Homework 9: NLP

Overview

Natural Language Processing is one of the most popular applications of machine learning. Top technical companies such as Google and Amazon have developed their online NLP tool for users to conduct automatic machine learning with designed algorithms. This project utilizes these two platforms to analyze text data from Amazon reviews.

Technique

Cloud Natural Language

In order to apply Cloud Natural Language API in colab, first I need to create a credential in the google cloud platform and import the key into colab. Then I conduct two kinds of text analysis, entities analysis and sentiment analysis on the given text. In entities analysis, the result gives the category of each word and corresponding confidence. In the sentiment analysis, the result gives a score showing the level and attribute of the sentiment. Negative values represent negative sentiment while positive values represent positive sentiment. The absolute values reflect the intensity of sentiment.

AWS Comprehend

Similar to Cloud Natural Language, I need to establish an IAM credential and call the account ID and key in the colab to connect with the Comprehend API. To perform text analysis, first I make two word clouds of negative reviews(those with a rating of only 1 star) and positive reviews(those with a rating of five stars). We can see from the following graphs that words such as "waste" and "avoid" are popular in the reviews of the negative one. Words like "great" and "love" are frequent in the reviews of the positive one.

Second I plot the top 30 frequent words in negative reviews and positive reviews respectively and get the bar charts below.

Top 30 (Uncleaned) Word frequencies in negative reviews

Top 30 (Uncleaned) Word frequencies in positive reviews
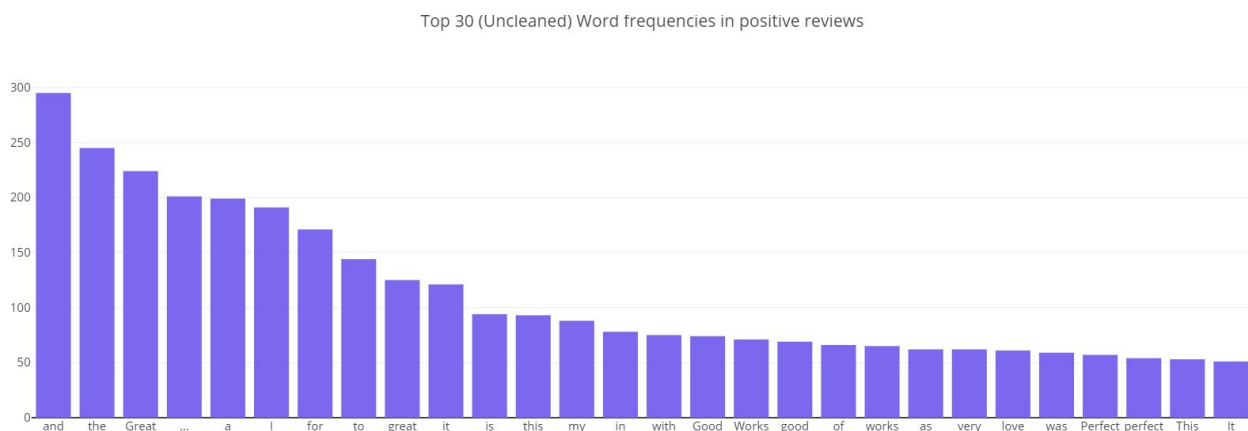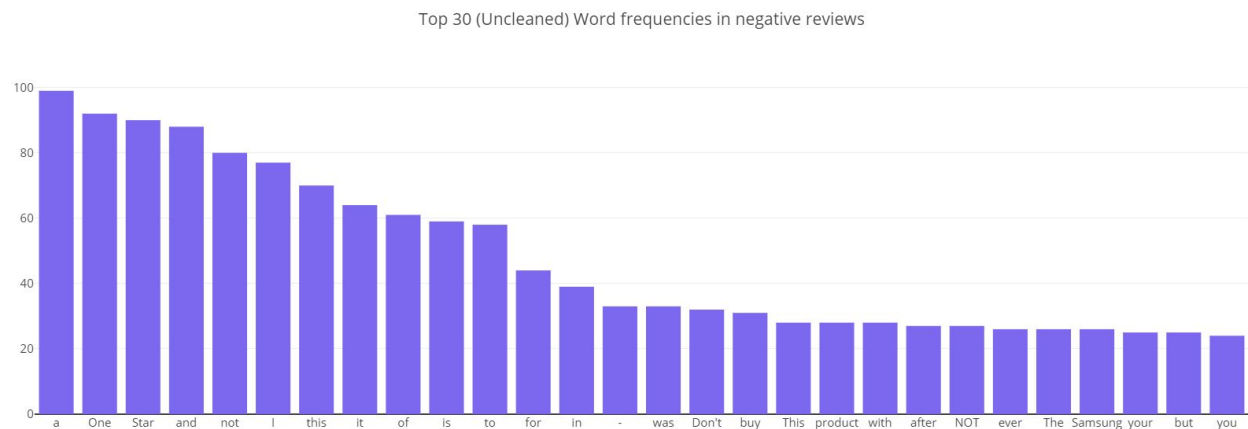
At last, using the Comprehend API, I conduct sentiment analysis on the review_body and divide them into three groups, positive, negative and neutral.

Conclusion

Amazon Comprehend and Google Cloud Natural Language both have powerful NLP tools including entities analysis, sentiment analysis, key phrases, language and syntax analysis. Once figuring out how to connect the API to each platform, the rest steps become very easy. These tools save us from model training and selecting, which is a great way to improve the efficiency of data analysis process.