ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

BERT: A Machine Learning Model For Online Fake News Detection

Chinnu Varghese[1], Chinnumol K V[2] , Amalu Jacob[3], Priya Joseph[4]

Assistant Professor, Department of Computer Science, Assumption College Autonomous, Changanacherry Kottayam[1]
chinnumaria0821@gmail.com

Assistant Professor, Department of Computer Science, Assumption College Autonomous, Changanacherry Kottayam[2]
chinnuvenu6@gmail.com

Assistant Professor, Department of Computer Science, Assumption College Autonomous, Changanacherry Kottayam[3]
amalujacob27@gmail.com

Assistant Professor, Department of Computer Science, Assumption College Autonomous, Changanacherry Kottayam[4]
priyajoseph053@gmail.com

Abstract: **Spreading fake news and spreading it on social media has become a big problem due to its devastating consequences. Various machine learning approaches can be used to identify fake news. However, I am wondering if the model dataset being used is not skewed as most of it focuses on a specific type of news (like politics). In our research, we looked at various research articles on fake news detection and found that BERT is the most accurate predictor of fake news, especially on very small datasets. Therefore, this model is a much better choice for languages with limited electronic content (for example, training data).**

Keywords: Machine Learning, Fake News Detection, Neural network, Deep learning, BERT, Natural language processing

I. INTRODUCTION

False stories can be defined as a form of yellow journalism or propaganda that contains deliberate false information or deceptive propaganda through traditional print and broadcast media or online media [1]. With the proliferation of online news sites, social networking sites, and other online media outlets, illegal online news is becoming more of a concern these days. But people are often unable to spend enough time looking for clues and being convinced of the reliability of the news. Therefore, given the proportion of users and providers of online news, the automatic detection of false news may be the only way to take corrective action, so from the research community; they are currently receiving a lot of attention.

Numerous research projects have been conducted on the discovery of fake news using traditional machine learning methods and deep learning methods over the years [7, 8, 10, 12, 13, 11, 9]. However, most of them focus on getting certain types of news (such as politics). Therefore, they designed models and functions designed for specific data sets to suit their interests.

II. RELATED WORKS

We go through various research papers worked on fake news detection and the findings of some of the researchers are given below:

1. [2], provides a comprehensive analysis of the performance of 19 different machine learning methods on three different databases. Of the 19 models, 8 are traditional learning models, 6 are traditional deep learning models, and 5 advanced language models such as BERT. They found that the BERT-based model works better than all other models in all data sets. Most importantly, BERT pre-selected models are able to withstand data size and can perform much better with very small sample sizes.

2. [3], provides an improved exBAKE model using previous training based on the BERT model to fully understand the content of the article. The results show that the model works very well in the FNC-1 database, which detects non-news items by analyzing the relationships between headlines and body texts of relevant headlines.
Their main contributions are summarized as follows:
• BERT [4] was used, which was the first to learn to detect non-existent news using the body text database. BERT includes pre-training language presentations developed by Google.
• [3], found that the data was unstable, so developing the BAKE model for data separation using weighted cross entropy (WCE).
• CNN and Daily Mail news data for BAKE pre-training is included. A large amount of news coverage is used to detect fake news.
• Finally, the performance of the proposed model exBAKE is examined, and showed that it works better than other models using FNC-1 data.

3. Of the existing methods [14, 15, 10], for false information detection, many useful methods have been introduced using traditional machine learning models. The main advantage of using deep learning model over existing content-based methods is that it does not require any handwritten features; instead, it identifies the best feature set by itself. CNN's powerful in-depth reading ability is mainly due to the use of multi-featured release sections that can automatically learn representations from the database. In the existing approaches [16, 17, 18], many ideas have been suggested to promote progress in the deep Convolutional Neural Networks (CNNs) such as the exploitation of temporary and channel information, the depth of structures, and graph-based multi-path information processing. The idea of using a layer block as a construction unit is also growing among researchers.
(Rohit Kumar Kaliyar et al., 2020), propose a comprehensive BERT-based (FakeBERT) deep learning approach by combining different CNN blocks with a single layer with Bidirectional Encoder representatives from Transformers (BERT). FakeBERT uses BERT as a sentence encoder, which can accurately detect the context of a sentence. Their work is compared with previous research activities [19] in which researchers look at text sequences in a unidirectional manner (either left-to-right or left-handed before prior training). Many existing and useful mechanisms have been transmitted [19, 20] by successive neural networks to encode relevant information. However, a with bidirectionally trained deep neural network can be a viable and precise solution for the discovery of false information. The proposed approach improves the performance of fake news acquisition with the ability to capture semantic and long-distance dependencies in sentences
4. Fake stories are especially prevalent in the current epidemic of COVID-19, leading people to believe in miracles and myths and potentially dangerous stories. Receiving false news quickly can reduce the spread of fear, turmoil and potential health risks. [6] has developed a two stage automated pipeline for the detection of COVID-19 false news using the machine learning models for natural language processing. The first model incorporates a reality facts checking algorithm that finds the most relevant facts regarding user claims regarding specific COVID-19 claims. The second model confirms the "true" level of the claim by incorporating a text link between the claim and the actual facts extracted from the COVID-19 manuscript. The database is based on a publicly available source of information containing more than 5000 COVID-19 false claims and verified explanations, a set of which was internally defined and validated for training and testing of their models. They tested a series of models based on text-based features to Transformer-based models and found that the BERT and ALBERT model pipeline in both stages respectively produced excellent results.

III. METHODOLOGY

BERT uses transformers, a viewing system that examines the contextual relationship between words (or sub-words) in a text. In its basic form, the Transformer consists of two different modes: an encoder, which reads the input text, and a decoder, which performs performance predictions. Since the purpose of BERT is to create a language model, it only requires an encoder method.

A. PRE-TRAINING BERT

For the BERT pre-learning algorithm, the researchers designed two unsupervised learning models. The first function is called Masked LM. It does this by accidentally masking 15% of documents and predicting the masked tokens. The second function is next-sentence prediction (NSP). They are motivated by tasks such as answering questions and completing them in natural language. These tasks require models that accurately represent the relationship between sentences. To do this, they are pre-trained in binary prediction

problems that can be easily generated from any corpus of language. For example: if the sentences are A and B, 50% of the time A is labelled "IsNext" and the other 50% of the time is a randomly selected sentence from the corpus and labelled "notNext". Pre-training on this task is useful for asking and answering questions and for natural language inference tasks.

1) Masked LM (MLM):

Before adding the input to BERT, 15% of the words in each sequence have been replaced by [MASK] tags. The model then attempts to predict the first value of the hidden word based on the context given to other unmasked words in the sequence. From a technical standpoint, predicting output names, we need:

● On the top of the encoder output, add a classification layer.

● Multiply the output vectors with an embedded matrix, converting it to vocabulary size.

● Calculate the chances of each word in the vocabulary with softmax.

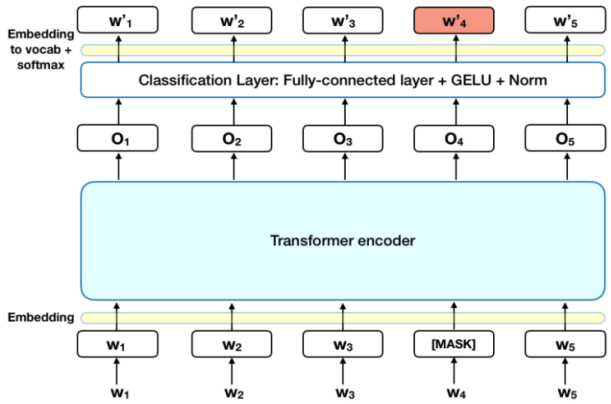

Fig. 1.

The diagram above provides an overview of the Transformer encoder. Input data is a sequence of tokens that are first embedded into vectors and then processed by a neural network. The result is a series of vectors of size H, each vector corresponding to an input character with the same index.

2) Next Sentence Prediction (NSP):

During training, the BERT model takes a pair of sentences as input and learns to predict whether the second sentence of the pair will be the next sentence in the original document. In the training process, 50% of the entries are pairs, where the second sentence is the next sentence in the first text, and the remaining 50% are random sentences from the selected corpus as the second sentence. Suppose any random sentence will be disconnected from the first sentence.

Before entering the model, we consider the introduction as follows so that the model can easily distinguish between the two sentences during training.

1. A [CLS] token is inserted at the beginning of the first sentence and a [SEP] token is inserted at the end of each sentence.
2. A sentence embedding indicating Sentence A or Sentence B is added to each token.
3. A positional embedding is added to each token to indicate its position in the sequence.

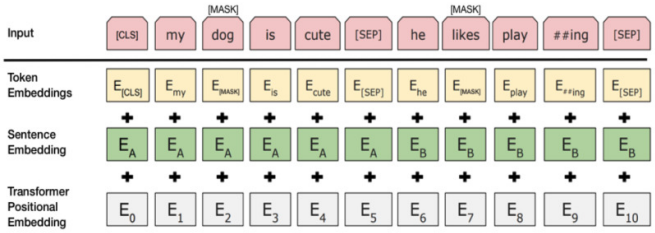

Fig. 2.

Source: BERT [Devlin et al., 2018], with modifications

To predict that the second sentence is really related to the first, the following steps are taken:

● All input sequences pass through the Transformer model.

● The output of the [CLS] token is converted to a 2 × 1 vector, using a simple classification layer.

● Calculates the chances of IsNextSequence with softmax.

When we train the BERT model, we train the Masked LM and Next Sentence Prediction together to minimize the join loss function of both strategies.

B. FINE-TUNING BERT

Fine Tuning BERT works by encoding text pairs in conjunction with self-attention. Self-attention is the process of checking the relationship between the current word and the previous word. It is very easy to use BERT for a specific task. BERT can be used for a variety of language functions and can only add a small layer to a basic model.

● Classification tasks such as sentiment analysis are performed in the same way as the Next Sentence classification, by adding a separating layer over the Transformer output for the [CLS] token.

● In Question Answering tasks (eg SQuAD v1.1), software detects a question about text sequence and is required to mark the answer in the sequence. Using BERT, a Q&A model can be trained by reading two additional vectors indicating the beginning and end of a response.

● In Named Entity Recognition (NER), software detects text sequences and is required to mark different types of

entities (Person, Organization, Date, etc.) from the text. Using BERT, the NER model can be trained by feeding the vector for the output of each token in the classification layer that predict the NER label.

In fine-tuning training, most hyper-parameters remain the same as in BERT training.

C. Performance Metrics

Various metrics are used to assess the effectiveness of the algorithm. Most of them are based on confusion matrices. A confusion matrix is a tabular representation of a test case classification model with four parameters: true positive, false positive, true negative, and false negative. (see Figure below).

Confusion Matrix

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

Fig. 3.

D. Accuracy

Accuracy is the most widely used indicator of the percentage of correctly predicted observations, true or false. The accuracy of a model can be calculated using the following equation:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FP}$$

In most cases, high fidelity values are good models, but in our case, when we train our classification model, articles predicted to be correct, when in fact false (false positives) can lead to negative results. Likewise, even if an article contains factual data, reliability issues can arise if the article is expected to be inaccurate.

IV. RESULTS AND DISCUSSION

One of the biggest problems with NLP is the lack of adequate training data. You usually have a lot of textual data, but to create a specific dataset, you need to break that dataset into pieces. And if they did, there would only be a few thousand or hundreds of thousands of examples of training courses labelled as human. Unfortunately, deep learning-based NLP models require much more data to function properly. To fill this data gap, researchers have developed a variety of methods for teaching universal linguistic representation models using a huge amount of memory-free text on the web (this is called pre-learning). These general-purpose pre-trained models can be fine-tuned on smaller task-specific datasets, e.g., when working with problems like question answering and sentiment analysis. This approach provides a significant improvement in accuracy over training from scratch on a smaller dataset for a specific task. BERT is a new addition to this NLP preparation technique. This is concern in a deep learning community as it presents modern results in a variety of NLP activities, such as answering questions.

The best thing about BERT is that it is free to download and use. You can use the BERT model to extract high-quality language features from text data, or to apply it to specific problems, such as answering questions with your own data and sentiment analysis, to produce state-of-the-art predictions.

V. CONCLUSION

Most of the data collected to detect fake news is in English. Since the spread of fake news has a negative impact on society, several studies have been conducted and several methods have been proposed to combat these fake texts. Readers should be able to distinguish real from fake news.

In our study, we looked at various research articles and found that the previously trained BERT model clearly understood the content of the articles. The results show that the model works best on all datasets that detect fake news by analyzing the headlines and body text of news articles and the relationship between these headlines and body text.

REFERENCES

[1].Leonhardt, D., & Thompson, S. A. (2017). Trump's lies. New York Times, 21.

[2]. Junaed Younus Khan.; Md. Tawkat Islam Khondaker.; Sadia Afroz.; Gias Uddin.; Anindya Iqbal. *A benchmark study of machine learning models for online fake news*
detection

[3]. Heejung Jwakey.; Dongsuk Oh.; Kinam Park.; Jang Mook Kang.; Hueiseok Lim *.exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT)*

[4]. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. *Bert: Pre-training of deep bidirectional transformers for language understanding*. arXiv 2018, arXiv:1810.04805.

[5]. Rohit Kumar Kaliyar.; Anurag Goswami.; Pratik Narang. *FakeBERT: Fake news detection in social media with a BERT-based deep learning approach*

[6]. Rutvik Vijjali.; Prathyush Potluri.; Siddharth Kumar.; Sundeep Teki. *Two Stage Transformer Model for COVID-19 Fake News Detection and Fact Checking*

[7]. Dai, E., Sun, Y., & Wang, S. (2020*). Ginger cannot cure cancer: Battling Fake health news with a comprehensive data repository. In Proceedings of the international AAAI conference on web and social media*, Vol. 14 (pp. 853–862).

[8]. Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). *Mvae: Multimodal variational autoencoder for fake news detection. In The world wide web conference (pp. 2915–2921).*

[9]. Rubin, V., Conroy, N., Chen, Y., & Cornwell, S. (2016*). Fake news or truth? Using satirical cues to detect potentially misleading news. In Proceedings of the second workshop on computational approaches to deception detection (pp. 7–17).*

[10]. Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019). *defend: Explainable fake*
news detection. In Proceedings of the 25th ACM SIGKDD international conference
on knowledge discovery & data mining (pp. 395–405).

[11]. Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). *Some like it hoax: Automated fake news detection in social networks*. arXiv preprint arXiv:1704.07506.

[12]. Wang, W. Y. (2017). *"liar, liar pants on fire": A new benchmark dataset for fake news detection.* arXiv preprint arXiv:1705.00648.

[13]. Zhou, X., & Zafarani, R. (2019). *Network-based fake news detection: A pattern-driven approach. ACM SIGKDD Explorations Newsletter*, 21(2), 48–60.

[14]. Ahmed H, Traore I, Saad S (2017*) 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. Springer, Cham, pp 127–138

[15]. Reema A, Kar AK, Vigneswara Ilavarasan P (2018*) Detection of spammers in twitter marketing: a hybrid approach using social media analytics and bio inspired computing. Information Systems Frontiers*
20(3):515–530695

[16]. Jwa H, Oh D, Park K, Kang JM, Lim H (2019) *exBAKE: Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT).* Appl Sci 9(19):4062

[17]. Kaliyar RK, Goswami A, Narang P, Sinha S (2020*) FNDNetA deep convolutional neural network for fake news detection*. Cognitive Systems Research 61:32–44

[18]. Munandar D, Arisal A, Riswantini D, Rozie AF (2018) *Text classification for sentiment prediction of social media dataset using multichannel convolution neural network.* In: 2018 International conference
on computer, control, informatics and its applications (IC3INA). IEEE, pp 104–109

[19]. De S, Sohan FY, Mukherjee A (2018) *Attending sentences to detect satirical fake news.* In: Proceedings of the 27th international conference on computational linguistics, pp 3371–3380

[20]. Karimi H, Roy P, Saba-Sadiya S, Tang J (2018) *Multi-source multi-class fake news detection*. In: of the 27th international conference on computational linguistics, pp 1546–1557

[21]. Gilda, S. (2017*). Evaluating machine learning algorithms for fake news detection.* In Research and development (scored), 2017 IEEE 15th student conference on (pp. 110–115). IEEE.

[22]. Gravanis, G., Vakali, A., Diamantaras, K., & Karadais, P. (2019). Behind the cues: *A benchmarking study for fake news detection. Expert Systems with Applications,* 128, 201–213.

[23]. Yang F, Liu Y, Xiaohui Y, YangM(2012) *Automatic detection of rumor on SinaWeibo*. In: Proceedings of the ACM SIGKDD workshop on mining data semantics, pp 1–7

[24].Liu, Y. *Fine-tune BERT for Extractive Summarization.* arXiv 2019, arXiv:1903.10318.

[25]. Rasool, T.; Butt, W.H.; Shaukat, A.; Akram, M.U. Multi-Label Fake *News Detection using Multi-layered Supervised Learning. In Proceedings of the 2019 11th International Conference on Computer and Automation*
Engineering, Perth, Australia, 23–25 February 2019; ACM: New York, NY, USA, 2019; pp. 73–77.