

Real Time Sentiment Analysis of Tweets

**CS 6301 – Practical Aspects of Data Science
Summer 2020**

Rajarshi Chattopadhyay
RXC170010

Introduction

This project is to extract real-time tweets from Twitter about a topic and use R packages to get sentence (tweet) level sentiments and overall sentiment during a period of time.

Data Collection

The *rtweet* R package has been used to download the tweets about a certain topic – in this case ‘Fauci’ for Dr. Fauci.

- A maximum of 1000 tweets have been set to be downloaded.
- Only English tweets are targeted.
- Recent tweets are fetched.
- Retweets are excluded.

Data Preprocessing

The sentences with URLs (http/https) are removed from consideration.

Sentiment Extraction

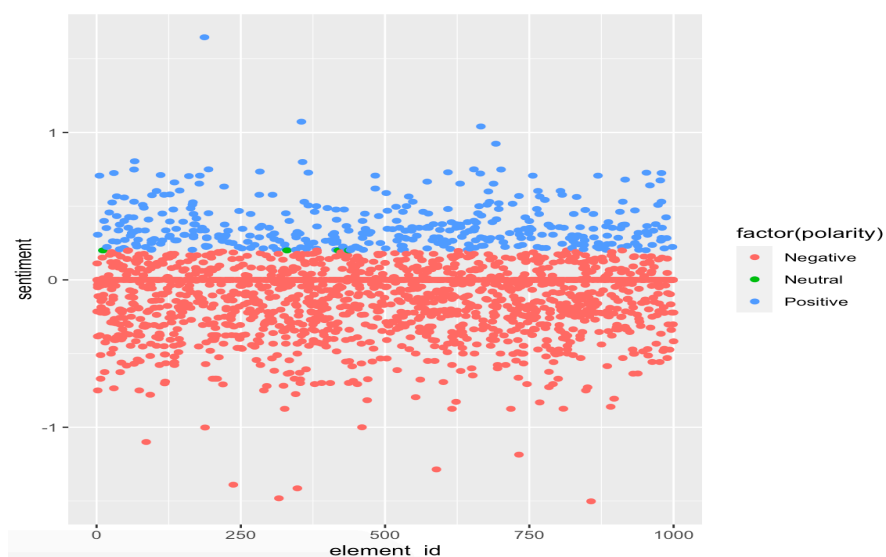
The *sentimentr* R package has been used for sentiment analysis.

The sentence/tweet level sentiment has been analyzed along with the overall sentiment during period of time.

The functions mainly used are:

- `get_sentences()`: To get sentences; returns a list of vectors of sentences.
- `sentiment()`: Approximates the sentiment (polarity) of text by sentence.
- `sentiment_by()`: Approximates the sentiment (polarity) of text by grouping variable(s).

On breaking each tweet into sentences, for the total of 1000 tweets fetched, there are 2360 different sentences. The sentiment of each sentence is obtained.



Sentence Sentiment

↑	sentence	element_id	sentence_id	word_count	sentiment	polarity
1	@GilpinPeri Peri I think Fauci should resign and be abl...	1	1	24	-0.21433035	Negative
2	He needs a big platform.	1	2	5	0.11180340	Negative
3	He can return to NIH in January.	1	3	7	0.00000000	Negative
4	Trump is divorcing Fauci .	2	1	4	-0.05000000	Negative
5	Next he will push a "vaccine" that he "produced" that ...	2	2	20	-0.03354102	Negative

2356	For instance, the guy in Sweden?	999	3	6	0.00000000	Negative
2357	That guy has all that blood on his hands.	999	4	9	0.00000000	Negative
2358	I literally do not understand what you expected Fauci ...	999	5	11	-0.30151134	Negative
2359	White House takes aim at Fauci as he disagrees with T...	1000	1	13	-0.22188008	Negative
2360	I am so sick of this administration attacking people.	1000	2	9	-0.41666667	Negative

Sentiment keywords

	element_id	negative	positive
1:	1	resign,prohibited	big
2:	2		big
3:	3		
4:	4		trump
5:	5	trump,exposed,fraud	

2356:	2356		
2357:	2357		
2358:	2358		understand
2359:	2359	disagrees,trump,virus	white
2360:	2360	sick,attacking	

This lists the different keywords present in the tweet sentences that determine their sentiment.

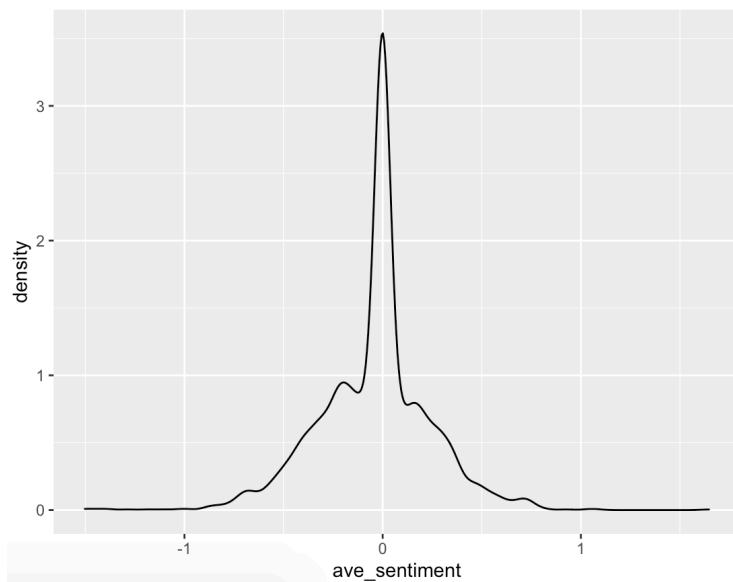
Highlighted sentences based on sentiment (polarity)

- 1: **-0.214**
 @GilpinPeri Peri I think Fauci should resign and be able to go on all the big networks which he is being prohibited from doing.
- 2: **+0.112**
 He needs a big platform.
- 3: **0.000**
 He can return to NIH in January.
- 4: **-0.050**
 Trump is divorcing Fauci .
- 5: **-0.034**
 Next he will push a "vaccine" that he "produced" that Fauci will not recognize because Trump exposed him as "fraud".
- 6: **-0.383**
 That would be the Trump and his enablers story .
- 7: **-0.750**
 Fraud in fraud out.
- 8: **+0.306**
 Dr Fauci is a national treasure .
- 9: **0.000**
 We must all rise to his defence.
- 10: **-0.012**
 @WISH_TV Ah since Fauci didn't suck up to Trump he's now being expedited.
- 11: **0.000**
 Makes sense as to what's going on here.

The red tweets have a negative sentiment associated.

Sentiment Distribution

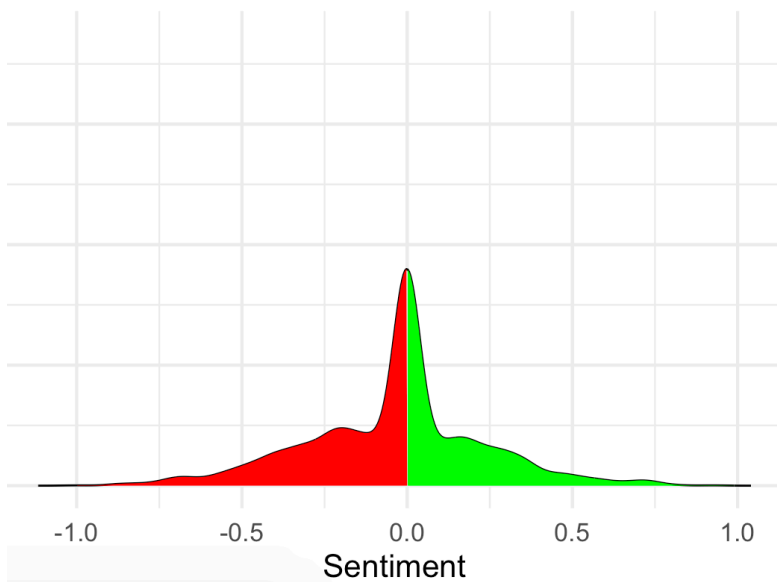
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-1.50241	-0.19368	0.00000	-0.03975	0.08333	1.64625



The tweets have a similar weight based on the average sentiment with the most positive one being 1.64 and the most negative being -1.50.

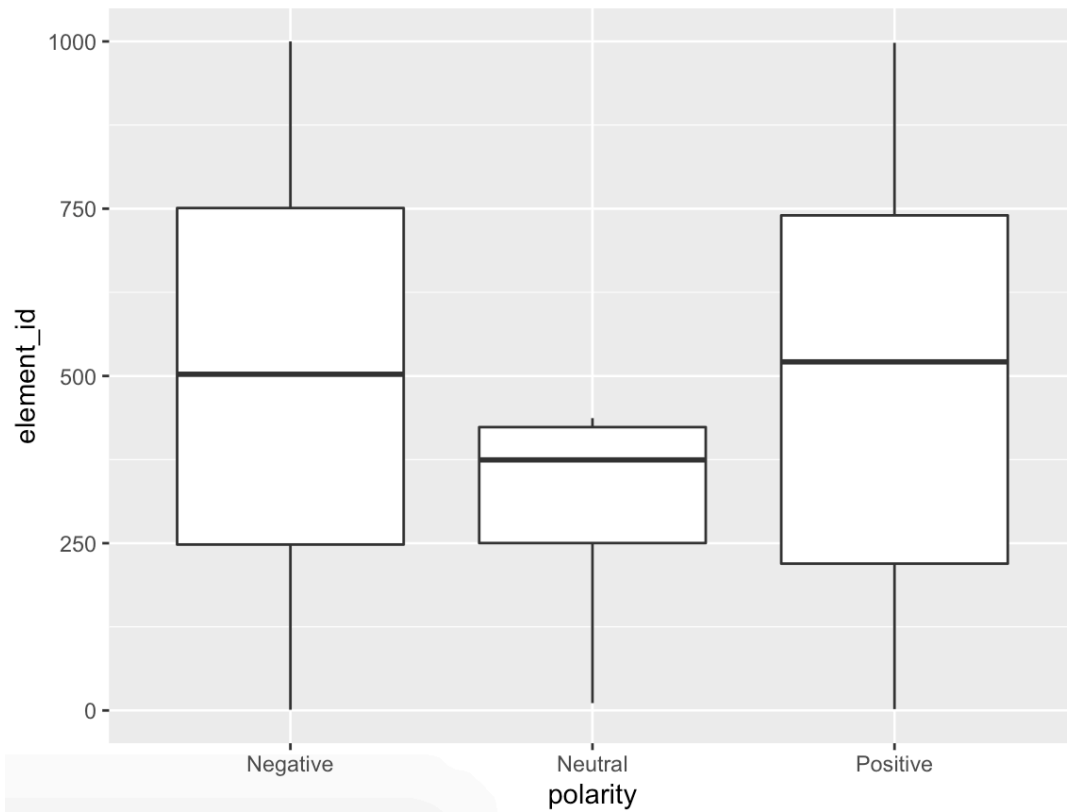
Outliers with sentiment values outside $[-1,1]$ are removed. This brings the number of sentences down to 2349.

Sentiment distribution across the tweets

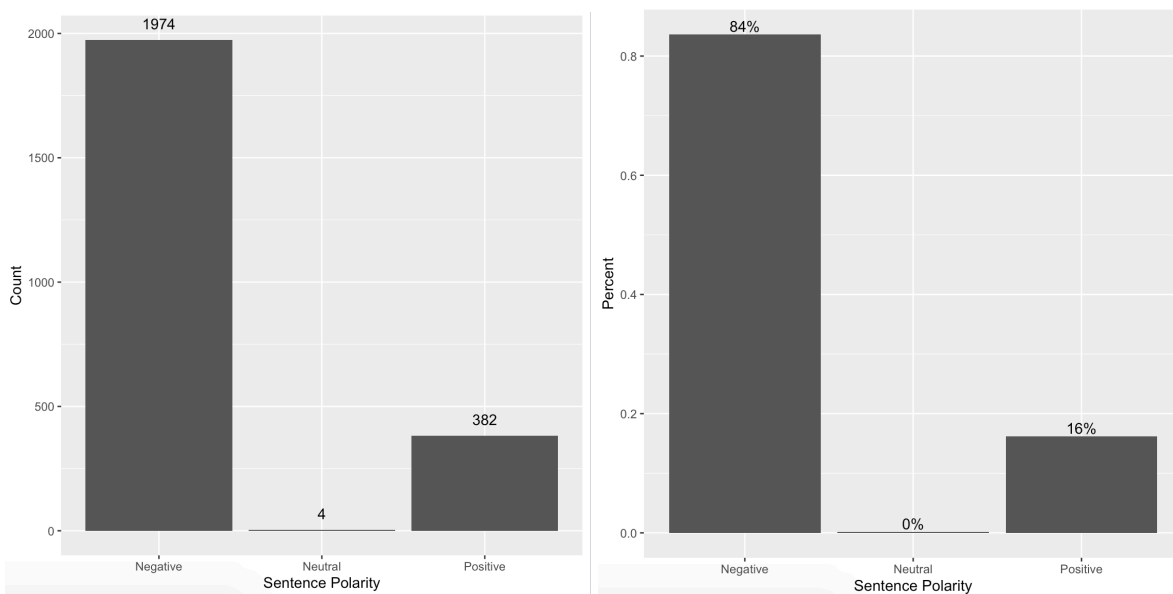


The positive sentiment is plotted green and negative red

Sentiment polarity over the tweets



Count of sentiment polarity



Summary and Conclusion

Each tweet is broken down into sentences and a sentiment score is calculated for each, thereafter, removing outliers.

Based on the 1000 tweets collected on the topic 'Fauci' corresponding to Dr. Fauci, 1974 (~84%) of the tweets have a negative sentiment associated with it.

References

sentimentr: <https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf>

rtweet: <https://cran.r-project.org/web/packages/rtweet/rtweet.pdf>

ggplot2: <https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf>