Advances in Parallel Computing Algorithms, Tools and Paradigms
D.J. Hemanth et al. (Eds.)

doi:10.3233/APC220015

Designing Deep Classifier for Effective Detection of Spoof News in Twitter

Naga Raju Mysore[a,1], Dr. K. Sankar[b], Dr. M. Islabudeen[c] and Dr. Sampath AK[d]
[a,b,c,d]*Associate Professor, Presidency University*

Abstract. A Spoof news is a fraud content meant to misguide the reader about the event with ill motive. In this article a reactive technique using deep learning is proposed to deal with it effectively. Spoof news are innumerable in number over microblog twitter and have wide range of bad effects overall. This is causing chaos and hoax among the readers about the issue. They are getting mislead about the issue a lot. As of now automatic locators of fake news are ineffective and few in number. This emphasized us to come up with smart locator with deep learning mechanism. One way of dealing with this issue is to make "blacklist" of origins and composers of counterfeit news. Here we need to examine all irksome instances of origins and creators in gradual manner. To cater this need we came up with a classifier based on deep learning mechanism that studies linguistic, network account aspects of twitter news and then distinguishes them into spoof and legitimate ones. We set up a deep learning model that takes both legitimate and spoof news elements as input and learns by analyzing their constructs. Then do the binary classification of news effectively thus avoiding the user not to misled by fake.

Keywords. Microblog, Counterfeit news, Deep Learning, Language Dialect Aspects, Interconnection Account Characteristics, Embedding layer and Model Parameters.

1. Introduction

Over a period of decade, huge amount of information availability in quick manner and exponential raise in social network users are noticed. This paved for opportunities for wrong doers intentionally dispense spoof news to users to mislead them with false information [1–11]. This is done to make real news hazy [2, 5, 7–11]. Another related term is, propaganda, meaning to publicize specific political intensions and their alternative plans [1, 8–11]. The counterfeit news usually uses sneaky language to influence users by misguiding them emotionally [5, 11] ("speech freedom"). Another aspect is to promote only one prospective in a highly bigoted way e.g., ("laud efforts of the Government."). Further, out of syntax, misspelling and restricted words usage are other characteristics of fake news language [7, 11]. Latest progress in natural language processing, data mining and machine learning leads to recognize language dialect aspects of spoof news and interconnection accounts aspects of their creators [1,2,4,5,7,8,11]. Counterfeit news identification is the capability to measure the rightness of information through the close study of their aspects [7, 11].

[1] Naga Raju Mysore, Presidency University, Bangalore, India; E-mail: nagaraju.m@presidencyuniversity.in

The disordered, tumultuous, ever changing nature and steep growth in number of social media users led to the demand for automated solutions for spoof news detection [1,2,6,8,10]. As they pose uncommon threats and so emerged as new research field with data, feature, model and application orientations .

2. Related Works

In [8] recent methods to distinguish news as fake based on nature of input data are given. In gave deep learning method framework for spoof news in social media and details 7 levels it to comprehend and manage. Work in [12] gives fake news definition and its handling using NLP and Deep Learning techniques. Work [4] presented detailed survey work on basic and advanced classifiers for fake news detection. Article considers linguistic aspects of spoof news and gives comparison of performance of basic and new techniques. Article [2] presented different features of fake news drawn through ensemble and reinforcement way. Their impacts on social and traditional media are given. In work [5] comes up with binary classification framework for LIAR and Kaggle data sets under consideration. It normalizes and tokenize those using tools and libraries of Python. Further SVM, CNN, LSTM, KNN, and Naïve Bayes classifier are tried with Bag of Words, Term Frequency-Inverse Document Frequency, and n-grams. It has achieved around 98%of accuracy with LSTM. In [6] Linguistic Inquiry and Word Count Method of handling fake news is proposed and its accuracy reaches to 82%. Approach given in [1] took dual classification, and SMOTE oversampling to manage imbalance in classification more suitable. RapidMiner Studio used and experiments are conducted. Then it achieved 99% accuracy. Article in [11] used qualitative data analysis tool to figure out fakes based on linguistic aspect.

2.1 Description of Data Bank

COVID 19 tweet data set with 6510 is considered. It is then categorized into factual and spoof tweets. Parts of tweet like caption and wording are taken into consideration. Thus, whole tweeter data is divided into four parts based on its class (factual/spoof) and type (caption/wording). They can be termed as, Factual Caption, Spoof Caption, Factual Wording, and Spoof Wording to balance it with the instances of majority class. This increases total number of tweets from 6510 to 6800.

Table 1. Tweets Data Bank

Category	Number
Factual Captions	5800
Spoof Captions	1000
Factual Wordings	5700
Spoof Wordings	1100

3. Proposed Approach

3.1. Traits Set

For each tweet, dialectal (Linguistic) and interconnection (Network) account traits are considered. The values of these traits usually not be evenly distributed. There are two techniques available, namely, normalization and standardization to make them to distribute uniformly. As we felt normalization method is enough to get that job done, used MaxAbsScaler Normalizer for the purpose. It considers magnitude of each data by ignoring its sign, determines maximum among them and then divides each of these values by this maximum. Interconnection account values of tweet gives its origin and whole path of its interconnection. The account traits such as user screen name, and in reply to user are discarded due to large missing values in them. Aspects like date and time are converted from character to numerical and 12h to 24hr format respectively. Dialectal traits describe syntactic and grammatical aspects of the language of the tweet. Table 2 depicts these details.

Table 2. Dialect and Interconnection Traits

Dialect Traits	Interconnection Account Traits
Average number of words	User IP address
Average number of syllables	Backer IP address
Flesch-Kincaid Measures	Backing IP address
Number of Large Words	Creation Date
Number of Large Sentences	Tweet Time
Number of Sentences	Number of Likes
Proportion of noun and verb modifiers	Number of Retweets
	Number of URLs

3.2. Inlay Layer

It is also called as embedding layer. It converts textual inputs into integers by taking parameters like vocabulary size and dimension of flatten vector. It finally produces representation of each sentence in the form of vector of uniform size in line with one-hot encoding scheme with padding option. In this paper the embedding layer size is taken as 18.

3.3. Deep Network Architecture and its functions

Spoof news identification is done in binary (Yes / No) fashion. Input is a Tweet content pair (TC1, TC2). The output will be then defined as below.

$$TCG = \begin{cases} 0 & \text{if TC1 is factual tweet and TC2 is also factual tweet} \quad \text{(No)} \\ 1 & \text{if TC1 is factual tweet and TC2 is spoof tweet} \quad \text{(Yes)} \end{cases} \tag{1}$$

Where TCG denotes label of tweet category, i.e., either factual or spoof. Tweet pair (TC1, TC2) is taken as input and normalized using MaxAbsScaler. Embedding vectors E1, E2 are created using embedding layer from TensorFlow of Keras. These two are united in concurrent manner to produce (E1, E2) vector. It is then fed to hidden layer as input. Hidden layer output then fed to fully connected network layer. The activation function used here is leky ReLU. Hidden layer produces matrix values with language dialects and interconnection traits as its dimensions. It generates output using maximum likelihood estimation. Class labels are finally generated at output layer using softmax(). The deep network schema is given in figure 1.

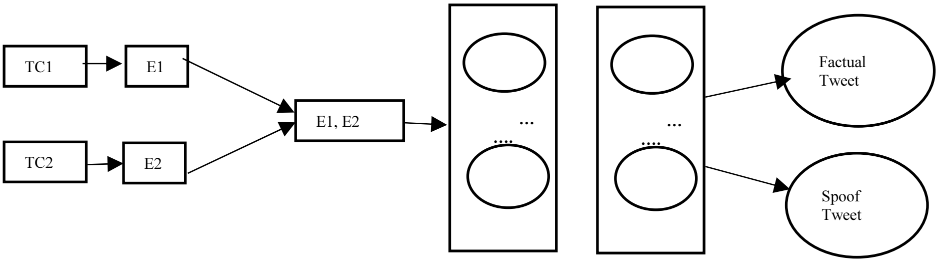

Figure 1. Deep Neural Network for detecting spoof news on twitter.

Input Layer → Embedding layer → Dense Hidden Layers → FC Layer + Leaky ReLU → Softmax

Acronyms: TC = Tweet Content, E= Embedding, FC =Fully Connected and ReLU=Rectified Linear Unit

4. Demonstration of working of proposed deep network architecture

The working of the proposed model is illustrated through conducting two experiments. In each experiment the pair (TC1, TC2) is fed as input to deep network whose scheme shown in figure 1 as given in sub-section 4.3. In experiment 1 TC1 being considered as factual caption and TC2 being taken as either factual wording or spoof caption or spoof wording. In experiment 2 TC1 being taken as factual wording whereas TC2 being considered as either spoof caption or spoof wording or factual caption. They are then normalized into embedded pair (E1, E2) by embedding layer of keras.
The other details of this architecture are as follows.

- Layers: two dense layers around 128 units and another dense layer with 1 unit
- Output Layer Function: Sigmoid curve of logistic regression for doing binary classification
- Activation Function used in dense layers: Leaky Rectified Liner Unit
- Dropout Value for regularization of deep network: 0.38
- Arguments passed to this proposed deep architecture and results obtained are tabulated in table 3.

Table 3. Arguments and contents given to the proposed model

Argument	**Content**
Performance Optimality Algorithm used	Adam - combination of Momentum and SMP optimization methods Learning
Accuracy Rate	0.006
Expiration Measuring Function	Binary cross entropy

7-fold cross validation is chosen in order to divide twitter data set under consideration into suitable training and testing data parts to obtain better classification finally.

5. Experiments Outputs and Patterns Noticed

This sub-section, first, presents results of two experiments and then draws conclusions on proposed system's performance from them. First confusion matrix is built, then precision and recall values have been calculated. From them F1 score is obtained as it gets final result as harmonic mean of both precision and recall. Finally, classification accuracy percentage is measured to gauge performance of the proposed deep network architecture.

5.1. Experiments Outputs

The following table presents results obtained in two experiments.

Table 4. Results obtained in two experiments

	Experiment 1		Experiment 2	
Tweet	Factual	Fraud	Factual	Fraud
Before SMOTE - on 6510 tweets				
Precision	97%	100%	98%	99%
Recall	96%	76%	97%	94%
Whole Accuracy	96%		95%	
Mean F1 value	98%			97%
After SMOTE - on 6800 tweets				
Precision	100%	100%	99%	100%
Recall	98%	97%	98%	95%
Whole Accuracy	97%	96%		
Mean F1 value	99%	98%		

Acronym: SMOTE - Synthetic Minority Oversampling Technique for addressing overfitting issue of minority class

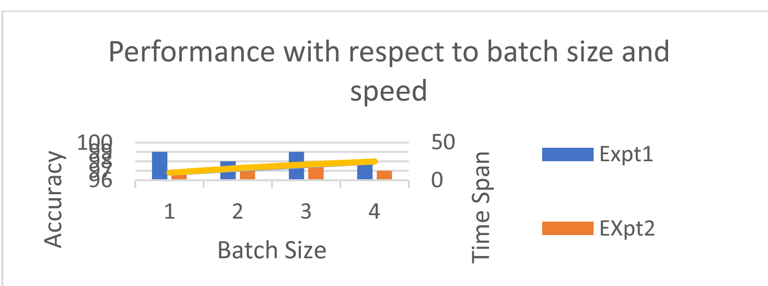

Figure 2 Accuracy based on batch size and learning speed aspects for both the experiments
Batch sizes 1.16 2. 32 3. 64 and 4. 128.

Figure 3 depicts proposed architecture model accuracy for both the experiments with and without user name aspect of interconnection trait.

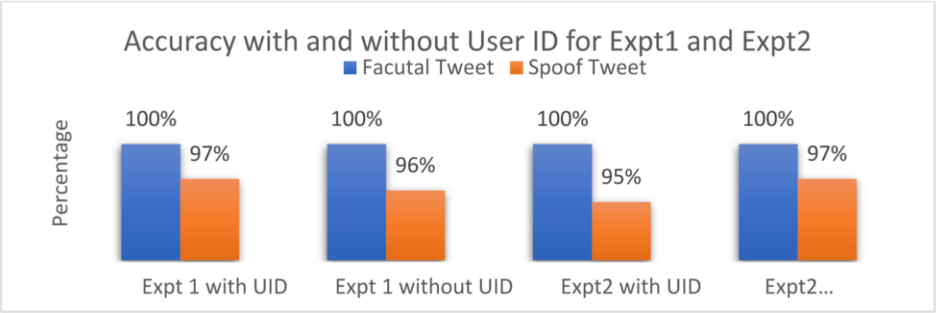

Figure 3. Accuracy measure with and without User ID (interconnection trait) for both the experiments

6. Directions of work in future

This work can be extended with other neural network configuration, like, Re-enforcement Neural Network. In a similar manner this work can be tried to resolve problems like spam and fraud detections. It can also be tried on data with other formats, like, images, audio, and video.

References

[1] Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. Proceedings of the National Academy of Sciences. 2016;113(3):554–559. Scheufele DA, Krause NM. Science audiences, misinformation, and fake news. Proceedings of the National Academy of Sciences. 2019;116(16):7662–7669.

[2] Kang, C., Goldman, A.: In washington pizzeria attack, fake news brought real guns. The New York Times (2016), https://www.nytimes.com/2016/12/05/business/media/cometping-pong-pizza-shooting-fake-news-consequences.html

[3] Reshma, R., Usha Naidu, V. Sathiyavathi, and L. SaiRamesh. "Stock Market Prediction Using Machine Learning Techniques." Advances in Parallel Computing Technologies and Applications 40 (2021): IOS Press, 331-340.

[4] Al-Rfou, R., Perozzi, B., Skiena, S.: Polyglot: Distributed word representations for multilingual nlp. In: Proceedings of the Seventeenth Conference on Computational Natural Language Learning. pp. 183–192. Association for Computational Linguistics, Sofia, Bulgaria (August 2018), http://www.aclweb.org/anthology/W13-3520

[5] Crestani, F., Rosso, P.: The role of personality and linguistic patterns in discriminating between fake news spreaders and fact checkers. In: Natural Language Processing and Information Systems: 25th International Conference on Applications of Natural Language to Information Systems, NLDB 2020, Saarbrücken, Germany, June 24–26, 2020, Proceedings. p. 181. Springer Nature.

[6] Jin, Z., Cao, J., Zhang, Y., Luo, J.: News verification by exploiting conflicting social viewpoints in microblogs. In: Thirtieth AAAI conference on artificial intelligence (2016)

[7] Kang, C., Goldman, A.: In washington pizzeria attack, fake news brought real guns. The New York Times (2016), https://www.nytimes.com/2016/12/05/business/media/cometping-pong-pizza-shooting-fake-news-consequefnces.html

[8] Rangel, F., Giachanou, A., Ghanem, B., Rosso, P.: Overview of the 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter. In: Cappellato, L., Eickhoff, C., Ferro, N., Névéol, A. (eds.) CLEF 2020 Labs and Workshops, Notebook Papers. CEUR-WS.org (Sep 2020)

[9] Sulthana, A. Razia, A. K. Jaithunbi, and L. Sai Ramesh. "Sentiment analysis in twitter data using data analytic techniques for predictive modelling." In *Journal of Physics: Conference Series*, vol. 1000, no. 1, p. 012130. IOP Publishing, 2018.
[10] Shu, K., Wang, S., Liu, H.: Understanding user profiles on social media for fake news detection. In: 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). pp. 430–435. IEEE (2018)
[11] Rashkin, H., Choi, E., Jang, J.Y., Volkova, S., Choi, Y.: Truth of varying shades: Analyzing language in fake news and political fact-checking. In: Proceedings of the 2017 conference on empirical methods in natural language processing. pp. 2931–2937 (2017)