Using a Transformer-based Language Model to Track Public Sentiment towards COVID-19 Vaccines

Ercong Luo, Daniel Cisek, Jiaying Zhuang

Introduction

The rise of social media has created a paradigm shift in the way that information is accessed and shared on an individual level, as well on a global scale [1]. Platforms like Twitter and Facebook provide easy access to a range of content that can create, augment, and proliferate both accurate and misleading information on a wide range of subjects. At no time has this been of more importance than the present, with no end in sight yet to the global Covid-19 pandemic. The fight to communicate accurate information about COVID-19 and dispel misinformation has never been more hotly contested, with researchers coining the term *infodemic* to capture the growing risks associated with the spread of misinformation [2,3]. Understanding public sentiment around COVID-19 is a necessity for governments, health organizations, and other relevant parties, as it provides invaluable insight into the spread of information during a pandemic, identifying the most effective avenues to disseminate verified information, provide updates, and advocate for broad social responsibility [4].

Not understanding this information comes at an exacting cost - the spread of misinformation has ramifications far beyond the social media sphere, including real-world impacts like speeding up the epidemic process, contributing to vaccine hesitancy, and undermining coordinated community response to the virus [5,6]. With that in mind, this study seeks to bring insight to one piece of this problem, assessing public sentiment towards COVID-19 vaccines [7]. This problem has a well-established theoretical background in the current research literature, but remains a challenge to solve given the evolving nature of the dialog surrounding COVID-19. As the conversation has evolved surrounding the efficacy of vaccines, and the recent concept of vaccine booster shots, the scope and language of the corresponding misinformation has shifted as well [8]. With this constant evolution comes the necessity of iterative development and the deployment of solutions to monitor and assess public perception over time.

Technology Approach

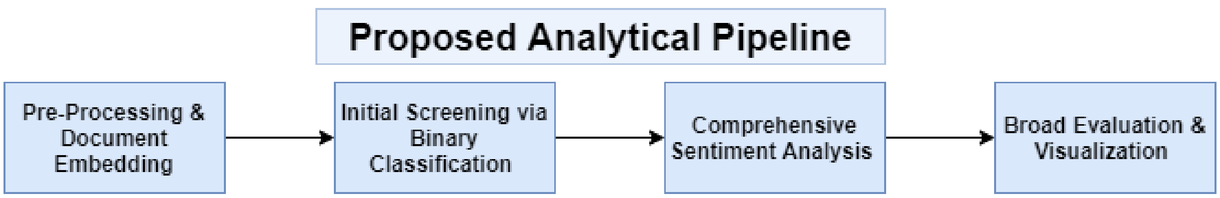
For the main part of the project, we would like to use a neural network model with attention (transformers) to implement a sentiment analysis tool for social media data. The input is a string of text about COVID-19 vaccines, and the output is the category(s) of sentiments that the text exhibits. Due to the scarcity of free labeled data and the lack of robust compute, we will not be training a transformer model completely from scratch. Instead, an already-trained transformer-based sentiment analyzer will be imported, and we will fine-tune the transformer model further with some of our own labeled data. This isn't to say that training a robust model is going to be easy: we anticipate that there will be significant challenges with data preparation in order to fine-tune the model. For example, a tweet can have overall negative sentiments but express positive sentiments towards the COVID-19 vaccine like the ones below:

It was scary to take a jab in the arm, but at least it's done! #Moderna #Vaxxed

Terrified of needles, but glad to have finally gotten vaccinated!

Sure, I'm finally vaccinated against COVID-19 now. But is it even gonna make a real difference if not enough people are going to take the vaccine?

The point that we're trying to make is that there are difficult edge cases such as sarcasm, the effect of negative words in a positive text and vice versa, rhetorical questions that make it difficult to have a robust model. But since this is how real people tweet on Twitter, we will strive to have a model robust enough to address some of these challenges. We have found datasets on Kaggle and GitHub repos that seem sufficient for both fine-tuning, and test accuracy evaluation, albeit further data engineering is needed. If further data is needed, we will pull tweets using the Twitter API and label some data on our own for training or testing purposes, or also make up texts for edge case detection. In addition to the main part of the project, we would like to implement a binary classifier that is a data preprocessing tool to filter out texts that are not related to COVID-19 vaccines, e.g. texts that mention COVID-19 but are not about the COVID-19 vaccines themselves. A flow-chart for our analytical tools can be found below:



In terms of hardware considerations, since the test and training datasets are not anticipated to exceed 5GB, our own laptops might be sufficient for the fine tuning of transformers. One of us has a Macbook Pro with an 8-core CPU and 16GB of RAM. If further compute resources are needed, Professor Larson said that he was willing to provide potential GPU compute resources from his own PC machine. In addition, cloud services such as AWS and Google Cloud provide limited but most likely sufficient free compute for students. The test dataset accuracy will be the main evaluation metric for our model.

Format of Deliverables

Visualizations would be utilized to demonstrate our results of the sentiment analysis. One is to treat it as time series data. The trend of emotions from Dec, 2020, when the first COVID-19 vaccine was made, until now, where everyone is recommended to be fully vaccinated before going to public places, would be visualized in an interactive way. The overall sentiment towards COVID-19 vaccines is highly correlated with the effectiveness of vaccine distribution and disease control, so we believe it serves real social value to study the pattern of how people's attitude towards the COVID-19 vaccines have changed over time, along with educated guesses about what prominent events caused these up and down shifts in public sentiment towards the COVID-19 vaccines.

The second deliverable will identify geographical aspects of the sentiment shifts. There might be an effect of agglomeration in the public perception of the COVID-19 vaccines resulting in what is known in the statistics community as the Simpson's Paradox. Therefore a visualization by demographics such as state location will be implemented, such as a series of heatmaps of sentiments by US states over time compiled into a gif. The geographical distribution of the sentiments might then be associated with other features of states, for instance, the population, GDP, landlocked vs coastal, and so on.Finally, the results of this study will be compiled in a white paper format to ensure the dissemination of our findings, and contribute to the broader research literature on this topic.

Most, if not all of the visualizations will be interactive, to ensure ease of use and quick interpretation. For the time pattern graphs, clicking on a specific period will give the trend of emotions in it. Also one can compare the sentiment change in different time periods in one graph to clearly see the difference between them. For the geographical analysis, click on a state and you will see the data of that state, including number of tweets in total, and number of positive and negative reviews to the vaccines. All of the visuals will be packaged in html files or in a Jupyter Notebook to ensure reproducibility.

References:

1. Babajide Osatuyi, Information sharing on social media sites, Computers in Human Behavior, Volume 29, Issue 6, 2013, Pages 2622-2631, ISSN 0747-5632, https://doi.org/10.1016/j.chb.2013.07.001.

2. Zarocostas, J. How to fight an infodemic. Lancet 395, 676 (2020).

3. WHO Situation Report 13. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf?sfvrsn=195f4010_6. Accessed: 2010-09-30.

4. Alamoodi AH, Zaidan BB, Zaidan AA, et al. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review. Expert Syst Appl. 2021;167:114155. doi:10.1016/j.eswa.2020.114155

5. Kim, L., Fast, S. M. & Markuzon, N. Incorporating media data into a model of infectious disease transmission. PLoS ONE 14, 1 (2019).

6. Cinelli, M., Quattrociocchi, W., Galeazzi, A. et al. The COVID-19 social media infodemic. Sci Rep 10, 16598 (2020). https://doi.org/10.1038/s41598-020-73510-5

7. Wilson SL, Wiysonge CSocial media and vaccine hesitancyBMJ Global Health 2020;5:e004206.

8. Zapan Barua, Sajib Barua, Salma Aktar, Najma Kabir, Mingze Li, Effects of misinformation on COVID-19 individual responses and recommendations for resilience of disastrous consequences of misinformation, Progress in Disaster Science, Volume 8, 2020, 100119,ISSN 2590-0617, https://doi.org/10.1016/j.pdisas.2020.100119.