Rumour Detection and COVID-19 Rumour Analysis

Jin Hong Yong
Student ID: 1198833
yongjy@student.unimelb.edu.au

Abstract

Since rumours has a significant social and economic effect, the detection of disinformation and misinformation has gained traction in recent years. Utilising several State-of-the-Art (SOTA) pre-trained language models based on the Transformer architecture, a binary content-based classification approach for detecting rumours is presented. The various experiments were based on the provided Rumours detection dataset, deploying Longformer, BERT, and RoBERTa, as well as different combinations of hyperparameters. Data was preprocessed via a multitude of preprocessing steps, including the concatenation of tweets and reply tweets. Transformers were chosen for the task due to their nature of being able to achieve imperative results in the absence of large datasets. Evaluation demonstrates that concatenating tweets and using a higher maximum sequence length in training the model yielded a 0.83 F1 score on the test set and a F1 score of 0.8187 on Codalab. Topic analysis, hashtag analysis and sentiment analysis further provides insight into the content of rumours for future development of better models.

1 Introduction

Various investigations fake news and rumours on various platforms has lead to the identification that the spread of rumours can have far-reaching and crippling consequences. For example, according to a study by (Rapoza, 2017) in the relationship between rumours and stock markets, it was pointed out that after a fake Associated Press's tweet alleged Barack Obama for being injured following an explosion in 2013, $130 billion in stock valuation was wiped out within minutes. Moreover, an investigation by (Gunther et al., 2018) has expressed concerns over fake news affecting the United States' 2016 Presidential election, which is further backed by evidence by (Allcott and Gentzkow, 2017), where their database indicated that 115 pro-Trump fake stories have been shared 30 million times on Facebook and 41 pro-Clinton fake stories have been circulated 7.6 million times.

In this paper, we leverage twitter data in the form of tweets and user replies on detecting rumours and fake news propagated on social media platforms such as twitter.

We focus on detecting false statements by looking at the public's reaction towards the claim. This approach taps into the joint knowledge of the public where users share erroneous information collectively, via applying natural language processing to comments directed at an argument.

According to recent SOTA models by (Ma et al., 2018) and (Kumar and Carley, 2019), the former had arranged the source argument and its consecutive replied tweets in a tree structure as shown in Figure **??**, whereas the latter leveraged LSTM on the former's work. The common ground between the two studies is that both had attempted to capture structural information within the rumour thread, via a tree structure. However, the downside is tree models do not specifically model interactions between nodes from other branches aside from the main branch, leaving out possible key information uncaptured, which is a severe limitation. Therefore, we leverage transformer achitecture with self-attention mechanisms to encapsulate structural information, thus modelling interactions between all tweets to excavate the underlying information from the agglomeration of replying tweets.

2 Methodology

2.1 Transformers

The discovery by (Vaswani et al., 2017) stated that attention mechanism enhances the modelling of long-range dependencies. Thus, it can be effectively applied to model interactions between tweets where many exchanges and discussions take place.

2.1.1 BERT

BERT was first introduced by Google in 2018, built to condition both left and right content in all layers to pretrain deep bidirectional representations from unlabeled text (Devlin et al., 2018). Among the few pretraining models, **BERT** was pretrained with Masked Language Modelling (MLM), where 15% of a sentence is replaced with a "[MASK]" token instead of the original tokens, allowing the model to learn the entire meaning of the sequence. As such, BERT will be able to capture underlying information in a tweet thread.

2.1.2 RoBERTa

Short for Robustly Optimized BERT Approach, RoBERTa had improved the results of BERT, introduced by (Liu et al., 2019). RoBERTa eliminates the Next Sentence Prediction (NSP) task from BERT's pre-training and replaces it with dynamic masking, in which the masked token changes over time, allowing the model to better contextualize a sequence. In addition, RoBERTa uses a substantial amount of training data during pretraining as compared to BERT, resulting in a better performance over BERT.

2.1.3 Longformer

Due to the self-attention operation scaling quadratically with sequence length, transformer-based models are unable to process long sequences. To overcome this constraint, Longformer was introduced by (Beltagy et al., 2020) which features an attention mechanism that scales linearly with sequence length, making it simple to process documents with substantial tokens. On long document tasks, Longformer consistently outperforms RoBERTa and sets new benchmarks on WikiHop and TriviaQA. Therefore, it is worth considering exploring Longformer's performance on the twitter thread rumour detection use case.

3 Experiments and Results

3.1 Data and Preprocessing

The data was cleaned by removing redundant information: URLs, emoji, mention-only and numbers-only tweets. As claimed by Khoo et al. (2020), the information carried by tweets can vary depending on the time they were created. Sceptical tweets towards the end of propagation could suggest a high likelihood that the source claim is a rumour. That being the case, we sort tweets in a thread according to time created, then concatenate them in chronological order. Finally, we truncate each block of string to length of `max_sequence_length`, composed of `max_sequence_length/2` starting tweets and `max_sequence_length/2` ending tweets.

Method	F1	Precision	Recall
BERT128	0.8289	0.8289	0.8289
BERT256	0.8201	0.8288	0.8115
BERT512	0.8164	0.8370	0.7967
RoBERTa128	0.8315	0.8181	0.8453
RoBERTa256	0.8065	0.7914	0.8222
RoBERTa512	0.8356	0.8289	0.8423
Longformer256	0.8324	0.7968	0.8713
Longformer512	0.8089	0.7807	0.8391
Longformer1024	0.4274	0.2834	0.8689

Table 1: Table showing F1 score, Precision and Recall for various approaches towards rumouor detection

3.2 Experimental Setup

For **BERT** and **RoBERTa**, we varied the `max_sequence_length` from 128, 256 to 512, with a fixed `batch_size` of 16 and 7 epochs. For the Longformer model, we varied the `max_sequence_length` from 256, 512 to 1024, and a `batch_size` of 8 due to memory limitations of the GPU. We have also incorporated weights for each class during training via dividing the size of unique classes by the size of largest class.

3.3 Results and Discussion

Results from Table 1 indicates that increasing the `max_sequence_length` pass 512 does not necessarily increase performance in the model, and might deteriorate the model's performance, possibly due to inability to capture important context. Various processing methods such as summarization and removing stop words might improve the performance when using greater sequence length. The models RoBERTa512 and Longformer256, being the highest F1 scoring model, were submitted to Codalab, returned with a F1 score of 0.8187 and 0.8023 respectively.

In summary, the RoBERTa512 model outperforms every other candidates, and will be chosen as the model to perform analysis on COVID-19 rumours in Section 5.

4 COVID-19 rumour Analysis

Using the optimal model from Section 4, rumours and non-rumours were predicted from the provided COVID-19 data. Data were first preprocessed as in Section 3, then the rumours and non-rumours threads were derived after utilising the predicted results. Subsequently, several analysis were performed to acquire insights into the distinction between rumours and non-rumours.

4.1 Topic Analysis: Topic Modelling using Latent Dirchlet Allocation (LDA)

Via LDA, underlying topics for each rumours and non-rumours thread were extracted, along with their respective top-10 keywords, as indicated in Table 2 and Table 3.

Topics	Keywords
Social Impacts	peace, corruption, rape, flight, anywhere, threaten, refund, travel, professor, coach
Law Violation	reportedly, owner, delete, mild, allegedly, defy, violate, woman, hannity, score
Carrier	claim, escape, madagascar, agent, carrier, belong, flee, separate, poison, false
Election	ballot, elderly, player, stock, facility, order, equipment, learner, african, nature
Politics	dem, employee, biden, distract, former, theory, billionaire, distraction, deflect, crook
COVID Death Cause	die, death, covid, cause, kill, count, would, list, symptom, condition
COVID Cases	test, positive, covid, family, symptom, day, get, negative, contact, member
COVID Reports	death, case, covid, number, report, new, confirm, state, rate, total
COVID Awareness	people, covid, get, go, think, virus, take, know, make, may
Politic statement	trump, say, covid, people, would, go, know, lie, pandemic, make

Table 2: Table showing topics in rumour threads and Top 10 keywords for each topic

It was observed that the difference in topics between rumours and non-rumours are that non-rumours involves discussion on Vaccines, Education and Fake News which is not seen in topics in rumours. On the other hand, rumours tend to discuss politics and conspiracy theories, as seen from keywords in the Politics topic. In addition, by investigating the similarity between keywords of identical topics in rumours and non-rumours, it was observed that rumours has a high probability of stemming from political statements, reports on COVID-19 cases, their transmission methods as well as awareness threads.

Topics	Keywords
Law	law, antibody, presidential, rewrite, herd_immunity, mark, bro, native, italian, cafe
News Announcements	transmission, extend, announcement, season, homeless, train, fever, oxygen, breathing, rare
Vaccine	vaccine, develop, discuss, tonight, animal, mostly, network, ride, therapy, director
Education	school, child, kid, student, teacher, parent, restriction, exam, open, online
Fake News	focus, scandal, important, abortion, post, newspaper, urge, imagine, social, investigation
Election	trump, covid, dead, die, people, vote, american, death, go, kill
COVID Transmission	virus, spread, call, come, stop, chinese, start, must, human, make
COVID Reports	death, case, covid, number, day, new, high, report, rate, today
COVID Awareness	covid, people, go, get, die, know, many, think, see, would
Politic statement	say, know, would, tell, take, lie, make, stop, could, need

Table 3: Table showing topics in rumour threads and Top 10 keywords for each topic

Topics progression over time can be seen from Table 4 and Table 5. In rumours, the topics evolved from safety measures of COVID-19 to worries of lockdown and finally to citing worries on unemployment. On the flip side, the non-rumours' topics progressed from transmission of COVID-19, to fighting pandemic via government supports and finally lockdowns and safety measures in public agaisnt COVID-19, which are similar to official news.

4.2 Hashtag Analysis

Hashtags used in each tweet thread were extracted and categorised into rumours and non-rumours. From Table 6, it was observed that most of the hashtags overlap between rumours and non-rumours. However, the usage of "breaking" hashtag is more frequent in the rumours thread. Thus, it might indicate rumours with hashtags including "breaking" have a high tendency towards being a rumour. From the overlapping hashtags, it can be inferred that hashtags are not very effective in deducing rumours.

Text Index	Keywords
9	mask, covid, wear, school, state, rule
108	inform, link, active, excess, detail, interview
110	virus, government, chinese, world, come
765	police, target, cancer, arrest, rip, lockdown
907	mail, fund, company, jail, fraud, criticise
999	doctor, treatment, use, treat, patient, cure
1099	trump, say, covid, people, would, go, know
1236	test, positive, covid, family, symptom, day
1566	job, lose, due, back, work, go, covid, new
1935	patient, home, hospital, send, care, money

Table 4: Table showing topics progression in rumour threads and keywords for each topic

Text Index	Keywords
76	transmission, extend, announcement, season, ho...
179	trump, covid, dead, die, people, vote, america...
1277	help, work, thank, due, need, family, take, co...
1435	money, pay, bill, fund, relief, business, comp...
3571	protest, black, riot, rally, police, protester...
4546	fight, pandemic, government, crisis, support, ...
4759	death, case, covid, number, day, new, high, re...
5222	focus, scandal, important, abortion, post, new...
5266	lockdown, government, pm, public, parliament, ...
6731	mask, wear, face, protect, wear_mask, public, ...

Table 5: Table showing topics progression in non-rumour threads and keywords for each topic

Rumours	Frequency	Non-rumours	Frequency
COVID19	282	COVID19	1834
coronavirus	151	coronavirus	761
Coronavirus	35	Coronavirus	180
BREAKING	35	Covid19	176
Covid19	23	covid19	115
covid19	19	BREAKING	54
COVID	7	CoronaVirus	48
COVID19PH	7	COVID-19	43
CoronaVirus	7	CoronavirusPandemic	42
Breaking	6	China	42

Table 6: Table showing top-10 hashtags in rumour and non-rumour threads

Type	Positive	Negative	Total
Rumour	180	2468	2648
Non-rumour	1971	12839	14810

Table 7: Table showing F1 score, Precision and Recall for various approaches towards rumour detection

4.3 Sentiment Analysis

A sentiment analysis model by `transformers` has been utilised to analyse the underlying emotions conveyed by rumours and non-rumours. According to the results in Table 7, both rumour and non-rumour has more negative sentiments compared to positive sentiments. However, the distribution of sentiments in non-rumours varies slightly from that of rumour, citing a 13.31% positive sentiments, a difference of 6.51% compared to 6.8% positive sentiments in rumours. Thus, it could be deduced that non-rumours tend to convey a slightly more positive sentiment, whereas rumours have a high inclination towards instigating negative emotions. By referring to Table 2, evidence shows that rumours have a higher count of negative sentiments.

5 Conclusion

RoBERTa having trained on a larger corpus, had an increased performance compared to BERT. Meanwhile incorporating long sequences do not always increase the accuracy of rumour detection, depicted by the results from utilising Longformer. Topic analysis provides insight into the content of rumours, which can be used to develop better models at detecting rumours. Contradictorily, hashtags were not an effective measure of detecting rumours. Lastly, Rumours were found to convey more negative sentiments, such as inducing panics.

References

Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives*, 31(2):211–36.

Iz Beltagy, Matthew E Peters, and Arman Cohan. 2020. Longformer: The long-document transformer. *arXiv preprint arXiv:2004.05150*.

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*.

Richard Gunther, Paul A Beck, and Erik C Nisbet. 2018. Fake news did have a significant impact on the vote in the 2016 election: Original full-length version with methodological appendix. *Unpublished manuscript, Ohio State University, Columbus, OH*.

Ling Min Serena Khoo, Hai Leong Chieu, Zhong Qian, and Jing Jiang. 2020. Interpretable rumor detection in microblogs by attending to user interactions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8783–8790.

Sumeet Kumar and Kathleen M Carley. 2019. Tree lstms with convolution units to predict stance and rumor veracity in social media conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5047–5058.

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*.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor detection on twitter with tree-structured recursive neural networks. Association for Computational Linguistics.

Kenneth Rapoza. 2017. Can ‘fake news’ impact the stock market? *by Forbes*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *arXiv preprint arXiv:1706.03762*.