Sentiment Analysis with Amazon Customer Reviews

Mir Tanvir Islam

1711730042

Department of Electrical & Computer Engineering

North South University

Date: 31st May 2020

Course Number & Name: CSE 445 Machine Learning

Instructor's Name: Mohammad Ashrafuzzaman Khan

Abstract

One of the biggest uses of Natural Language Processing in Machine Learning is sentiment analysis from text. In this paper I attempt to predict customer sentiment using Amazon product reviews and their corresponding ratings. I have experimented with traditional bag of words model, as well as Machine Learning models such as Naïve Bayes, Support Vector Machine etc. By analyzing these texts, I found it possible with high accuracy to predict the customer sentiment from their reviews.

Link to project on Github: https://github.com/mirtanvirislam/NLP-Sentiment-Analysis. The file github.com/mirtanvirislam/NLP-Sentiment-Analysis/blob/master/notebooks/Sentiment%20Analysis%20v3.ipynb contains all the machine learning algorithms

Table of Contents

Table of Contents

List of Figures

List of Tables

1 Introduction

One of the emerging fields of machine learning is Natural Language Processing (NLP). NLP can be used for a variety of cases, such as: document classification, emotion analysis, sentiment analysis. Sentiment Analysis is a method of analyzing text polarity (positive, neutral or negative) from a text, document, sentence or paragraph.

2 Background

Amazon is a popular online shopping site, which receives thousands of orders. Their product listings have an option for buyers of the product to rate the product and write a review. This data is a perfect source of data for building a sentiment analysis model. There is a review in text with a corresponding rating on a scale of 1 to 5. With the help of machine learning, the review text as the input and rating as the target, we can build a model that predicts sentiment from reviews.

The dataset used is openly available on Amazon's website as well as Kaggle. Link to the data: https://www.kaggle.com/bittlingmayer/amazonreviews. The data comprised of 28332 reviews in total. The data was of some of Amazon's own products for until May 2019.

2.1 Sentiment Analysis

Analysis of sentiment is the perception and classification of emotions (positive, negative , and neutral) using text analysis techniques within the data. Interest analysis helps companies to detect customer sentiment in online interactions and feedback against goods, brands or services.

There are many types of sentiment analysis, such as:

i. Fine-grained Sentiment Analysis

ii. Emotion detection

iii. Aspect-based Sentiment Analysis

iv. Multilingual Sentiment Analysis

The Main ways of conducting sentiment analysis are:

i. Rule-based system

ii. Automatic system

iii. Hybrid system

2.1.1 Rule-based System

Rule based systems do not use any machine learning, instead they use algorithms that simply check certain rules against the text. This approach is not dynamic and can quickly become obsolete, this approach works, but not with remarkable accuracy.

Approaches in Rule-based systems include:

i. Stemming, tokenization, part-of-speech tagging, parsing

ii. Lexicons

iii. Counting no. of positive and negative words

2.1.2 Automatic System

Automatic methods rely not on manually crafted rules but on machine learning techniques. Usually a task of sentiment analysis is modeled as a classification problem, by which a classifier is fed a text and returns a category, e.g. positive, negative, or neutral.

Machine learning algorithms take various variables as input and a metric to test it against. In the case of sentiment analysis, we use review text as input and review rating as the metric.

Approaches in Automatic systems include:

i. Feature Extraction

ii. Bag of words/ Bag of n-grams

iii. Text vectorization

iv. Naïve Bayes, Linear Regression, Decision Trees, Random Forest Support Vector Machines etc.

2.1.3 Hybrid System

Hybrid structures incorporate under one framework the appropriate characteristics of rule-based and adaptive techniques. The tremendous advantage to these programs is that the tests are always more reliable than only Rule based or only Automatic approaches.

3 Methodology

In this endeavor to train a model to perform sentiment analysis, I took two approaches:

i. Rule based sentiment analysis

ii. Automatic Machine Learning approach

Before the main sentiment analysis part, I analyzed the data and observed it. And found the following:

```
Dataset size: 28332
```

Columns in the dataset:

```
#   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   id                    34660 non-null  object
 1   name                  27900 non-null  object
 2   asins                 34658 non-null  object
 3   brand                 34660 non-null  object
 4   categories            34660 non-null  object
 5   keys                  34660 non-null  object
 6   manufacturer          34660 non-null  object
 7   reviews.date          34621 non-null  object
 8   reviews.dateAdded     24039 non-null  object
 9   reviews.dateSeen      34660 non-null  object
 10  reviews.didPurchase   1 non-null      object
 11  reviews.doRecommend   34066 non-null  object
 12  reviews.id            1 non-null      float64
 13  reviews.numHelpful    34131 non-null  float64
 14  reviews.rating        34627 non-null  float64
 15  reviews.sourceURLs    34660 non-null  object
```

```
 16  reviews.text           34659 non-null  object
 17  reviews.title          34655 non-null  object
 18  reviews.userCity       0 non-null      float64
 19  reviews.userProvince   0 non-null      float64
 20  reviews.username       34658 non-null  object

Total product names : 28332
Total unique product names : 65

Total product ASINs : 28332
Total unique product ASINs : 65
```

Counting the unique ASINs and product names I found out that there are 65 unique products. Examples of some product names are :

```
product_names : ['AmazonBasics AAA Performance Alkaline Batteries (36 Coun
t)'
 'AmazonBasics Nylon CD/DVD Binder (400 Capacity)' 'Amazon Echo ,Äì White'
 'Amazon Echo Show - Black' 'Echo Spot Pair Kit (Black)'
 'Fire TV Stick Streaming Media Player Pair Kit'
 'AmazonBasics AA Performance Alkaline Batteries (48 Count) - Packaging Ma
y Vary'
 'AmazonBasics Ventilated Adjustable Laptop Stand'
 'AmazonBasics Backpack for Laptops up to 17-inches'
 'AmazonBasics 11.6-Inch Laptop Sleeve']
```

Distribution of the reviews revealed that most of the reviews were 5 star reviews:

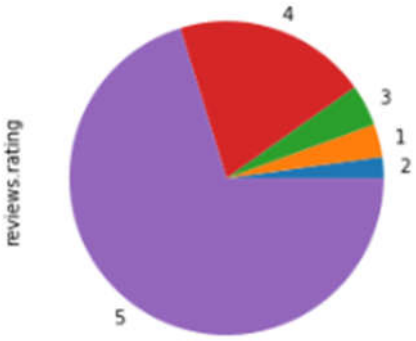

Figure 1: Review rating distribution pie chart

Review rating statistics:

```
2      616
1      965
3     1206
4     5648
5    19897
```

Next, visualizing review rating per ASIN. Bar plot shows no. of reviews. Point plot shows the rating and variance in the rating. Products with many reviews have rating with low variance. Products with few reviews have rating with high variance.

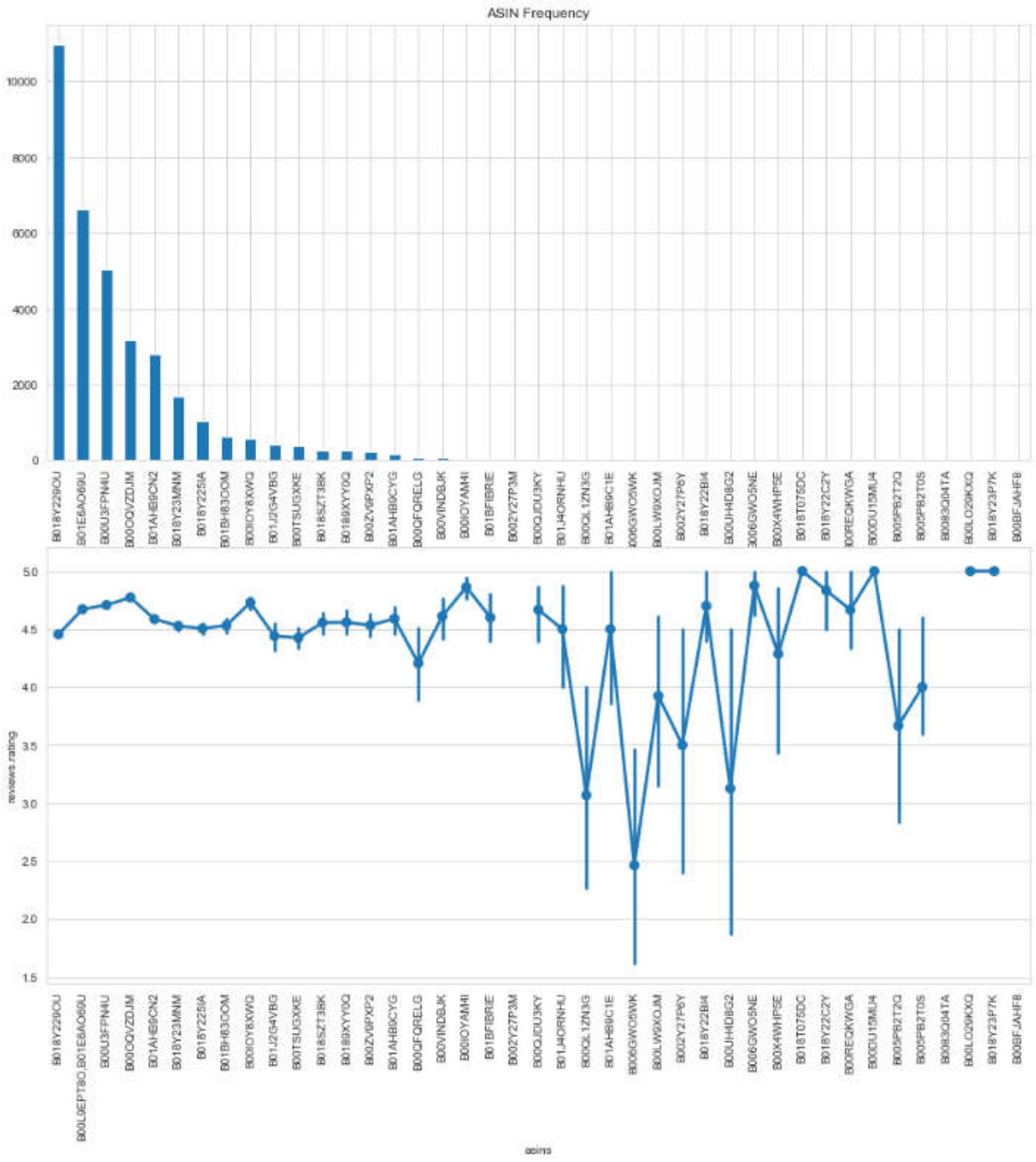

Figure 2: Visualizing review rating per ASIN

Before starting sentiment analysis, I checked for null values and found there to be none.

Approach 1: Counting positive and negative word frequency

One way of sentiment analysis is counting the no. of positive words in the text (example – good, great, A+, awesome) and counting the no. of negative words in the text (example- worst, bad, awful). Then we can assign positive sentiment if there are more positive words or we can assign negative sentiment if there are more negative words. This can be done by using positive word and negative word dictionaries. This process gave the results:

```
Positive reviews:  0
Negative reviews:  23
Neutral reviews:  28309
```

This result is unsatisfactory. This approach is not a good way of conducting sentiment analysis.

Approach 2: Sentiment Analysis based on rating

By assigning higher ratings (4, 5) as positive sentiment, rating 3 as neutral sentiment and lower ratings (1, 2) as negative sentiment, we get the results:

```
Sentiment label statistics:
Neutral      1206
Negative     1581
Positive    25545
```

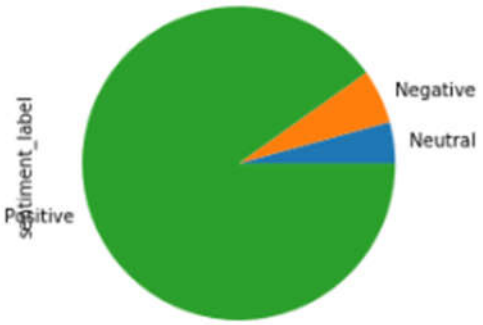

Figure 3: Distribution of positive, neural and negative sentiment in data

Which is a much better representation. Now, we can apply machine learning on the review text to build a sentiment analyzer that can predict sentiment from text.

I divided the dataset into two parts and ensured that the distribution of ratings remained same in both sets:

```
Train data size: (22665, 4)
Test data size: (5667, 4)
```

Next, I used Count vectorizer and TF-IDF and made a bag of words model.

Then, I trained some machine learning models with the data. I made the following models:

Multinomial Naïve Bayes, Linear Regression, SGD Classifier, SVC Classifier, Decision tree and Random forest classifier.

4 Results and discussion

The results acquired from different machine learning algorithms are stated below:

Table 1:
Naive Bayes classifier accuracy report:

	precision	recall	f1-score	support
1	0.80	0.03	0.05	316
0	0.00	0.00	0.00	241
-1	0.90	1.00	0.95	5110
accuracy			0.90	5667
macro avg	0.57	0.34	0.33	5667
weighted avg	0.86	0.90	0.86	5667

Table 2:
Linear Regression accuracy report:

	precision	recall	f1-score	support
1	0.76	0.54	0.63	316
0	0.61	0.11	0.19	241
-1	0.94	0.99	0.97	5110
accuracy			0.93	5667
macro avg	0.77	0.55	0.60	5667
weighted avg	0.92	0.93	0.91	5667

Table 3:
SGD classifier accuracy report:

	precision	recall	f1-score	support
1	0.81	0.55	0.65	316
0	0.00	0.00	0.00	241
-1	0.94	1.00	0.97	5110
accuracy			0.93	5667
macro avg	0.58	0.52	0.54	5667
weighted avg	0.89	0.93	0.91	5667

Table 4:
SVC classifier accuracy report:

	precision	recall	f1-score	support
1	0.79	0.73	0.76	316
0	0.72	0.30	0.43	241
-1	0.96	0.99	0.98	5110
accuracy			0.95	5667
macro avg	0.82	0.67	0.72	5667
weighted avg	0.94	0.95	0.94	5667

Table 5:
Decision tree classifier accuracy report:

	precision	recall	f1-score	support
1	0.69	0.66	0.67	316
0	0.64	0.49	0.55	241
-1	0.96	0.98	0.97	5110
accuracy			0.94	5667
macro avg	0.76	0.71	0.73	5667
weighted avg	0.93	0.94	0.94	5667

Table 6:
Random forest classifier accuracy report:

	precision	recall	f1-score	support
1	0.94	0.53	0.68	316
0	0.97	0.43	0.60	241
-1	0.95	1.00	0.97	5110
accuracy			0.95	5667
macro avg	0.95	0.65	0.75	5667
weighted avg	0.95	0.95	0.94	5667

The Random forest classifier gave the best accuracy 95%, which is a satisfactory result and a good sentiment analyzer.

5 Conclusions and recommendations

From the experiment results it can be concluded that machine learning algorithms such as Random forest, Decision Tree and SGD are highly effective in learning a sentiment analyzer from Amazon customer reviews. Recurrent Neural Networks have begun to make their way into the field of natural language processing with promising results. For future research, experimenting with building a sentiment analyzer with Recurrent Neural Network is suggested.

6 References

This is the last section of the report, prior to any appendices. The references should not be double-spaced, but single-spaced. For a technical report, use the CSE style.

[1] Wanliang Tan, Xinyu Wang, Xinyu Xu. Sentiment Analysis for Amazon Reviews.

[2] Mick Zhang. Amazon-Reviews-using-Sentiment-Analysis
https://github.com/mick-zhang/Amazon-Reviews-using-Sentiment-Analysis/blob/master/Amazon%20Project%20Github.ipynb

[3] Ashutosh Mahawar. Amazon-Reviews-Sentiment-Analysis
https://github.com/ashu20031994/Amazon-Reviews-Sentiment-Analysis