MAIS 202 – Project Deliverable 2

Preliminary Results

Problem Statement:

The project attempts to do aspect based sentiment analysis on the reviews of 400,000 unlocked phones on amazon. The machine learning model will take in a review on a phone. Then, it will output the overall sentiment of the review. If the review makes reference to the ease of use, battery life, appearance, or overall quality, the model will predict the sentiment of the user on that aspect too.

Data Preprocessing:

I am using the Amazon Reviews: Unlocked Mobile Phones dataset (PromptCloud, 2016). The data had information on the product name, brand, price, rating, review and review vote. I choose to extract the rating as it is a good indicator on the sentiment of the reviews. I classified ratings of 4-5 to be positive, ratings of 3 to be neutral and ratings of 1-2 to be negative. The only preprocessing that was done for the ratings was to change them from strings to integers. Then to represent them in a three classes format - class 0: negative, class 1: neutral, class 2: positive.

Most of the preprocessing work went into the reviews. I had to try different layouts of cleaning the reviews and choosing the vocab. At first I tried stripping away anything but letters and whitespaces. Then, I noticed that there were often words that were stuck together, such as: goodthen. That could be attributed to bad punctuation, reviewers might have forgotten to insert a space after commas, periods and etc. Thus, I had to replace characters such as -,.: with whitespaces and then strip away all the other characters. After doing so, I noticed that some of the most popular vocab was only one word, such as ‘g’. I then discovered that was because I was stripping away numbers and the users would make references to 4G, 3G, and GB. Thus, I had to exclude stripping away numbers in cleaning the reviews.

After cleaning the reviews, I had to appropriately filter the vocabulary. I spoke before about making sure to translate the words in reviews. However, googletrans has a limit on characters and that wasn’t going to work with such dataset. As the dataset is huge, we don’t need the information from non-English words, there are enough English reviews to make a good analysis. To filter the vocabulary, I made sure to exclude all stop words from the vocab. Moreover, I excluded all numbers as most number references were referring to phone names such as iPhone 6 or Samsung Galaxy Note 3. There were also lots of random single letters as a result of replacing the full stops (ex: i.e. / e.g) so I choose to filter them out. I then took the 4000 most commonly used in the filtered vocabulary as my vocab of choice to do the bag of words methods. I choose 4000 to make the model quick enough at the same concise.

Machine Learning Model:

As I alluded to in the previous section, I represented the reviews using the bag of words method. As the dataset was huge, around 400,000 entries, I choose a random sample of 150,000 entries to make computation easier. Of those samples, I split the data into 80% for training and 20% for testing. As I had a huge dataset, I didn’t feel the need to be extra careful on the splitting ratio, all works!

This time around, I have only done the overall sentiment analysis of the reviews. I have tried using Naive Bayes (NB), Random Forest Classifier (RFC) and Support Vector Machines (SVM) to classify the data. On first glance, NB and RFC had an accuracy of 75% for both the train and test data, while SVM had an accuracy of 85% for the test and 90% for the training. However, when I was entering sample reviews myself, NB and RFC were not doing well at all, with both classifying “terrible” and “bad” as positive reviews. However, SVM ended up working really well. The relatively high accuracies of the NB and RFC could be explained by the fact that more than 68% of the reviews were of class 2, so when the models were predicting class 2 for almost everything, they were able to get away with it.

From now on, SVM will be my model of choice when predicting the overall sentiment of the upcoming reviews!

Preliminary Results:

I am pretty happy about the SVM results. I have tried some reviews of my own and I am satisfied with the results. It got "the battery life is bad, but the phone is good overall" as a positive review but "the battery life is bad” as a negative review which I took to mean that the model makes good predictions on complex sentences.

Here’s what the confusion matrix (train first, test second) looks like:

Precision, recall and the f1-score were great for class 2, good for class 0 but really low for class 1. This could be explained by the fact the less than 10% of all reviews are of class 1 so the model wasn’t able to train well on them. I would also say that it is even difficult to classify neutral tone in text in real life, so I understand why the model wasn’t able to predict them correctly!

	precision	recall	f1-score	support
0	0.846	0.840	0.843	18721
1	0.785	0.262	0.393	6231
2	0.903	0.973	0.937	55048
accuracy			0.887	80000
macro avg	0.845	0.692	0.724	80000
weighted avg	0.880	0.887	0.872	80000

	precision	recall	f1-score	support
0	0.796	0.787	0.792	4677
1	0.490	0.177	0.259	1473
2	0.892	0.956	0.923	13850
accuracy			0.859	20000
macro avg	0.726	0.640	0.658	20000
weighted avg	0.840	0.859	0.843	20000

Next Steps:

I want to make the model more complex and work on the aspect based part of it!