**NLP and Sentiment Analysis of Amazon Reviews**

Paul Thomas

901231546

**Abstract:**

This paper is an introduction to the basics of Natural Language Processing. It describes some of the preprocessing techniques used in NLP and why they are used. Then the paper goes on to describe uses for NLP and why it is important today in our society. Finally, the paper goes on to cover the use of NLP for sentiment analysis in python. The sentiment analysis is based on reviews from Amazon on multiple Amazon made products such as the kindle. Two separate models, VADER and RoBERTa, are used quantify the sentiment analysis score of each review. The results are compared to each other and the original ratings given by the writer of the review to analyze the accuracy of each model and how they produce each score.

**Introduction:**

I have always had an interest in AI and Machine Learning, so while researching for this project I came across the subject of Natural Language Processing (NLP). NLP is a subject that falls under the combined umbrella of Data Mining as well as Machine Learning that I personally find very interesting. Elizabeth Liddy of Syracuse University defines NLP as, “A theoretically motivated range of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications.” Just as any other subject, what we understand about NLP has massively advanced since it’s inception. Over the past few decades especially, NLP has been incorporated into many facets of our daily lives without us knowing. Some examples of NLP include email spam detection, language detection and translation and even ad placement. These advances in NLP helped by the large amount of text on the internet and advances in computer hardware, have even allowed for the generation of text by computers.

Amazon is a large corporation known for their online store and delivery services. Like most online retailers it is difficult to know exactly what you are getting when you order a product online. As such many retailers have a place where customers can review the product they have received to hopefully entice further customers to trust and purchase their products. I used two NLP methods to analyzes a set of reviews of Amazon’s physical products such as the kindle or fire stick and compare and contrast their results and how the analyze the reviews.

**Description & Analysis:**

When it comes to NLP there are common steps and methods used in different models to analyze a given text. Like other data mining and machine learning techniques first the data must be preprocessed to put the data in a predictable form. The first step in preprocessing is tokenization. A tokenizer takes in the natural language text and separates it into discrete bits of information suitable for machine learning. Depending on how complex the tokenizer is, it can split the text by whitespace, punctuation and even emojis. Unfortunately, this step works best for languages where words are separated by whitespace and sentences are separated by punctuation. Depending on the situation another step for preprocessing would be stop word removal. This step focuses on removing words that do not add meaningful data or is extremely common. However, for uses of NLP like Sentiment Analysis, removal of the stop words could massively change the sentiment of the text. Lemmatization is the process of finding the root of a word. It is similar to another preprocessing method called stemming, but stemming only chops of the beginning or end of the word to find the stem. Lemmatization would be able to tell that better is derived from good as such it is more accurate. However, with the added accuracy comes more intense resource requirement to use as it requires access to something like a dictionary. Lastly, another way of preprocessing data is called Parsing. Parsing works by taking the data and separating it into different phrases or sentences by figuring out which words naturally go together. There are many versions of parsers, some look for grammar used to separate sentences while others find a subject and separate the data by the words with a relationship to that subject. Parsing is similar to tokenization, but focuses more on grouping words together rather than completely separating them.

Once the data has been processed, NLP has multiple uses that are seen everywhere in our lives today. These uses include ad placement on websites based on your search history or words said around your phone, chatbots and virtual assistants where the response is computer-generated, auto correct on applications predicting words used or correcting grammar and spelling and much more. However, for my code I focused on sentiment analysis as a way to provide an example of NLP.

**Demo:**

The VADER model: a sentiment analysis model used to analyze text. Vader works by mapping each word to a sentiment score. Sentiment score is measured on from -1 to +1. The center, 0, is labeled as a neutral sentiment. The creators of VADER polled a large number of people to determine the sentiment score for words, punctuation, degree modifiers and capitalization. The model takes the sentiment score of each word, adds it up and then normalizes it to a value between -1 and 1. This model’s deficiency are mostly in its failure to understand sarcasm or when words that are labeled as positive are used to convey a negative tone. The VADER model represents the sentence as four normalized values, neg, neu, pos and compound. Were neg is the negative, neu is neutral, pos is positive and compound is the final sentiment score of the inserted text.

The RoBERTa Model: is a self-supervised NLP model that was trained using over 100 million different tweets to better determine sentiment analysis. Similar to VADER it provides normalized values neg, neu and pos to represent the inserted text. However, RoBERTa has an overall more accurate sentiment analysis score as it has been trained to recognize intent rather than analyzing each individual word. This also has the added effect of having a limit to the size of the text that can be analyzed.

For my data I used a data set containing Amazon reviews for different products created by Amazon such as the kindle or the fire stick. Out of the almost 1600 reviews from the data set, I chose the first 500 to initially send through my code.

After reading the first 500 tuples into python, I retrieved the review text and its rating for each tuple. After storing the tuples, they are then sent through the VADER and RoBERTa model for sentiment analysis and insert into a data frame. While each tuple is being analyzed I remove the tuples that would cause problems when it is time to show the results. The number of tuples is decreased to 260 after removing data entries where the customer did not submit a rating out of 5 or where the review was too much for the RoBERTa model to handle. After analyzing the tuples, I created a set of graphs to visualize the relationship between the VADER sentiment scores and the RoBERTa sentiment scores. Also I found the two most positive 1 star reviews based on RoBERTa and VADER and then I did the same for the two most negative 5 star reviews.

**Results:**

Based on my demo, I found that VADER ended up being more precise as it usually had a much higher neutral value than RoBERTa. While RoBERTa was better at catching sarcasm or a negative sentiment made our of positive words, it didn’t catch when there was a positive or negative sentiment towards another product in competition with the one discussed in the review. Because VADER put an emphasis on including punctation and capitalization in determining its sentiment score. However, RoBERTa ended up being more accurate even though this model bigger outliers, there are less of them overall. Even still the outliers could be attributed to the model being trained on tweets instead of product reviews. The table below maps the VADER and RoBERTa scores to one another where each point is colored based on how many stars the associated with the review.

Chart, scatter chart

Description automatically generated

Figure : Blue-1, Orange-2, Green-3, Red-4, Purple-5

**Conclusion**

Natural Language Processing is far more prevalent in the real world than I ever thought. This combination of data mining and machine learning is fascinating and will keep growing as computer become more advanced. Regarding my demo, I found that while VADER was often closer in its sentiment analysis score to the review’s actual rating, RoBERTa provided a better sentiment analysis based on the intent of each sentence. For example, when RoBERTa would give a high negative score for a 5 star review, it was often because much of the review would be negative about a previous experience with competing project as to why Amazon’s product was good. In the end enjoyed learning about NLP and wish to further my knowledge about other models and uses for NLP.

**Sources:**

<https://surface.syr.edu/cgi/viewcontent.cgi?article=1043&context=istpub>

<https://books.google.com/books?hl=en&lr=&id=72yuDwAAQBAJ&oi=fnd&pg=PR5&dq=natural+language+processing+introduction&ots=gWaP403in1&sig=PWO5Kk1LsQfoHDFj5zVjKIRnXR0#v=onepage&q&f=false>

<https://ieeexplore.ieee.org/abstract/document/8947771?casa_token=dU9kGlBK8qoAAAAA:8Tn1_wRyoKyTqgNuHKSQaibFECwdHDSrvtkDllkLX1wskVcuiaXgRjjtJgk1o5tugYf3ysSFWjkPQw>

<https://medium.com/@piocalderon/vader-sentiment-analysis-explained-f1c4f9101cd9>

<https://towardsdatascience.com/lemmatization-in-natural-language-processing-nlp-and-machine-learning-a4416f69a7b6>

<https://towardsdatascience.com/text-pre-processing-stop-words-removal-using-different-libraries-f20bac19929a>

<https://neptune.ai/blog/tokenization-in-nlp>