

Explore How the U.S. General Public Responds to Breaking News in COVID-19 Pandemic on Twitter

Introduction After the official declaration of the pandemic on March 11th, 2020, the significant social impact of COVID-19 begins to show signs. As the vast majority of people are encouraged to stay at home, Twitter, one of the largest social media, is where countless interpersonal communication happens. In the United States, the country has the highest number of confirmed cases. The government has announced many policies, while President Donald Trump also broadcasts his opinions towards the pandemic. In this study, we attempt to analyze how the overall sentiment of COVID-19 related tweets changes with heated news using machine learning tools. We believe that this project can serve as a starting point for understanding how the development of COVID-19 affected people's feelings and expressions and how distinct groups of Twitter users react differently to COVID-19.

Data The dataset comes from the COVID-19 Twitter dataset (Release 12.0). After subsampling 10,000 tweets for each day from May 19th to June 1st and limiting the user scope to the United States, we got 52,341 tweets with ten selected variables that we hypothesized to be valuable for the analyses. We developed a text-processing pipeline for tweets to ensure that they are comparable to natural language, which a machine can encode. This is achieved by replacing the emojis by plain texts, expanding the abbreviations and contractions, and removing special Twitter components (URLs and username mentions). To train and evaluate our model, we used the SemEval-2017 Task 4A dataset, in which tweet texts are labelled by three classes: negative, neutral, or positive. This dataset also went through the text-processing pipeline.

Model To help us evaluate emotions, we developed our sentiment prediction model precisely optimized for Twitter texts. We experimented with multiple model architectures. For text embeddings, we tried GloVe (Global Vectors for Word Representation), pre-trained BERT (Bidirectional Encoder Representations from Transformers) and RoBERTa (Robustly Optimized BERT Pretraining); for the output layer, GRU (Gated Recurrent Units) and LSTM (Long Short-Term Memory) layers were tested in combination with a multilayer perceptron. Meanwhile, we grid-searched the best set of hyperparameters on over 400 different setups. Our final model builds two GRU layers and a linear output layer on top of the RoBERTa base, and it achieves 70.2% accuracy on the validation dataset.

Analysis The predicted sentiment is represented by three integer scores: 0 for negative, 1 for neutral, and 2 for positive. Using our model to predict a score for each of the tweet texts in our COVID-19 dataset, we did two analyses and interpretations.

Firstly, we plotted the daily average sentiment score against date (March 19th to June 1st), and found headlines of the corresponding day on some of the interesting peaks and valleys. We found that there are more troubling feelings towards worrying news and more positive expressions towards bright news.

Furthermore, we split the tweets using previously selected variables and studied reactions across groups. We found significant differences in tweets' sentiment grouped by retweet count ranges as well as users' follower count, in the sense that popular tweets and users express more depression than others. Time spent on twitter of a user also contributes to varying in their expressions. However, whether a username contains emoji or the users' locations does not appear to influence the sentiment scores. (Full details can be found in our presentation dashboard.)

References

- [1] Duong, Viet & Pham, Phu & Yang, Tongyu & Wang, Yu & Luo, Jiebo. (2020). The Ivory Tower Lost: How College Students Respond Differently than the General Public to the COVID-19 Pandemic.
- [2] Liu, Yinhan & Ott, Myle & Goyal, Naman & Du, Jingfei & Joshi, Mandar & Chen, Danqi & Levy, Omer & Lewis, Mike & Zettlemoyer, Luke & Stoyanov, Veselin. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.
- [3] Devlin, Jacob & Chang, Ming-Wei & Lee, Kenton & Toutanova, Kristina. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.