Amazon alexa Sentiment analysis Machine Learning for Natural Language Processing 2020

Melan Sekoua

`melan.sekoua@ensae.fr`

April 2020

Abstract

My project talk about the feedback of amazon customer on some products ,their feeling that express on amazon products and identify the products with good or bad mark and suggest to the same group of people with the same taste. In my work, firstly, I cluster feedback of customer following their rating using two approaches (kmeans ++ and pca, kmeans et tnse) for identifying those approach have the best representation of feedback and product, secondly I build an embedding model to predict the most products they like and lastly, I suggest an approach to recommend some products.

1 Problem Framing

In the context, I highlight the main point that's developedd in my through this differents bullets.

- Visualize data ans show the main topic use in the feedback

- Explanatory data and describe each feature of data

- Preprocessing and analysis

- Build and evaluate my model

- Recommendation

2 Experiments Protocol

I just describe my idea in the different topic mentionned upper in order to have an inderstanding of my work.

2.1 Visualization data

In this part, I drew the the most common word use in customer feedback in my notebook and I observed some describe the feeling of customer such as love etc.

2.2 Preprocessing

This part consisted to clean feedback customer , removed the stopwords, the punctuation and had the most useful vocabulary to train and evaluate the model.

2.3 Training and evaluation

I used two approaches for understanding my data, one using the cluster method to find the pattern of feedback and the sub-products of amazon alexa, other to build an embedding model.

2.4 Model used

In my work, the models used are a pre-trained BERT, an embedding model with some layers, random forest with a tfidf model and gridsearch using the tfidf approach

2.5 Implementation

This implementation is available in my notebook

3 Results

The work presents an approach to predict a rating according the customer feedback and can recommend some product if their ranking is near and avoid other with low ranking.