

Sentiment analysis of Twitter comments related to COVID-19

Group Member

Sijia Han, Dai Dong, Frank Shi

1. Problem Statement

Twitter has one of the most significant microblogging social media in the world. It has 330 million registered users up to the first quarter of 2019 (Tankovska, 2021). For Twitter users, it has become a norm to use tweets and comments limited by the number of characters to express their lives and emotions. During the COVID-19 pandemic that started in December 2019, a significant number of related comments are spread every day through different social media and Twitter. With the trend of the epidemic, analyzing the public sentiment data in the comments can also help understand the public's emotional changes during the epidemic, which can help understand the public's perception of the epidemic. So, in our project, We plan to make a sentiment analysis of tweets based on the monthly comparison. By applying the trend of the number of infected people in the COVID-19 pandemic, we could analyze and determine the relationship between the number of infections and people's emotions.

2. Possible Approach

One of the possible approaches is rules-based sentiment analysis. It is a simple model with a bunch of preset words labelled positive or negative. Then the model would use the word lists to fit in the data and calculate the total of positive, negative and neutral. Then it would determine the sentiment of the sentence (L.I. Tan 2015).

Apply a sentiment dictionary. Since there are some already-done sentiment dictionaries, it would be a better approach than the previous one since we do not need to define the present word list ourselves, and these words have already been given a better score for sentiment analysis (Reagan 2017).

Combine the sentiment dictionary with machine learning. Since the sentiment dictionaries may not perfectly fit with all circumstances, we might need to reset the score of different words in the dictionary. To shoulder the load, LSTM-RNN based ML model could be a good solution (Agrawal, 2020).

In 2017, the International Workshop on Semantic Evaluation opened several tasks on semantic evaluation for participants to work out. Task 4 included one subtask that "Given a tweet, decide whether it expresses the POSITIVE, NEGATIVE or NEUTRAL sentiment." We planned to use and improve the model with the best f1-score in the competition, provided with open code. It is an LSTMs and CNNs model (Cliche, 2017).

3. Plan

- 1) First step will be to modify our dataset and finish our data preprocessing in the first week of March.
- 2) Tokenizing, model training during the middle of March.
- 3) Test result concluding and report writing the end of March and the first week in April.

References

- Agrawal, S. (2020, January 10). *Sentiment Analysis using LSTM (Step-by-Step Tutorial)*. Medium. <https://towardsdatascience.com/sentiment-analysis-using-lstm-step-by-step-50d074f09948>
- Cliche, M. (2017, April 20). *BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs*. ArXiv.Org. <https://arxiv.org/abs/1704.06125>
- Reagan, A.J., Danforth, C.M., Tivnan, B. *et al.* Sentiment analysis methods for understanding large-scale texts: a case for using continuum-scored words and word shift graphs. *EPJ Data Sci.* 6, 28 (2017). <https://doi.org/10.1140/epjds/s13688-017-0121-9>
- L. I. Tan, W. S. Phang, K. O. Chin and A. Patricia, "Rule-Based Sentiment Analysis for Financial News," *2015 IEEE International Conference on Systems, Man, and Cybernetics*, Hong Kong, China, 2015, pp. 1601-1606, doi: 10.1109/SMC.2015.283.
- Tankovska, H. (2021, January 27). Twitter: Monthly active users worldwide. Retrieved March 04, 2021, from [https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/#:~:text=How%20many%20people%20use%20Twitter,daily%20active%20users%20\(mDAU\).](https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/#:~:text=How%20many%20people%20use%20Twitter,daily%20active%20users%20(mDAU).)