Detection and Exploratory Analysis of COVID-19 Rumor Tweets

Authors

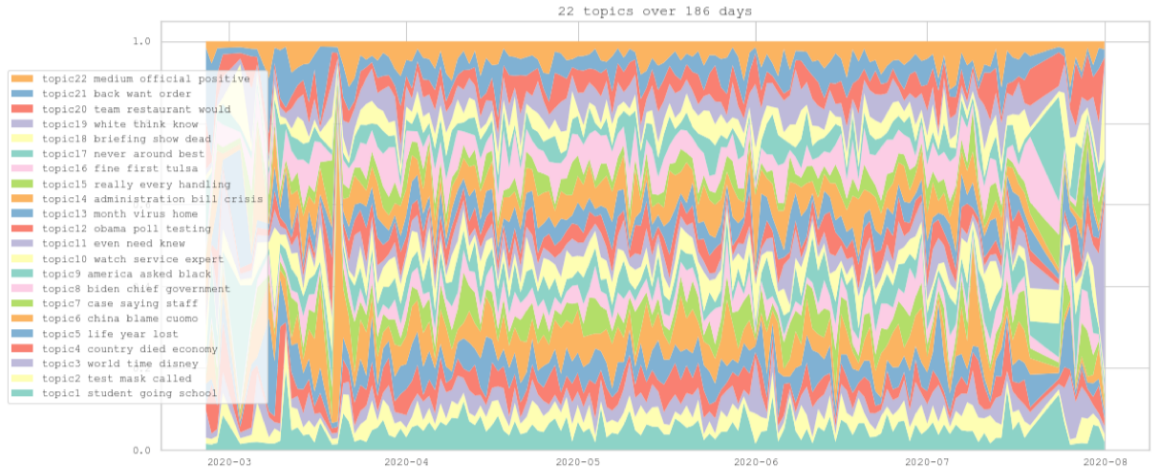

Figure 3: Evolution of latent rumor topics from February 27th to August 1st. [5]

1 Introduction

In recent years, rumor detection has received more and more attention. Devlin et al. (2018) proposed BERT which is widely used in the NLP field and has achieved good results in rumor detection. Nguyen et al. (2020) proposed the BERTweet model to further improve the performance of Bert on English Tweets. Bian et al. (2020) adopted BiGCN, taking the structure of wide dispersion into account.

Furthermore, growing research interest has emerged in studying covid tweets and users. (Tian et al., 2020) studied the user features from metadata and analyzed rumor topics from hashtags; (Wicke and Bolognesi, 2021) analyzed the trendings of covid topics in accordance with social events.

Inspired by these works, the current paper aims to provide rumor detection solutions, identify rumors from a set of covid tweets and study the characteristics of covid rumors. In particular, we are interested to learn: (i) what unique features differentiate rumor users? (ii) lexically, what are salient hashtags/tokens of rumor tweets (iii) topically, what are major rumor categories? How do they evolve? (iv) finally, does the sentiment and emotion distribution on rumors and retweets vary from truth? We present our task-based answers as follows.

2 Task I: Covid Rumor Detection

2.1 Datasets

2.2.1 Machine Learning Approaches

With the given tweet ids, 1565 and 527 sequences of tweets (source and replies) were collected, which constitute our labeled train and development datasets. We consider these datasets highly imbalanced where rumors only take up 20%. Note that a small number of retweets

were ignored on purpose as their source tweets were no longer traceable. The unlabeled test set on Kaggle consists of 558 sequences.

2.2 Methodology and Justification

2.2.2 BERT-based Approaches

2.3 Results and Discussion

2.3.1 Machine Learning Approaches

2.3.2 BERT-based Approaches

2.3.3 Ensemble and Final Evaluation

We adopted the majority voting ensemble to improve model robustness, leading to an F1 score = 0.916.

The ensemble and the BERTweet model as our final submissions achieved F1 scores of 0.916 and 0.918 on the public set, and 0.862 and 0.871 on the private set respectively. We also noticed that some models that we had rejected earlier could have perform even better on the private set. This inconsistency may result from the limited amount of test data as well as lack of model robustness.

3 Task II: Exploratory Analysis of Rumor Tweets

Applying our model on the covid dataset, we recognized 2060 rumors and 13896 replies in total, with their counterparts being 22802 and 107150 correspondingly. Tweets were produced by 6644 unique users, between January 9th and August 1st, as figure 1 suggests.

3.1 Comparison on User Level

3.1.1 Discussion on metadata features

3.2 Comparison on Lexical Level

3.2.1 Hashtags and bigrams

3.2.2 Results & discussion

3.3 Comparison on Topical Level

3.3.1 LDA modeling

3.3.2 Results & discussion

3.3.3 Temporal analysis of rumor topics

3.4 Comparison on Sentiment and Emotions

3.4.1 Distribution across corpus and time

3.4.2 Results & discussion

Tables and Figures

	BoW	TF-IDF	NLTK
NB	0.707/0.283	0.833/0.798	**0.845**/0.770
LR	0.789/0.601	**0.897**/0.772	0.82/0.765
SVM	0.862/0.686	**0.895**/0.766	0.838/0.786

Table 1: Classifier performance of ML models on only source tweets / the whole event tweets.

BERT	Roberta	XLNet	BERTweet
0.917	0.878	0.875	**0.934**

Table 2: Performance of BERT related models.

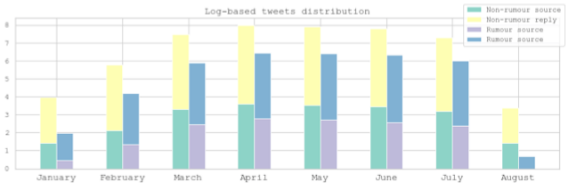

Figure 1: Log-based tweets volume in 8 months

	rumor median	non-rumor median	rumor mean	non-rumor mean
name_len	11	11	11	11
acct_age	4354	4164	3895	3773
#followers	181k	82k	1693k	938k
#friends	1716	1326	8819	5863
#favorites	9992	7447	40641	32098
#posts	37947	27521	86717	70172
#retweeted	831	529	3574	3123
#favorited	2289	1385	14060	11313

Table 3: Analysis of user characteristics.

Top words of 22 latent rumor topics	
1 student going school fall action need week	12 obama poll testing comment question clear threat
2 test mask called much texas wearing message	13 month virus home fact everyone feel anyone
3 world time disney china reopen seen announced	14 administration bill crisis kill republican tell pelosi
4 country died economy great pres shut friend	15 really every handling reopening talk church voter
5 life year lost died return tested claimed	16 fine first tulsa former plan doctor emergency
6 china blame cuomo democrat call report governor	17 never around best something word time human
7 case saying staff supporter ventilator joke contact	18 briefing show dead penny member kung asks
8 biden chief government fighting adviser senator mcconnell	19 white think know come public outbreak asking
9 america asked black thousand border life vote	20 team restaurant would federal care arizona fight
10 watch service expert working rate department mortality	21 back want order making look came true
11 even need knew stop still believe failed	22 medium official positive personal thought story belief

Table 4: Salient words under each topic (N=22)

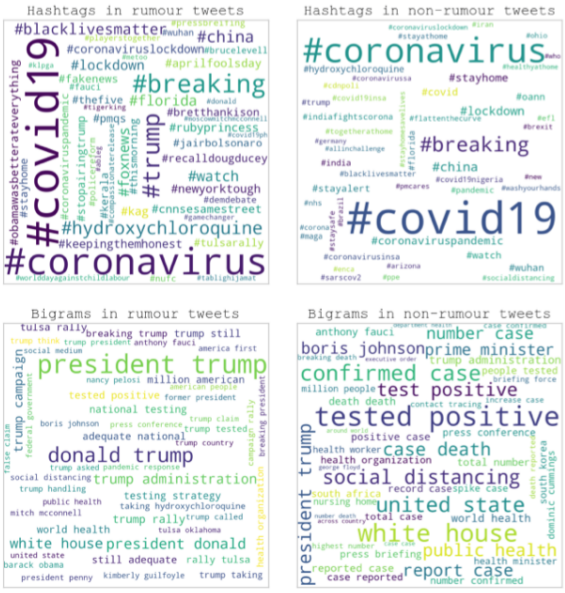

Figure 2: Word cloud of hashtags and bigrams.

Figure 4: Coherence score across different N. Rumor curve peaks at N=5; Non-rumor curve slows down after N=60

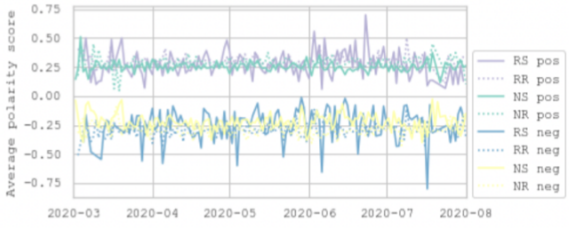

Figure 5: Comparison of temporal polarity score ("RS" for rumor source, "NR" for non-rumor reply)

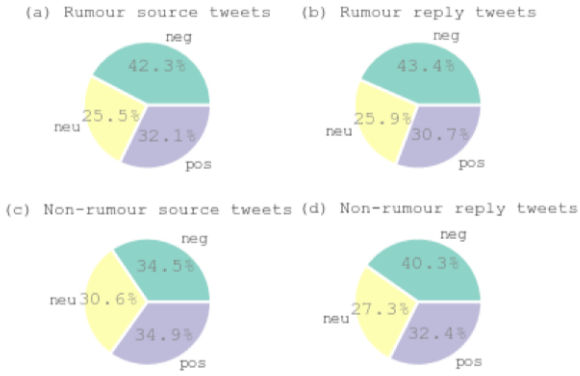

Figure 6: Sentiment distribution

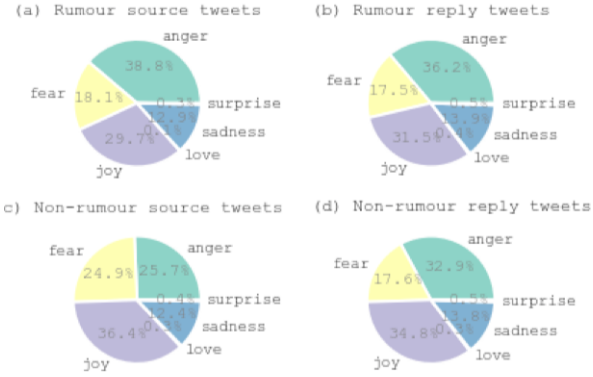

Figure 7: Emotion distribution