Content analysis of the Twitter posts after ex-Pres. Trump called COVID-19 virus(SARS-CoV-2) as Chinese Virus.(1497 words)

Himani Anil Deshpande

Data Science, Indiana University Bloomington

February 21, 2022

Abstract

The aim of the project is to analyze the reactions of population using statistical and topical analysis. I have tried to find the reactions of people on the statement made by ex-President Trump on March 16th 2020 where he referred to SARS-CoV-2 as Chinese Virus. And, analyze the engagement that the tweets have received using the number of likes, retweets and comments.

1 Introduction

With the COVID-19 pandemic, there has not only been an increase in fatalities but also an increase in hate-crimes against already marginalized Asian- American population especially ethnic groups with origins in Southeast Asia, East Asia, and the Indian subcontinent.[1,2] The most common forms of the crimes are shunning, taunting, economic boycotts of Asian businesses, harassing and attacking.[3] The Stop AAPI Hate was launched in March 2020 as an initiative to track and respond to incidents or crimes against Asian Americans and Pacific Islanders in United States which were rising alarmingly after the pandemic.[4]

The organization has reported a surge in racial hate crimes and incidents based on the data from March 19th 2020 to March 31st 2020. The report shows that 65% of the incidents were verbal harassment with 18% of bias in the form of shunning, which is avoiding physical contact with someone due to their race as the second highest form of discrimination. The organization suggests that the increase could be due to lifting of restrictions, which allowed people to let a year's worth of pent-up vitriol and hatred to flow through in the form of hate crimes and xenophobic activities.[3]

This also corroborates with another study which shows that social media discussions and posts gives a strong motivation or incitement for real world crimes.[5] In this project I am looking at 2 **research questions**
1. How are the responses/reactions of people after Trump calling corona novel virus as Chinese virus?
2. ~~Analyze~~ how the tweets were engaged by other users using Likes and Retweets?

2 Methods

I have used Twitter which has up to 397 million users to manually collect the tweets. Twitter generates a huge corpus of linguistic data where each user has a limit on content size with which they have to voice opinions. This micro-blogging site makes it easier for searching on a topic using case insensitive hashtags.

So, I have collected 150 tweets in English language from March 12th to March 19th 2020.

Hashtags have been demonstrated to predict the creation of hate organizations and the likelihood of hate crimes.[6]

The data was collected manually using snowball sampling method, as the probability of people tweeting in response to another's tweet were higher. The dataset contains details of the tweet content, the number of likes, the retweets, and discussions on that tweet where people engaged and the date the Tweet were posted.

I have used positive and negative co-occurring hashtags which was compiled using other research papers and news articles.[7,8]

Negative: #WuhanVirus, #ChineseVirus , #ChinaDidThis, #ChinaLiedPeopleDied,

#AsianVirus, #ChinaVirus, #ChineseBioterrorism, #FuckChina, #KungFlu, #MakeChinaPay, #wuhanflu, #jap, #Bongbong [7,8]

Positive: #StopAAPIHate, #IAmNotAVirus, #WashTheHate, #RacismIsAVirus, #IAmNotCovid19, #BeCool2Asi, #AllLivesMatter [7,8]

2.1 Analysis

Using terms like “Chinese Virus” and “Wuhan Virus” is associating the disease with a nation or ethnic groups sparks fear, resentment, anxiety, and disgust towards people who belong to that ethnic background. [3]

The tweets were tagged or coded using categories like Positive, Negative, Neutral/Irrelevant, Dismissing.[7,8]

To identify **Negative** tweets I tried finding content which explicitly agreed with using the racial Hashtags, other racial slurs or blamed the culture, religion, country or ethnic groups.

Example :

I agree wholeheartedly with Donald Trump!! This is a China Virus !!! #COVID2019AU #ChinaVirus

To identify **Positive** tweets I looked at content of the tweet rather than just Hashtags as they are used to attract attention. Some tweets used the negative Hashtags for Positive tweets or for neutral topics on which they were expressing their views. So I tagged tweets positive when they explicitly called out or criticized ex-President Trump’s handle(@realDonaldTrump) or people on verbally attacking an Asian entity or expressed support towards people or group which were targeted.

Example: Trump showing his true colours once again. Vile man. It’s not a virus defined by a nation nor its people #chinesevirus

The Tweets were tagged **Dismissive** when they tried to justify using the Negative hashtags or justifying that its not racism if people belonging to the same ethnic group, country or culture used the same tags like #WuhanVirus for referring to COVID 19.

Example: In Taiwan and HK, COVID-19 is still called #WuhanVirus or #ChineseVirus. If someone posts about "COVID-19" on social media, the comments are usually like "oh you mean the Wuhan virus" or "oh you mean the Chinese virus". Just saying /shrug

The tweets where tagged **Neutral/Irrelevant** when they referred to other political issues or topics which were connected to COVID 19 like the conditions of medical professionals all over the world or the fatalities due to pandemic, news, advertisements.

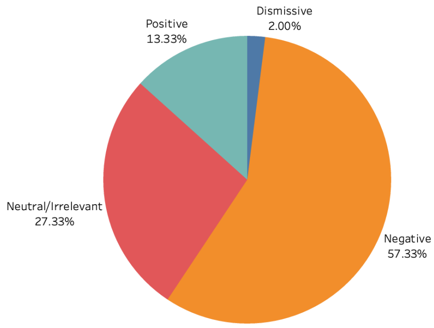

Figure 1: Pie Chart of Distribution of Code/Tags in Dataset

Example: By now, all states should be allowing medical professionals to work across state lines. The virus, while officially in all states, is very localized in terms of its impact. Important to get medical personal from areas with little affect to those with major problems. #WuhanVirus

3 Results

Information visualization is a powerful way for decision-making and getting our findings across.

3.1 Pie-Chart of Code/Tags in Dataset

Figure 1 shows a Pie-chart which shows the distribution of the 4 codes that I have used for identifying the tweets. 57.33% of my dataset contains Negative tweets, which is closely followed by Neutral/Irrelevant tweets at 27.33%. Positive tweets are just 13.33%.

3.2 WordCloud of Tweet Content

Using the content of the tweet, I have created a wordcloud of words with count above 50. From Figure 2 we can see that Virus, Chinese, Wuhan, Wuhanvirus, chinavirus and covid19 have been repeatedly used in the tweets. We can see that words like CCP, communist, responsible, Kungflu, government, racist, disgusting, originated have been used frequently. On the other hand positive words like AllLivesmatter, coronaviruspandemic, iamnotavirus,rascismisavirus have barely been used.

3.3 Word Cloud of codes using Tweet Engagement

I have calculated the mean engagement by summing up number of retweets and likes then dividing by the number of tweets. On comparing the

Figure 2: Wordcloud of Tweet content with occurrence above 50

Figure 3: Word cloud of Tags/codes using mean engagement

Figure 3 and 5, we can observe that the Negative tweets have shown higher engagement when we take into account the number of retweets and likes rather than just comments. Even the positive and dismissive tweets have shown a higher user engagement.

3.4 Bar chart and wordcloud of Codes using Number of Comments

We can see that the Negative tweets have a higher contribution in terms of count. However, I analyzed using the count of comments. From Figure 4 and Figure 5, we can see that surprisingly Neutral/Irrelevant tweets have been the topic of discussion rather than the negative comments.

Going back to the dataset, I can see that 3 tweets have shown to have more than average user engagement. These tweets make the distribution for Neutral/Irrelevant tweets skewed.

3.5 Line Graph of Code/Tags over Date Distribution

Figure 6 shows a Line-chart which shows the distribution of the 4 codes that I have used for identifying the tweets. We can see that the Negative tweets have sparked after 15th March 2020. And after 14th March 2020, the decreasing positive tweets have been increasing and the Neutral/Irrelevant tweets are showing a linear increase.

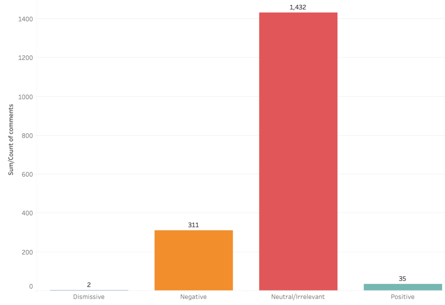

Figure 4: BarChart of Tags/codes using count of Comments

Figure 5: Wordcloud of Tags/codes using count of Comments

4 Conclusions

The project is based on a small dataset of 150 tweets, which has been collected using snowball sampling method. The research question was to identify the reactions of population and to find the user engagement.

I have used tweet content to analyze the reac-

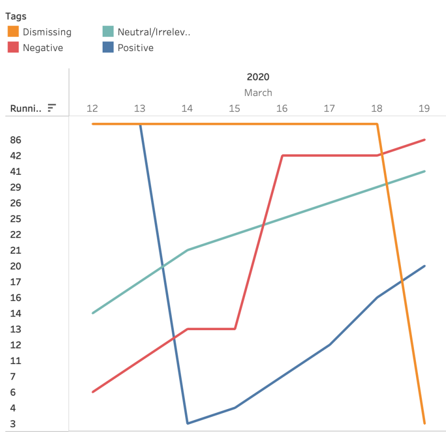

Figure 6: Line Chart of Trend of Code/Tags Over Date Distribution

tions of population, which shows that the racist words have shown to be used more frequently than the positive or neutral words. Using the tweet content for analysing the intention or reactions of user aligns with the sophisticated research methods like Natural Language processing which shows that the negative animosity has increased during pandemic.[2,3,4,7,8]

From figure 6 which takes into account the Running Total of the tweets shows that the Negative tweets have spiked after President Trumps statements.The analysis done on the basis of user engagement and count of comments shows that the population is more neutral than negative.

5 Limitations and Future Works

The research shows that Negative reactions have increased after 16th March 2020. But there are a few limitations like not using a detailed codebook, which could capture diverse set of sentiments and using posts of English language. I have also neglected to use other media platforms like articles, Facebook groups, Reddit, etc.

Another Limitation is that I didn't take the tweets which have Links for analysing the reactions. Using the Twitter API would have made it easier to get other details like the followers, following, age group, and the geographical location of each user.

In future, I would like to use sophisticated methods like Natural Language processing and machine learning for analyzing these added parameters of demographics to make better contribution to the project.

6 References

1. Kang, Y. E.,(2020, Mar 31) "Asian Americans Feel the Bite of Prejudice During the COVID-19 Pandemic," NPR. Retrieved from: URL

2. Rogin, A., Nawaz, A.,(2020, Jun 25)'We Have Been Trough Tis Before.' Why Anti-Asian Hate Crimes are Rising Amid Coronavirus," PBS. Retrieved from: https://www.pbs.org/newshour/nation/we-have-been-through-this-before-why-anti-asian-hate-crimes-are-rising-amid-coronavirus

3. Yam Kimmy (2021, May 11). Anti-Asian hate incident reports nearly doubled in March, new data says. NBC News. Retrieved from: https://www.nbcnews.com/news/asian-america/anti-asian-hate-incident-reports-nearly-doubled-march-new-data-n1266980

4. AAPI Equity Alliance (AAPI Equity), Chinese for Affirmative Action (CAA), and the Asian American Studies Department of San Francisco State University(2020, Mar 19) Stop AAPI Hate Retrieved from: www.stopaapihate.org/about/

5. Relia, K., Li, Z., Cook, S. H., Chunara, R. (2019). Race, Ethnicity and National Origin-Based Discrimination in Social Media and Hate Crimes across 100 U.S. Cities. Proceedings of the International AAAI Conference on Web and Social Media, 13(01), 417-427. Retrieved from: URL

6. Hswen, Y., Xu, X., Hing, A., Hawkins,J B. , Brownstein, J.B., Gee, G C.,(2021 May 1) "Association of "covid19" Versus "chinesevirus" With Anti-Asian Sentiments on Twitter", American Journal of Public Health 111, no. 5. Retrieved from: https://doi.org/10.2105/AJPH.2021.306154 PMID: 33734838

7. He, B., Ziems, C., Soni, S., Ramakrishnan, N., Yang,D., Kumar, S., (2020, May 25). Racism is a Virus: Anti-Asian Hate and Counter speech in Social Media during the COVID-19 Crisis. Retrieved From:https://arxiv.org/abs/2005.12423

8. An, J., Kwak, H., Lee, S. C., Jun, B., Ahn, Y. (2021, Sep 15). Predicting Anti-Asian Hateful Users on Twitter during COVID-19. Retrieved from:https://arxiv.org/abs/2109.07296