

Siva at WNUT-2020 Task 2: Fine-tuning Transformer Neural Networks for Identification of Informative Covid-19 Tweets

Siva Sai

Birla Institute of Technology & Science, Pilani Campus

Pilani-333031

f20170779@pilani.bits-pilani.ac.in

Abstract

Social media witnessed vast amounts of misinformation being circulated every day during the Covid-19 pandemic so much so that the WHO Director-General termed the phenomenon as “infodemic.” The ill-effects of such misinformation are multifarious. Thus, identifying and eliminating the sources of misinformation becomes very crucial, especially when mass panic can be controlled only through the right information. However, manual identification is arduous, with such large amounts of data being generated every day. This shows the importance of automatic identification of misinformative posts on social media. WNUT-2020 Task 2 aims at building systems for automatic identification of informative tweets. In this paper, I discuss my approach to WNUT-2020 Task 2. I fine-tuned eleven variants of four transformer networks - BERT, RoBERTa, XLM-RoBERTa, ELECTRA, on top of two different preprocessing techniques to reap good results. My top submission achieved an F1-score of 85.3% in the final evaluation.

1 Introduction

In today’s highly connected world, social media assumes a vital role during pandemics like Covid-19. Social media platforms have been used positively by international health organizations and Governments to disseminate information about the pandemic and precautions regarding the same. Unfortunately, the world has also seen the misuse of social media for achieving cheap ends, severely damaging the physical and mental health of individuals, along with the increase in societal distrust. The false information is dangerous during a crisis when the mass panic can only be controlled with information(Lancet, 2020). A study(Wilson and Chen, 2020) identifies that panic in social media regarding Covid-19 spread faster than the Covid-19

virus spread. The proportionately large volume of misinformation compared to credible information can be seen from the very less engagement generated by posts from WHO (in several thousands) as compared to false news(over 52 million)(Mian and Khan, 2020).

The direct damage of misinformation is clearly evident. In Iran, hundreds of people died after consuming methanol, which is propagated on social media as one of the cures¹. There is even concern from authorities that even if a vaccine is discovered, anti-vaccination groups on social media may prevent people from getting vaccinated. False and conspiracy information about leading health organizations like WHO and leading scientists can quickly decrease the public trust on the correct agents, thus causing self-damage to individuals. The ill-effects of misinformation during pandemics can reach such extremes where they can even affect governmental policies. For example,(Mian and Khan, 2020) reminds us how the rejection of AIDS by the South African government in the early 2000s cost more than 300,000 lives.

In this context, proper identification and control of misleading information on social media assume a pivotal role. However, manual identification and control of the false posts is a Herculean task. For example, a 900 percent increase of fact-checkers during Covid-19 could not handle the flow of misinformation(Shahi et al., 2020). So developing automated methods and techniques to identify misinformation is essential in the current context. The WNUT-2020 Task 2 take steps in this direction. Recently, Transformer neural architectures achieved State-of-the-art results in several NLP tasks, including text classification. In this work, I make use of eleven variants of four different Transformer archi-

¹<https://www.who.int/news-room/feature-stories/detail/immunizing-the-public-against-misinformation>

lectures to reap better results on the identification of informative Covid-19 tweets.

The rest of the paper is organized as follows. Section 2 discusses work related to the Covid-19 misinformation spread. Section 3 provides a brief description of the task and dataset. Later, in section 4, I discuss the training procedure followed by my submission's internal and official results in section 5. Finally, I conclude my work in section 6.

2 Related works

(Shahi et al., 2020) performs an exploratory study on Covid-19 misinformation. Their work focuses on content, authors, and propagation of misinformation of Covid-19 related tweets. Their work shows that false claims spread more rapidly than partially false claims on Twitter. Further, they revealed that verified Twitter handles like celebrities and organizations are also involved in propagating misinformation. (Pennycook and Rand, 2020) in their work, discuss the damaging effects of misinformation regarding Covid-19. Their findings reveal that misinformation increases fear in society, creates disaccord, and can even lead to direct damage. The direct damage can be due to harmful medical advice, overreaction to the situation like hoarding and underacting such as deliberate engagement in risky behavior. However, (Gallotti et al., 2020) maintains an optimistic tone regarding the spread of misinformation. The authors claim that false information is quickly replaced with reliable information when the epidemic hits a particular area. Their analysis is based on 100 million tweets in 64 languages on the Covid-19 topic.

The quantitative analysis, done on 673 tweets by (Kouzy et al., 2020) shows that around 25% of tweets are misinformative, and 17% of the tweets have unverifiable information. The authors also report that misinformation rate is higher than informal individual accounts and that certain tags like “@2019_nconv” and “Corona” are associated with higher misinformation than tags like “COVID-19”. (Yang et al., 2020) reports that most of the misinformation spreads via retweets and that social bots are involved in amplifying and posting low-credibility information. Furthermore, they analyze that the amount of misinformation on Covid-19 is more than the total volume of New York Times articles. There is no research in the automatic identification of misinformative posts on social media so far.

3 Brief description of Task and Dataset

In this section, I present a brief description about the task and the dataset provided. Interested readers can refer the task description paper (Nguyen et al., 2020) for further details.

3.1 Task

The objective of WNUT-2020 Task 2 is to develop systems that automatically identify an English tweet related to Covid-19 as informative or not informative. Hence, this is a binary classification task.

3.2 Dataset

The complete dataset provided by the organizers of the task, consists of 10K tweets related to Covid-19 of which 4719 tweets are informative and 5281 tweets are uninformative. Table 1 provide samples of informative and uninformative tweets from the dataset.

4 System Description

4.1 Preprocessing

The tweets provided by organizers are anonymized. All the user mentions are replaced with ‘@USER.’ Basic preprocessing like URL removal is also done beforehand by. In my work, I experimented with two different preprocessing techniques to analyze the impact of preprocessing on the system's final performance. Later, I fine-tuned the models with final tweets obtained using both the preprocessing techniques. One of the techniques does minimal preprocessing, which can help preserve additional semantics in tweets. The other one does complete preprocessing along with the introduction of additional features for emojis.

4.1.1 Minimal preprocessing

In this step, I perform the following steps:

- Removal of user mentions and numbers.
- Removal of the hash symbol in hashtags. Here, I am not deleting hashtags as they can have some useful information, and removing them completely may lead to loss of information. Few examples are - “#IndiaFightsCOVID”, “#SummerVibes”. The first hashtag is more promising to be from an informative tweet than the second one.
- Removal of url symbol - “HTTPURL”.

Text	Label
#COVID19 Updates. #SothSudan's fist positive patient is a 29 yo Dutch, who arrived in the country from Netherlands via Addis Ababa on 28 February. #SSOT HTTPURL	INFO
Democrats somehow managed to fight ebola without calling it ""the African virus."" A cluster of COVID-19 cases has emerged in New York City's Hassidic neighborhood, so it's only a matter of time before the local Trump Klux Klan starts talking about ""the Jew virus.""	UNINFO

Table 1: Sample tweets from the dataset.

- Removal of Emojis.

Often, proper-casing of letters and the presence of punctuation helps in improving the embeddings extracted by Transformer architectures. Few transformer architectures' versions like Bert-base-cased and Bert-large-cased are expected to perform well with this technique.

4.1.2 Complete preprocessing

In this technique, first three steps remain the same as that of minimal preprocessing. Additional steps followed are:

- Lower casing of letters.
- Conversion of emojis to text².
- Removal of extra white spaces, punctuation, and all non-alphabetical characters.

This kind of preprocessing is best suited for 'uncased' versions of the Transformer networks like Bert-large-uncased.

4.2 Transformer architectures

Recent advancements in several NLP tasks, particularly text classification, are made possible by Transformer architectures. At their core, Transformers use attention mechanism, which helps in representing the contextual information. They have outperformed all the traditional methods like Bag of Words models, N-grams features, and dictionary-based techniques in most of the NLP domain tasks. These architectures had shown good performance in cross-lingual and multilingual contexts as well (Conneau et al., 2019). Due to these architectures' success, I focused entirely on fine-tuning transformer models in my work to identify informative Covid-19 tweets. I used the following Transformer models for fine-tuning:

- BERT based models(base uncased, base cased, large uncased, large cased).

- RoBERTa-based models(base, large).
- XLM-RoBERTa-based models(base, large).
- ELECTRA-based models(base discriminator, small discriminator, large discriminator).

A brief description of each of the architectures is provided below:

BERT BERT(Bidirectional Encoder Representations from Transformers)(Devlin et al., 2018) is a multi-layer bidirectional encoder transformer. The model is pre-trained on two tasks - Masked Language Modeling(MLM) and Next Sentence Prediction. There are two broad variants of the architecture available - BERT Base with 12 Transformer layers & 110 million parameters and BERT Large with 24 Transformer layers & 340 million parameters. This architecture is a path-breaking invention in the field of NLP, and several other variants of this are developed later. Few of them are Multilingual BERT(mBERT), A Lite BERT(ALBERT), and A Distilled version of BERT(DistilBERT).

RoBERTa RoBERTa(Liu et al., 2019) builds on BERT's masked language modeling task, and this architecture does not have the sentence prediction task in its training objectives. The authors of RoBERTa modified key hyper-parameters of BERT and trained the model with larger data and larger mini-batches. Also, the masking pattern applied to training data is changed dynamically in RoBERTa. This helped RoBERTa to perform better on MLM objective and leading to good down-stream task performance. The base version of RoBERTa has 125M parameters, and the large version has 355M parameters.

XLM-RoBERTa XLM-RoBERTa(Conneau et al., 2019) is a multilingual model trained on 2.5 TB data from CommonCrawl. This model is released by the Facebook AI team as an extension to the XLM-100 model, with the biggest update being the large amount of training data

²<https://pypi.org/project/emoji/>

used in the former. This architecture showed improved performance on several NLP tasks of low-resource languages and outperformed other transformer models like mBERT on cross-lingual benchmarks. XLM-RoBERTa’s base version has 250M parameters, while its large version has 560M parameters. Both versions have a vocabulary size of 250K.

ELECTRA ELECTRA(Efficiently Learning an Encoder that Classifies Token Replacements Accurately)(Clark et al., 2020) matched the performances of its predecessor models like RoBERTa and XLNet with less than one-fourth of computational budget and achieved State-of-the-art performance on SQuAD benchmark. The general design of GANs highly inspires this architecture design. The model used a new pre-training objective - Replaced Token Detection. Three variants of ELECTRA are available: ELECTRA-Small with 14M parameters, ELECTRA-Base with 110M parameters, and ELECTRA-Large with 335M parameters.

4.3 Training

In this work, I fine-tuned all the aforementioned Transformer models using the FARM framework³. Minimal hyper-parameter tuning is performed. Before the test data was made available by the organizers, I have used the provided validation data as test data, and 10% of the training data for dev set. A batch size of 32 is used, and the class weights are incorporated in loss function(Binary Cross Entropy) to upweight the loss of minorities as the dataset is imbalanced. AdamW optimizer is used for optimization. Early stopping with a patience of 5 is used targeting positive-class F1-score, the evaluation metric specified by organizers. A maximum epochs of 50 are specified. Dropout of 0.2 is used to prevent over-fitting. Sequence length of 70 is used as around 90% of the tweets have fewer than 70 tokens. I have evaluated the model once for every 100 batches during fine-tuning. So, the model is evaluated once in 1.25 epochs. All the experiments are performed on Google Colaboratory.

5 Results

5.1 Internal Evaluation Results

As mentioned in 4.3, validation data is used as test data for internal evaluation, and the final evaluation is done by organizers using gold test labels.

³<https://github.com/deepset-ai/FARM>

Model	Prepro	F1	Acc
BERT-base-uncased	comp	0.86	0.81
BERT-base-uncased	comp	0.86	0.86
BERT-base-cased	min	0.86	0.87
BERT-large-uncased	comp	0.86	0.86
BERT-large-cased	min	0.83	0.85
RoBERTa-base	min	0.87	0.87
RoBERTa-large	min	0.89	0.89
XLM-RoBERTa-base	comp	0.85	0.85
XLM-RoBERTa-large	comp	0.88	0.88
ELECTRA-base	min	0.83	0.83
ELECTRA-small	min	0.82	0.83
ELECTRA-large	min	0.85	0.85

Table 2: Internal Evaluation Results.

Run	F1-score	Acc
Run-1	0.85	0.85
Run-2	0.83	0.83

Table 3: Official Results.

Internal evaluation results are provided in Table 2. From results, it can be inferred that minimal preprocessing, which preserves casing and punctuation achieves better results with fine-tuning.

5.2 Official Results

In the organizers’ official results, my two submissions(best performing ones in internal evaluation results) obtained 85% and 83% positive class F1-score. The detailed results are provided in Table 3.

6 Conclusion

In this work, I presented my system submission details for WNUT-2020 Task 2. My submissions, using eleven variants of four different fine-tuned Transformer architectures achieved F1-score of 85.3% and 83.2% during official evaluation. The ill-effects of misinformation and the importance of automatic identification techniques regarding Covid-19 posts are also highlighted in the paper. My work also emphasizes that minimal preprocessing works better than complete preprocessing for fine-tuning State-of-the-art Transformers.

I make my source code publicly available to facilitate further experimentation in the field⁴.

⁴<https://github.com/SivaAndMe/Fine-tuning-Transformer-architectures-for-identification>

References

- Kevin Clark, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. *arXiv preprint arXiv:2003.10555*.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Riccardo Gallotti, Francesco Valle, Nicola Castaldo, Pierluigi Sacco, and Manlio De Domenico. 2020. Assessing the risks of “infodemics” in response to covid-19 epidemics. *arXiv preprint arXiv:2004.03997*.
- Ramez Kouzy, Joseph Abi Jaoude, Afif Kraitem, Molly B El Alam, Basil Karam, Elio Adib, Jabra Zarka, Cindy Traboulsi, Elie W Akl, and Khalil Baddour. 2020. Coronavirus goes viral: quantifying the covid-19 misinformation epidemic on twitter. *Cureus*, 12(3).
- The Lancet. 2020. Covid-19: fighting panic with information. *Lancet (London, England)*, 395(10224):537.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Areeb Mian and Shujhat Khan. 2020. Coronavirus: the spread of misinformation. *BMC medicine*, 18(1):1–2.
- Dat Quoc Nguyen, Thanh Vu, Afshin Rahimi, Mai Hoang Dao, Linh The Nguyen, and Long Doan. 2020. WNUT-2020 Task 2: Identification of Informative COVID-19 English Tweets. In *Proceedings of the 6th Workshop on Noisy User-generated Text*.
- Gordon Pennycook and David G. Rand. 2020. Who falls for fake news? the roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, 88(2):185–200.
- Gautam Kishore Shahi, A. Dirkson, and T. A. Majchrzak. 2020. An exploratory study of covid-19 misinformation on twitter. *ArXiv*, abs/2005.05710.
- Mary E Wilson and Lin H Chen. 2020. Travellers give wings to novel coronavirus (2019-ncov).
- Kai-Cheng Yang, Christopher Torres-Lugo, and Filippo Menczer. 2020. Prevalence of low-credibility information on twitter during the covid-19 outbreak. *arXiv preprint arXiv:2004.14484*.