Fake News Detection

Progress Report

Debjyoti Roy
112070373

Priyanka Bedarkar
112046765

Siddarth Harinarayan
112026390

Abstract

Fake news, deliberate disinformation, hoaxes, parodies and satire are various ways to mislead people in order to damage an agency, entity, or person, and/or gain financially or politically. Of late, fake news has been in the spotlight of mainstream journalism and general public because of how it can have an effect on the political scenario of a country(Fake News). The primary channel for spreading such content is the social media and it sometimes finds its way into the mainstream media as well. Day by day it is becoming increasingly important to detect and classify fake news as such, because of the grave impact it can have on the political results of an election. A good amount of research is being done in this regard, yet we do not have any state of the art technology to do so. In this project, we try to improve upon the existing model of Wang(2017), inculcating various ideas we learned during the course of the subject.

1 Introduction

Our main aim of the project is detection and classification of fake news. The fabrication of falsified eye-catching and intriguing statements are made to capture audience's attention to sell the information is very dangerous especially when it is used as a weapon to shape the politics . Hence the problem of fake news detection needs much more attention than it currently receives. The primary challenge for solving the issue of fake news is how loose the definition of the term Fake news is. For e.g. fake news can be classified into various categories: a statement which is known to be completely false, or a speech stating some statistics as facts for which no real analysis has been done, or a piece of text which is satirical. Several attempts have been made but we do not have a robust solution for a reliable verification of fake news yet Figueira and Oliveira(2017).

1.1 Literature Review

- Hanselowski et al.(2018) uses fake news challenge dataset which classifies the news based on four classes namely, agree, disagree, discuss, and unrelated. The model is trained using two stacked LSTM for embedded token sequence and three layered neural network to estimate the probability of which class it belongs to. This paper uses only LSTM based model to get the test predictions.

- Kim(2014) prominently discusses the idea of sentence classification using CNN and max-over-time-pooling. They have utilized dropout on the penultimate layer with l2-norm constraint of weight vectors for regularization.

- Wang(2017) proposes a solution that involves convoluted neural network for news statement, and bidirectional LSTM for other features of the news such as speaker, location, etc. CNN, like any other neural network, consists of an input layer, an output layer and multiple

hidden layers. The hidden layers of a CNN typically consists of , pooling layers, fully connected layers, and normalization layers. We are going with only one hidden layer, max pooling layer, as implemented in the paper. Bidirectional LSTM is essentially a neural network that has a caching mechanism that stores relevant information and discard irrelevant ones. Using these together as a hybrid model, their results show that CNN model has given best accuracy of 27 percent.

Hanselowski et al.(2018) uses LSTM based implementation to train the model, Kim(2014) used CNN to do sentence classification, and Wang(2017) used a hybrid of CNN and LSTM for training and prediction. After a thorough study on these implementations described in above research papers, we came up with an idea of further enhancing the training hypothesis that includes CNN, Bi-LSTM, and classical Machine Learning techniques like Gradient Boosted Decision Trees.

1.2 Current issues

We have numerous features from the dataset on which we train our model. In the deep learning techniques, we are not sure about the weightage given to each of the features. As the size of our dataset is small enough, we plan on using classical machine learning algorithms along with the deep learning techniques. We think this will help us in capturing better temporal behaviour of the sentences and provide us a better control over hyper-parameter tuning and alter the model design accordingly. As the training for machine learning models can be quickly done, we can try on different types of models and pick the best performer.

1.3 Our approach

We plan to train an efficient Machine Learning Model to draw the relationship between similarity-based features and the output labels, and use this as a prior (input) to the deep learning hybrid model mentioned in the Wang(2017).We can do this as our current dataset is small enough to handle computationally intensive Feature Engineering on the data. These similarity-based features (inputs to ML model) could be similarities between the word count, 2-grams and 3-grams; similarities after transforming these counts with TF-IDF, and few other features. ML model we could experiment with includes XGBoost because the model is robust; no normalization is needed and it can be regularized in several different ways to avoid over fitting.

1.4 Evaluation

In this project we are going to predict the authenticity of news using the Liar Dataset. The dataset contains a decade-long, 12.8K manually labeled short statements in various contexts from PolitiFact.com, which provides detailed analysis report and links to source documents for each case, along with the statment, subject, speaker, speaker's job title, party affiliation and more. The output is one among the six valid classes that classifies the news content: *pants-fire, false, barely-true, half-true, mostly-true, true*. To understand the dataset better, we examined the distribution of the labels. We noted that the labels are uniformly distributed as seen in Figure 2. We will be classifying test data set based on the class under which the news has been predicted upon. Further, we will verify several results by tuning the hyper-parameters and identify the best configuration for the model.

2 Current Progress

We implemented the algorithm mentioned in Wang(2017) which uses a hybrid model of CNNs and bi-LSTMs. We received an accuracy of 20.49% using this baseline implementation. The confusion matrix generated from our code run for the resultant output is shown in Table 1.

3 Expected Results

According to Wang(2017), Bi-LSTM model gives an accuracy of 23% and CNN model has

Table 1: Confusion Matrix

predicted -> actual	true	mostly true	half true	barely true	false	pants fire
true	26	2	31	58	94	0
mostly true	17	2	42	76	112	0
half true	19	2	45	86	114	1
barely true	13	1	40	68	92	0
false	15	0	43	69	122	1
pants fire	8	1	12	21	50	0

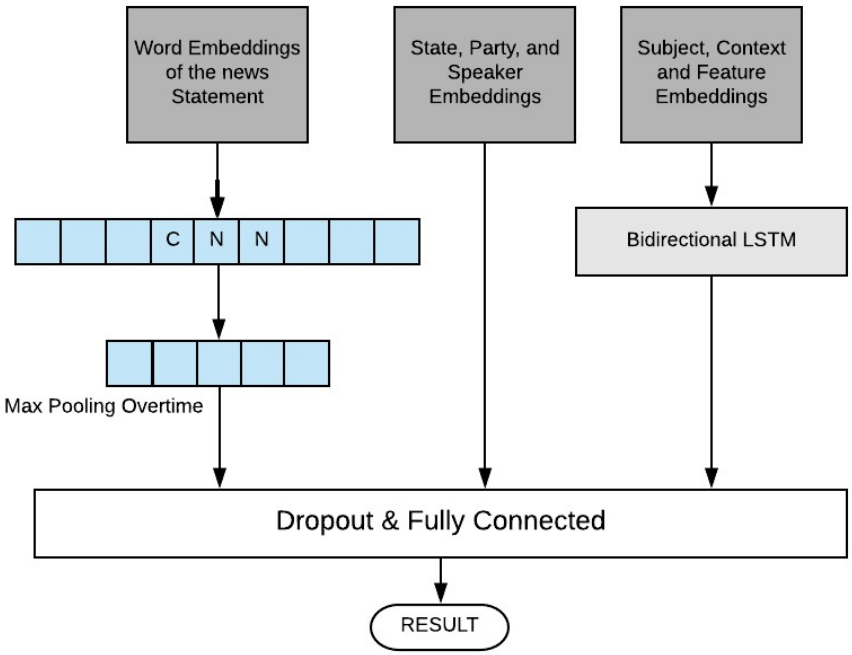

Figure 1: Model Architecture

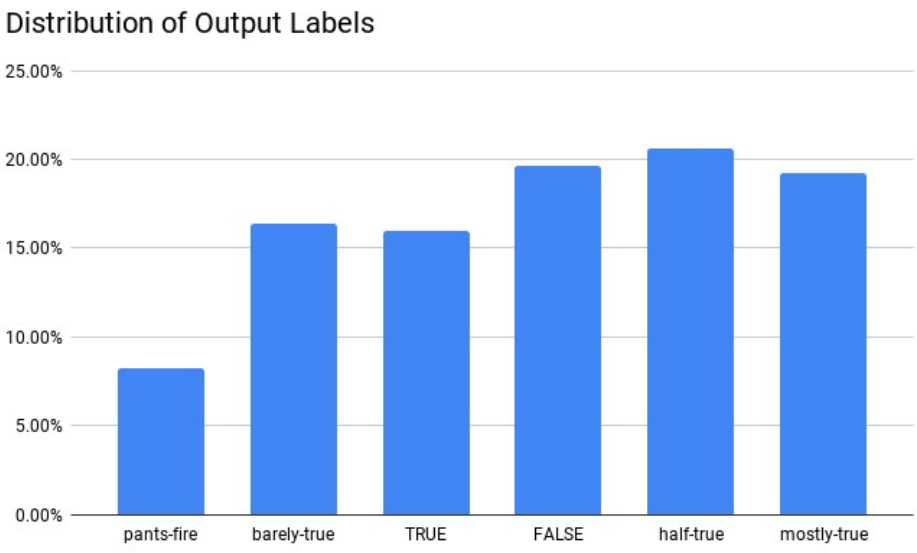

Figure 2: Output Label Distribution

an accuracy of 27% which is the best result of the author. After incorporating our machine learning model approach to the hybrid model of CNN and Bi-LSTM, we are expecting to see the accuracies in the range of 28% to 35%.

4 Questions we have

- We are currently training our model based on the Liar dataset. We have procured another dataset (Politifact Dataset), which we are planning on testing our model upon. Is it advisable to proceed in this manner or utilise the second dataset for training the model as well?

References

Fake News. Fake news — Wikipedia, the free encyclopedia.

Álvaro Figueira and Luciana Oliveira. 2017. The current state of fake news: challenges and opportunities. *Procedia Computer Science*, 121:817–825.

Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M Meyer, and Iryna Gurevych. 2018. A retrospective analysis of the fake news challenge stance detection task. *arXiv preprint arXiv:1806.05180*.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.

Politifact Dataset. Politifact fake news dataset.

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