

main

April 25, 2018

1 'China and US Trader War' Twitter Analysis

1.1 Website Link: <http://keran.pythonanywhere.com>

1.2 Model Link: <http://keran.pythonanywhere.com/input>

1.3 1. Project Purpose:

Donald Trump and its trade war has incresing effect all over the word. President Donald Trump threatened to impose 35% to 45% tariffs on Chinese imports to force China into renegotiating its trade balance with the U.S. The immediate result of that would be a fierce trade war. The purpose of our project is to analyze the 'trader war' comments from twitter, including web scraping, sentiment analysis, LDA topic model and business insights based on data. ##### Twitter is a popular app with many politic related topics and users. For example, Donald Trump personally is a user with many follower on twitter.

1.4 2. Data Scraping, Preprocessing and Models:

1. Twitter Data Scraping: We used twitter API to extract twitter streaming sample including words 'tariff', 'trade war' and '#TradeWar' and save it as a .txt file. Please apply a twitter API id through the twitter website.

Please find the code in the file: lib/get_twitter.ipynb

2. Data Preprocessing: We extracted twitter data into json and transform it to a dataframe for further machine learning model.

Please find the code in the file: lib/cleaning.ipynb

The outcome data in the data/clean_data.csv

3. Sentiment Anlysis Model (Machine Learning Model): We used vader sentiment analysis to do analysis on our data set, which is a pre-trained deep learning model. And we combine this with the 'opinion leaders', the top 20 retweeted users among this topic, to extract their sentiment score.

Please find the code in the file: lib/get_twitter.ipynb

The outcome data in the data/data_with_vadersentiment.csv

4. LDA Topic Model (Machine Learningmodel): We applied LDA Model to extract twitter topics under negative attitude data and positive attitude data.

Please find the code in the file: lib/lda.ipynb

5. LSTM(Long Short Term Memory) (Recurrent Neural Network): We transform the text data into word vectors based on the Word2vec technique using the google news data, then applied LSTM Model to extract twitter topics under negative attitude data and positive attitude data. Since the 'support' and 'not support' are too general, we decide to show Vader Model result on our website.

2 3. Visulization and Web Developing:

D3 Visulization: We used d3 to show the twittee and retwitte relationship among all the users. Please find the detail code in the file: lib/Flask/statis/js/jquert-1.12.1.js

R Visualization: We mainly used ggplot and plotly package to show all the results. Please find the detail code in the file: lib/Visulization.R

HTML Frontend Development: We have two html files for our website:
index.html: our website homepage
input.html: our sentiment analysis model web page
Please find the detail code in the file: lib/Flask/templates

Flask Backend Web Development: We used flask to build our website framework and insert sentiment analysis.
Please find the detail code in the file: lib/Flask/routes.py

3 4. Insights:

1. Around 60% tweets are negative for the 'Trader War'.

2. LDA Topic: After we get key-words for negative and positive topics we draw the word cloud on the left-hand side. For Negative topics: The large "Trade War" and "Trump" represent people's concerns about the President Trump's announcement about the tariffs which triggered the Sino-US trade war. Many people expressed the concerns regarding the stock market. Dow Jones plunges more than 724 points. This trade war put Wall Street into panic. Another concerns is regarding the agriculture, To fight back, China impose 25% duties on imports of the commodities. As an agricultural exporter, U.S. suffered from the increasing tariff of exporting "soybean" to China. It threatened the food consumption which is inelastic demand in China. However, "Mexico" may benefit from the trade war since it can export more soybeans and pork to satisfy China's demand of agricultural products. That's not good news for "farmers" since they may face higher price but lower sales. For Positive topics: On the other side, as a main "steel" importer, U.S. increased the tariff of steel to 25% and limited the investment of technology corporations from China. The possible "manufacturing" "job" became the hot discussion topic. China's technology and steel industry faced a lot of problem for entering the U.S. market.

3. Busines Insights: Sentiment Analysis among all twittes has strong correlation with Dow Jones Index: Please find detail throuhg our website.

4 5. Group Members:

Keran Li: kl2993@columbia.edu

Mengqi Chen: mc4396@columbia.edu

Wenyi Tao: wt2271@columbia.edu

Xinlei Cao: xc2420@columbia.edu

Xinrou Li: xl2685@columbia.edu