PROJECT TITLE

Customer Review Analytics

Student Mentor: Anurag Tyagi
Team Number: 2

Shikha Jain (201301004)
Vivek Vishnoi (201505504)
Chanakya Malireddy (201325031)

Problem Statement

To understand the user reviews for a given product and extract the sentiment for each aspect of the product mentioned in the review.

Abstract

With the advent of the Internet it has become very easy for a user to provide their feedback. With the massive proliferation of e-commerce, the number of reviews is growing exponentially. For a popular product, the number of reviews can range from hundreds to thousands. This has made managing customer reviews difficult for the manufacturers.

Basically, a review consists of sentiments about various aspects of a product. Our system receive as input a set of texts (product reviews) discussing a particular entity and detects the main aspects of the entity and estimate the average sentiment of the text per aspect.

Proposed Approach

1. Crawl user reviews for predefined competitor and brand websites to fetch a master list of products across pre-defined categories

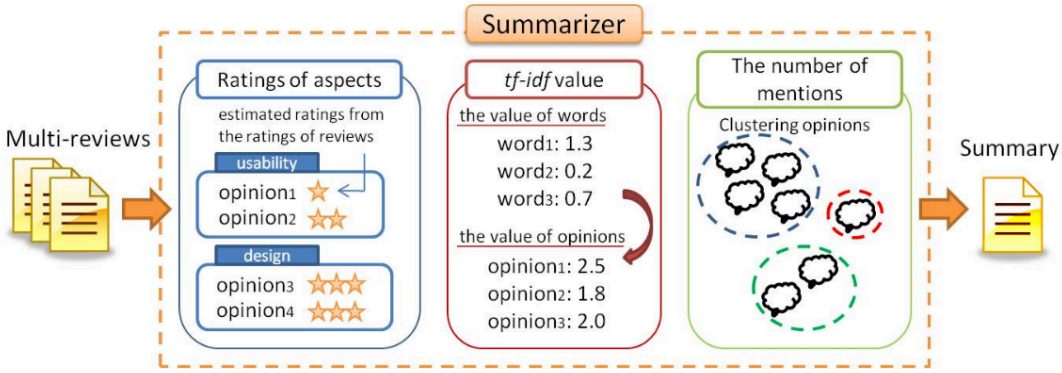
2. Identify review parameters for each product.

3. Assign weights to each review parameter based on the number of reviews it appears in.

4. Extract the sentiment for each review parameter mentioned in the user review.

5. Create a hierarchy of parameters for each product and map them with their corresponding sentiments.

6. Create an API that returns the sentiment score for a given aspect of a given product.



Applications

Parameter-based sentiment analysis of user reviews would allow us to give a detailed feedback to the manufacturer. Such a feedback would help them understand if the general public is unhappy with a certain aspect of their product and hence can help them modify it accordingly. For example, users may be unhappy with the screen resolution in the new iPhone 6s mobile). It can also be used to develop new products with emphasis on those particular parameters.

Such an analysis also helps us provide a targeted recommendation system for the users. For example, we can

provide suggestions for products with good sentiment on screen resolution to users who might have complained about the same in their previous reviews.

Challenges

1. Crawling user reviews from various websites, which block crawlers.

2. Problems with identifying the aspect, which is being talked about in a given review and its corresponding sentiment. For example – *"The new iPhone has a bad camera but a long lasting battery."* (The sentiment for camera is negative while that of the battery is positive)

3. If the review is comparative in nature. For example – *"I recently bought an iPhone. But my Samsung Grand has a better processor."* (iPhone was the original product but the review does not talk about it directly).

4. Problems with anaphora resolution. For example – *"The new iPhone has a lithium battery. It is really bad.* "(What does the "it" refer to?)

Major Project - Second Deliverables Details

Dataset containing a weighted hierarchy of parameters for each product, generated from raw user reviews crawled from various websites.

- Create a master product list.
- Create a hierarchy of parameters for each product.

- Crawl user reviews from various websites for each product in the master list.
- Weight the parameters of each product.

Major Project - Third Deliverables Details

Create an API that returns the sentiment score for a given aspect of a given product.

- Identify the aspects being talked about in each review.
- Identify the sentiment for each aspect in a given review.
- Assign a final score to each aspect of every product.
- Create an API to provide easy access to the above dataset.

Tools to be used

- **Language** – Python
- **Crawlers** – Scrapy, BeautifulSoup
- **NLP** – NLTK
- **Classification & Machine learning** - Scikit, Pandas
- **Web framework** – Django
- **Database** - SQLite
- **APIs** - REST

References

1. “Mining and summarizing customer reviews” [Minqing Hu and Bing Liu]
2. “Mining opinion features in customer reviews” [Minqing Hu and Bing Liu]
3. “A Survey of Text Classification Algorithms” [Charu C. Aggarwal and ChengXiang Zhai]
4. An Unsupervised Aspect-Sentiment Model for Online Reviews [Samuel Brody and Noemie Elhadad]
5. Opinion digger: an unsupervised opinion miner from unstructured product reviews. [Samaneh Moghaddam and Martin Ester]