Xuhe Lu

Professor Li

Problem Solving & Software Design

November 6, 2017

Sentiment Analysis of Stranger Things and Insidious

Overview:

The second season of Stranger Things was just released on Netflix around a week ago. Personally, I am not a fan of the horror and thriller genre so I haven't seen it until last week. When watching it, I got a strong sense that the show was strongly based on the movie Insidious. Thus, I thought it would be cool to compare the show and movie with sentiment analysis of its reviews on IMDB.

Implementation

Since I was trying to look at the general sentiment of the reviews, I first had to pull reviews from the IMDB website. This was done using the imdbpie package. Using the imdbpie package, I was able to scrape the reviews from the website and write them into a notepad file. Originally I wanted to scrape every single review that was available however the imdbpie package was only able to scrape the first page of reviews on IMDB. Since IMDB allows you to show 10 reviews at a maximum, I was only able to use the top 10 most "useful" reviews.

When the reviews are scraped and written into a notepad file, they are at first their own separate files named review_0 to review_9, I then take all of those separate files and combine them into a singular notepad file called strangerthings_reviews for Stranger Things and insidious_reivews for Insidious.

Finally, the two combined files were run through sentiment analysis. This was done using a package called 'nltk' which is used for natural language processing. With this package, I was able to conduct the sentiment analysis on the two files.

Results

The results for the sentiment analysis were as follows:

Stranger Things:

This is the output that for the sentiment analysis on Stranger Things:

As shown by the output of the sentiment analysis, among the most “useful” reviews on IMDB, there is a generally positive sentiment on the TV show. This means that overall there were more positive words used to describe the show compared to negative words.

In contrast, this was the sentiment analysis for Insidious:

```
{'neu' : 0.655, 'compound' : 0.999, 'pos' : 0.092, 'neg' : 0.253}
```

In this sentiment analysis, it shows an overwhelmingly larger negative sentiment towards the movie. This was a surprise to me given the stark similarities of the show and the movie. In this case, there were many more negative words used to describe the movie and there was also less neutral words that was used.

Reflection

My original goal of this project was to show that two similar sources of media would garner two similar review sentiments. This was obviously proven to be false. Since I worked in a single person group, I had to limit the scope of the project so I would not be tasked with too much work I could handle. However, I believe that if I had more time, I would have included word frequency analysis along with this to perhaps have another view point at proving my point. Another thing I wish I could have done differently was perhaps find another package that could scrape every single review on IMDB instead of just the top 10 reviews. This was not a huge factor though because the top 10 most “useful” and highly community rated reviews also reflects the overall sentiment of the community very well. It will filter out many of the outlier garbage reviews so only having those 10 reviews is not a huge detriment. Overall, I am happy with what I accomplished with my project.