Sentiment Analysis of Trump's Tweets on China

Course Code: MY459
Candidate Number: 36388
Word Count: 2943

June 1, 2020

Abstract

This essay analyzes the features of Trump's China-related tweets sentiments. The China-related tweets are from Jan 27, 2011 to May 30, 2020. I use dictionary methods to obtain sentiment scores for each tweet. The average sentiment score of Trump's tweets on China is slightly positive, which is much lower than that of the other three Asian countries: North Korea, South Korea and Japan. The time series of average sentiment scores for China display a pattern correlated with Trump's election campaign. Besides, the scores have continued dropping since early 2020. In regard to topics of negative China-related tweets, one of the five topics is the U.S. trade with China. The proportion of this topic over time is also affected by Trump's campaign.

1 Introduction

Twitter is a political tool that helped get Donald Trump elected. Since his inauguration, he has fully integrated Twitter into the fabric of his administration, reshaping the nature of the presidency and presidential power(Shear et al. (2019)). China is one of the countries that Trump's tweets frequently mention. It's worth exploring the features of Trump's China-related tweets sentiments. What are Trump's tweets sentiments for China like? Is the sentiment score high compared to other Asian countries? Are there any patterns for the scores in time series? What are the topics for a specific sentiment tweets? To conduct a sentiment analysis of Trump's tweets on China, I use dictionary methods to obtain average sentiment scores from all China-related tweets until May 30, 2020. The average sentiment score of Trump's tweets on China is slightly positive, which is much lower than that of the other three Asian countries: North Korea, South Korea and Japan. The time series of average sentiment scores for China display a pattern correlated with Trump's election campaign. Besides, the scores have continued dropping since early 2020. To further explore the content, I use LDA to model the probability distribution of topics for the negative tweets. One of the five topics is the U.S. trade with China. The proportion of this topic over time is also affected by Trump's campaign.

2 Motivation

When Mr. Trump, the current president of the United States, entered office, Twitter was a political tool that had helped get him elected and a digital howitzer that he relished firing. In the years since, he has fully integrated Twitter into the very fabric of his administration, reshaping the nature of the presidency and presidential power. From 2009 to 2020, he has posted more than 48,000 tweets on his @realDonaldTrump Twitter account, which more than 66 million accounts follow. According to David Robinson, some words of his tweets tend to lead to unusually many retweets, or unusually few. The most retweeted words include some emotional words e.g. "fake"(Robinson (2017)).

There's no denying that when it comes to international politics, China is one of the countries that he likes mentioning. Compared with other Asian countries like Japan, North and South Korea, from May 4, 2009 to May 30, 2020, the number of Trump's tweets about China is more than twice the sum of the others(Table 1).

Table 1: the number of Trump's tweets in regard to certain Asian countries

Country	the Number of Tweets
China	969
North Korea	199
Japan	112
South Korea	50

What's more, there are several interesting features regarding the trend of his tweets mentioning China over time. Firstly, between July 2011 and November 2012 (Obama's reelection), a full 7 percent of Trump's tweets mentioned China, which is a high frequency compared to the average frequency he mentions China. Secondly, after he was inaugurated in 2017, there are several months each year when over 5 percent of his tweets mention China, as shown in the Figure 1. Last but not least, as the Figure 2 shows, after 2016, the percentage of Trump's tweets that mention China keeps increasing.

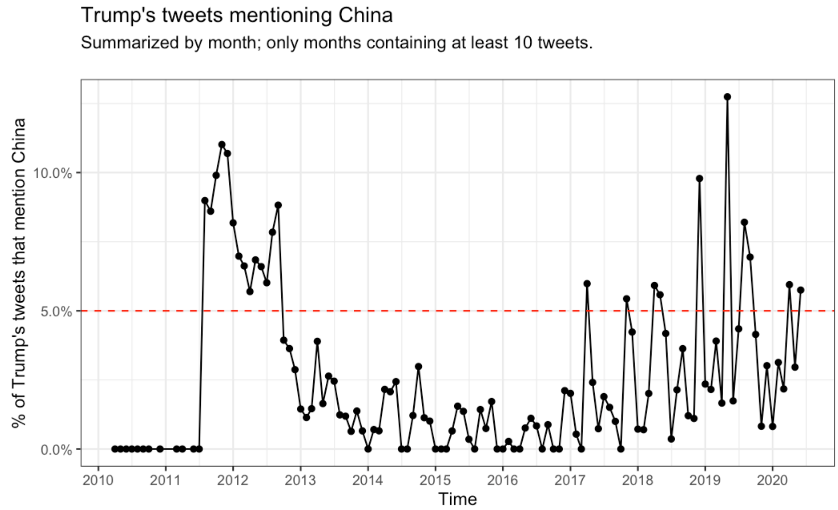

Figure 1: After Trump was inaugurated in Jan 2017, there are several months each year when over 5 percent of his tweets mention China. The tweets are summarized by month and only months containing at least 10 tweets are considered.

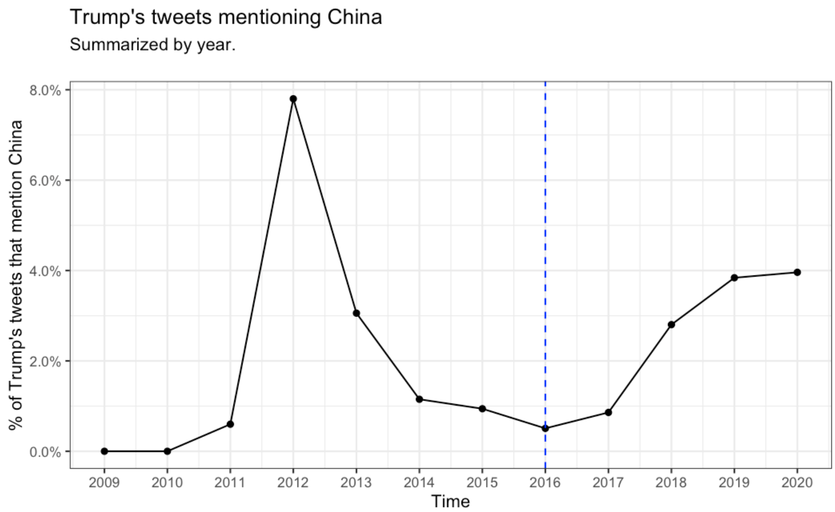

Figure 2: After 2016, the percentage of Trump's tweets that mention China keeps increasing. The blue dashed line indicates 2016. The tweets are summarized by year.

Another fact is that Trump has different tones of China over time. In the summer of 2019, he announced increased tariffs on 300 billion dollars worth of Chinese goods, using a tweet to deepen tensions between the two countries. For example, he claimed that ”China is doing very badly, worst year in 27 - was supposed to start buying our agricultural product now - no signs that they are doing so. That is the problem with China, they just don't come through. Our Economy has become MUCH larger than the Chinese Economy is last 3 years...." on July 30, 2019. During the Covid-19 outbreak, his tweets described the virus as "China Virus". However, he sometimes praises China. For instance, on Jan 25, 2020, he said "China has been working very hard to contain the Coronavirus. The United States greatly appreciates their efforts and transparency. It will all work out well. In particular, on behalf of the American People, I want to thank President Xi!". Occasionally, he shows a confusing attitude towards China as shown below.

Figure 3: Donald Trump's latest tweet about China.

In 2016, David Robinson analyzed Trump's Android and iPhone tweets. He confirms the hypothesis that Trump writes only the angrier Android half(Robinson (2016)). In 2019, Nick Cochrane uses the topic modeling with LDA and VADER sentiment analysis to parse through the noise of president's Twitter feed and isolate his sentiment toward trade with China(Cochrane (2019)). However, there is no complete sentiment analysis of Trump's tweets on China.

3 Data

The dataset used in this project is from `The Trump Archive` by Brendon Brown, which contains all tweets from the @realDonaldTrump Twitter account from 2009 through 2020. The information of the dataset includes source, log id, texts, created time, the count of retweets and the count of favorite. I use the data from 2009-05-04 to 2020-05-30.

I define China-related tweets as the texts containing any word in the set "China", "Chinese", "President Xi", "Beijing", which ignores case. The extracted China-related tweets are between Jan 27, 2011 and May 30, 2020. Prior to analysis, I have the following preprocessing steps. Tweets are tokenized and the documents are bags of words. Words are made lowercase. When constructing a sparse document-feature matrix, I remove punctuation, numbers, English stopwords and urls. Besides, words and symbols which are not of interest like 'rt', 'amp', etc are removed. The constructed document-feature matrix is of 969 documents and 3480 features.

The figure below is the wordcloud of the most frequent 50 features for the document-feature matrix. China is, of course, most frequently mentioned. Other frequent words include trade, currency, jobs, tariffs, President Xi, Korea, etc.

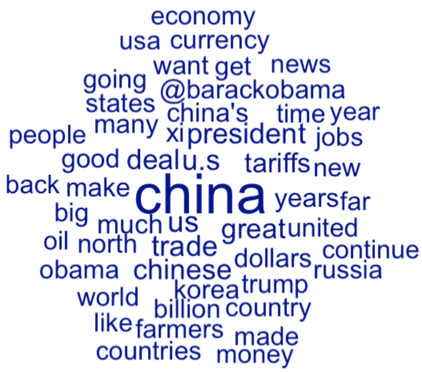

Figure 4: the wordcloud of the most frequent 50 features for the China tweets document-feature matrix.

I also compared Trump's sentiments for China with other Asian countries like North Korea, South Korea and Japan. These countries' tweets are defined as tweets containing any country-specific words. For example, the tweets on Japan should include the words "Japan" or "Japanese", regardless of case.

4 Methods

To obtain sentiment scores for each tweet, I use dictionary methods. Dictionary methods are reliable because there is no human decision making as part of the text analysis procedure. Besides, it bridges qualitative and quantitative text analysis. Dictionary construction involves a lot of contextual interpretation and qualitative judgment, which means, in my case, the sentiment categories of words are identified. I will use the positive and negative categories in the augmented General Inquirer dictionary to measure the extent to which Trump adopted a positive or negative tone. What's more, as González-Bailón and Paltoglou (2015) show, for Twitter data, the dictionary accuracy in document classification could be close to or even higher than some machine learning approach. There are several papers using dictionary methods for social media data classification. For example, Kramer, Guillory, and Hancock (2014) use dictionary methods to identify positive and negative Facebook texts before testing emotional contagion hypothesis.

Because the sentiments of each tweet are not known, supervised machine learning methods are not suitable. Though unsupervised methods are possibly appropriate to classify the documents, the fundamental problem is that other dimensions like language, rhetoric style are likely to be more predictive. The reason is that the first dimension in unsupervised scaling will capture main source of variation, whatever that is. Besides, the validation comes after estimating. For discriminant construct validity, the estimated positions should match other existing measures where they should match, and depart where they should depart. Unfortunately, in my dataset, there are no other existing measures of sentiment. It would be questionable to validate that a construct of interest is measured.

After finding sentiment scores for each tweet, I uses the Latent Dirichlet Allocation (LDA) topic model to assign negative-sentiment tweets topics. The model is for discovering the main topics in an unstructured corpus. It's appropriate for my analysis because it requires no prior information, training set, or human annotation. Besides, LDA is a probabilistic model, which means that each document is considered to be about a mixture of topics. Thus, the average probability that each tweet is about a particular topic over time could be computed. In 2019, Nick Cochrane uses LDA to assign each tweet a topic when explore whether a SP 500 trading strategy can take advantage of President Trump's trade-related tweets and predict short-term moves in the market(Cochrane (2019).

5 Results

For sentiment sentiment detection using dictionary methods, I use the positive and negative categories in the augmented General Inquirer dictionary to measure the extent to which Trump adopted a positive or negative tone. After using the dictionary when constructing the document-feature matrix, I obtain sentiment scores for each China-related tweet. The average sentiment score of Trump's tweet on China is 0.32, a score just over zero. On an average level, Trump tweets have slightly positive sentiment for China. The tweets with positive sentiment scores are classified as positive tweets and vice versa. The tweets with zero score are neutral. The number of positive, neutral, and negative tweets is as below. The number of negative tweets is close to that of the positive. The most negative tweet is "We are winning, big time, against China. Companies amp; jobs are fleeing. Prices to us have not gone up, and in some cases, have come down. China is not our problem, though Hong Kong is not helping. Our problem is with the Fed. Raised too much amp; too fast. Now too slow to cut....". The most positive tweet is "Just had a long and very good call with President Xi of China. Deal is moving along very well. If made, it will be very comprehensive, covering all subjects, areas and points of dispute. Big progress being made".

Table 2: the number of positive, neutral, and negative tweets. The number of negative tweets is close to that of the positive.

Negative	**Neutral**	**Positive**
304	283	382

Then, I compare the average sentiment score of Trump's tweet on China with other Asian countries. After normalizing words by text length, I obtain the sentiment scores for the four countries: China, North Korea, South Korea and Japan(Table 3). What's impressive is that the sentiment score for China is much lower than that of the others. China's score is only 0.89, while the sentiment score for North Korea is about four times that of China and the other two countries obtain scores over 5.

Table 3: the average sentiment scores for the four countries: China, North Korea, South Korea and Japan.

Country	Sentiment Score
China	0.89
North Korea	3.16
South Korea	5.48
Japan	5.95

Afterwards, I look at how average sentiment scores each month for China change over time. The red dashed line in Figure 5 represents a 'neutral' sentiment average.

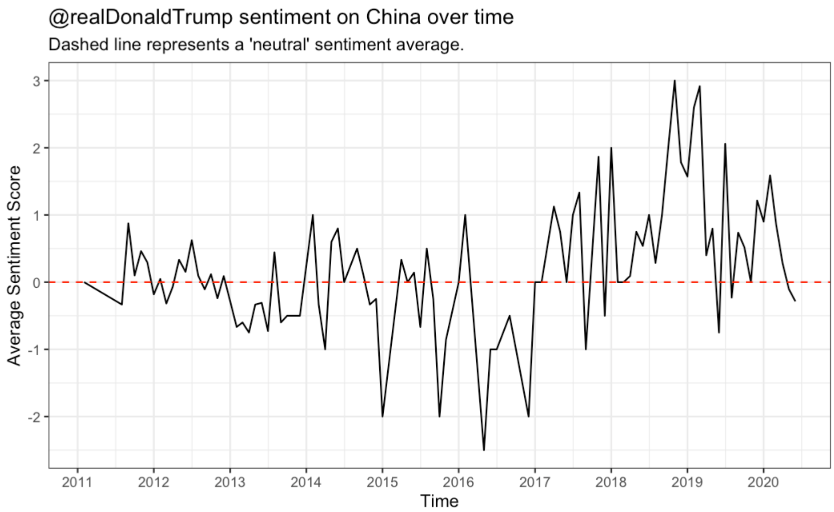

Figure 5: how average sentiment scores each month for China change over time. The red dashed line represents a 'neutral' sentiment average.

The red dashed line in Figure 6 is when Trump launched his campaign. The blue line is inauguration day and the grey is 2020-02. It can be shown that during his campaign for presidency, most of Trump's tweets display negative sentiment for China. However, after inauguration day on Jan 20, 2017, his sentiment turns towards positive. In most months after 2017, the average sentiment scores are positive, though there are some fluctuations towards negative. It should be noted that since Feb, 2020, the average sentiment scores continue decreasing. In recent two months, the sentiment scores have turned negative.

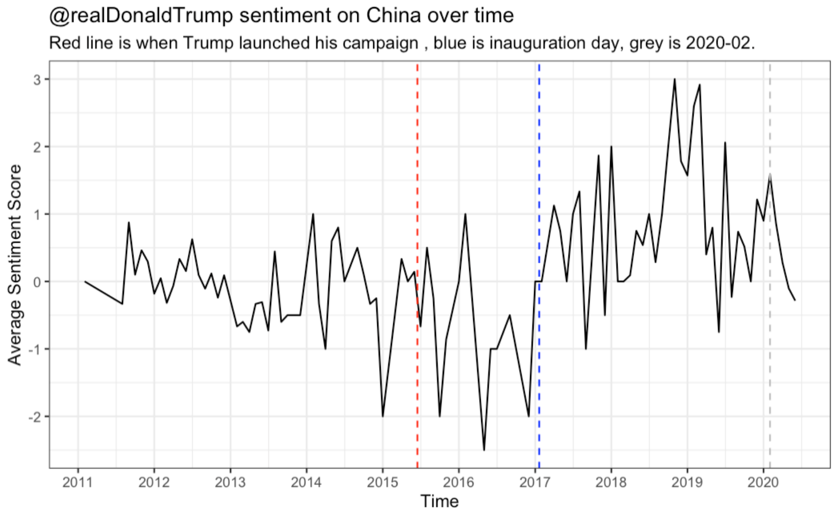

Figure 6: Red line is when Trump launched his campaign , blue is inauguration day, grey is 2020-02.

To further explore the contents of negative sentiment tweets, I run a LDA model to assign topics to them. When constructing the document-feature matrix, I also remove features that often only appear in one or two documents, in addition to remove stop words, numbers, urls and punctuation. After experimenting with different numbers of topics, I choose the number as 5. Topic 2 captures terms relevant to trade with China. The words most associated with the topic 2 are "trade, chinese, us, president, china's, billion, jobs, war, world, great, people, leaders, lost, going, business". I plot the estimated proportion of topic 2, trade with China, over time(Figure 7). I find a spike in the lead-up during Trump's campaign.

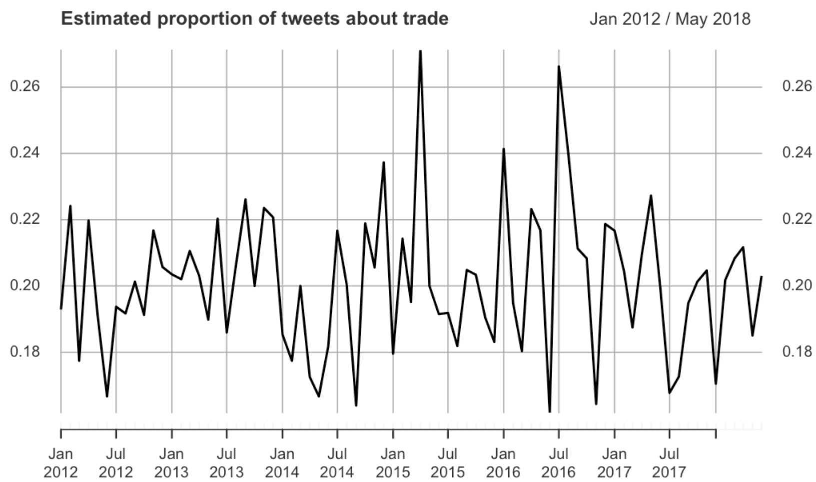

Figure 7: estimated proportion of topic 2, trade with China, over time.

6 Conclusions

In conclusion, the sentiment of Trump's tweets on China is slightly positive and fluctuates in different periods. The sentiment score for China is much lower than that of the other three Asian countries: North Korea, South Korea and Japan. The time series of average sentiment scores for China display a pattern correlated with Trump's campaign. Besides, the scores have continued dropping since early 2020. What's more, one of topics for negative China-related tweets is the U.S. trade with China. The proportion of this topic over time is also affected by Trump's campaign. To increase the accuracy of sentiment detection, supervised machine learning methods, which require manual coding of a subset of the tweets, are worth pursuing.

References

Cochrane, N. (2019). Trump, tweets, and trade. *URL:https://towardsdatascience.com/trump-tweets-and-trade-96ac157ef082(visited on May 30, 2020)*.

González-Bailón, S., & Paltoglou, G. (2015). Signals of public opinion in online communication: A comparison of methods and data sources. *The ANNALS of the American Academy of Political and Social Science*, *659*(1), 95–107.

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790.

Robinson, D. (2016). Text analysis of trump's tweets confirms he writes only the (angrier) android half. *Variance explained*, *9*, 2016.

Robinson, D. (2017). Trump's android and iphone tweets, one year later. *URL: http://varianceexplained. org/r/trump-followup/(visited on April 18, 2018)*.

Shear, M. D., Haberman, M., Confessore, N., Yourish, K., Buchanan, L., & Collins, K. (2019). How trump reshaped the presidency in over 11,000 tweets. *The New York Times*.