Aspect-based Sentiment Analysis using ML

Ishaan Buch

18BCE079 5CSEB

Nirma University

Ahmedabad, 382481

Table Of Contents

Abstract

Sentiment analysis is used by businesses to understand the polarity of the reviews and comments which helps them make decisions. Aspect based sentiment analysis divides the texts into smaller fragments in order to understand the polarity better. In this paper we will see how to perform aspect based sentiment analysis.

Key words- Aspect Based Sentiment Analysis, Natural Processing Language, Lemmatisation

1. Introduction

Sentiment analysis also known as opinion mining, is a technique that detects polarity e.g. positive or negative opinion within a given text, whether it is from a document, paragraph or sentence. Sentiment analysis uses natural processing language (NLP), text analysis, computational linguistics and biometrics to identify and study texts. Sentiment analysis is usually applied to materials like reviews and survey responses, online and social media and healthcare for applications which vary from marketing to customer service.

It is estimated that almost 80% of the data is unstructured or unorganised. Huge amounts of data is created everyday in the form of emails, chats, conversations etc and it is hard to analyse, understand and sort through all of them and it is also time consuming. Sentiment analysis helps businesses makes use of the unstructured data by automatically understanding and analysing. There are many benefits to sentiment analysis one of them is that it can analyse large amounts of data easily I an efficient and cost effective way. Sentiment analysis is also beneficial for realtime analysis and can also give us an conclusion right away.

2. Types of Sentiment Analysis

Sentiment analysis models focus on the polarity of the data i.e. positive or negative but is also focuses on feelings and emotions and even on intentions. There are many popular sentiment analysis models. Some of them are:

1) Fine grained Sentiment Analysis

2) Emotion Detection

3) Aspect based Sentiment Analysis

4) Multilingual Sentiment Analysis

3. Aspect Based Sentiment Analysis

Aspect Based Sentiment Analysis (ABSA) use a set of texts as its input for eg. product reviews, messages which discuss a particular entity. The model attempts to detect all the main aspects i.e. features of the entity and try to estimate the sentiment of the given statement. Many different methods have been used for ABSA but there is no established task decomposition for it and there are no specific evaluation methods for the tasks ABSA models have to perform.

We will be understanding the process of performing aspect based sentiment analysis with the help of an dataset and we will see the outcomes.

3.1 Dataset

We are using a dataset which consists of tweets from different accounts on twitter. The dataset is a .csv file which consists of 7 different columns. They are tweet_id, author_id, inbound, created_at, text, response_tweet_id and in_response_tweet_id. Every tweet has a unique tweet id and it is stored in tweet_id. The account name is stored in the account_id column. The column inbound states whether the tweet is the first tweet of a thread. The text column includes the text which is shared in the tweet. The date and time when the tweet is published is given in created_id. If the tweet is a response to any of the tweets in the data set then it tweet_id is stored in response_tweet_id which is the tweet id of the tweet it is in response to. And if there is another response to the tweet another tweet the it is stored in in_response_tweet_id. The text is not ready for direct processing so the next step is text cleaning.

	tweet_id	author_id	inbound	created_at	text	response_tweet_id	in_response_to_tweet_id
0	1	sprintcare	False	Tue Oct 31 22:10:47 +0000 2017	@115712 I understand. I would like to assist y...	2	3.0
1	2	115712	True	Tue Oct 31 22:11:45 +0000 2017	@sprintcare and how do you propose we do that	NaN	1.0
2	3	115712	True	Tue Oct 31 22:08:27 +0000 2017	@sprintcare I have sent several private messag...	1	4.0
3	4	sprintcare	False	Tue Oct 31 21:54:49 +0000 2017	@115712 Please send us a Private Message so th...	3	5.0
4	5	115712	True	Tue Oct 31 21:49:35 +0000 2017	@sprintcare I did.	4	6.0

Table 1: These are the first five entries of the dataset

3.2 Text Cleaning and Processing

The text contains lots of special characters and the username of the tweet handle. So the text is not ready for processing. Text cleaning is really important stage for the aspect based analysis as the text has to be in a structured format. Using the python library re, we remove the urls from the the text tweets. We use the compile function given in the re library.

After text cleaning we make a list of all the stop words in the tweet using the ntlk library. Stop words are the most commonly used words which are filtered out before or after natural language data is processed. Stop words are the words which do not add meaning to the sentence and they can be ignored without changing the meaning of the sentence. They are basically function words which are used for better sentence formation. It is not necessary to remove stop words but for performing sentiment analysis as they are present in abundance and might change the outcome of the model. Removal of stop words help in increasing the classification accuracy of the model.

tweet_id	author_id	inbound	created_at	text	response_tweet_id	in_response_to_tweet_id	textclean
1	sprintcare	False	Tue Oct 31 22:10:47 +0000 2017	@115712 I understand. I would like to assist you. We would need to get you into a private secured link to further assist.	2	3.0	I understand. I would like to assist you. We would need to get you into a private secured link to further assist.
2	115712	True	Tue Oct 31 22:11:45 +0000 2017	@sprintcare and how do you propose we do that	NaN	1.0	and how do you propose we do that
3	115712	True	Tue Oct 31 22:08:27 +0000 2017	@sprintcare I have sent several private messages and no one is responding as usual	1	4.0	I have sent several private messages and no one is responding as usual
4	sprintcare	False	Tue Oct 31 21:54:49 +0000 2017	@115712 Please send us a Private Message so that we can further assist you. Just click 'Message' at the top of your profile.	3	5.0	Please send us a Private Message so that we can further assist you. Just click 'Message' at the top of your profile.
5	115712	True	Tue Oct 31 21:49:35 +0000	@sprintcare I did.	4	6.0	I did.

Table 2: These are the first five entries of the dataset after text cleaning and we have added the cleaned text in the column textclean

3.3 Lemmatisation

With the help of the function WordNetLemmatizer() provided by the ntlk library we perform lemmatisation. Lemmatisation is the process of reducing the word to its root form to understand its actual meaning. For example if the word "better" is used in the text then it is changed to its root word "good" as it helps in the analysis of the words. It is similar to stemming but in the case of lemmatisation the context of the sentence is also taken into consideration which might help in deciding the root word. Lemmatisation is essential in understanding the meaning of the sentence but its only disadvantage is that it is very time consuming. Lemmatisation is a slow algorithm as it derives the root word from a dictionary so it takes time in traversing through it.

3.4 Aspect Mining

After all the process the next step is to identify the aspects of the text. Aspects are the subjects on which the texts are based. During aspect mining we identify the aspects of each tweet.

Table

With the help of the space library we are able to identify the nouns which are supposed to be the subjects of the sentence. Spacy helps us identify all the nouns, verbs, adjective and adverb. We consider the nouns to be the aspects of the text and we store by creating a new column with the name "Aspects".

For eg, for this text "Please send us a Private Message so that we can further assist you. Just click 'Message' at the top of your profile" we get the aspects "message" and "profile".
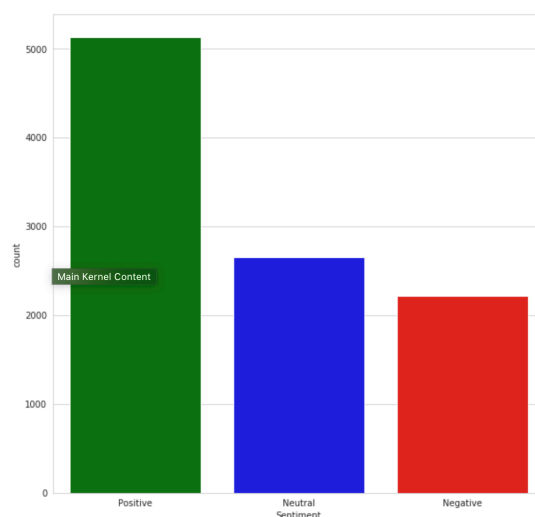
3.5 Sentiment Analysis

VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, and works well on texts from other domains. With the help of this tool we are able to figure out the polarity of the text. After determining the polarity of the text, if the polarity is greater than or equal to 0.05 then we consider the text as positive, if it is less than -0.05 then we consider it as negative and if it is between that range then the sentiment is considered Neutral.

reated_at	text	response_tweet_id	in_response_to_tweet_id	textclean	Polarity	Sentiment	Aspects
ue Oct 1 2:10:47 0000 017	@115712 I understand. I would like to assist you. We would need to get you into a private secured link to further assist.	2	3.0	I understand. I would like to assist you. We would need to get you into a private secured link to further assist.	0.6369	Positive	[link]
ue Oct 1 2:11:45 0000 017	@sprintcare and how do you propose we do that	NaN	1.0	and how do you propose we do that	0.0000	Neutral	[]
ue Oct 1 2:08:27 0000 017	@sprintcare I have sent several private messages and no one is responding as usual	1	4.0	I have sent several private messages and no one is responding as usual	-0.2960	Negative	[messages, one]

Table 3: These are the first three entries of the dataset and we have determine the polarity and the sentiment of those texts

3.6 Result

After performing sentiment analysis on the texts we come across the the dataset is dominated by positive tweets and the number of neutral and negative tweets are fairly close to each other.

4. Conclusion

Aspect Based Sentiment Analysis (ABSA) is one of the techniques used to perform sentiment analysis and it identifies certain keywords from the texts by dividing the texts into smaller parts in order to efficient identify the sentiment of the text. Various methods can be used to identify the sentiment, it may be clustering, classification etc. Many different algorithms can be used and each have their own accuracy. We used the Vader tool to perform sentiment analysis whereas many algorithms like random forest classifier etc can be used too.