Fine-Grained Fake News Detection System

Harsh Goyal
harsh20562@iiitd.ac.in

Ayush Sharma
ayush20042@iiitd.ac.in

Anupam Narayan
anupam20030@iiitd.ac.in

Shivam Jindal
shivam20125@iiitd.ac.in

Anas Ahmad
anas20023@iiitd.ac.in

March 13, 2023

1 Motivation

False or misleading information presented as news is known as fake news. Most of the time, fake news aims to hurt someone or something's reputation or make money from advertising. Of late, fake news has been in the spotlight of mainstream journalism and the general public because it can affect a country's political scenario. In this project, we attempt to detect the authenticity of specifically political news. Social media is the primary channel for disseminating such content, though it occasionally makes its way into mainstream media. Because fake news can significantly impact an election's political outcome, it is becoming increasingly important to identify and classify it as such. The loose definition of "fake news" presents the primary obstacle to resolving the problem. For instance, fake news can be broken down into subcategories: a statement known to be completely false, a speech that presents statistics as facts that have not been thoroughly investigated, or satirical writing. Our main aim of the project is to detect and classify fake news.

2 Problem Statement

Fake news is false or misleading information presented as news. Multiple strategies for fighting fake news are currently being actively researched, for various types of fake news. Our task is to determine whether a news statement/speech is fake or not given an input statement using Natural Language Processing and associated language models. Our aim is to design some novel and hybrid models using pre-trained large language models to integrate the original news statement/speech with the metadata. Fake News is a text-classification task. Since the news statements are very short in length and the text from the speech is noisy text containing grammatical errors, this makes the task more difficult and interesting.

3 Literature review

3.1 Liar, Liar Pants on Fire

The author proposed a hybrid Convolution Neural Networks Framework for integrating text and metadata. Hybrid CNN consists of two parallel Convolution Layers in which the input to the first Convolution Layer is the word embeddings for the given statement, followed by the MaxPooling, while the input to the second convolution layer is the metadata like speaker name, subjects of the speech,/statement, speaker's party, etc., followed by a Bidirectional-LSTM layer. The output of both layers is then concatenated and fed into a fully-connected layer with softmax on the output layer.

3.2 A Retrospective Analysis of the Fake News Challenge Stance Detection Task

The author uses a fake news challenge dataset which is a 4-class classification task. The author used two stacked LSTMs with 50-dimensional GloVe word embeddings as input and a three-layered neural network with 600 neurons each. The output from a three-layered neural network is fed into the output layer consisting of 4 neurons to estimate the class-wise probabilities to estimate to which class it belongs.

3.3 Exploring Text-transformers in AAAI 2021 Shared Task: COVID-19 Fake News Detection in English

The author proposed two solutions to the problem. The first solution consists of the use of RNNs (also called Bidirectional-LSTMs) as the LSTMs are based on previous text information while the RNNs are based on both previous and later text information, which helps in getting a better context of the sentence. The second solution consists of using 3 different techniques - namely, a Five-Fold Single-model ensemble, a Five-Fold Five-model ensemble, and a Pseudo Label algorithm. In the Five-Fold Single-model ensemble, the author used the same transformer-based pre-trained models (like BERT, Roberta, etc.) on all the five-folds of the dataset. The author used different pre-trained models for fine-tuning each fold in the Five-Fold Five-model ensemble. In the Pseudo Label algorithm, the author proposed that if the test data sample is classified to some class with a probability greater than 0.95, then the author proposed to use that test data sample as the training data sample for future test samples.

3.4 On the Benefit of Combining Neural, Statistical and External Features for Fake News Identification

The author proposed how using three different word embeddings in parallel and then concatenating them can benefit fake news detection. The author combined neural embeddings, statistical features, and external features to similar type of models(with a little difference in activation functions) in parallel and concatenated them to get the combined features. The author defined neural embeddings as the skip-thought vectors which encode the given sentences to vector embedding of length 4800. The statistical features are defined as the vectors obtained from the text using Bag-of-words(BOW), TF-IDF, and n-grams techniques. The external features include heuristics such as the similarity between headline-body pairs, the number of similar words in the headline and the body, etc. All these three types of features are then combined using pre-trained models and then fed into dense layers to predict the correct label for the given input sentence.

4 Models and Evaluations

We plan to use Machine Learning algorithms like Logistic Regression, Support Vector Machines, Random Forests, etc., and Deep Learning based pre-trained models like BERT and other such state-of-the-art language models and fine-tune them for our purpose and aim to create a novel architecture from them to get desirable results. We will use the standard metrics such as F1-score for evaluation and also explore benchmarks made explicitly for this task which have been proposed by papers that we came during the literature survey.

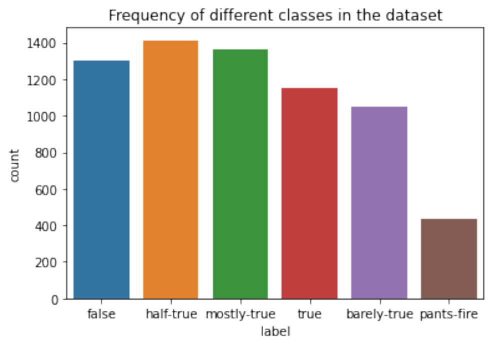
5 Dataset Details

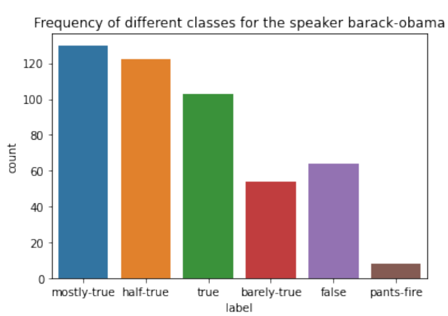
Dataset Statistics	
Training set size	10,269
Validation set size	1,284
Testing set size	1,283
Avg. statement length(tokens)	17.9

The 6-classes of the dataset are defined as:-
(These definitions were taken from PolitiFact's "truth-o-meter" methodology page).
1. true – The statement is accurate, and there's nothing significant missing.
2. mostly-true – The statement is accurate but needs clarification or additional information.
3. half-true – The statement is partially accurate but leaves out important details or takes things out of context.
4. barely-true – The statement contains an element of truth but ignores critical facts that would give a different impression.
5. false – The statement is not accurate.
6. pants-fire – The statement is inaccurate and makes a ridiculous claim. a.k.a. "Liar, Liar, Pants on Fire!"

6 Result And Analysis

The results presented show the performance of various machine learning models on the given dataset. The evaluation metric used is the weighted F1 score, which is a measure of the model's accuracy in predicting both the positive and negative classes.





Models	Weighted F1 Score
CV + MNB	0.20
TF-IDF + LR	0.25
CV + LR	0.24
TF-IDF + SVM	0.24
CV + DT	0.23
TF-IDF + RF	0.23

Looking at the results, the highest performing model is the TF-IDF + LR (logistic regression) model with a weighted F1 score of 0.25. (TF-IDF stands for Term Frequency-Inverse Document Frequency). This indicates that this model achieved the best balance between precision and recall for both the positive and negative classes.

The CV + MNB (Multinomial naive Bayes) model has the lowest performance with a weighted F1 score of 0.20. This indicates that the model's accuracy in predicting both classes is not very high.

The other models, including CV + LR (logistic regression), TF-IDF + SVM (support vector machines), CV + DT (decision trees), and TF-IDF + RF (random forest), all have relatively similar performances with weighted F1 scores ranging from 0.23 to 0.24.

7 References

[1] William Yang Wang. 2017. "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pages 422–426, Vancouver, Canada. Association for Computational Linguistics.

[2] 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. In Proceedings of the 27th International Conference on Computational Linguistics, pages 1859–1874, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

[3] Li, X., Xia, Y., Long, X., Li, Z., Li, S. (2021). Exploring Text-transformers in AAAI 2021 Shared Task:

COVID-19 Fake News Detection in English. ArXiv. https://doi.org/10.48550/arXiv.2101.02359

[4] Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B., Mittal, A. (2017). On the Benefit of Combining Neural, Statistical and External Features for Fake News Identification. ArXiv. https://doi.org/10.48550/arXiv.1712.03935

8 Contributions

Motivation - Anas Ahmad, Shivam Jindal
Problem Statement - Anupam Narayan, Ayush Sharma
Literature Review - Harsh Goyal, Anas Ahmad
Models and Evaluations - Harsh Goyal, Ayush Sharma, Shivam Jindal, Anupam Narayan