Sentiment Analysis for Ipad 2

Sida Ye

Prof. Aldous

Department of Statistics

Introduction to sentiment analysis

Whenever we need to make a decision, we often seek out the opinions of others. In the past, we seek opinions from friends and family or companies would use surveys, focus groups, opinion polls, and consultants. Now, customer reviews on the Internet has risen exponentially over the last decade. It is an important resource for buying products or attending events. In these situations, we would like to see what others are saying about them. Also it is a significant aspect for companies making decisions about their products or services.

Sentiment analysis, the computational study of how opinions, attitudes and emotions are expressed in natural language, provides techniques for extracting some emotional words, subjective information from large datasets of customer reviews and summarizing it. It can thus be vital to service providers, product producers, moviemakers, allowing them to quickly assess how new products and features are being received.

Methods to implement sentiment analysis

1. First step is collection data from amazon. I used the statistical software, R, to write down the function, which can automatically retrieve customer reviews and rates from Amazon. I only need to input the product ID in amazon and the function will output the customer reviews and rates in a csv format file.

2. Second is going to find the list of words, which contains both positive words and negative words. Then, the next step is to tokenization. Tokenizing (splitting a string into its desired constituent parts) is fundamental to all Neuro-linguistic programming tasks. It is very complex if you want to do it extremely accurate, because sometime it is uniquely represented a single cluster of punctuation like :-( might already represent the whole opinion. Hence, I did in a sample way. I use the function split in R in order to split every word with a space. For example, “I love ipad. It is awesome.” By applying the function, we get eight units.

3. Then we are moving to the next step. I try to match the words from word list with words in the reviews. Here, I use a Bayesian classifier, which means one positive match counts one point, and one negative match counts negative one point.

4. Split all the sentence into single, lower letter words. This is a preparation for the further sentiment analysis.

5. Based on the method above, I can get a sentiment score finally.

Results

This data frame contains the review ID, amazon star rate, sentiment score, number of positive words, number of negative words, length of reviews and the context of customer reviews

Conclusions

From the results, we can conclude that ipad 2 has a very positive customer reviews. But I still want to know how popular it is. Thus, I decided to compare it with its competitors, Samsung Galaxy Tab 10 and HP Touchipad. I did the same steps for the other two products. Then, I got these pie plots below.

Acknowledgments

I want to appreciate Professor Aldous for providing this opportunity to me to share my ideas about sentiment analysis and always providing encouraging guidance.

References

- The code used in this analysis is available on GitHub: https://github.com/sida-ye/sentiment-analysis-ipad-2
- For more information about sentiment analysis, please visit: https://www.streamingreview.com/sentiment-analysis