

BIA-660 Web Analytics

Midterm Report - Analysis of Smartphone reviews

Motivation:

As the Apple Inc launched iPhone 8 in WWDC this year, which draw people attention to the latest smartphones with the most advanced technologies. Our group is very interested in analyzing smart iphone industry specially by observing people's attitudes toward some popular products on social media and online reviews.

Background:

Apple released iPhone 8 along with iPhone 8 Plus this year at its worldwide developer conference. In addition to these two new phones, Apple also released iPhone X to commemorate iPhone 10th year anniversary. In the industry of smart phones, **iPhone 8** has many competitors. Samsung **Galaxy S8** and **Note 8** also have very big market shares. Our group is very interested in analyzing smart iphone industry specially by observing people's attitudes toward these three products and customers' reviews regarding these products on Best Buy website. Because of customers can give numerical scores to the products, our group wants to categorize reviews based on these scores and understand why customers give these scores. Moreover, we are also going to analyze some technical review articles from top tech websites and find out how they are similar or different from public opinions.

Our approach:

We decided to build website scrapers to get customers' reviews from Best Buy website. The website does not allow web crawler to get users data. Therefore, we have to use Selenium webdriver which can automate any web browsers in order to access the information we want. Then, all the data will be saved locally for further analysis. Next, we need to search for professional tech reviews about these products on some websites such as **Cnet**, **TechCrunch**, and **Techradar** and save them into local text files.

After collecting data, we will use natural language processing skills such as tokenization and lemmatization to analyze those review comments and determine attitude from customers and public toward these products by doing sentiment mining. Then, we will be using data visualization technique to explicitly reflect our results.

Data sources:

Our group has several data sources:

1 Bestbuy website:

<https://www.bestbuy.com/>

Best Buy is a very great website for shopping for electronics, computers, and cell phones. We use this website as one of our data sources is because it is a very popular website and many customers choose to leave their reviews in Ratings & Reviews section. We can easily use web scrapers to retrieve scores and reviews of certain products we want to analyze.

https://www.bestbuy.com/site/reviews/apple-iphone-8-64gb-gold-verizon/6009932?page=2&sort=MOST_HELPFUL

This is a typical URL which shows the name of the phone and page number.

2 Top tech websites:

<https://www.cnet.com/>

<https://techcrunch.com/>

<http://www.techradar.com/>

These three websites are the top leading technology media which have many news and reviews about current products. Many experts publish reviews on these three websites. We can directly search the product by typing the name within these three websites. The reviews will be shown.

For now, our group has already collected customer's' reviews for *apple-iphone-8-64gb-gold-verizon* and *apple-iphone-8-256gb-space-gray-at-t*. They were saved into a file called *iphone8_review.txt*. Titles, ratings, and reviews are separated by commas. We also collected a tech article from techradar and saved into a file called *iphone8_article_1.txt*.

Methodology:

Alternative 1:

- We will group the reviews obtained by web crawler based on ratings.
- Next, we will process the data by removing the stopwords.
- For each group, we will find which words appear the most times and which words are the most important or having the most weight in reviews.
- We will also find customers' attitude toward these products by doing sentiment mining.
- Lastly, we will use the same methodology to analyze reviews from tech websites and compare the results with those of customers'.

Alternative 2:

- When we get the clean data, we will focus on important features of a smartphone such as camera, screen, processor, etc.
- We are interested in finding how people describe these features of a smartphone by collecting adjective of these features.
- Next, we can calculate score of each feature for a product and compare with what we get from the articles.
- Lastly, comparing the features of each product and we can find out the customer's attitude for those products.

Challenges and open questions:

We have not yet learned sentiment mining. After finishing to collect data, our group cannot continue to do the project without knowing sentiment mining.

1. How to get useful information from the reviews from tech websites.
2. How to remove the irrevelent adjective from the articles.
3. How to allocate the describe words to each feature in a customer's review.