Evaluating Different Learning Approaches for COVID-19 Fake News Detection

Weiqi Wu
School of Information Science and Technology
Shanghaitech University
Shanghai, SH 200120
wuwq@shanghaitech.edu.cn

Yufei Zhang
School of Information Science and Technology
Shanghaitech University
Shanghai, SH 200120
zhangyf5@shanghaitech.edu.cn

Abstract

Social media platforms enable people to access the latest news during the COVID-19 outbreak. However, the flooding fake news has become a serious issue. Various machine and deep learning approaches have been proposed to perform fake news detection with reasonable accuracy. We evaluate different supervised text classification algorithms on Contraint@AAAI 2021 COVID-19 Fake news detection dataset. The machine learning approaches are SVMs with different kernels, Logic Regression and Decision Tree. For deep learning, we fine-tune the pre-trained Bidirectional Encoder Representations from Transformers (BERT) model as the base model and add CNN layers and BiLSTM layers on the top of it. Furthermore, a voting classifier is proposed to explore whether ensemble learning can further improve the accuracy of prediction. The result shows that SVM model with rbf kernel performs best among all the machine learning models and BERT model with BiLSTM layer performs best among deep learning ones. The ensemble learners by voting method have shown an overall better score on all performance metrics as compared to the individual learners.

1 Introduction

Social platforms such as Twitter, Facebook, Weibo, are becoming the most popular way for people to get the latest information as the number of users increases. Since the outbreak of the Coronavirus disease 2019 (COVID-19), those platforms update news on the pandemic every day, helping individuals and social medias in sharing their opinions and information with a much wider audience. However, it has become equivalently easy for people with malicious intent to spread fake news on social platforms, which may create a political, social, and economic bias in the minds of others. Hence, detection of fake news has become a topic that the society needs to pay attention to during the ongoing COVID-19 crisis.

Fake news detection has been given considerable attention by researchers in natural language processing. Fake news detection on traditional media mainly relies on news content, while on social media, extra social context auxiliary information can be used to as additional information to help detect fake news, like users' profiles, social networks [1] and comments to the news[2]. In practice, it is difficult to efficiently collect and process the social context due to a sea of posts and articles.

Preprint. Under review.

Therefore, it is of great significance to accurately perform fake news detection using the features of news text without any other relevant metadata.

Various machine and deep learning approaches have achieved reasonable accuracy in fake news detection. The study by Ahmed et al. [3] included extracting linguistic features such as n-grams from textual articles and training multiple machine learning models as classifiers, including K-nearest neighbor (KNN), support vector machine (SVM), logistic regression (LR), linear support vector machine (LSVM), decision tree (DT), and stochastic gradient descent (SGD). Bangyal, W. H. et al. [4] applied more machine learning methods such as Naive Bayesian and Adaboost, together with four deep learning approaches Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Recurrent Neural Networks (RNN), and Gate Recurrent Unit (GRU). Omar Sharif et al. [5] combined CNN+BiLSTM with tf-idf and Word2Vec embedding techniques, finding that deep learning does not necessarily perform better than SVM on particular datasets. Bidirectional Encoder Representations from Transformers (BERT) is used as a classifier in the work of Apurva Wani et al. [6] and combined with different parallel blocks of the single-layer CNN in the FakeBERT proposed by Rohit Kumar Kaliyar et al. [7].

The existing methods mainly rely on the learning of a single model or algorithm. Since each model has its own emphasis on learning, it's reasonable that the accuracy may be further improved by ensemble learning. In this project, we compare several machine and deep learning approaches and select the top classifiers to explore ensemble learning techniques. Experiment is conducted to solve the Contraint@AAAI 2021 COVID-19 Fake news detection shared task [8]. The task aims in improving the classification of the news based on COVID-19 as fake or real, formulated as a text classification problem.

2 Methodology

To make what have been learned in class into practice, we propose three kinds of learning methods including seven models to evaluate the performance of fake news detection. Figure 1 shows the pipeline of our detection system.

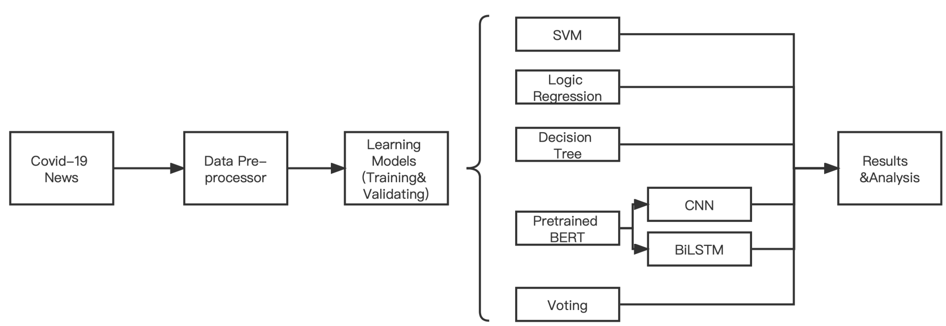

Figure 1: Processing Methods

2.1 Machine Learning Method

SVM Support Vector Machine (SVM) is a supervised learning, mapping training examples to points in space so as to maximize the width of the gap between the real tweets and fake ones. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall. By using different kernel tricks, SVMs can efficiently perform a non-linear classification besides a linear one. In this paper, linear, polynomial and rbf kernels are used.

Logistic Regression Logistic Regression can be seen as a generalized linear model using a logistic function (sigmoid),

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

to model a binary dependent variable. Then the predicted values are probabilities and are therefore restricted to $(0, 1)$ through the logistic distribution function.

Decision Tree A decision tree is a nonparametric hierarchical model for supervised learning whereby the local region is identified through a sequence of recursive splits in a small number of steps – divide-and-conquer approach. Sklearn package is used to construct the tree to predict the tweet is real or not.

2.2 Deep Learning Method

BERT The pre-trained Bidirectional Encoder Representations from Transformers (BERT) model is used as a base model. It contains 12 transformer blocks, 12 self-attention heads, and a hidden size of 786. We use 2e-5 as the learning rate and 4 epochs to train the model. The structure is borrowed from [9].

BERT+CNN Convolutional Neural Network (CNN) is added on the top of the BERT model with frozen parameters. It consists of two convolutional layers with kernel size (1,768) and (2,768) and ReLU activation function. A max pooling and dropout layer with rate 0.2 are also added. Finally, we apply a linear layer with softmax activation layer. The learning rate is set to 5e-5 and the epoch number is 4.

BERT+BiLSTM Similarly, Long short-term memory(LSTM) layer is added on the top of the BERT model with frozen parameters. To increase the amount of information available to the network, a Bidirectional LSTM (BiLSTM) layer is used, which is a sequence processing model that consists of two LSTMs: one taking the input in a forward direction, and the other in a backwards direction. Similarily we apply a linear layer with softmax activation function after the BiLSTM. The learning rate for this model is 5e-5 and the epoch number is 4.

2.3 Ensemble Learning Method

By the No Free Lunch Theorem[10] , there is no single learning algorithm that in any domain always induces the most accurate learner. Ensemble learners tend to have higher accuracies, as more than one model is trained using a particular technique to reduce the overall error rate and improve the performance of the model.

Voting Since we have different models listed above, voting is one way to combine multiple classifiers, which is a linear combination of the learners trained on the whole dataset. Here we select three best performing models, that are SVM model with rbf kernel, BERT models with CNN layers and BiLSTM layer according to Figure 1. The prediction is produced by majority voting.

3 Experiment

3.1 Dataset

The Contraint@AAAI 2021 COVID-19 Fake news detection dataset [8] has a predefined split of training set (6,420 entries), validation set (2,140 entries) and test set (2,140 entries). Each data entry is made up of three parts: 'id' is the number of the tweet; 'tweet' is its detailed context; 'label' indicates whether the tweet is real or fake. Training set contains 3060 fake tweets and 3360 real ones while validation and test data contain 1020 fake tweets and 1120 real ones each.

3.2 Preprocessing

Data preprocessing is necessary in text classification. HTML tags, special characters, stop-words and noises are removed as these things do not offer useful information. For machine learning methods, we next transform the clean texts to vectors using the bag-of-words technique with the TF-IDF method for calculating the score of each word.

For BERT and BERT-based methods, we adopt the same tokenization and encoding method as Wang et al.[9]. Use both BertTokenizer and manually defined tokenizer to tokenize the tweets. [CLS] and

[SEP] tokens are respectively inserted before the first sequence of tokens and between each sequences. Next, map tokens to ids and apply pad and truncation. BERT requires that all tweets must have the same fixed length and the max length of 512 tokens per sentence, since only 10 out of 8560 rows has length that is over 512. Last, use Attention Masks to avoid incorporating the padded tokens into the interpretation of the sentences. The above preprocessing is done by concatemerizing the training, validation and test dataset. After preprocessing, three sub-datasets were separated according to the original order to carry on the following experiment.

3.3 Result

Evaluation Metrics We use two metrics to test our models' classification results on fake news detection, accuracy and F1 score. Accuracy is defined as:

$$Accuracy = \frac{number\ of\ correctly\ classified\ news}{total\ number\ of\ news}$$

The other metric is F1 score, which is defined as:

$$F1\ score = 2 \times \frac{precision \times recall}{precision + recall}$$

where the precision is the ratio of the number of true positive (TP) results and the number of all positive ones. The recall is the ratio of the number of true positive (TP) results and the number of all samples that should be judged as positive.

Table 1: Results using seven models

Methods	Models	Accuracy	F1 Score
Machine Learning	SVM (Linear)	93.97	93.98
	SVM (poly3)	94.11	94.12
	SVM (rbf)	**94.25**	**94.26**
	Logistic Regression	92.76	92.76
	Decision Tree	87.29	87.29
Deep Learning	BERT	94.95	95.25
	BERT+CNN	95.14	95.47
	BERT+BiLSTM	**95.37**	**95.55**
Ensemble Learning	Voting	**95.98**	**96.11**

Analysis From results shown in Figure 1, voting model has the highest accuracy and F1 score. In machine learning, SVM model with rbf kernel performs best. Machine learning models often calculate faster than deep learning models. However, their performance differences are not significant. In deep learning, BERT model with BiLSTM layer performs best. It proves that ensemble learning performs better indeed.

4 Conclusion and Discussion

In this project, we discuss the task of detecting COVID-19 fake news using machine learning models, deep learning ones and ensemble techniques. Experiment is conducted on the Contraint@AAAI 2021 COVID-19 Fake news detection shared task [8], which is formulated as a text classification problem. SVM model and BERT model with BiLSTM layer achieve better performance than others in their domain. The ensemble learners by voting method have shown an overall better score on all performance metrics as compared to the individual learners.

Though these models have had improved performance, they require a large amount of labeled data, which needs much manual annotation. Therefore, a reinforcement learning model may be researched in the future.

References

[1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective, 2017.

[2] Shi W. Tian J. Qi Z. Li B. Hao H. Xu B. Zhou, P. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, page 207–212, 2016.

[3] Hadeer Ahmed, Issa Traore, and Sherif Saad. Detection of online fake news using n-gram analysis and machine learning techniques. In *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, pages 127–138, Cham, 2017. Springer International Publishing.

[4] Qasim R. Rehman N. U. Ahmad Z. Dar H. Rukhsar L. Aman Z. Ahmad J. Bangyal, W. H. Detection of fake news text classification on covid-19 using deep learning approaches. *Computational and mathematical methods in medicine*, 2021.

[5] Omar Sharif, Eftekhar Hossain, and Mohammed Moshiul Hoque. Combating hostility: Covid-19 fake news and hostile post detection in social media. *CoRR*, abs/2101.03291, 2021.

[6] Apurva Wani, Isha Joshi, Snehal Khandve, Vedangi Wagh, and Raviraj Joshi. Evaluating deep learning approaches for covid19 fake news detection. In Tanmoy Chakraborty, Kai Shu, H. Russell Bernard, Huan Liu, and Md Shad Akhtar, editors, *Combating Online Hostile Posts in Regional Languages during Emergency Situation*, pages 153–163, Cham, 2021. Springer International Publishing.

[7] Anurag Goswami Rohit Kumar Kaliyar and Pratik Narang. Fakebert: Fake news detection in social media with a bert-based deep learning approach. *Multimedia Tools and Applications*, 80(8):11765–11788, 2021.

[8] Parth Patwa, Shivam Sharma, Srinivas Pykl, Vineeth Guptha, Gitanjali Kumari, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. Fighting an infodemic: Covid-19 fake news dataset. *Communications in Computer and Information Science*, page 21–29, 2021.

[9] Yuxiang Wang, Yongheng Zhang, Xuebo Li, and Xinyao Yu. Covid-19 fake news detection using bidirectional encoder representations from transformers based models. *arXiv preprint arXiv:2109.14816*, 2021.

[10] Ethem Alpaydin. *Introduction to machine learning*. MIT press, 2020.