Apple Watch Tweets Analyzer

Qichen Jing
Department of ECE
Boston University
Boston, MA 02215
jingqc@bu.edu

Abstract

Apple Watch becomes very popular recently, and many people may consider whether to buy an Apple Watch or not. My project focus on analyzing people's attitudes about Apple Watch. Twitter is an online social networking service that enables users to send and read about short 140-character messages called "tweets". People are very likely to express their feeling through tweets.

1 Problem Overview

I scrape tweets using hashtag#applewatch to analyze people's sentiment. Starting from April 13 to today, I have gotten about 26,000 tweets about apple watch using twitter streaming API. Here is the link for twitter Streaming APIs:
https://dev.twitter.com/streaming/overview/request-parameters
This API could help to scrape one tweets about a specific key word, here I use "Applewatch."

2 Data Set

Till now, the data set have 26, 000 tweets from different users of different countries. This number will still increase since my computer will keep scraping these days. There is many information of a tweets scraped by streaming API, here I only saved the most important information: "The 140 words text".

3 Clean data

Here are some examples about the tweets:

Will the **Apple Watch** eclipse the classic Swiss watch? on.wsj.com/1b6MkNb

So THIS is where they manufacture the #**AppleWatch**? #FAQinHell

11 surprises I learned about the **Apple Watch** cnet.co/1PYWboR.

Before using sentiment analyzing methods, I need to clean the data firstly. Here are some basic processes:

a. Lower Case: convert the tweets to lower case.

b. URLS: urls do not help to get sentiment information, so I can eliminate all the urls

using regular expression matching.

c. @username: usernames also do not provide sentiment information, we can eliminate them using regular expression matching.

d. Punctuations and additional white spaces: remove punctuation at the start and ending of the tweets.

e. Stop words such as a, is, the, with etc do not indicate any sentiment and could be removed.

f. Repeating letters: some tweets may be “Apple Watch is soooooo amazing”, we can just change it to “Apple Watch is so amazing”

g. Words must start with an alphabet: we can remove words starting with number or not alphabet.

h. Because tweets I collected are from different languages, I need to group them by language but I did not figure how to group it now.

4 Analyzing part

After cleaning data, I will use some sentiment analyze api or something trained by researchers to analyze my data and find the trend of popularity of apple watch. Comparing this trend with the stock price of price in those days, try to get something. Another part is to analyze tweets of different language to figure out attitudes of people from different country. But I do not know whether there is a sentiment analyzer for other languages, I will figure it out later.