Sentimental Analysis on Climate Change Tweets

Code

Annie Liang

1/30/18

Hide

```
# load twitter library - the rtweet library is recommended now over
twitteR
library(rtweet)
# plotting and pipes - tidyverse!
library(ggplot2)
library(dplyr)
library(ggraph)
library(igraph)
```

```

Attaching package: 'igraph'

The following object is masked from 'package:tidyr':

    crossing

The following objects are masked from 'package:dplyr':

    as_data_frame, groups, union

The following objects are masked from 'package:stats':

    decompose, spectrum

The following object is masked from 'package:base':

    union
```

Hide

```
# text mining library
library(tidytext)
```

Obtain 10000 most recent tweet on #climatechange and store under climate_tweets

Hide

```
climate_tweets <- search_tweets(q = "#climatechange", n = 10000,
                                      lang = "en",
                                      include_rts = FALSE, token= tw
itter_token)
```

```
Searching for tweets...
Finished collecting tweets!
```

Let’s take a look at the tweets!

Hide

```
head(climate_tweets$text)
```

```
[1] "WINNING STRATEGY FOR #democrats2020 &amp; #Bernie2020 1) HONEST
TALK ABOUT NAFTA, Overseas Competition, Tariffs; Immigrant Labor; 2)
Address Security; &amp; #BlueCollar Job Fears; 3) Up Federal $ For I
nfrastructure, Flooding &amp; #ClimateChange, 4) Help BlueCollar Ada
pt To Service Economy. https://t.co/nfL7LqOieJ"
[2] "Coal is a health issue (via air pollution &amp; climate change)
, which is why we've signed this letter to the Victorian EPA. You ca
n too! #climatechange #coal https://t.co/LiCsCTm687"
[3] "\"Paris Agreement 2015/Art.2.1(a):Holding the increase in globa
l average temperature to well below 2°C.\" https://t.co/nb9jeqOkzU #
climatechange #climateaction\n#environment\n#energy https://t.co/zlp
HhDK6Ff"
[4] "#renewable #RenewableEnergy #climatechange #ClimateAction #wind
#coal High quality coal used with new technology is better for clima
te than wind power, see https://t.co/Dt5PTtY6V0"
[5] "the 7-day Philly forecast has three days in the 60s:  60, 66, 6
7\nand talk of the 70s for the middle of next week\n#ClimateChange i
s REAL"
[6] "What Happens When Arctic Glaciers &amp; Permafrost Diminish? Le
arn more here: https://t.co/m6e7W99FYE #climatechange #climateaction
\n#environment\n#energy https://t.co/TcwkyLCXS2"
```

Convert the dataset into tidyverse data_frame for better data manipulation.

Hide

```
library(dplyr)
climate_tweet_df<- data_frame(line= 1:9335, text=climate_tweets$text
)
head(climate_tweet_df)
```

line <int>
1
2
3
4

5
6

6 rows | 1-1 of 2 columns

There are url’s in the tweets. They don’t not provide useful information in telling what are the words that people use when talking about climate change. One way to remove is by using regular expression.

Hide

```
climate_tweet_df$text <- gsub("http.*","",  climate_tweets$text)
climate_tweet_df$text <- gsub("https.*","", climate_tweet_df$text)
head(climate_tweet_df$text)
```

```
[1] "WINNING STRATEGY FOR #democrats2020 &amp; #Bernie2020 1) HONEST
TALK ABOUT NAFTA, Overseas Competition, Tariffs; Immigrant Labor; 2)
Address Security; &amp; #BlueCollar Job Fears; 3) Up Federal $ For I
nfrastructure, Flooding &amp; #ClimateChange, 4) Help BlueCollar Ada
pt To Service Economy. "
[2] "Coal is a health issue (via air pollution &amp; climate change)
, which is why we’ve signed this letter to the Victorian EPA. You ca
n too! #climatechange #coal "
[3] "\"Paris Agreement 2015/Art.2.1(a):Holding the increase in globa
l average temperature to well below 2°C.\" "
[4] "#renewable #RenewableEnergy #climatechange #ClimateAction #wind
#coal High quality coal used with new technology is better for clima
te than wind power, see "
[5] "the 7-day Philly forecast has three days in the 60s:  60, 66, 6
7\nand talk of the 70s for the middle of next week\n#ClimateChange i
s REAL"
[6] "What Happens When Arctic Glaciers &amp; Permafrost Diminish? Le
arn more here: "
```

Tidytext allow me to do all of the following in one call-

1. Tokenization (restructure the dataset as one-token-per-row format), convert text to lowercase and remove punctuation using unest_tokens(). Another package that does tokenization is library(tm). Tm doesn’t offer much of the all-in-one-go function like tidytext.
2. Remove stop words

3. Count the frequency use of the most common words

Hide

```
library(wordcloud)
```

```
Loading required package: RColorBrewer
```

Hide

```
library(RColorBrewer)
climate_tweet_clean<- climate_tweet_df %>%
#first arug is the name of col where the unique word will be store f
ollows by the col name from the data_frame that I want to pull uniqu
e words from.
  unnest_tokens(word, text) %>%
  anti_join(stop_words)
```

```
Joining, by = "word"
```

Hide

```
head(climate_tweet_clean)
```

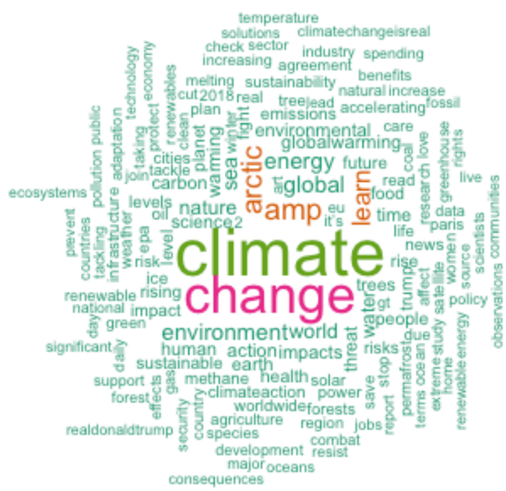
line <int>	word <chr>
1	winning
1	strategy
1	democrats2020
1	amp
1	bernie2020
1	1

6 rows

Wordcloud displaying the top frequent words on what people tweet about climate changes.

Hide

```
climate_tweet_clean %>%
  count(word, sort= TRUE) %>%
  with(wordcloud(word, n, max.words=150, random.order=FALSE,colors=b
rewer.pal(8,"Dark2"), rot.per=0.35))
```

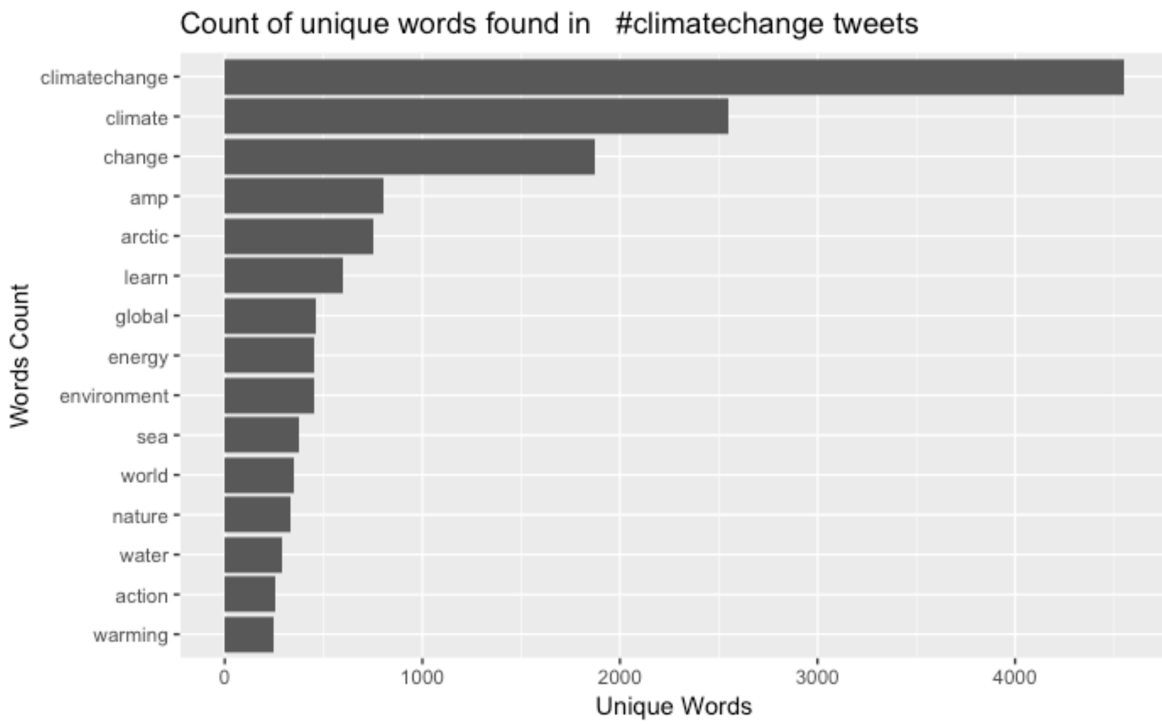


Let’s get a better look up at the top 15 words used.

Hide

```
climate_tweet_clean %>%
  count(word, sort= TRUE) %>%
  top_n(15) %>%
  mutate(word =reorder(word,n)) %>%
  ggplot(aes(x= word, y=n)) +
  geom_col() + coord_flip() + labs(y="Unique Words", x="Words Count"
, title= "Count of unique words found in   #climatechange tweets")
```

```
Selecting by n
```



Sentiment Analysis

Tidytext package includes 3 sentiment lexicon in the sentiments dataset. Here, I use `bing lexicon` to find a sentiment score for each word. The `bing lexicon` (By Bing Lu) builds the lexicon based on positive verse negative classification. `Nrc` lexicon classifies words into a wider range of ten sentiments. The last one is `afinn lexicon` (By Finn Arup Nelson) that assigns numerical score (from -5 to 5) to imply the negative or positive impact of certain words.

Here is glimpse of what bing lexicon classifies as negative and positive words.

Hide

```
get_sentiments("bing")
```

word	sentiment
<chr>	<chr>

2-faced	negative
2-faces	negative
a+	positive
abnormal	negative
abolish	negative
abominable	negative
abominably	negative
abominate	negative
abomination	negative
abort	negative

1-10 of 6,788 rows Previous **1** 2 3 4 5 6 ... 100 Next

Most common positive and negative words

Hide

```
climate_tweet_bing <- climate_tweet_clean %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()
```

```
Joining, by = "word"
```

Hide

```
head(climate_tweet_bing)
```

word <chr>	sentiment <chr>	n <int>
threat	negative	218
sustainable	positive	177

trump	positive	175
risks	negative	169
risk	negative	121
sustainability	positive	111

6 rows

There are more negative words than positive words on climate change tweets which may not be a strong surprise as the result of recent coal policy.

Hide

```
library(tidyr)
climate_tweet_bing_tot <-climate_tweet_bing %>% spread(sentiment,n)
climate_tweet_bing_tot
```

	word <chr>	negative <int>	positive <int>
1	abomination	2	NA
2	absence	1	NA
3	absurd	3	NA
4	abundance	NA	3
5	abundant	NA	2
6	abuse	2	NA
7	abused	18	NA
8	abuses	2	NA
9	abusive	1	NA
10	abyss	4	NA

1-10 of 1,366 rows Previous 1 2 3 4 5 6 ... 100 Next

Hide

```
summary(climate_tweet_bing_tot)
```

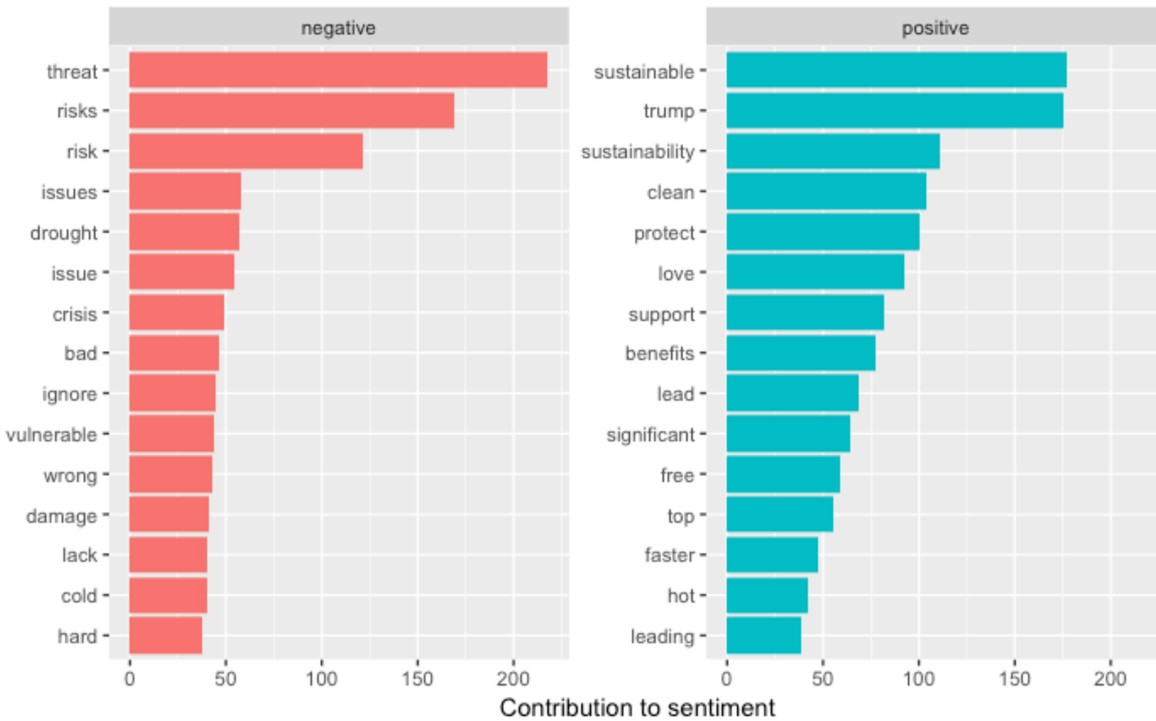
```
     word              negative          positive      
 Length:1366        Min.   :  1.000   Min.   :  1.000  
 Class :character   1st Qu.:  1.000   1st Qu.:  1.000  
 Mode  :character   Median :  2.000   Median :  2.000  
                    Mean   :  5.185   Mean   :  6.786  
                    3rd Qu.:  4.000   3rd Qu.:  5.000  
                    Max.   :218.000   Max.   :177.000  
                    NA's   :518       NA's   :848      
```

Bar graph is a great way to see which negative and positive related words are most commonly used.

Hide

```
climate_tweet_bing %>%
  group_by(sentiment) %>%
  top_n(15) %>%
  ungroup() %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n, fill = sentiment)) + geom_col(show.legend = FA
LSE) +   facet_wrap(~sentiment, scales = "free_y") + labs(y = "Contr
ibution to sentiment", x = NULL) + coord_flip()
```

```
Selecting by n
```



Explore words that occur together in tweets.

Hide

```
climate_tweet_paired_words <- climate_tweet_df %>%
  unnest_tokens(paired_words,text, token = "ngrams", n = 2)
head(climate_tweet_paired_words)
```

line	paired_words
<int>	<chr>
1	winning strategy
1	strategy for
1	for democrats2020
1	democrats2020 amp
1	amp bernie2020
1	bernie2020 1

6 rows

Hide

```
climate_tweet_sep_words <- climate_tweet_paired_words %>%
  separate(paired_words, c("word1", "word2"), sep = " ")
climate_tweets_clean2 <- climate_tweet_sep_words  %>%
  filter(!word1 %in% stop_words$word) %>%
  filter(!word2 %in% stop_words$word)
head(climate_tweets_clean2)
```

line <int>	word1 <chr>	word2 <chr>
1	winning	strategy
1	democrats2020	amp
1	amp	bernie2020
1	bernie2020	1
1	1	honest
1	honest	talk

6 rows

Hide

```
#bigrams count
climate_words_count <- climate_tweets_clean2 %>%
  count(word1, word2, sort = TRUE)
head(climate_words_count)
```

word1 <chr>	word2 <chr>	n <int>
climate	change	1737
sea	level	169
sea	levels	138
level	rise	135

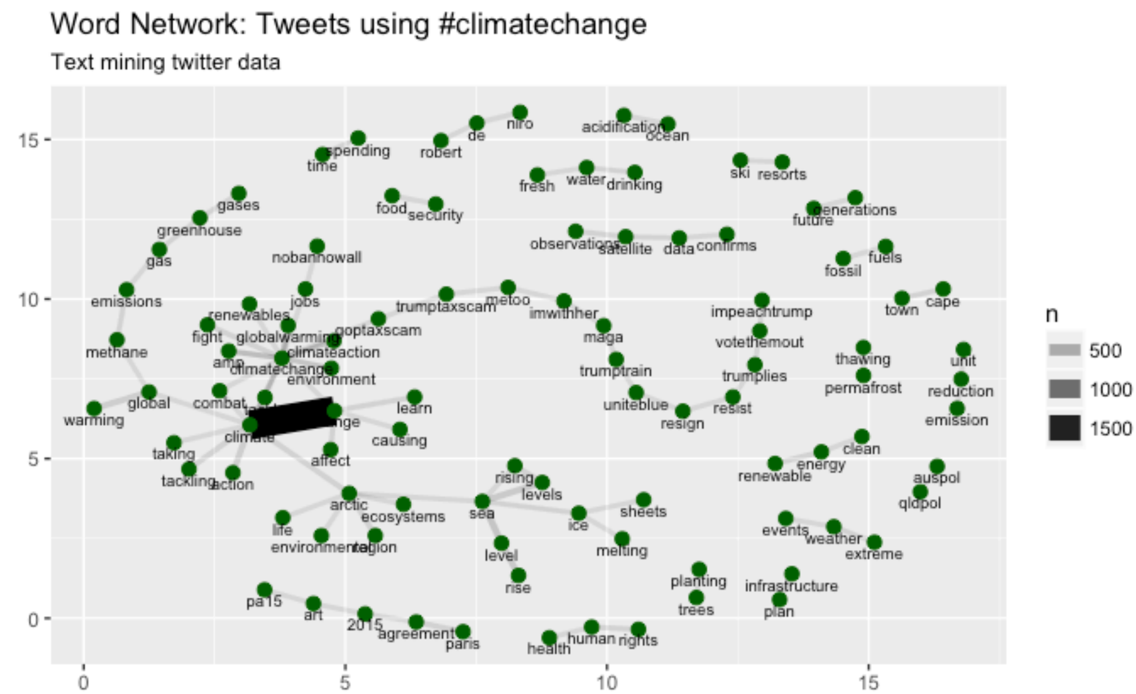
climate	action	116
global	warming	116

6 rows

Word network is useful when wanting to get an idea of how words are associated with each other for a certain topics. Here, it's interesting to see food and security as a pair when people tweet about climate change. If I never had done this analysis, I'd probably associate food and shortage or security and home together. Word network can potentially reveals information about up-to-date topics that would allow companies to make decisions.

Hide

```
set.seed(12431)
climate_words_count %>%
        filter(n >= 26) %>%
        graph_from_data_frame() %>%
        ggraph(layout = "fr") +
        geom_edge_link(aes(edge_alpha = n, edge_width = n)) +
        geom_node_point(color = "darkgreen", size = 2.6) +
        geom_node_text(aes(label = name), vjust = 1.4, size = 2.5) +
        labs(title = "Word Network: Tweets using #climatechange",
             subtitle = "Text mining twitter data ",
             x = "", y = "")
```



Next steps: Integrate the dataset into a shinny app that allow user to enter text in a text box and show the next word prediction on a table or word cloud.

References David Robinson and Julia Silge, 2017.Text Mining with R: A Tidy Approach Obtaining and using access tokens on rtweet, https://cran.r-project.org/web/packages/rtweet/vignettes/auth.html (https://cran.r-project.org/web/packages/rtweet/vignettes/auth.html)