Online Social Network Analysis - CS579
Fall 20

Intermediate Report

November 11, 2020

Fake New Classification

Tushar Nitave (tnitave@hawk.iit.edu)
Varun Shanbhag (vshanbhag1@hawk.iit.edu)

Abstract

Fake news and rumours are present before the existence of the internet. A news is categorized as fake when news articles are deliberately modified to deceive readers. In this project we aim to propose a method to classify such fake news using machine learning techniques. We propose to generate features which will be leveraged by the machine learning algorithm. We also plan to evaluate the results of our model using several evaluation metrics like precision, recall, f1-score.

Introduction

Overall idea of the project it to classify the given the title of a fake news article A and the title of a coming news article B, into three categories:

Agreed: B talks about the same fake news as A.

Disagreed: B refutes the fake news in A.

Unrelated: B is unrelated to A.

We try to understand the semantic meaning of the given text article and then try to classify in one of these 3 categories. We have discussed how we plan to go about this in the method section.

Related Work

Tremendous amounts of research work has been carried out in order to precisely classify the fake news in the last decade. In [1] an research survey was conducted on state-of-the-art technologies for classifying fake news. A hybrid approach which uses linguistic cue and machine learning was found to be useful. A thorough review of fake news detection on social media, data mining algorithms and evaluation metrics was carried out in [2]. In [3] authors proposed Event Adversarial Neural Network (EANN) which uses even-invariant features for fake news detection. An sentence-comment co-attention-sub-network was developed to leverage news contents as well as comments for classification of the fake news. A novel approach which combines string similarity features along with a deep neural network (bi-directional Recurrent Neural Network (RNN)) was proposed. This proposed model outperformed the previous state-of-the-art models.

The provided dataset is highly imbalanced which makes predictions very inaccurate. There are several methods and strategies proposed in literature for making the dataset balanced. In [6] authors showed that synthetic minority oversampling technique (SMOTE) is beneficial for low-dimensional data.

Method

In this section, we talk briefly about feature extraction, exploratory data analysis, data cleaning, approaches for training the classification model.

Exploratory Data Analysis (EDA)

The provided train.csv dataset consists of 256442 data points along with 4 features namely *t_id1, t_id2, title1_en, title2_en.* The label consists of three classes namely *agree, disagree* and *unrelated.* The provided dataset is highly imbalanced which can be seen in the figure 1. So in order to make the dataset balanced we used RandomUndersampler to

reduced the frequency of ***unrelated*** class to that of 75000 data points. On the other hand we upsampled the ***disagree*** class to 75000 data points.

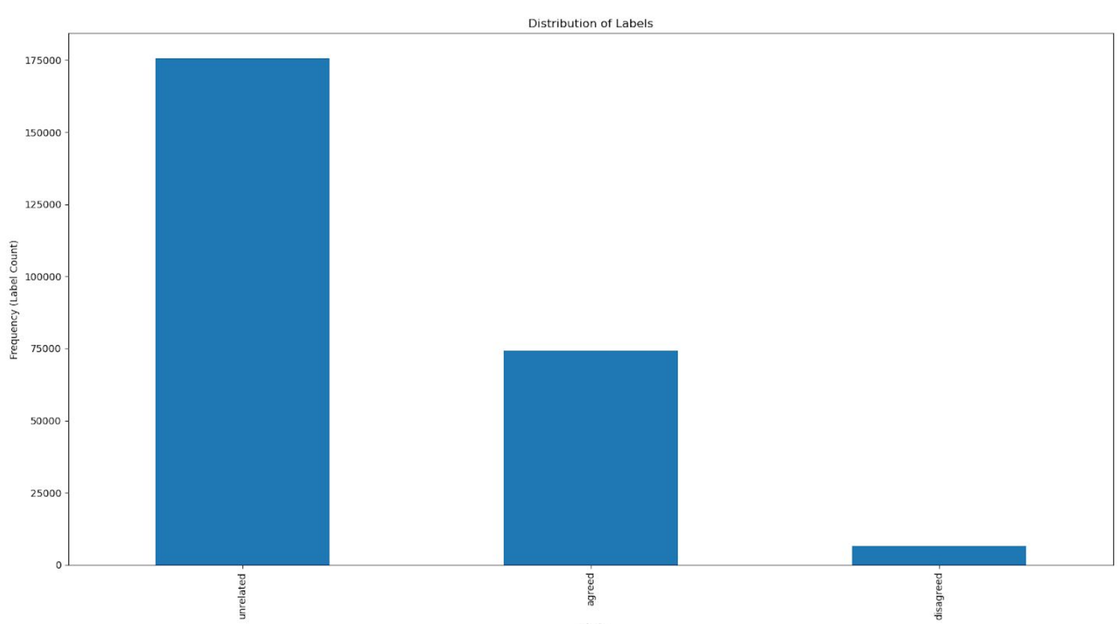

Fig.1 Class frequency in train.csv

Pre-Processing

We scan through text to remove stop words(A stop word is a commonly used word (such as “the”, “a”, “an”, “in”) that a search engine has been programmed to ignore, both when indexing entries for searching and when retrieving them as the result of a search query.) and do stemming(Stemming is the process of producing morphological variants of a root/base word.) All this processed text is used for feature extraction. We also remove null values if any(none in this case)

Feature Extraction

For generating the embeddings of the given news titles we used Google Universal Sentence Encoder [7] which generates 512 bit long embeddings for the given input sentence. It basically encodes text into high-dimensional vectors which can be used for text classification, semantic similarity etc. We also calculate the cosine similarity of the two sentences to and fro to measure the similarity and use it as an additional feature. Apart from this there are several other features which we plan to incorporate into this data for training the model.

Model Training

As of now we are investigating various machine algorithms for training our model. We experimented by using the fully-connected dense network with 3-hidden layers and 64 units in each layer, followed by a dropout layer for regularization. The train set was divided into validation and training where validation consists of 10000 samples shuffled from the original training set.

As of now we cannot claim anything about the performance and evaluation of the model. We plan to work with basic evaluation metrics like F1-score, precision, recall, accuracy.

Future Work

1. Add more features for training the model
    a. Number of common words between ***title1_en*** and ***title2_en***
    b. Refutation features between two titles
    c. Word n-grams common between two headlines
    d. CIDEr similarity score between two headlines
2. Experiment with different neural network architecture
    a. Simple LSTM followed by a dense network with softmax activation.
    b. Bi-directional LSTM
    c. CNN followed by a fully connected dense layer. CNN will allow us to extract features and reduce the high-dimensional vectors.
3. Hyperparameter tuning for neural network model
4. Provide a visual interface for inference. Model deployment on the Flask server.

References

1. Conroy, N. K., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology, 52(1), 1-4.
2. Shu, Kai, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. "Fake news detection on social media: A data mining perspective." ACM SIGKDD explorations newsletter 19, no. 1 (2017): 22-36
3. Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., ... & Gao, J. (2018, July). Eann: Event adversarial neural networks for multi-modal fake news detection. In Proceedings of the 24th acm sigkdd international conference on knowledge discovery & data mining (pp. 849-857).
4. Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019, July). defend: Explainable fake news detection. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 395-405).
5. Borges, L., Martins, B., & Calado, P. (2019). Combining similarity features and deep representation learning for stance detection in the context of checking fake news. Journal of Data and Information Quality (JDIQ), 11(3), 1-26.

6. Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. Journal of artificial intelligence research, 61, 863-905.
7. https://tfhub.dev/google/universal-sentence-encoder/4