Anamitra Bhattacharyya
PREDICT-452 Sect. 55
Individual Assignment 2,Option 2
Social Media Data Collection

Executive Summary

The aim of this project was to extract tweets from Twitter information related to the Chevrolet Bolt. The latter will be used in my team project related to 'Nature & Environment', to perform extraction of sentiment analysis of the Chevy Bolt compared to other competitor electric vehicles. A Twitter account was created using developer tools, authentication credentials were utilized to automate data collection of a Twitter stream to extract tweets referencing the 'Chevrolet Bolt'. The extracted information was not parsed for detailed analysis, but the raw collected data was output into a JSON formatted file, and as two separate text files to review the tweets.

Research design and methods

The strategy used for this exercise was to use a Python program (using the Twitter package) designed to access Twitter data related to mentions of the 'Chevrolet Bolt'. Specifically, a Twitter account was established in order to obtain Twitter credentials (*via* https://apps.twitter.com) for the purposes of deriving Twitter authorization to access Tweet data. The example keyword argument was a search string related to the Chevy Bolt, which was passed using a method that used the Twitter API to derive 200 tweet results. The data was output in JSON file format (Javascript Object Notation). Each of the 200 dictionary records represented in the JSON output file were parsed, to first display each entire tweet record and output to a plain text file, and secondarily parsed to a plain text file containing the text part of the tweet.

Implementation and programming

The Python package 'Twitter' was used to develop a Python program (AB_run_twitter_jump_start_show.py) that used the Twitter API to provide access by a registered Twitter user *via* the OAuth method to obtain tweets for a specific keyword search (*e.g.* 'Chevrolet Bolt'). The tweets retrieved using the keyword search were output as dictionary objects to a JSON formatted file. The

latter was parsed to display each tweet dictionary object, in text format to a plain text file, as well as to display the text portion of the tweet in a separate text file.

Main program: **AB_run_twitter_jump_start_show.py** (main program that calls several methods to pass twitter user access credentials, perform search using user-defined keyword arguments, write the data relating to 200 tweets to a JSON file, which is parsed to extract each full tweet record as a dictionary item printed to a text file, and also parsed to display just the text data to a second text file). The results of the Twitter data acquisition were captured in 3 files: 1 JSON-format file and 2 text files (*e.g.* **AB_tweet_file.json, AB_tweet_review_file.txt, AB_tweet_text_file.txt**).

Output files:

AB_tweet_files.json: JSON format of the 200 tweets derived using the Twitter API

AB_tweet_review_file.txt: Text file of the parsed JSON file to display each of the 200 dictionary items representing each of the tweets collected.

AB_tweet_text_file.txt: Text file displaying just the text portion of the 200 tweets collected in the JSON file.

Conclusions

Automated data acquisition of 200 tweets related to the "Chevrolet Bolt" electric car was successfully obtained in the project, using the Twitter API. The extracted information from the Twitter feed was stored in JSON format for further analysis, and only parsed to extract the complete tweet record for each of the 200 tweets, as well as the text portion of the tweets.