Final Project Report

Weige Guo

12/18/2017

Introduction

Disputes between Capitalsim and Communism, two of the most adopted economic systems, have long existed. Given limited time these systems have been adopted as well as different measurements, it is hard to say which is a better system that could benefit a country to the most. However, exploring social data and finding out people's perception about the two systems in a society that allows freedom of speech can shed somne lights on the topic for future studies.

This project explored the online perpection, twitter exclusively, of these two systems and generated some interesting insights. The project contains four parts: Data Extraction, Text Analysis, Mapping, Emoji Analysis and Shiny Application.

Data Extraction

Data were extracted two times in this project with two methods: twitterR and Stream. 5000 tweets containing "capitalsim" and "communism" respectively from December 6th, 2017 to December 13th, 2017 were extracted by twitterR to be used in text analysis, emoji analysis and Shiny application. Tweets sent from the United States that contain "communism" are extractd by Stream for 30 seconds to be used in mapping.The reason that two methods are employed lies in that data extracted by twitterR do not include location information that would be used in mapping. As a result, one should notice that data used in mapping are different than the ones used in other analysis.

```
capdata <- searchTwitter('capitalism', since = '2017-12-6', until = '2017-12-13'
                         , lang = "en", n = 5000)
comdata <- searchTwitter('communism', since = '2017-12-6', until = '2017-12-13'
                         , lang = "en", n = 5000)
capdata <- twListToDF(capdata)
comdata <- twListToDF(comdata)
```

```
library(streamR)
filterStream("tweets_cap.json", track="capitalism",
             locations = c(-125, 25, -66,50),  timeout = 30,
             oauth = my_oauth)
```

```
filterStream("tweets_com.json", track="communism",
             locations = c(-125, 25, -66,50),  timeout = 30,
             oauth = my_oauth)
```

Text Analysis

In this section, sentimental analysis and wordcloud are used to underline people's perception of the two economic systems.

Step 1

First, I split the text into seperate words, eliminated stop words and irrelevant words such as “rt”, “tweet”,“http”, counted the occurance of words and plot the ones that occur more than 10 times in all 800 tweets. Than I repeated for communism data.

For capitalsim, “anti”, “homelessness”, “harassment” and “mordern” are among the top ones and for communism, “people”, “fighting”, “revolutionary” are frequently repeated.

```
tidycap <- capdata %>%
  dplyr::select(text) %>%
  unnest_tokens(word,text)
data("stop_words")
tidycap <-  tidycap %>%
  anti_join(stop_words)

tidycap[!grepl("rt", "http", "tweet", tidycap$word),]


tidycap %>%
  count(word, sort = TRUE) %>%
  filter(n>120) %>%
  mutate(word = reorder(word, n)) %>%
  ggplot(aes(word, n)) +
  geom_col(fill = "skyblue") +
  xlab(NULL) +
  coord_flip()
```

Step2

Next,I compared the words that are frequently used in two sets of data(capitalsim and communism). The more close the word is to the top right, the more frequently it is used in both datasets. “socialism”, “failure”, “free” are some of the words that are relatively equally used in both data sets. (These words are close to the line in the graph)

From this graph we get to see what are the common words that people mention when talking about “capitalsim” and “communism”, and what are the words that are mentioned more specifically when talking about either “capitalism” or “communism”.

```
frequency <- bind_rows(mutate(tidycap, ideology = "capitalism"),
                       mutate(tidycom, ideology = "communism")) %>%
  mutate(word = str_extract(word, "[a-z']+")) %>%
  count(ideology, word) %>%
  group_by(ideology) %>%
  mutate(proportion = n / sum(n)) %>%
  dplyr::select(-n) %>%
  spread(ideology, proportion) %>%
  gather(ideology, proportion, `communism`)
frequency


library(scales)

ggplot(frequency, aes(x = proportion, y = `capitalism`,
                      color = abs(`capitalism` - proportion))) +
```

```
  geom_abline(color = "purple", lty = 2) +
  geom_jitter(alpha = 0.1, size = 2.5, width = 0.3, height = 0.3) +
  geom_text(aes(label = word), check_overlap = TRUE, vjust = 1.5) +
  scale_x_log10(labels = percent_format()) +
  scale_y_log10(labels = percent_format()) +
  scale_color_gradient(limits = c(0, 0.001), low = "purple", high = "darkslategray4") +
  theme(legend.position="none")
```

Step3

Here I would like to see the differece of people's sentiments toward two economic systyems. I split the text into lines, set line number as row number, counted sentiment scores for each row, and compared the results of two data sets with plots.

Although twitter is more frequently used in capitalist countries such as the United States than communist countries such as China, the two sentimental graphs show that people are more positive about communism than about capitalism because communism has more positive values than capitalsim. This may due to the fact that people are more familiar with the economic system they are living in, and they are more easily to notice and get unsatisfied by the drawbacks of the system their country is adopting. Moreover, people tend to beautify the things that are mysterious to them. Hence the perception of communism is more positive than capitalsim. It can be boldly inferred that the same analysis of communist countries' social media platforms can draw an opposite result.

```
tidycap2 <- capdata %>%
  dplyr::select(text) %>%
  mutate(linenumber = row_number()) %>%
  ungroup() %>%
  unnest_tokens(word, text)

tidycom2 <- comdata %>%
  dplyr::select(text) %>%
  mutate(linenumber = row_number()) %>%
  ungroup() %>%
  unnest_tokens(word, text)


tidycap2 <- tidycap2%>%
  mutate(ideology = "capitalsim")
tidycom2 <- tidycom2%>%
  mutate(ideology = "communism")

com2cap2 <- rbind(tidycap2, tidycom2)

nrcjoy <- get_sentiments("nrc") %>%
  filter(sentiment == "joy")

nrcjoy

joycap <- com2cap2 %>%
  inner_join(nrcjoy) %>%
  count(word, sort = TRUE)

com2cap2 %>%
```

```
  filter(ideology == "capitalism") %>%
  inner_join(nrcjoy) %>%
  count(word, sort = TRUE)

com2cap2 <- com2cap2 %>%
  inner_join(get_sentiments("bing")) %>%
  count(ideology, index = linenumber %/% 30, sentiment) %>%
  spread(sentiment, n, fill = 0) %>%
  mutate(sentiment = positive - negative)


ggplot(com2cap2, aes(index, sentiment, fill = ideology)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ideology, ncol = 2, scales = "free_x")
```

Step4

Next, I selected sentimental words in 5000 tweets mentioning "capitalism" and "communism" respectively, plotted the top 10 most used positive words and negative words. The results can be seen in the Presentation.

The graphs tell us what specific positive and negative sentiments people have towards the two economic systems.

```
tidycap2 <- capdata %>%
  dplyr::select(text) %>%
  mutate(linenumber = row_number()) %>%
  ungroup() %>%
  unnest_tokens(word, text)

tidycom2 <- comdata %>%
  dplyr::select(text) %>%
  mutate(linenumber = row_number()) %>%
  ungroup() %>%
  unnest_tokens(word, text)


tidycap2 <- tidycap2%>%
  mutate(ideology = "capitalsim")
tidycom2 <- tidycom2%>%
  mutate(ideology = "communism")

com2cap2 <- rbind(tidycap2, tidycom2)

nrcjoy <- get_sentiments("nrc") %>%
  filter(sentiment == "joy")

nrcjoy

joycap <- com2cap2 %>%
  inner_join(nrcjoy) %>%
  count(word, sort = TRUE)

com2cap2 %>%
```

```
  filter(ideology == "capitalism") %>%
  inner_join(nrcjoy) %>%
  count(word, sort = TRUE)

com2cap2 <- com2cap2 %>%
  inner_join(get_sentiments("bing")) %>%
  count(ideology, index = linenumber %/% 30, sentiment) %>%
  spread(sentiment, n, fill = 0) %>%
  mutate(sentiment = positive - negative)


ggplot(com2cap2, aes(index, sentiment, fill = ideology)) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ideology, ncol = 2, scales = "free_x")
```

Step5

Last I used wordcloud to demonstrate the mosted used sentimental words in 800 tweets mentioning“capitalsim” and “communism”. I distinguished the negative ones from the positive ones by using different colors.

```
library(reshape2)
library(wordcloud)


  tidycap %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("darkgrey","red"),
                   max.words = 100)
```

Mapping

From this section I would like to find out people in which parts of the United States are tweeting about contents related to “capitalism and”communism“. I aqcuired tweets sent from the United States that contain”communism" for 20 seconds, plot the map of the United States, and locate these tweets in the map.

According to the result, the majority of the tweets come from either east coast or west coast. Some scattered in the middle. The result is predictable because east coast and west coast are where people with higher educational backgound gather.

```
library(streamR)
filterStream("tweets_cap.json", track="capitalism",
             locations = c(-125, 25, -66,50),  timeout = 30,
             oauth = my_oauth)

tweets.cap <- parseTweets("tweets_cap.json", verbose = TRUE)

library(ggplot2)
library(grid)

map.data <- map_data("state")
```

```
cappoints <- data.frame(x = as.numeric(tweets.cap$lon),
                        y = as.numeric(tweets.cap$lat))

cappoints <- cappoints[cappoints$y > 25, ]
ggplot(map.data) +
  geom_map(aes(map_id = region),
           map = map.data,
           fill = "white",
           color = "grey20", size = 0.25) +
  expand_limits(x = map.data$long, y = map.data$lat) +
  theme(axis.line = element_blank(),
        axis.text = element_blank(),
        axis.ticks = element_blank(),
        axis.title = element_blank(),
        panel.background = element_blank(),
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        plot.background = element_blank(),
        plot.margin = unit(0 * c( -1.5, -1.5, -1.5, -1.5), "lines")) +
        geom_point(data = cappoints,
        aes(x = x, y = y), size = 3,
        alpha = 1/2, color = "skyblue")
```

Emoji

From this section I would like to explore what are the most used emojis when people are talking about capitalsim and communism. Results can be seen in the presentation.

Shiny Application

Results can be seen in the presentation.