Social Pulse - Understanding Brand Perception using Social Media Analytics
ECE 229 Project

Murali M K D, Hao Qiu, Pai Tong, Liyang Ru, Chris Light, Yen-Ju Tseng
A59004607, A59012849, A59004757, A59012876, A15694506, A59005785

Abstract

Listening to customers' thoughts is one of the most important aspects for any brand growth. Social media analytics provides a powerful way to understand how they are perceiving your brand, industry and competitors. Here we are proposing 'Social Pulse', a platform for Brand Marketers based on statistical text/NLP analytics. To achieve this, we plan to utilize Reddit (PRAW) and Twitter API data, along with auxiliary sentiment datasets for the model building process. The models are aimed at extracting sentiment polarities and detecting emotions. We firmly believe that social media analytics will have immeasurable market prospect on product feedback and expansive application scenarios.

1 Introduction

According to a study, two-thirds of brand marketers agree that social listening has increased in value in the last year. It is really important for variety of reasons - performance measurement, brand reputation, product development, competitive insights, bench marking etc. With huge data scattered over the internet, we are focusing on data scraping, consolidated and diverse text analytics and action-oriented user stories which are detailed in the later sections.

2 Data

The data is a combination of primary Social Media Platform API responses and auxiliary datasets to train models for sentiment analysis.
For primary data, we are considering Twitter and Reddit, however each have subtle nuances. Reddit comments are often larger bodied with a character limit of 10,000, where Twitter is limited to 280. Twitter, on the other hand, offers geo-location of valid tweets. Size of this data is affected by the rate limits set per company.
The auxiliary datasets in consideration are ones that offer trinary positive/negative/neutral labels as well as n-ary emotional labels. 'Twitter and Reddit Sentiment analysis' [1] is such a trinary dataset of size 37k Reddit and 163k Twitter comment entries. 'GoEmotions' [2] is a dataset with emotion labels and 58k comments from Reddit.

3 User Story

Knowing your customer is the key to maximizing profits in a competitive market. Having a strategy for Social Listening is what I use to understand my customers. That's why, as a Brand Marketer, I want to utilize 'Social Pulse' so that I can understand issues with my customers, track trending concepts, and measure the brand sentiment. With Chronological plots of brand sentiment and Geoinformatic graphs, I can locate the time and place that has the most potential for growth in my company. These tools are further utilized by time filtered data, keyword searching, and the data choice of singular or mixed social media channels. The benefit of Social Pulse is two fold when used to cross-analyze with competitors.

4 Data Science Product

The user need of brand sentiment will be addressed showcasing the sentiment polarity, and emotions radar based on the models built. Also, statistical techniques like LDA topic modelling will be used to identify key themes about the brand. Time series sentiment plots will show the trends over time suggesting important events. Sentiment word clouds and LDA-topics (combined with sentiment) will provide key themes of the brand. Geo-heatmaps will provide geography based brand perception. Combining sentiment and themes across different dimensions will enhance the usability and decision making of the user. Competitor analysis will help the user compare similar brands in the industry and help in product development decisions. Multiple selection filters and drilldowns to view extreme posts will help the user's interpretability of brand perception. Glossary will be provided for any definitions, plot explanations and usage documentation.

References

[1] Charan Gowda, Anirudh, Akshay Pai, and Chaithanya kumar A. Twitter and reddit sentimental analysis dataset, 2019.

[2] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. Goemotions: A dataset of fine-grained emotions, 2020.