Using Twitter to estimate the impact of COVID-19 onto public opinion regarding space exploration

Sven Berberich
Computer science department, Radboud University, Nijmegen, The Netherlands
svberberich@gmail.com

ABSTRACT

In this paper we will perform sentiment analysis on data from Twitter to estimate the change in public opinion on space exploration since the onset of COVID-19. The results show that, since the start of the pandemic, public opinion on space exploration seems to have slightly worsened but has most notably become more intense in either direction (negative or positive).

1 INTRODUCTION

Progress in space exploration has always been very dependent on funding and policies of governments. It is therefore quite important to consider governmental policies when trying to predict the progress in the aerospace industry. Many of these policies are undergoing significant changes since the onset of the COVID-19 pandemic, as it is changing the lives of humans everywhere. Which leads to the question of how COVID-19 impacts governmental space exploration policies. Which in turn are likely mostly related to public opinion about space exploration.[5]

In this paper we are going to use Twitter to estimate the change in public opinion on space exploration since the onset of the COVID-19 pandemic.

This estimation can then be used as indicator for potential policy changes regarding space exploration. The code base and used data for this paper can be found on Github:
github.com/NoxDecima/Space-Opinion-Mining

2 BACKGROUND

The classical way to poll the public opinion over a certain topic is to perform a large scale questionnaire over the population that you want to poll. This process costs a lot of time and effort, but the rise of social media platforms (also refereed to as micro-blogging platforms) such as twitter and Facebook provide us with abundant data about opinions shared on these platforms. This data can be used to acquire representative approximation about public opinions on a specific topic by analysing the data via sentiment analysis. [2]

Prior research has shows that there seems to be a strong correlation between the public opinion of a topic and the topic related policies and funding provided by government. [4] In space exploration specifically, public opinion seems to be directly influence policy making. [5] This makes the new possibilities of quickly gaining representative approximations about the public opinion on a certain topic with micro-blogging platforms even more valuable, since it could provide information about possible future policy changes.

A specific example of using such micro-blogging data to measure interest in space policy has been performed by Wendy N.Whitman Cobb [6]. This paper uses Twitter and Google Trends to estimate interest in space policy within the American public.

TxMM'21-'22, Radboud University, Nijmegen, Netherlands

3 APPROACH

3.1 Data

For our analysis we will be using tweets from Twitter related to the Aerospace industry. To acquire the data from Twitter, we use the open source tool Twint [7]. We collect all tweets within a certain time frame that contain the word "Aerospace", we use this filter to only receive tweets that are relevant for our analysis.

Using this approach we collect data for two separate time frames, namely from the 01.10.19 until the 31.12.19, about three months before the COVID-19 pandemic affected most of the world, we will refer to the data acquired from this time frame as the Pre-COVID dataset. Additionally we collect data from the 01.19.20 until 31.12.20, which is half a year after most countries got affected by COVID-19, this dataset will be referred to as In-COVID dataset.

Both datasets encompass the same three months while being a year apart. This is done to reduce the impact of external confounding factors on the public opinion. The data we collected contains 69869 tweets for the Pre-COVID dataset and 66725 tweets for the In-COVID dataset.

3.2 Preprocessing

Before we can apply our sentiment analysis on our tweets we first have to clean up our data.

- **English language**
  Twitter is a multilingual platform which means that our data contains a multitude of different languages. Although our approach in principle works for any language, we have to make sure that our sentiment analysis method is compatible with the languages present in the data. For this paper we will restrict ourselves to the English language to get the most representative results. This means in practice that all tweets that are not in English are filtered our of our data sets.

- **Duplicate tweets**
  Since we want a representative approximation of public opinion we have to attempt to filter out tweets written by bot accounts. We filter out all the tweets that have the exact same content and leave only the first occurrence of these tweets in the datasets.

All the above mentioned steps are applied to both, the Pre-COVID and In-COVID datasets. After filtering out all these tweets we ended up with 61118 tweets for the Pre-COVID dataset and 59110 tweets for the In-COVID dataset.

3.3 Analysis

Now that our data is cleaned we can perform our sentiment analysis. For this paper we used the VADER sentiment analysis. [3] This sentiment analysis method was specifically created for micro-blogging data and on average outperforms human raters. The output of this method consists out of 4 values, a negative score, a neutral score, a positive score and a compound score which is a combination of the three previous ones. The negative, neutral and positive scores are all on a scale of 0 to 1 indicating the intensity of the corresponding sentiment. The compound score on the other hand is on a scale of -1 to 1 where values close to -1 indicate a overall negative sentiment and values close to 1 indicate a overall positive sentiment.

We apply this method by first tokenizing each tweet into sentences then apply the sentiment analysis to each sentence and average over all sentences for each of our 4 scores, so that we end up with one negative, one neutral, one positive and one compound average score for each tweet.

Finally we want to distinguish whether there is a significant difference in public opinion since before the pandemic and during the pandemic we perform Wilcoxon rank sum tests, also known as Wilcoxon–Mann–Whitney tests [1], to gain statistical results for our question. We apply the test to each of our 4 scores and evaluate the differences in them between the Pre-COVID and the In-COVID datasets.

4 RESULTS

Figure 1 shows the distribution of the compound scores for both datasets. With the human eye these two distributions seem almost identical however they are in fact different. The Pre-COVID dataset has a average compound score of approximately 0.158 while the In-COVID dataset has a average compound score of approximately 0.151.
If we now perform the Wilcoxon rank sum test on the two sets of compound scores we in fact see that there is a significant difference between the two groups of compound scores ($p - value \approx 0.0029$) and that the In-COVID dataset has a on average lower score.

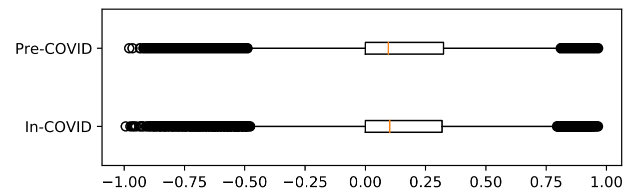

Figure 1: Boxplot of compound scores of the Pre-COVID and In-COVID datasets

Figure 2 shows the the distribution of the negative scores for both datasets. The huge majority of scores is very close to 0 which makes the means difficult to see but what is visible is that the In-COVID dataset seems to have more extreme outliers than the Pre-COVID dataset. The average negative score of the Pre-COVID dataset is approximately 0.017 while the average for the In-COVID dataset is approximately 0.024. Performing the Wilcoxon rank sum test on these sets see that there is significant difference between the two sets of negative scores ($p - value \approx 2.04 \cdot 10^{-60}$) the negative scores are thus significantly higher in the In-COVID dataset than in the Pre-COVID dataset.

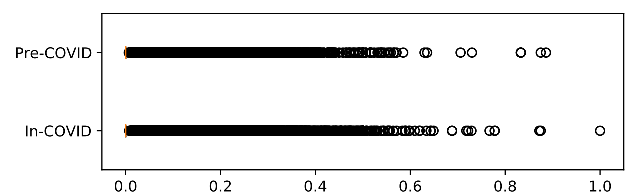

Figure 2: Boxplot of negative scores of the Pre-COVID and In-COVID datasets

Figure 3 shows the the distribution of the neutral scores for both datasets. In the plot we can see that the In-COVID distribution has a higher variance and a lower mean than the Pre-COVID distribution. The mean neutral score for the Pre-COVID dataset is approximately 0.891 while the average neutral score for the In-COVID dataset is approximately 0.879. The results of the Wilcoxon rank sum test indicate yet again a significant difference between the two sets of scores. ($p - value \approx 2.92 \cdot 10^{-70}$) Therefore we can conclude that the neutral scores for the In-COVID dataset are significantly lower than the neutral scores for the Pre-COVID dataset.

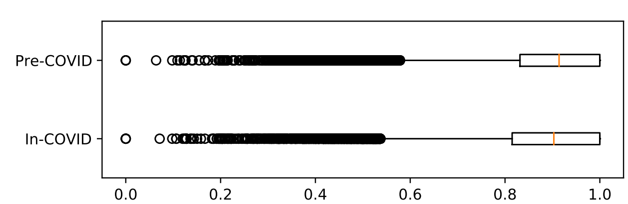

Figure 3: Boxplot of neutral scores of the Pre-COVID and In-COVID datasets

Figure 4 shows the the distribution of the positive scores for both datasets. The plot shows that the mean and variance of the In-COVID dataset are larger than the Pre-COVID dataset for the positive scores. The average positive score for the Pre-COVID dataset is approximately 0.086 and the average positive score for the In-COVID dataset is approximately 0.092. Once again using the Wilcoxon rank sum test we see a significant difference between the two sets of scores ($p - value \approx 6.61 \cdot 10^{-19}$), we can thus conclude that the positive scores are significantly larger in the In-COVID dataset than the Pre-COVID dataset.

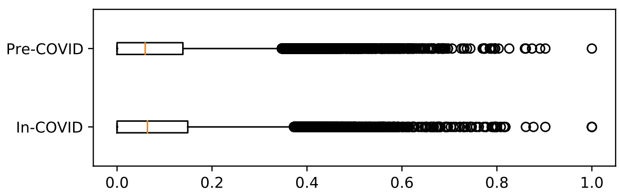

Figure 4: Boxplot of positive scores of the Pre-COVID and In-COVID datasets

Summing up our results regarding each of our 4 scores, we have found out that the compound and neutral scores for the In-COVID dataset are, although only slightly, significantly lower than these scores for the Pre-COVID dataset. We also saw that the values for the positive and negative scores in the In-COVID dataset are significantly larger than the values for these scores in the Pre-COVID dataset.

4.1 Job related tweets

Additionally to the public opinion regarding space we also inspected the relations between the two datasets for the subset of tweets containing the words "job" or "hire" to get an insight into to the public opinion regarding jobs in the Aerospace industry. The number of such tweets in the Pre-COVID dataset is 2228 while the number of such tweets in the In-COVID dataset is 2640.

Figure 5 shows the the distribution of the compound scores for the subset of job tweets of both datasets. We can clearly see that the mean of the In-COVID dataset is larger than the one for the Pre-COVID dataset. The average for the Pre-COVID dataset is approximately 0.128 while the average for the In-COVID dataset is approximately 0.164. Using the Wilcoxon rank sum test we indeed see that the In-COVID dataset has significantly larger compound scores than the Pre-COVID dataset in their respective job related tweets. ($p - value \approx 6.47 \cdot 10^{-9}$)

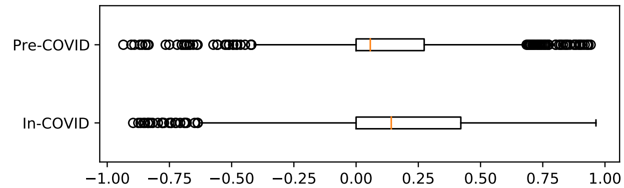

Figure 5: Boxplot of compound scores of the Pre-COVID and In-COVID datasets for the subset of tweets containing either "job" or "hire"

Figure 6 shows the the distribution of the negative scores for the subset of job tweets of both datasets. Here we can see that the subset of job tweets for the In-COVID dataset has a much larger variance over the Pre-COVID dataset. The average value for the Pre-COVID dataset is approximately 0.022 while the average score for the In-COVID dataset is approximately 0.030. Performing another Wilcoxon rank sum test reveals that the In-COVID dataset has significantly larger negative score values than the Pre-COVID dataset within the subset of job related tweets. ($p - value \approx 6.99 \cdot 10^{-6}$)

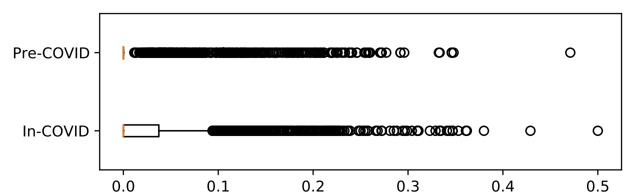

Figure 6: Boxplot of negative scores of the Pre-COVID and In-COVID datasets for the subset of tweets containing either "job" or "hire"

Figure 7 shows the the distribution of the neutral scores for the subset of job tweets of both datasets. In the plot we can clearly see that the mean as well as variance in the In-COVID distribution is smaller than in the Pre-COVID dataset. The average neutral score for the Pre-COVID dataset is approximately 0.893 while the average neutral score for the In-COVID dataset is approximately 0.862. The Wilcoxon rank sum test confirms our assumption that the In-COVID dataset has significantly lower neutral values compared to the Pre-COVID dataset within the job related tweets. ($p-value \approx 9.23 \cdot 10^{-29}$)

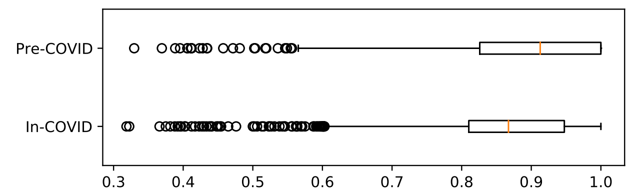

Figure 7: Boxplot of neutral scores of the Pre-COVID and In-COVID datasets for the subset of tweets containing either "job" or "hire"

Figure 8 shows the the distribution of the positive scores for the subset of job tweets of both datasets. Inspecting the plot we can see that the In-COVID dataset has a larger mean and variance than the Pre-COVID dataset. The average positive score for the Pre-COVID dataset is approximately 0.082 while the average positive score for the In-COVID dataset is approximately 0.102. The Wilcoxon rank sum test confirms that the In-COVID dataset is significantly larger than the Pre-COVID dataset within the subset of job related tweets. ($p - value \approx 2.87 \cdot 10^{-18}$)

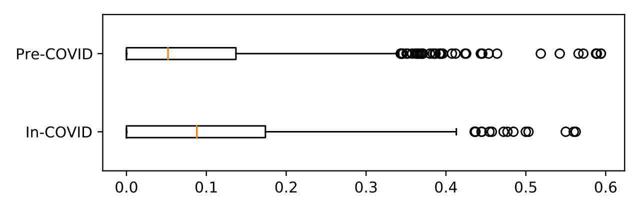

Figure 8: Boxplot of positive scores of the Pre-COVID and In-COVID datasets for the subset of tweets containing either "job" or "hire"

In summary, for the subset of tweets that are job related, we saw that the neutral score values where significantly smaller for the In-COVID dataset than the values for the Pre-COVID dataset. In contrast to that we found out that the scores for the compound, negative and positive scores where significantly larger in the In-COVID dataset than in the Pre-COVID dataset.

5 DISCUSSION

Looking at our results for the whole dataset there are two interesting conclusions we can make. First, it seems that public opinion regarding space exploration has gotten slightly but notably worse since the start of COVID-19. We can see this by the significant decrease in the compounding score values since the onset of COVID-19.

Secondly, it seems that rather then the overall opinion changing, the intensity of opinions seems to have increased. This means that people who had not as strong of an opinion on space exploration before the onset of COVID-19 seem to have become more extreme in their opinion since COVID-19 started. This is indicated by the decreases in neutral score values since the onset of COVID-19 as well as the simultaneous increases in negative as well as positive score values.

Important to mention here is that there are a couple of potential problems with the approach and results described in this paper which would need further research to make sure that the results mentioned in this paper are accurate. These potential problems are:

- **Potential confounding factors**
  It is entirely possible that in between the time of the Pre-COVID and In-COVID dataset another unknown factor had a large impact on public opinion regarding space exploration which could be completely unrelated from COVID-19.
- **Search term bias**
  All collected tweets had to include the term "Aerospace" which means that we in fact estimated public opinion towards this specific term. It could be possible that this specific term has a no linearly related opinion towards general space exploration which could invalidate any connection drawn.
- **Estimation of public opinion**
  Although research has shown that Twitter can be a great tool at estimating public opinion it is still an estimation, so it could be the case that actual public opinion about space exploration differs from this estimate.

5.1 Job related tweets

For the subset of tweets that are job related we also see the same trend of opinions becoming more intense. Also indicated by decreasing neutral score values since the onset of COVID-19 while the positive and negative values increase.

Different however is that we see a significant and notable increase in in the compound score values since the onset of COVID-19. This could imply that although the public opinion as a whole has gotten worse, the job opportunities and industry is doing better than before COVID-19. This does however by no means imply a causation.

On potential problem that should be considered for the job related tweets specifically is the **Filter term bias**. The words "job" and "hire" could have inherent biases which could have influenced these results.

6 CONCLUSION

We have used Twitter data to estimate the change in public opinion on space exploration since the onset of COVID-19 and have identified that opinion seems to have slightly worsened and opinions seem to have become more extreme.

Additionally we have identified that public opinion on job related topics in the space exploration seems to have improved.

REFERENCES

[1] Venita DePuy, Vance W. Berger, and YanYan Zhou. 2005. *Wilcoxon–Mann–Whitney Test.* John Wiley Sons, Ltd. https://doi.org/10.1002/0470013192.bsa712

[2] Balakrishnan Gokulakrishnan, Pavalanathan Priyanthan, Thiruchittampalam Ragavan, Nadarajah Prasath, and AShehan Perera. 2012. Opinion mining and sentiment analysis on a Twitter data stream. In *International Conference on Advances in ICT for Emerging Regions (ICTer2012)*. 182–188. https://doi.org/10.1109/ICTer.2012.6423033

[3] C.J. Hutto and Eric Gilbert. 2015. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014*.

[4] Benjamin I. Page and Robert Y. Shapiro. 1983. Effects of Public Opinion on Policy. *The American Political Science Review* 77, 1 (1983), 175–190. http://www.jstor.org/stable/1956018

[5] Alan Steinberg. 2011. Space policy responsiveness: The relationship between public opinion and NASA funding. *Space Policy* 27, 4 (2011), 240–246. https://doi.org/10.1016/j.spacepol.2011.07.003

[6] Wendy N. Whitman Cobb. 2015. Trending now: Using big data to examine public opinion of space policy. *Space Policy* 32 (2015), 11–16. https://doi.org/10.1016/j.spacepol.2015.02.008

[7] C. Zacharias and F. Poldi. 2021. TWINT - twitter intelligence tool. https://github.com/twintproject/twint.