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 work through this differents bullets mentionned.

- 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

This step consisted to draw or display the most common word use in customer feedback in my notebook when the customer expressed their feeling on amazon product. I generally note lovely lexical has been used to describe customer feeling, that's expressed amazon product is very useful for the customer.

2.2 Preprocessing

This part consisted to clean feedback customer , removed the stopwords, the punctuation, find the distribution of the word used in feedback, check if some theorem is true in order to gave us a tool to train and evaluate my model.

2.3 Training and evaluation

Here, I used two approaches for understanding my data, one of them is related to the clustering method(kmeans++ and pca and kmeans ans tsne) and the other is related to use an embedding model pre trained and my own embedding model. For the evaluation, I used common metrics used for the kind of work, which it highlighted in my notebook/

2.4 Model used

To build my models, I need to transform each feedback in vector and had a matrix ,that's is useful and understanding for the algorithm. To do, it used two approaches, one used the tfidf method and the other used a simple vocab from the data. After mentionning, the pre trained, random forest and grid search had been used for my utilities.

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.