

Name : RESHMA MANOJ KUMAR

CWID : 20007266

Course : CS 541 - Artificial Intelligence

Email : rmanojku@stevens.edu

Sentimental LIAR: Extended Corpus and Deep Learning Models for Fake Claim Classification

Project Proposal

Abstract

The rampant integration of social media in our every day lives and culture has given rise to fast and easier access to the flow of information than ever in human history. However, the inherently unsupervised nature of social media platforms has also made it easier to spread false information and fake news. Furthermore, the high volume and velocity of information flow in such platforms make manual supervision and control of information propagation infeasible. This paper aims to address this issue by proposing a novel deep learning approach for automated detection of false short-text claims on social media. We first introduce Sentimental LIAR, which extends the LIAR dataset of short claims by adding features based on sentiment and emotion analysis of claims. Furthermore, we propose a novel deep learning architecture based on the BERT-Base language model for classification of claims as genuine or fake. Our results demonstrate that the proposed architecture trained on Sentimental LIAR can achieve an accuracy of 70%, which is an improvement of ~30% over previously reported results for the LIAR benchmark.

Information about dataset :

The Sentimental LIAR dataset is a modified and further extended version of the LIAR extension introduced by Kirilin et al. In this dataset, the multi-class labeling of LIAR is converted to a binary annotation by changing half-true, false, barely-true and pants-fire labels to False, and the remaining labels to True. Furthermore, the speaker names are converted to numerical IDs in order to avoid bias with regards to the textual representation of names. The binary-label dataset is then extended by adding sentiments derived using the Google NLP API. Sentiment analysis determines the overall attitude of the text (i.e., whether it is positive or negative), and is quantified by a numerical score. If the sentiment score is positive, then the sample is tagged as Positive for the sentiment attribute, otherwise Negative is assigned. A further extension is introduced by adding emotion scores extracted using the IBM NLP API for each claim, which determine the detected level of 6 emotional states namely anger, sadness, disgust, fear and joy. The score for each emotion is between the range of 0 and 1.

Authors : Bibek Upadhayay, Vahid Behzadan.

Training dataset size : 3.78 MB.

Testing dataset size : 470 KB.

Link for code repository and Dataset :

Dataset Link (Testing and Training Data) :

<https://github.com/UNHSAILLab/SentimentalLIAR>

GitHub Repository Link : github.com/reshmamanoj2408/AI-Project.