Aspect Sentiment Classification Project Report

Saikrishna Kalahasti Karthik, Malavika Ramprasad

Abstract:

Sentiment analysis is one of the most important tools in natural language processing. It involves capturing the sentiment from a natural text. Aspect based sentiment analysis is an extension of the usual sentiment analysis capturing both sentiment and semantic details encoded in a text. Performing aspect sentiment analysis to find the polarity for the aspect term helps to capture nuances about the objects of interest in the given text.

In this project, we experimented aspect sentiment analysis by building multiple models to classify the polarity of the aspect terms in the given text including Random Forest, Multinomial NB, Linear SVM.

Introduction:

Sentiment Analysis is an important task in natural language understanding since it opens wide range in understanding people's opinion on diverse topics. It aims to analyze an individual's opinion or emotions. It is gaining importance both from a business perspective and academically.

Most of the analysis method tend to classify a given sentence as positive, negative or neutral. However, the need for a more fine-grained approach, such as Aspect-Based Sentiment Analysis (ABSA) is increasing, which would analyze the emotion attached to the sentence in a more deeper fashion.

For example, consider the sentence "The price is reasonable although the service is poor.". This sentence has a positive polarity with respect to price and negative polarity with respect to the service. If we classify this sentence as a neutral, it would ignore that the price was unreasonable, or the service was poor. This is the challenge faced by Aspect Based Sentiment Analysis.

The input to aspect-based sentiment analysis will pertain to millions of available customer reviews in various domain. There have been various approaches to tackle this problem. The tasks usually associated with aspect-based sentiment analysis is given below:

- Aspect Extraction
- Aspect Sentiment Classification

In this project, we have been given review sentences from two domains - laptops and restaurants along with the aspect term for each review. Classifiers has been built with these reviews and aspect term and then it is used to predict the sentiment analysis of all the review belonging to test data.

Data Pre-processing:

The dataset used for this analysis consisted of 3600 restaurant reviews and around 2200 laptop reviews in tab separated format:

- example_id: unique identifier for the review
- text: the review (laptop and restaurant)
- aspect_term: the aspect term for the review
- term_location: the start and end position for the aspect term
- class: the polarity for the review based on the aspect term

In our model we split the data into approx. 70% for training and 30% for testing.

Data preprocessing steps:

- Removal of [comma] with ‘,’.
- Removal of stop words.
- Removal of punctuations and special characters.
- Convert sentence to lower case.
- Tokenization and POS Tagging.

Features:

- Unigram, Bigram and Trigrams were used as features.
- Window of words around the aspect term was taken as features. Window size of 20 was used.
- 7-9 words to the left and 7-9 words to right along with the aspect in the middle contributed to a single feature vector.

Classification Models:

- Multinomial Naïve Bayes
- Linear SVM with sentiWordNet
  - SentiWordNet: It is a lexical resource for opinion mining. It assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity.
- Random Forest Model
- Linear SVM

Validation:

K-fold Cross Validation (k = 10) was used to evaluate the models on the training and test data set.

Evaluation and result:

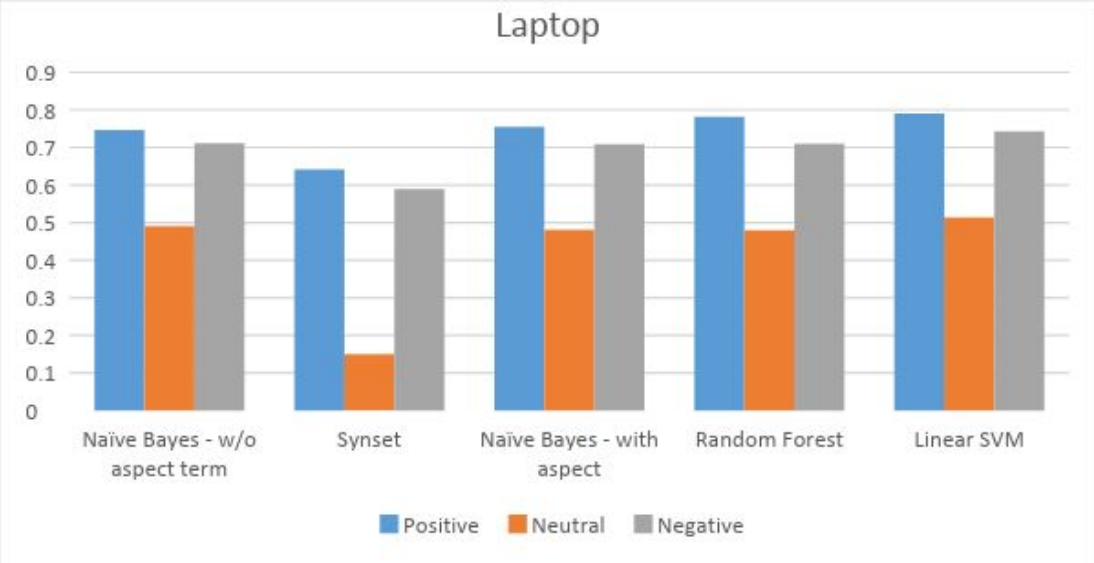
Dataset - Laptop

Model	Positive class			Neutral class			Negative class			Accuracy
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	
Naïve Bayes – without aspect term	0.6932	0.8079	0.7462	0.5851	0.4231	0.4911	0.7295	0.6949	0.7118	**67.18%**
Synset	0.5729	0.7285	0.6414	0.3333	0.0077	0.050	0.5500	0.6339	0.5890	**50.81%**
Naïve Bayes – with aspect term	0.6846	0.8411	0.7548	0.6410	0.3846	0.4808	0.7302	0.6881	0.7086	**69.54%**
Random Forest	0.7912	0.7644	0.7823	0.7856	0.3433	0.4801	0.6223	0.8312	0.7100	**69.90%**
Linear SVM	0.7858	0.8050	0.7910	0.5679	0.4840	0.5140	0.7340	0.7572	0.7430	**72.22%**

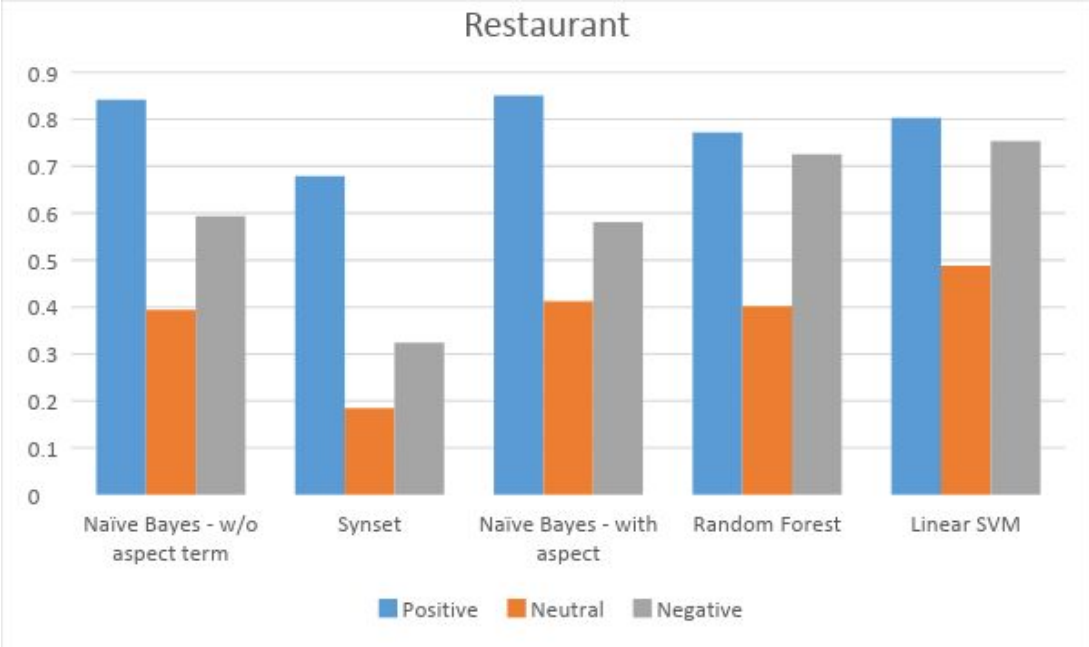
Dataset - Restaurants

Model	Positive class			Neutral class			Negative class			Accuracy
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1	
Naïve Bayes – without aspect term	0.7790	0.9166	0.8422	0.5191	0.3178	0.3942	0.6557	0.5430	0.5940	**68.81%**
Synset	0.6627	0.6968	0.6793	0.2045	0.1682	0.1846	0.3230	0.3242	0.3236	**51.37%**
Naïve Bayes – with aspect term	0.7933	0.9179	0.8511	0.5103	0.3458	0.4123	0.6415	0.5312	0.5812	**71.40%**
Random Forest	0.7861	0.7602	0.7722	0.6861	0.2348	0.4012	0.6360	0.8465	0.7258	**69.90%**
Linear SVM	0.7864	0.820	0.8031	0.5912	0.4200	0.4878	0.7220	0.7915	0.7541	**72.99%**

Laptop Dataset:



Restaurant Dataset:



Conclusion:

In our project we built various classifiers and each of these gave us different result and performance. Based on the accuracy obtained, Linear SVM classifies the reviews and assigns the correct sentiment better than the other models.

References:

1. Bo Wang - "Deep Learning for Aspect-Based Sentiment Analysis"
2. Wei Xue - "Aspect Based Sentiment Analysis on Review Data"
3. https://scikit-learn.org/
4. Bhoi, Joshi - "Various Approaches to Aspect-based Sentiment Analysis"