Team 19: Identification of fake news in an online news media

Tathagata Raha, 2018114017
Aayush Upadhyaya, 2019202010
Jeevesh Kataria, 2019201058
Pramud Bommakanti, 2020900011

Abstract

This document describes the project outline for the term project as part of the IRE course.

1. Problem Statement

With the advent of the internet, social media and online news media, there is an increasing number of news, posts that we get to see on a daily basis. However, sometimes some news, posts or articles do not clearly represent the factual information, but represent some form of biasing, factual incorrectness or hate speech, which clearly differs way from the actual information. Since the digital data is growing now at ever so fast rate, it calls for a automated digital mechanism to identify the veracity of such news articles or social media posts. Here we aim at experimenting with various models and coming up with the best approach that can help detect fake news in online news or social media posts.

2. Relevant Reading

Several studies have been done to find out the whether AI can be used to effectively detect fake news.One such study finds that AI can effectively be used for such purpose [2], with accuracy over 90%. The most used method for automatic fake news detection is not just one classical machine learning technique, but instead a amalgamation of classic techniques coordinated by a neural network [1]. Recent studies have shown that fake and real news spread differently on social media, forming propagation patterns that could be harnessed for the automatic fake news detection. Propagation-based approaches have multiple advantages compared to their content-based counterparts, among which is language independence and better resilience to adversarial attacks [10].

1. **BERT**: BERT stands for Bidirectional Encoder Representations from Transformers.BERT is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers.BERT adopts to wide range of tasks, such as question answering and language inference [7].

2. **XLNet**: The researchers from Carnegie Mellon University and Google have developed a new model, XLNet, for natural language processing (NLP) tasks such as reading comprehension, text classification, sentiment analysis, and others. XLNet is a generalized autoregressive pretraining method that leverages the best of both autoregressive language modeling (e.g., Transformer-XL) and autoencoding (e.g., BERT) while avoiding their limitations. [9]

3. **RoBERTa**: We find that BERT was significantly undertrained and propose an improved recipe for training BERT models, which we call RoBERTa, that can match or exceed the performance of all of the post-BERT methods [8].

3. Dataset

We have selected a few datasets for our fake news detection task. They are as follows:

Nela-GT 2019 This dataset is a improvement in the Nela-GT 2018 dataset which is a multi-labelled dataset which contained 713k articles from 194 news and media outlets including mainstream, hyper-partisan, and conspiracy sources.[4]

Fake News Challenge This dataset was built using 49972 news articles and was built for the purpose of fake news detection using stance classification.[5]

RumourEval This dataset was built for the SemEval 2019 Task 7: RumourEval which focused on predicting the veracity of tweet rumour on the basis of retweets and replies.[3]

BanFakeNews We also want to extend our project to Indic languages. This dataset contains around 50k Bengali articles.[6]

4. Implemetation

Our goal would be to implement an end-to-end fake news detection model and provide comparisions of the different models we tried. We will do the implementation in stages:

Qualitative analysis of the dataset In this stage, we will perform qualitative analyses of the datasets mentioned above. Apart from that, we will look into the recent advancements in fake news detection and plan our approach to tackle the problem. We will also try some simple classification models like naive bayes, SVM, logistic regression, etc.

Neural classification methods In this stage, we will try different neural classification methods to classify fake news. We will use different SOTA embedding and classification models like BERT, XLNet, Elektra, etc and finetune them for our task. For the Bengali task, we will use the XLR 100 langs multilingual model as well as BengaliBERT model and check which performs better.

Stance detection and classification In this stage, we will classify, detect and classify stance for the news i.e. we will try to find a way to support or refute the article on the basis of other articles.

5. Timeline

- Project Outline - 27th Sep 2020
- Interim deliverable - 25th Oct 2020
- Final Deliverable - 5th Nov 2020
- Report and Presentation - 12th Nov 2020

6. Interim Deliverables

As part of the interm deliverable, we will try to perform qualitative analyses of the datasets and develop the neural network classification models.

1. The model(s) used
2. The scripts for preprocessing the data
3. A brief report on the qualitative analysis of the datasets
4. A brief report covering the details of the models
5. Code which we wrote to replicate the models

7. Final Deliverables

As part of the final deliverable, we shall try to improve the above models and also develop models for stance detection and classification.

1. The model(s) used
2. The scripts for preprocessing the data

3. The final report
4. Code which we wrote to replicate and develop novel models

References

[1] Rafael Garcia Ana Cristina Cardoso Durier da Silva, Fernando Vieira. Can machines learn to detect fake news? a survey focused on social media. 2019.

[2] Murat GoksuNadire Cavus. Fake news detection on social networks with artificial intelligence tools: Systematic literature review. 2019.

[3] Genevieve Gorrell, Elena Kochkina, Maria Liakata, Ahmet Aker, Arkaitz Zubiaga, Kalina Bontcheva, and Leon Derczynski. SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 845–854, Minneapolis, Minnesota, USA, June 2019. Association for Computational Linguistics.

[4] Maurício Gruppi, Benjamin D. Horne, and Sibel Adalı. Nela-gt-2019: A large multi-labelled news dataset for the study of misinformation in news articles. 2020.

[5] Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M. Meyer, and Iryna Gurevych. 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, August 2018. Association for Computational Linguistics.

[6] Md Zobaer Hossain, Md Ashraful Rahman, Md Saiful Islam, and Sudipta Kar. BanFakeNews: A dataset for detecting fake news in Bangla. In *Proceedings of The 12th Language Resources and Evaluation Conference*, pages 2862–2871, Marseille, France, May 2020. European Language Resources Association.

[7] Kenton Lee Kristina Toutanova Jacob Devlin, Ming-Wei Chang. Bert: Pre-training of deep bidirectional transformers for language understanding. 2019.

[8] Naman Goyal Jingfei Du Mandar Joshi Danqi Chen Omer Levy Mike Lewis Luke Zettlemoyer Veselin Stoyanov Yinhan Liu, Myle Ott. Roberta: A robustly optimized bert pretraining approach. *arXiv:1907.11692*, 2019.

[9] Yiming Yang Jaime Carbonell Ruslan Salakhutdinov Quoc V. Le Zhilin Yang, Zihang Dai. Xlnet: Generalized autoregressive pretraining for language understanding. *arXiv:1906.08237*, 2020.

[10] Meghana Moorthy Bhat Justin Hsu Zhixuan Zhou, Huankang Guan. Fake news detection via nlp is vulnerable to adversarial attacks. *arXiv:1901.09657*, 2019.