Climate Change Misinformation Tweet Detection using Transfer Learning with pre-trained BERT family models

Nhan Kiet To – 1043668 – nhankiett@student.unimelb.edu.au

Semester 1 2020 – COMP90042 Natural Language and Processing

The University of Melbourne

Abstract

Misinformation is a type of fake news, in the age of digital world, fake news is booming rapidly in scale and depth (Wong, 2019), misguide the understanding, knowledge of the reader, many attempts has been made in order to apply Artificial Intelligence (AI) – Machine Learning/Deep Learning to detect the fake news. In this paper, a solution to detect Climate Change Misinformation Tweets will be provided along with the analysis, using the RoBERTa (Yinhan Liu, 2019), ALBERT (Zhenzhong Lan, 2020) pretrain model from the renown BERT model which is often regard as AlexNet (Alex Krizhevsky, 2012) of NLP.

1. Introduction

While the fake news term itself is rather large and cover up many aspects, from "high-quality" fake news which are very hard to detect to "low-quality" ones which can be easily distinct in just a glance. Misinformation is a type "high-quality" fake new, as it method is generally to derive reader by first providing a "true" statement then subtly come to a consequence which is "untrue", e.g.: "Chicago has been experienced the coldest weather of all time in its history, yet people still talking about Climate Change", the first part of the sentence talks about the cold winter in 2019 which Chicago experienced, which is true to be one of the coldest winter in its history (These Are the Coldest Temperatures of All Time in Chicago — And We Could Get Even Colder, 2019), but for the second part, thanks to the first part it come to a conclusion that is "there is no climate change", by the nature of human, if one does not know about the second part, yet know about the first one and they find the first one is true by their knowledge, they will quickly succumb to the second part and believe that there is no climate change.

As AlexNet brought Deep Neural Network or Deep Learning to become a star in the AI field by demonstrated it superior performance in the Visual Recognition Challenge, BERT

(Bidirectional Encoder Representations from Transformers) is a type of language model developed and published in 2018 by Google is often regard as "AlexNet" of Natural Language Processing (NLP).

2. Dataset and files

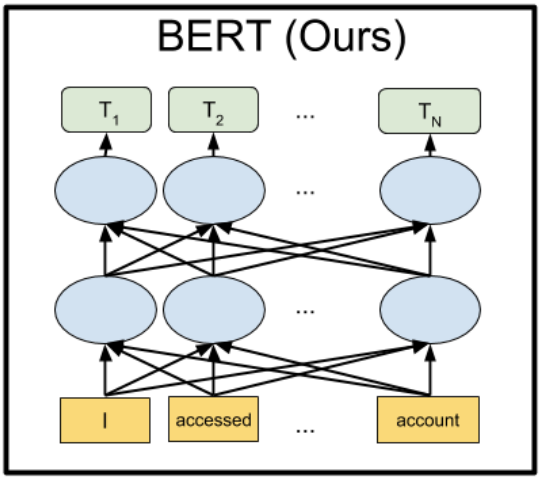
- **train.json**: contains 1168 samples of "true" climate change misinformation tweet with the label of 1, with a very broad range of related topics in the tweet sample from politics involve to Greta Thunberg – a young and renown activist in Climate Change and Environment, thus is it a big challenge, even for a human to distinct whether these tweets "misinformation" or not if there was no label given. This is the training data for the detector.
- **dev.json**: contains 100 sample of both "true" and "false" climate change misinformation tweet, with the label of 1 and 0 respectively. This is the evaluation data for the detector.
- **dev-baseline-r.json**: the baseline result provided, with the F1 score of about 0.53,

to help understand more about the expected format of the test output.

- **scoring.py**: use to calculated the score of the outputs against the ground truth dev.
- **test-unlabelled.json**: contains 1410 samples with no label, use for predicting the label for the codalab competition.
- **train_not_misinfo**: contains 2440 samples with "False" label which is 0, multiple sources of tweets was downloaded and wrap up into this file with respective format Although there are more data on the csv resource file, yet too much data on one label will brings unbalance data. This file weights only 348KB.
- **not_misinformation_preprocessing.py**: this contain a python script to processing the .csv tweet source and make the "train_not_misinfo" file.
- **dataproprocessing.py**: this file contains the script to load the train.json, dev.json and test-unlabelled.json to format which is used in this solution.
- **training.py**: this file contains the python script to load the training and dev file and put it to the model to train, then evaluate. The model output or weight will be put to directory "\outputs" in the same directory with this training.py file. With the first parameter has the option of "0" to start the training anew and another character to continue training.
- **predicting.py**: after the weights or model is available, it will be used to predict the unlabelled test data or dev data for performance benchmark.
- External data source: These csv files are process by the "not_misinformation_preprocessing.py" and put together and become the "train_not_misinfo.json"
    - 1377191648_sentiment_nuclear_power.csv (CROWDFLOWER, Emotions About Nuclear Energy, 2017)
    - administration_tweets.csv (OCHOA, 2017)
    - Apple-Twitter-Sentiment-DFE.csv (CROWDFLOWER, Apple Twitter Sentiment, 2017)
    - BarackObama.csv (FIVETHIRTYEIGHT, 2018)
    - newsarticles_tweets.csv (OCHOA, 2017)
    - progressive-tweet-sentiment.csv (CROWDFLOWER, Progressive Issues Sentiment, 2017)
    - socialmedia-disaster-tweets-DFE.csv (CROWDFLOWER, Disasters on Social Media, 2017)
    - text_emotion.csv (CROWDFLOWER, Sentiment Analysis in Text, 2017)
    - womenmarch.csv

3. Approach

Deep Learning is famous for its "data" consuming trait (Yuji Roh, 2019), while the external training data was limited to be below 10MB, thus, transfer learning (Learning, 2009) from a pre-trained model is suitable in this situation, for this case, BERT was used.



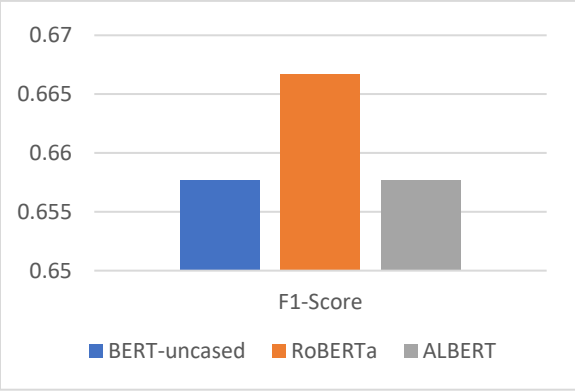
(Google, 2018)

Not only the "quantity" of data is needed, the "quality" of data is also no less important, sometimes, the data itself is even more important than the algorithms, "A dumb algorithm with lots and lots of data beats a clever one with modest amounts of it." (Domingos, 2012). If a tweet is not "climate change misinformation" is can belong to many other topics, from "true climate change information" to politics, random thoughts, covid-19, funny, other campaigns, technology, etc., hence, many type of topics has been attempted to covered in this solution.

For the case sensitive, uppercase or lowercase was kept in this solution, because for some pair of words such as "IT" – Information Technology in short or "it" in – "it is a banana", will not be able to distinct from one another if all words were to be convert to lowercase. Fortunately, BERT also cover the uncase and case scenario.

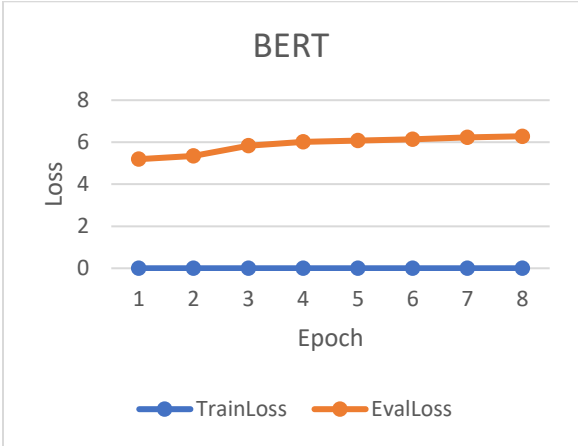
4. Result and Analysis

The result from the three pretrain model in the BERT family with the F1 score on the dev dataset and ground truth. RoBERTA has a better performance than BERT and ALBERT, but very little difference.

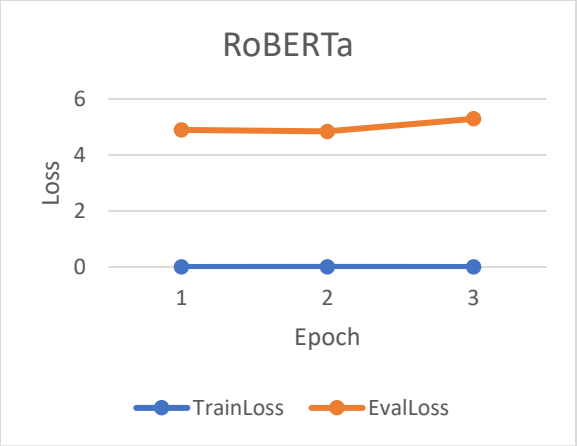


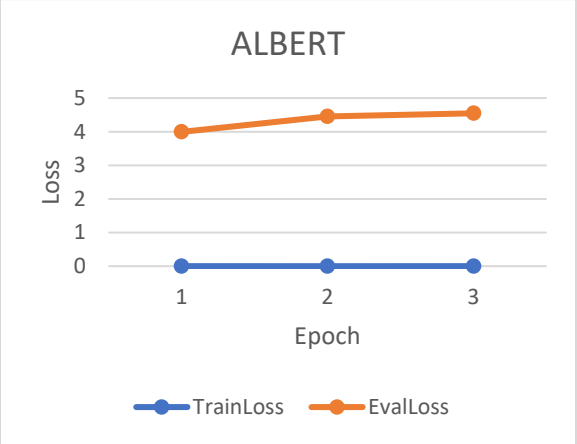
The data is very noisy, generally a climate change misinformation tweet is very well invested in the "quality", unlike a non-climate change misinformation tweet which is relatively normal in quality. The length of an average tweet which is label as True could be as long as 100 words or up to thousands of words. On top of that, if a tweet is not climate change misinformation it could also be a tweet related to climate change, or related to any different kind of topics which was not covered in the dataset.

For the BERT model, after 8 epochs, the curve shows that the model is underfit. Although the lost of train data is very small, smaller than zero, yet the loss on evaluation dataset is large, with the mean around 5 and increasing overtime. Clearly, the training dataset is not enough in-depth complexity, as many other cases were not cover.



RoBERTA and AlBERT model shows the same symptom, due to the heavy load and time consuming, these models are only trained at 3 epochs.





5. Implementation

Please read the "README.txt" file for implementation.

6. Future Improvements

The dataset would be even more carefully select, long tweet should be priority and even manually label thousands sample should also be implemented.

Use other pretrained model or different parameters, learning rates, etc.

Spending more time to train the model, maybe up to days for it.

7. Conclusion

Overall, a solution to detect Climate Change Misinformation has been proposed, using the BERT family model: BERT, RoBERTa, ALBERT. Although the performance is not as desired, it has been analyzed and the way to improvements for better performance is cleared.

References

Alex Krizhevsky, I. S. (2012). *ImageNet Classification with Deep Convolutional.* Retrieved from University of Toronto: http://www.cs.toronto.edu/~hinton/absps/imagenet.pdf

CROWDFLOWER. (2017). *Apple Twitter Sentiment*. Retrieved from data.world: https://data.world/crowdflower/apple-twitter-sentiment

CROWDFLOWER. (2017). *Disasters on Social Media*. Retrieved from data.world: https://data.world/crowdflower/disasters-on-social-media

CROWDFLOWER. (2017). *Emotions About Nuclear Energy*. Retrieved from data.world: https://data.world/crowdflower/emotions-about-nuclear-energy

CROWDFLOWER. (2017). *Progressive Issues Sentiment*. Retrieved from data.world: https://data.world/crowdflower/progressive-issues-sentiment

CROWDFLOWER. (2017). *Sentiment Analysis in Text*. Retrieved from data.world: https://data.world/crowdflower/sentiment-analysis-in-text

Domingos, P. (2012, 10). A Few Useful Things to Know About Machine Learning.

Communications of the ACM. doi:10.1145/2347736.2347755

FIVETHIRTYEIGHT. (2018). *Twitter Ratio*. Retrieved from data.world: https://data.world/fivethirtyeight/twitter-ratio/

Google. (2018, 2 11). *Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing*. Retrieved from Google AI Blog: https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html

Learning, A. S. (2009). *A Survey on Transfer Learning.* IEEE. doi:10.1109/TKDE.2009.191

OCHOA, J. A. (2017). *travel_ban*. Retrieved from data.world: https://data.world/alarcj/travel-ban

These Are the Coldest Temperatures of All Time in Chicago — And We Could Get Even Colder. (2019, 1 30). Retrieved from NBC Chicago: https://www.nbcchicago.com/news/local/cold-chicago-weather-illinois/2278/

Wong, Q. (2019, 6 5). *Fake news is thriving thanks to social media users, study finds*. Retrieved from CNet: https://www.cnet.com/news/fake-news-more-likely-to-spread-on-social-media-study-finds/

Yinhan Liu, M. O. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. Retrieved from https://arxiv.org/pdf/1907.11692.pdf

Yuji Roh, G. H. (2019, 8 12). *A Survey on Data Collection for.* Retrieved from arxiv: https://arxiv.org/pdf/1811.03402.pdf

Zhenzhong Lan, M. C. (2020, 2 9). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. *ICLR 2020.* Retrieved from https://arxiv.org/abs/1909.11942