ASPECT-BASED SENTIMENT ANALYSIS

Tarushi Gupta,[1] Saumya Chauhan[2]
tarushi2506@gmail.com[1]
saumyachauhan987@gmail.com[2]
Department of Electronics and Communication Engineering,
IGDTUW, New Delhi

Vivekananda Journal of Research
October, 2021, Vol. 10, Special Issue, Pg No. 35-41.
ISSN 2319-8702(Print)
ISSN 2456-7574(Online)
Peer Reviewed Refereed Journal
© Vivekananda Institute of Professional Studies
http://www.vips.edu/vjr.php

ABSTRACT

Sentiment Analysis is the task of identifying sentiments like positive, negative, or neutral in a text using natural language processing techniques. It is used in multiple domains like medical, education, media, and entertainment, corporate businesses, etc to help discern emotions reflected in a piece of text. The text could include medical reports, student feedback, reviews of customers, emails of employees among a few. The advances in technology have led to the use of machine learning models and AI software to perform the task of sentiment analysis. The problem is that these models generally consider the entire piece of text to report an overall sentiment, giving a compound polarity of either positive, negative, or neutral. It does not help in identifying the cause for such sentiments.This project focuses on studying a novel approach of aspect-based sentiment analysis. The approach has been explained and illustrated its use in modern world applications by taking an example of finding the aspects of concern from the Amazon reviews collected from an open social media platform, Twitter.

1. INTRODUCTION

Sentiment refers to a view or opinion that is held or expressed. Sentiment Analysis is the process of detecting and understanding the sentiments expressed majorly in textual language via natural language processing techniques. It finds use in a diverse set of sectors like finance, retail, education, marketing, medical, media, etc. Over the years, various studies have been conducted to experiment with natural language processing techniques and machine learning models to extract sentiments from text data. A basic task in sentiment analysis is to identify and classify the polarity of a sentence or document into positive, negative, or neutral. These classifications can be extrapolated to advanced emotion terms like “happy”, “sad”, “angry” which find use most prominently in social media text analytics and hate speech detection.

Here only subjective comments are to be studied. The most common method used to identify opinions in a text is to classify the text as subjective or objective. Subjective text is based on personal opinions, interpretations, points of view, emotions, and judgments. Objective text is fact-based and doesn’t involve personal opinions/emotions. Example of a subjective sentence: “This product has been the best purchase of my life.” Example of an objective sentence: “The maximum retail price of this product is $100.” To determine the subjectivity score of the reviews scraped from the AmazonHelp Twitter page, the python library called TextBlob has been used which returns a subjectivity score. The classification of the score is as objective, neutral, and subjective [-1.0,1.0]. 1.0 represents very subjective. According to Liu, subjective/objective text identifications find their applications in business, advertising, sports, and

social science like online review classification, stock price prediction, social media analysis, document summarising, email analysis, etc.

Existing approaches to sentiment analysis can be grouped into 3 main categories: knowledge-based, statistical methods, and hybrid approaches [2]. Knowledge-based techniques are used for classifying text using unambiguous emotion words like happy, sad, afraid, bored [3]. Statistical methods leverage elements from machine learning like latent semantic analysis, support vector machines, "bag of words", and deep learning. Hybrid approaches leverage both machine learning and elements from knowledge representation such as ontologies and semantic networks in order to detect semantics that is expressed in a subtle manner. While these approaches employ computer-based models and established techniques, they often fail to capture the mixed sentiments that can be represented in a sentence. For example, in the sentence: "The movie had a great storyline, but the cinematography was so poor it made me dizzy". It can be classified as either positive, negative, or neutral depending on the model used as the storyline is good but the cinematography is bad. To effectively capture the aspects of the sentence and the specific sentiments referring to them, a more sophisticated approach of aspect-based sentiment analysis is used.

In this paper, an attempt to use this advanced approach to sentiment analysis to improve the customer experience review for Amazon's customers by extracting the aspects and their sentiments has been made. Aspect-based sentiment analysis is the task of identifying the fine-grained opinion polarity towards a specific aspect associated with a given target. A feature or an aspect is an attribute or component of an entity, like "price", "quality", "customer experience" etc. This approach provides the advantage of capturing the nuances of objects of interest.

2. METHODOLOGY

Data Collection

The dataset for training and analysis was scraped from Twitter using the Twitter RESTful API. The Tweepy library was used for the extraction of tweets using python language. The steps involved in the process were building an application to gather the following credentials of a developer account.

- consumer_key<-' xxxxxxxxxxxxxxx'
- consumer_secret<-'xxxxxxxxxxxxxx'
- access_token<- 'xxxxx-xxxxxxx'
- access_secret<-'xxxxxxxxxxxxxxxxxxxx'

Data Processing Steps

The data preprocessing is performed to clean the data and make it uniform. The extracted tweets are then preprocessed for sentiment analysis. The preprocessing of data was performed after loading the data into a dataframe and involved lowering the case of all tweets, removing URLs, hashtags, emojis, extra spaces, HTML tags, removing stop words, lemmatization, and tokenization.

POS Tagging

Tagging is the process of assigning a tag to a word in a corpus. The process of classifying words

into their parts of speech and labeling them accordingly is known as part-of-speech tagging, POS-tagging. There are various libraries and tools available for the task of POS-tagging, the most basic being the nltk library. The TextBlob library is used for generating these tags. Some of them are represented in the table.

Table 1: Part of Speech Tags Set used in the TextBlob library.
(taken from - https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html)

POS Tag	Description	Examples from the English Language
JJ	Adjective	great, new, high, big, local
NN	Noun, singular, or mass	home, order, price, time
NNS	Noun, plural	deliveries, services
VB/VBD	Verb/ Verb, past participle/ Verb, past tense	is, make, said, agreed, placed
RB	Adverb	already, really, still, early, now, yet
IN	Preposition or subordinating conjunction	with, in, over, of, by

Aspect Term Extraction

Aspect terms are nouns that can be extracted using a simple loop and the TextBlob library. After extracting the nouns and saving them in a separate column, we ran the Counter() function from the collections package/library to find the frequency of each aspect term in sorted order. From this, the top 10 aspect terms were selected and based on domain knowledge 5 aspect keywords were extracted. These are the aspect terms that will be focussed on for aspect-based sentiment analysis.

They were identified as - *customer service*, *delivery*, *account*, *time*, and *order*.

Sentiment Analysis

Natural Language Processing (NLP) is a unique subset of Machine Learning which cares about real-life unstructured data. Although computers cannot identify and process the string inputs, libraries like NLTK have found a way to process string mathematically. Twitter is a platform where most people express their feelings towards the current context. As humans, we can guess the sentiment of a sentence whether it is positive or negative. Similarly, we developed a simple Twitter Sentiment Analysis supervised learning model using python and NLP libraries.

VADER (Valence Aware Dictionary for Sentiment Reasoning) is a model used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion. It is available in the NLTK package and can be applied directly to unlabelled text data.

VADER sentiment analysis relies on a dictionary that maps lexical features to emotion intensities known as sentiment scores. The sentiment score of a text can be obtained by summing up the intensity of each word in the text.

The NLTK library attributes the polarity after performing Tokenization on a given string of tweets. Tokenization is the process of breaking down words into smaller tokens. The lexicon contains sentiment scores that are summed up to give the polarity of the tweet. VADER's SentimentIntensityAnalyzer() takes in a string and returns a dictionary of scores in each of four categories:

- negative

- neutral

- positive

- compound (computed by normalizing the scores above)

sentiment_analyzer_scores("@AmazonHelp It's received today itself.....Thanks for the prompt response.....got best of services as always")

Polarity----------{'neg': 0.0, 'neu': 0.326, 'pos': 0.674, 'compound': 0.735}

Sentiment Metric	Score
Positive	0.674
Neutral	0.326
Negative	0.0
Compound	0.735

Figure 1: Sentiment Metric for a sample tweet from the collected data.

The Positive, Negative, and Neutral scores represent the proportion of text that falls in these categories. This means our sentence was rated as 67% Positive, 33% Neutral, and 0% Negative. Hence, all these should add up to 1.

The Compound score is a metric that calculates the sum of all the lexicon ratings [4] which have been normalized between -1(most extreme negative) and +1 (most extreme positive). In the case above, lexicon ratings for *thanks* and *best* are 1.9 and respectively 3.2. The compound score turns out to be 0.735, denoting a very high positive sentiment.

1. **positive sentiment:** `compound` score >= 0.05
2. **neutral sentiment:** (`compound` score > -0.05) and (`compound` score < 0.05)
3. **negative sentiment:** `compound` score <= -0.05

Figure 2: Sentiment classification based on Compound Score.

OBSERVATIONS AND CONCLUSION

The sentiment analysis performed using the VADER library on the initial scrapped data shows approximately 40% neutral sentiment, 32% positive sentiment, and 28% negative sentiment. The distribution in Figure 3. helps in identifying that, tweets reflecting positive and negative sentiments don't differ by a large percentage (only 4%) and the primary goal for any business is to work on converting the negative sentiments to neutral and positive in order to minimize their percentage. While this distribution gave an overall idea of sentiments, it is critical to understand which aspects are being disliked by the public. This is where aspect-based sentiment analysis comes in.

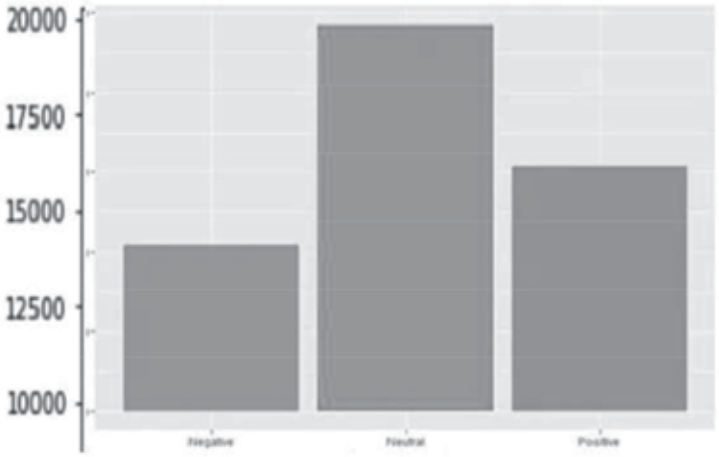


Figure 3: Bar chart representing the count of tweets reflecting the negative, neutral, and positive sentiment

After the extraction of aspect keywords, a plot of their distribution in tweets was plotted (shown in Figure 4.). The plot shows the distribution of different aspects identified.

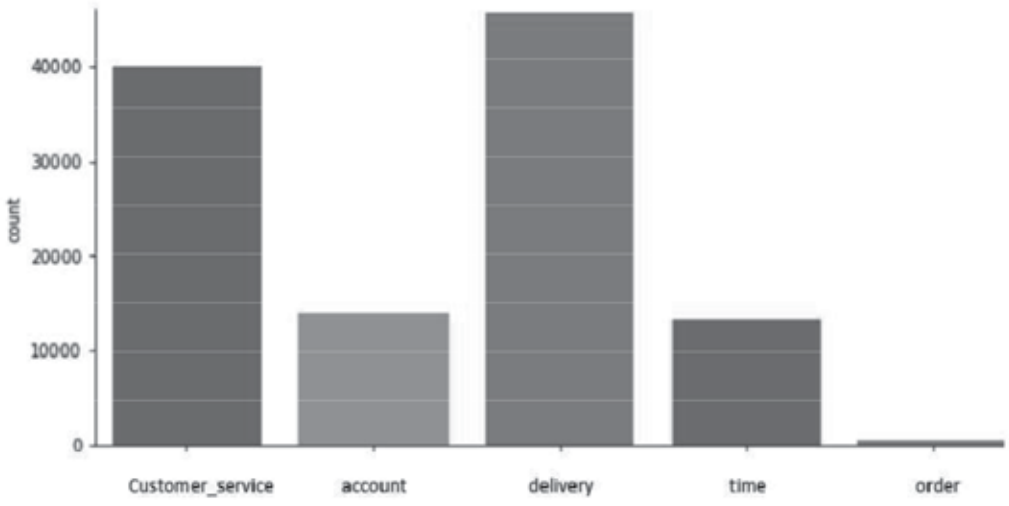


Figure 4: Bar chart representing the count of tweets reflecting the identified aspect keywords.

After identifying the aspect keywords, performing sentiment analysis on the adjectives of these aspects produced the following results which are plotted in Figure 5. It shows the sentiment polarity for each of the aspects. While customer service shows both positive and negative sentiments, people spoke negatively about the time aspect of Amazon services. Further investigation into these aspects will help in analyzing the root cause for negative sentiments which an organization can then work on improving.

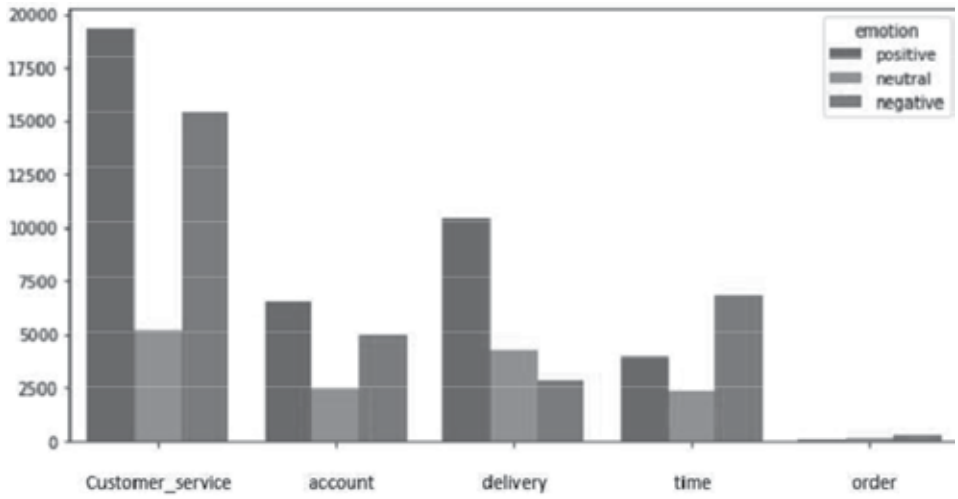


Figure 5: Bar chart representing the count of tweets reflecting the polarity of the individual aspect keywords.

The applications and study of opinions and feedback on social media are vast and this study has tremendous future scope. Machine learning can be used for a deeper analysis of the customer psyche. The various techniques can then be contrasted and compared to identify the most efficient method for conducting a comprehensive customer analysis of a product or a company, based on reviews.

REFERENCES

[1] *Liu, Bing (2012-05-23). "Sentiment Analysis and Opinion Mining". Synthesis Lectures on Human Language Technologies.* ***5*** *(1): 1—167. doi:10.2200/S00416ED1V01Y201204HLT016. ISSN 1947-4040.*

[2] Cambria, E; Schuller, B; Xia, Y; Havasi, C (2013). "New avenues in opinion mining and sentiment analysis". IEEE Intelligent Systems. ***28*** *(2): 15—21. CiteSeerX 10.1.1.688.1384. doi:10.1109/MIS.2013.30. S2CID 12104996.*

[3] Ortony, Andrew; Clore, G; Collins, A (1988). *The Cognitive Structure of Emotions* (PDF). Cambridge Univ. Press. Archived from the original (PDF) on 2015-11-23.

[4] Silvana Aciar(2010), Mining Context Information from Consumer'sReviews,http://ids.csom.umn.edu/faculty/gedas/cars2010/Aciar -CARS-2010.pdf

[5] A. Gural Vural, B. Barla Cambazoglu, Pinar Senkul (2012), Sentiment Focused Web Crawling, CIKM'12, October 29–November 2, 2012, Maui, HI, USA, ACM Transactions on the web, DOI: 10.1145/2396761.2398564

[6] Zhang Hailong, Gan Wenyan, Jiang Bo(2014), Machine Learning and Lexicon Based Methods for Sentiment Analysis: A Survey, Published in: 2014 11th Web Information System and Application Conference, DOI: 10.1109/WISA.2014.55

[7] Abhishek Kaushik, Anchal Kaushik, Sudhanshu Naitani(2015), A Study on Sentiment Analysis: Methods and Tools, International Journal of Science and Research (IJSR), Volume 4 Issue 12, December 2015

[8] Joseph Mei, Richard Frank (2015), Sentiment Crawling: Extremist Content Collection through a Sentiment Analysis Guided Web-Crawler, IEEE/ACM International Conference,2015, DOI: 10.1145/2808797.2809373

[9] Shivaprasad T K, Sentiment Analysis of Product Reviews: A Review, ICICCT 2017

[10] vaderSentiment Github Repository at https://github.com/cjhutto/vaderSentiment/blob/master/vaderSentiment/vader_lexicon.txt