Fake News Detection using RoBERTa

Tejal Shanbhag
Department of Computer Science
Rochester Institute of Technology
Rochester, NY 14623
ts8583@rit.edu

Arjun Ravikumar
Department of Computer Science
Rochester Institute of Technology
Rochester, NY 14623
ar4038@rit.edu

Saloni Shah
Department of Computer Science
Rochester Institute of Technology
Rochester, NY 14623
ss7170@rit.edu

1 Project Details

1.1 Type of research project

We intend to do a Application Study based on RoBERTa to implement a "Fake News" classifier. The inspiration to do the same was from the paper "Natural Language Contents Evaluation System for Detecting Fake News using Deep Learning" [1]. The mentioned paper uses BERT which is Bidirectional Encoder Representations from Transformers is a natural language processing pre-training technique developed by Google in 2018. We propose to conduct an Application study on the Fake News dataset using the more advanced RoBERTa which is Robustly Optimized BERT pretraining Approach developed by Facebook in 2019.

1.2 Motivation

Fake News has been an issue forever and is not something new. But recently the spread of fake news through social media has been one of the biggest issues discussed. Due to the nature of the current internet and the popularity of Social Media it has become more easy that ever to publish and spread Fake News. During the time of COVID-19 there were fake news of how the virus originated, how the virus can be cured using household remedies and especially fake news regrading why it is not recommended to wear masks during the pandemic. These news had affected societies and people in general as they looked and seemed genuine. This is our main motivation for doing this project.

1.3 Problem Description

We plan to train our RoBERTa model using the dataset [2] and validate it with the same using an 80-20 approach. The dataset consists of the title, text of the news and a label field which denotes weather the news is FAKE or REAL. We plan to use a pre-trained model for the same as the data set is relatively small for an accurate prediction. We then intent compare the results received with the existing BERT results from the paper and fine tune the application for better accuracy.

1.4 Related Work

A previous research on the same topic was conducted in the paper "Natural Language Contents Evaluation System for Detecting Fake News using Deep Learning" [1]. We intend to upgrade the BERT implementation of the paper to RoBERTa.

32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

1.5 Timeline of the Project

Distribution of Time		
Task	Description	Time Required
Study RoBERTa	Learn how to implement RoBERTa in python	2 weeks
Pre-process dataset	Cleaning and removal of unusable values from dataset	1 weeks
Train model	Training of model on the dataset	4 weeks
Optimization	Optimise the model on the dataset	2 weeks

Table 1: Project Timeline

1.6 Overall Project Description

The project will use a RoBERTa model for Natural Language Processing to classify fake news from real news. In the project we expect to study if there is any major increase in accuracy when using RoBERTa instead of the BERT model. We also intent to study if there are specific optimisation methods which can be used to improve the accuracy of the model.

References

[1] Y. Ahn and C. Jeong, "Natural Language Contents Evaluation System for Detecting Fake News using Deep Learning," 2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE), Chonburi, Thailand, 2019, pp. 289-292, doi: 10.1109/JCSSE.2019.8864171.

[2] Real and Fake news dataset. https://www.kaggle.com/nopdev/real-and-fake-news-dataset

[3] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V. (2019). "Roberta: A robustly optimized bert pretraining approach.", arXiv preprint arXiv:1907.11692.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805, 2018