for LSTM methodology is provided in Figure 4 to better visualize the concept of LSTM. Here,
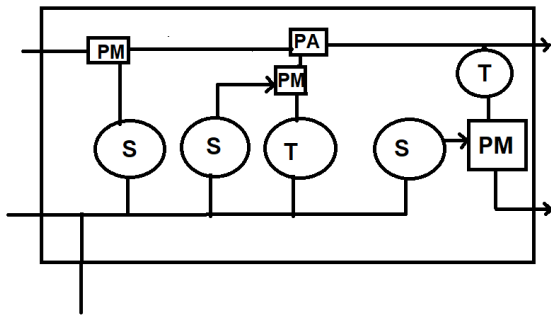

Fig. 4.   Working Process of LSTM

—Pointwise multiplication is denoted by "PM".

—Pointwise addition is denoted by "PA".

—Sigmoid activation is denoted by "S".

—tanh activation is denoted by "T".

—cell state is gien by horizontal line.

—Concatenation of vector is denoted by arrow sign.

Ideally, a forget gate, an inclusive part of LSTM is implemented before-hand to foresee whether or not an information would be kept in sequence or would be thrown out which is later fed into the LSTM network. for classification purposes, we have used binary cross entropy as loss function depicted in equation 5.

$$BCE(l_f) = -(l_f log(p) + (1 - l_f) log(1 - p)) \quad (5)$$

3.5 Proposed Approach

Our proposed approach can be categorized as follows:

(1) Collect the data from dataset [6]

(2) Divide data into train-test 80-20 split

(3) Pre-process the data
   (a) Eliminate null rows
   (b) Use stop-words elimination technique
   (c) Stem words with porter stemmer

(4) Perform one-hot encoding for word embedding

(5) Build model by adding embedding layer

(6) Add LSTM layer

(7) Add dense layer for output purpose

(8) Train on 80% data with validation

(9) Test on 20% data

(10) Save the trained model weights

4. EVALUATION AND RESULT ANALYSIS

4.1 Model Training

A total of 373,501 trainable paramaters were used in model training purposes of which word embedding contanied 315000 parameters, Long Short Term Memory(LSTM) layer carried 58400 parameters and Dense layer carried 101 parameters. The model has been trained on 8160 samples and was validated on 2041 samples thus maintaining the 80-20 train-test split ratio in our experimentation process.

4.2 Performance Analysis

The performance analysis has been visualized through generating the graph for accuracy and loss for COVFake model depicted in Figure 5 and Figure 6. As Depicted in Figure 1, the skewness and maximum of the data being classified towards Class 0 has created a substantial effect on training and testing. Figure 5 has training and validation accuracy of COVFake model where we can see a fairly good training accuracy and validation accuracy within 0.96 to 1.00 range. As for loss graph depicted in Figure 6 that has the training and validation loss of COVFake model, we can observe a steady loss curve for training loss albeit validation loss getting higher as epochs tends to increase.
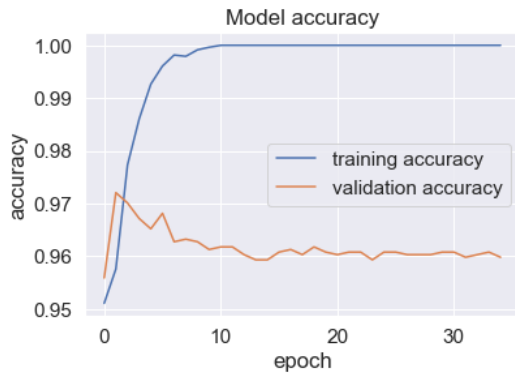
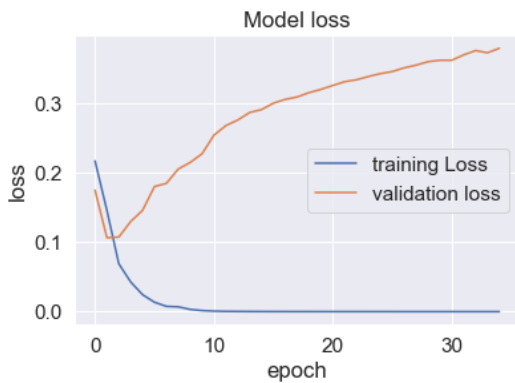Fig. 5. Training and Validation Accuracy for COVFake

Fig. 6. Training and Validation Loss for COVFake

4.3 Result Analysis

The evaluation and result analysis is performed in our experimentation through generating the Confusion matrix as depicted in Figure 7.
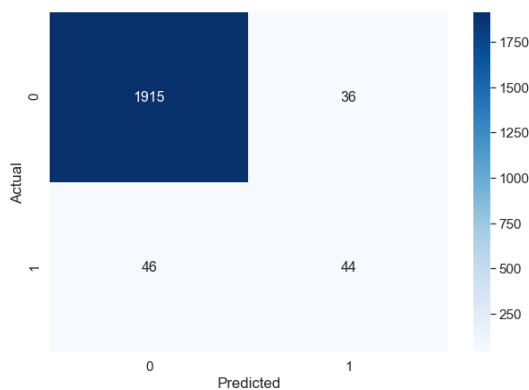

Fig. 7. Confusion Matrix for COVFake

The confusion matrix suggests that, despite the skewness of data and problem of over-fitting, the model performed pretty good. The

performance of the model has also been summarized from the classification score of Precision, Recall and F1-score generated to visualize our result as observed in Table 1. In Table 1, we can observe that, although, Class 0 had good results of precision, recall and F1-score, Class 1 has suffered which is more likely to be the result of disparity of data being non-uniform in nature. Furthermore, The dataset is also run-through logistic regression, naive bayes classifier, support vector machine and compared with COVFake to visualize the performance of our model. The Figure 8 has comparative analysis between LR, NB, SVM and COVFake which shows that our model COVFake that has utilized word embedding and LSTM has out-performed the traditional machine learning algorithms with a better performance. Finally, we can conclude that, our model with
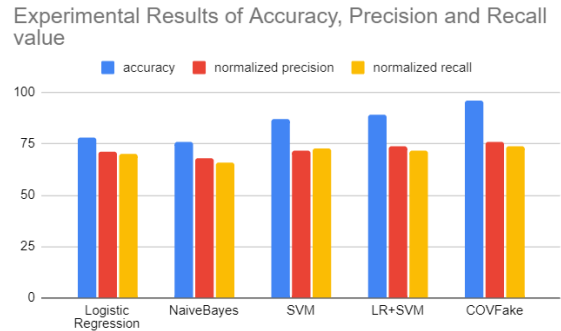

Fig. 8. Comparative analysis

Accuracy of 96% out-performs the results of other classifiers and shows promising result albeit skewed dataset with metrics such as precision, recall and F1-score.

5. CONCLUSION

This proposed model has presented a better performance of classification on fake news data on COVID. The dataset achieved 96% accuracy on the dataset. As a result, we are able conclude the hypothesis of using LSTM and word embedding as a viable model for natural langugae processing, specifically on fake news detection related to COVID. Sometimes the proposed model confused to understand the fake news due to overfitting as well as skewness of data. Future work would include increasing the dataset size and making the dataset uniform of 50-50 classes which will help to perform the model better. Also, tweaking the hyper parameters would be a viable future research work to obtain better accuracy.Furthermore, LSTMs can be further exploited to Bi-directional LSTMs and Gated Recurrent Units (GRUs) to understand the effect of these models.The implementation of the investigation and experiment is available in Github [1]

6. REFERENCES

[1] Iftikhar Ahmad, Muhammad Yousaf, Suhail Yousaf, and Muhammad Ovais Ahmad. Fake news detection using machine learning ensemble methods. *Complexity*, 2020, 2020.

---

[1]COVFake implementation.

Table 1. Precision, Recall, F1-score and Support for COVFake

Class Label/ Average	Precision	Recall	F1-score	Support
Class 0	0.98	0.98	0.98	1951
Class 1	0.55	0.49	0.52	90
Macro Average	0.76	0.74	0.75	2041
Weighted Average	0.96	0.96	0.96	2041

[2] Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In *International conference on intelligent, secure, and dependable systems in distributed and cloud environments*, pages 127–138. Springer, 2017.

[3] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36, 2017.

[4] Oberiri Destiny Apuke and Bahiyah Omar. Fake news and covid-19: modelling the predictors of fake news sharing among social media users. *Telematics and Informatics*, page 101475, 2020.

[5] Marina Azzimonti and Marcos Fernandes. Social media networks, fake news, and polarization. Technical report, National Bureau of Economic Research, 2018.

[6] Sumit Banik. Covid fake news dataset, November 2020.

[7] Sahil Chopra, Saachi Jain, and John Merriman Sholar. Towards automatic identification of fake news: Headline-article stance detection with lstm attention models. In *Stanford CS224d Deep Learning for NLP final project*, 2017.

[8] Sahar Ghannay, Benoit Favre, Yannick Esteve, and Nathalie Camelin. Word embedding evaluation and combination. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 300–305, 2016.

[9] Shalini Ghosh, Oriol Vinyals, Brian Strope, Scott Roy, Tom Dean, and Larry Heck. Contextual lstm (clstm) models for large scale nlp tasks. *arXiv preprint arXiv:1602.06291*, 2016.

[10] Yoav Goldberg and Omer Levy. word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*, 2014.

[11] Mykhailo Granik and Volodymyr Mesyura. Fake news detection using naive bayes classifier. In *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKR-CON)*, pages 900–903. IEEE, 2017.

[12] Wahab Khan, Ali Daud, Jamal A Nasir, and Tehmina Amjad. A survey on the state-of-the-art machine learning models in the context of nlp. *Kuwait journal of Science*, 43(4), 2016.

[13] Anuradha Khattar, Priti Rai Jain, and SMK Quadri. Effects of the disastrous pandemic covid 19 on learning styles, activities and mental health of young indian students-a machine learning approach. In *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 1190–1195. IEEE, 2020.

[14] Phong Le and Willem Zuidema. Quantifying the vanishing gradient and long distance dependency problem in recursive neural networks and recursive lstms. *arXiv preprint arXiv:1603.00423*, 2016.

[15] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[16] Salman Bin Naeem, Rubina Bhatti, and Aqsa Khan. An exploration of how fake news is taking over social media and putting public health at risk. *Health Information & Libraries Journal*, 2020.

[17] Usha M Rodrigues and Jian Xu. ¡? covid19?¿ regulation of covid-19 fake news infodemic in china and india. *Media International Australia*, 177(1):125–131, 2020.

[18] Max Roser, Hannah Ritchie, Esteban Ortiz-Ospina, and Joe Hasell. Coronavirus pandemic (covid-19). *Our World in Data*, 2020.

[19] Natali Ruchansky, Sungyong Seo, and Yan Liu. Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 797–806, 2017.

[20] Jacques Savoy. A stemming procedure and stopword list for general french corpora. *Journal of the American Society for Information Science*, 50(10):944–952, 1999.

[21] Kazuki Shimizu. 2019-ncov, fake news, and racism. *The lancet*, 395(10225):685–686, 2020.

[22] Richard Socher, Yoshua Bengio, and Christopher D Manning. Deep learning for nlp (without magic). In *Tutorial Abstracts of ACL 2012*, pages 5–5. 2012.

[23] Edson C Tandoc Jr, Zheng Wei Lim, and Richard Ling. Defining fake news a typology of scholarly definitions. *Digital journalism*, 6(2):137–153, 2018.

[24] Muhammad Umer, Zainab Imtiaz, Saleem Ullah, Arif Mehmood, Gyu Sang Choi, and Byung-Won On. Fake news stance detection using deep learning architecture (cnn-lstm). *IEEE Access*, 8:156695–156706, 2020.

[25] Sander van der Linden, Jon Roozenbeek, and Josh Compton. Inoculating against fake news about covid-19. *Frontiers in Psychology*, 11:2928, 2020.

[26] Michela Del Vicario, Walter Quattrociocchi, Antonio Scala, and Fabiana Zollo. Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web (TWEB)*, 13(2):1–22, 2019.

[27] William Yang Wang. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648*, 2017.

[28] Liang Wu and Huan Liu. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In *Proceedings of the eleventh ACM international conference on Web Search and Data Mining*, pages 637–645, 2018.

[29] Jia Xue, Junxiang Chen, Ran Hu, Chen Chen, ChengDa Zheng, and Tingshao Zhu. Twitter discussions and concerns about covid-19 pandemic: Twitter data analysis using a machine learning approach. *arXiv preprint arXiv:2005.12830*, 2020.

[30] SM Zobaed, Mohsen Amini Salehi, Albert Zomaya, and Sherif Sakr. Big data in the cloud., 2019.