



Sentiment analysis for Product Reviews

Anil B.

School of Computing and Information Technology,
REVA University, Bangalore. India

Aditya Kumar Singh

School of Computing and Information Technology,
REVA University, Bangalore. India

Aman Kumar Singh

School of Computing and Information Technology,
REVA University, Bangalore. India

Aman

School of Computing and Information Technology,
REVA University, Bangalore. India

Surekha Thota

School of Computing and Information Technology, India
REVA University, Bangalore. India

Abstract: Sentiment analysis or opinion mining is the study of extracting opinions and providing polarity to the pieces of text using DMT and NLP techniques respectively. Nowadays internet is used as a source of learning, getting reviews for various products or services, getting ideas. Millions of reviews are generated on the internet every day for a product. Because of the huge number of reviews, it is very difficult to handle and understand the reviews. Sentiment analysis is the research area which is used to extract the opinion from given review and classifying the polarity of the opinion using the process of NLP, computational linguistics, text analytics. There are many algorithms which are used to tackle NLP problems. In this paper, we have discussed different methods and we will concentrate on Vader for classifying and analyzing the reviews. We have collected the data for product reviews from Amazon, Flipkart, and Snapdeal. Perform the sentiment analyse on the reviews related to a particular product, compute the polarity score and finally provide a single line review of the product that helps the customer to easily decide whether to purchase the product.

Keywords: sentiment, analysis, DMT, NLP

I. INTRODUCTION

Sentiment is a thought, attitude or judgment that is expressed by feeling. Sentiment analysis is also known as opinion mining which studies people sentiment towards an entity. Sentiment analysis is the process of understanding and classifying the sentiments defined. The Internet is the best place for getting the data for sentiment analysis. From a user point of view, people express their views on various social media through blogs, forums or social networking sites. As we are concentrating on product reviews the user post their views about a certain product on Amazon, Flipkart, Snapdeal in the form of reviews. Per day millions of reviews are posted on social media or online shopping website, classifying and analyzing these reviews is a very difficult task to perform. We are using Vader lexicon which is a part of Natural language processing which is used to determine whether the sentiment if the review is positive, negative or neutral. Sentiment analysis does not depend on any platform or domain. Sentiment analysis is not only restricted to the product review or twitter tweets and can be used in social media network, healthcare, management and can be used by various organizations for their growth. Opinion mining has two methodologies namely sentiment analysis and sentiment classification.

There are several flaws that hinder the process of sentiment analysis, some general flaws are people freely post their content so the quality of opinion cannot be guaranteed, online spammer post spam on websites some spams are meaningless at all while others are irrelevant, known as the fake opinion, the Ground truth is also not available. Ground truth refers to a tag of a certain opinion indicating whether it is positive, negative or neutral.

Vader is a lexicon and sentiment analysis tool, vader.txt is a text file which is a data set used to give both polarity and intensity of sentiments which will solve the problem of ground truth.

Manually creating sentiment lexicon is very intensive and error-prone process so it is no wonder the researcher and practitioner rely heavily on existing lexicon as a primary resource. Vader.lexicon.txt contains words and emoticons, each of which has been assigned with a polarity score, which will decide it is positive, negative or neutral.

We have designed a web-based application which works on this sentiment analysis process. Our aim is to provide the genuine review of the product from certified buyers. we have provided the users with two options, first the user can select the product from our database and if the product is not present in our database the user has the option to copy paste the URL of product, using the URL we will do web scrapping and fetch

the reviews of that particular product, and we will analyze and prompt the result with our generic review and the percentage of acceptance. If at all the polarity of the service review is in negative we will voluntarily inform by sending an email to the respective seller as well as the website hosting that seller about the negative performance.

II. LITERATURE SURVEY

Sentiment analysis is done in two ways, one is machine learning and other is lexicon based approach. In lexicon based approach, includes dictionary based approach, in this there's a predefined dictionary which contains the polarity for each word. Here in the analysis every word is taken into account and polarity is calculated by using dictionary included. Biggest disadvantage of dictionary based approach is sentiment polarity is calculated based on words but not on the context level [1], which doesn't completely satisfy for the analysis.

In Corpus based approach, which comes under lexicon based approach, this approach solves the problem of finding opinion words with context orientation. This method depends on the syntactic patterns which occur along side of opinion words [3]. Syntactic patterns are defined with the constraints like AND, OR, EITHER, HOWEVER, BUT and other such words.

Statistical approach, as it uses statistical techniques, used to find co-occurrence pattern or opinion words. Klenner and Fahrni [4] proposed of extracting posterior polarity using the adjectives occurrence. Bose and Hue [6] considered as a fact that the reviews will be in completely random style as the customers were from different backgrounds.

Semantic approach, values are computed on the principles of similarity between the words. This method is presented by Vossen and Marks [2], which analyses the relationship between the actors and subjects by the description of verbs, nouns and adjectives.

Natural language processing techniques with lexicon approach, this is used to find the syntactical structures and help in finding semantic relationships. Park and Min [5] proposed on their paper that they used these NLP techniques to recognize time and tense expressions using few other mining techniques. They captured time expressions related to use of products and with time period in which they purchased.

Romeo and Moreo [7] used lexicon based approach sentiment analysis algorithms after using Natural Language Processing techniques. They used specific modules of taxonomy lexicon which are specifically used for news sentiment analysis for the analysis of customer opinions on the topics of news items.

III. METHODOLOGY

The proposed methodology has below important steps.

1. Database of products reviews
2. Sentiment Sentences extraction and POS tag:
3. Polarity Score computation:
4. Web scrapping

1. Database of products reviews :

Our aim was to provide flexibility to the customers, so that they can choose the reviews specified either on Amazon,

Flipkart or Snapdeal websites. Hence gathering of reviews from these websites plays a crucial role. Using web scrapping, we have gathered the product reviews that belong to 5 major categories like books, laptop, mobile, wristwatch, and TV. These reviews are maintained in a database for faster access. Number of product reviews that are stored in the database is shown in Figure 1.

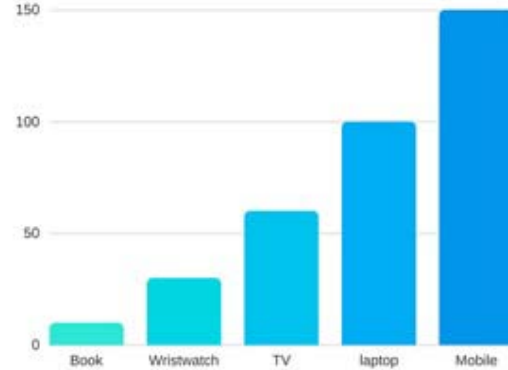


FIGURE 1 NUMBER OF REVIEWS IN DATABASE

Each review we extracted contains the following information:

1. Reviewer name
2. Product id, product name, and product images
3. Date of the review posted
4. Rating
5. Review text

For the particular product, each of the retrieved review undergoes sentiment extraction and POS tagging.

2. Sentiment Sentences extraction and POS tag:

We consider a sentence as sentiment sentence if it contains at least one positive or negative word. All the word present in the sentence is first tokenized into separate English words.

Every word in the sentence have some role which describes how the word is used, these roles are known as part of speech i.e., verb, noun, pronoun, adjectives, adverbs, preposition, conjunction and interjection. POS Tagger plays a very important role in sentiment analysis because of the following reason:

1. Noun and pronoun do not have any sentiment usually
2. It is also used to distinguish the word that can be used in different POS

3. Polarity Score computation:

Using Vardar lexicon, we will assign a polarity to each of the tokenized English words. Then it will calculate the polarity for the whole review text, which we consider as sentence polarity. The sentence polarity can either be positive or negative. The range of positive polarity is between 0 to +1 and range of negative is between 0 to -1. Now we will calculate the compound polarity using the mathematical formula:

$$CP = \frac{\sum \text{Sentence polarity}}{\text{Number of Sentences}}$$

This compound polarity is used to depict the bar graph present in our result page. Figure 2 depicts how bar graph varies based on CP.

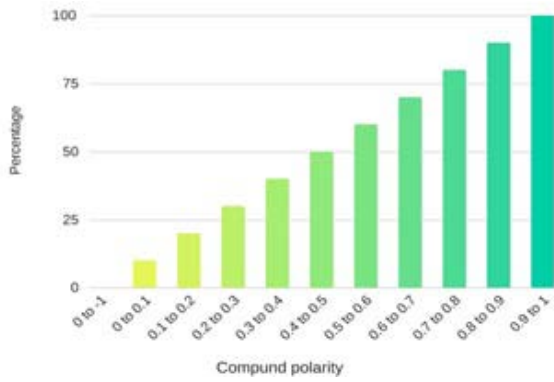


FIGURE 2 CP AS BAR GRAPH

Our resultant rating will rate the product between 0 to 100 percent, which is graphically represented by a bar graph containing one generic review.

4. Web scrapping:

If the product is not found in our database the user will get a prompt, where the user has to provide the URL of the product from any of the three websites namely Amazon, Flipkart, and Snapdeal. In the back end, an algorithm will scrap the reviews of the product from the provided URL. The flow chart of web scrapping is shown in Figure 3.

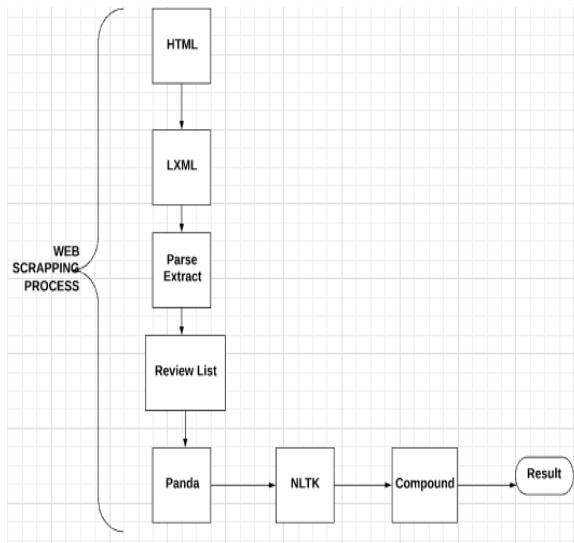


FIGURE 3 WEB SCRAPPING PROCESS

IV. System design:

We have designed our system using platforms such as:

1. Python
2. HTML, CSS
3. JavaScript, Bootstrap

For data collection, we have used webharvy tool to extract the reviews in CSVformat. These CSV files are then read by panda module which will, in turn, provide dataset, which is a collection of reviews. These datasets are the processed by NLTK Module to give a compound polarity.

We have used NLTK module of python for sentiment analysis and CGI module for web development whose flow is depicted in Figure 4

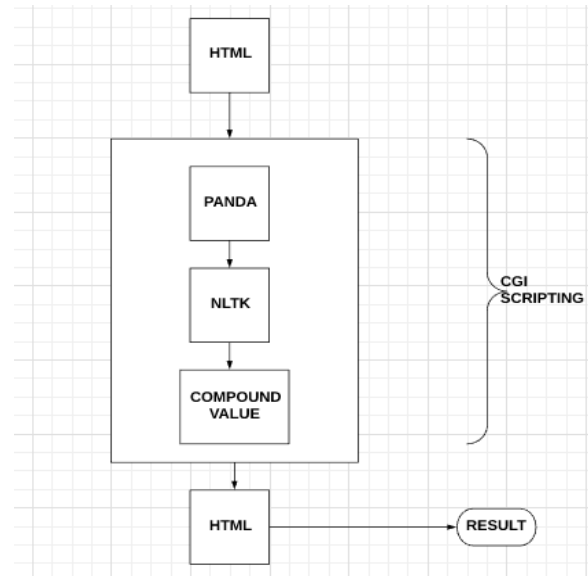


FIGURE 4 CGI SCRIPT

The overall flow of sentiment analysis is depicted in Figure 5

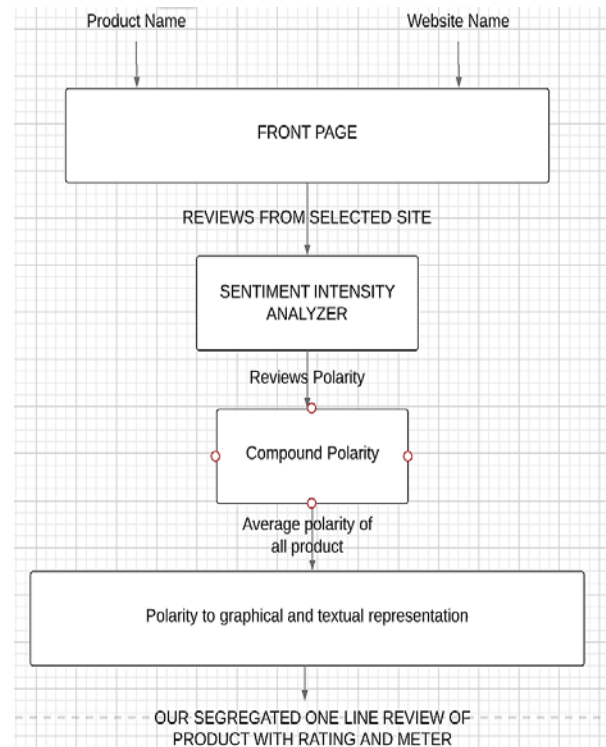


FIGURE 5SYSTEM FLOW OF PRODUCT REVIEW

V. RESULT

The main aim to develop this system is to provide the user with genuine review and to reduce the time for the user to decide whether the product is worth or not.

Advantages of using sentiment analysis for reviews:

If at all the user is going to check the reviews from many websites, it will consume more time for them to understand

and decide whether the product is worth enough for them to buy or not.

It is also not possible to read and analyze each and every review from different sites, reading few of the reviews to decide whether the product is worth or not may end up in taking a wrong decision.

By using sentiment analysis the user will get the promising result without investing much time in searching for the reviews which is depicted in Figure 6. It will save time and effort and can decide just by looking at the compound polarity and generic one line review.

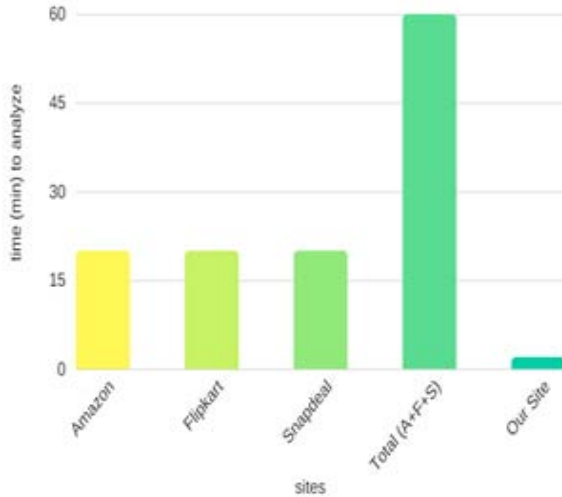


FIGURE 6 REVIEW ANALYSIS OF PROPOSED METHODOLOGY

VI. CONCLUSION

In this paper, we have conducted a schematic study on sentiment analysis on product reviews. The task of sentiment analysis, especially in the field of product reviewing is in nurture state and would take some time to become fully fledged into its domain. Right now, we have worked with only the very simplest of algorithms among many like Naive Bayes, SVM etc..., we can enhance our model by adding extra information like the closeness of the word to all its superlative degrees to make our result more precise and effective.

After analyzing some of the research domain in SA, it is believed that SA algorithms and SA techniques are still in an open field of research. The later model could be considered as reference model which can be used to compare with the proposed model.

Sentiment analysis uses information from micro-blogs, e-commerce, newsgroup etc. for market research to build a business intelligence to understand the subjective reason why the consumer is or not responding to something (ex. why the particular group of people buying only certain company product? What companies should do to tackle this situation?). Collected reviews from the different site could be the solution to this and can help companies to think in right directions.

Customers often use sentiment analysis to automatically sort their opinion, to identify the best product for their use. As customer becomes more or more automated through Machine

Learning understanding the sentiment of given case to its root level becomes very important.

As per developers and Business perspective view it creates a boom in the tech market, where developers create an understanding of public interest and business leaders try to produce 360 views of their brand products.

Our work relies on all these parameters; its aim is to give one line segregated review of the user-selected product. As with advancement in many other fields, it is believed that introduction of deep learning in SENTIMENT ANALYSIS could produce a cutting-edge application to ease the task.

REFERENCES

- [1] Qiu, Guang, et al. "DASA: dissatisfaction-oriented advertising based on sentiment analysis." *Expert Systems with Applications* 37.9 (2010): 6182-6191.
- [2] Maks, Isa, and Piek Vossen. "A lexicon model for deep sentiment analysis and opinion mining applications." *Decision Support Systems* 53.4 (2012): 680-688.
- [3] Hatzivassiloglou, Vasileios, and Kathleen R. McKeown. "Predicting the semantic orientation of adjectives." *Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the european chapter of the association for computational linguistics*. Association for Computational Linguistics, 1997.
- [4] Fahrni, Angela, and Manfred Klenner. "Old wine or warm beer: Target-specific sentiment analysis of adjectives." *Proc. of the Symposium on Affective Language in Human and Machine*, AISB. 2008.
- [5] Min, Hye-Jin, and Jong C. Park. "Identifying helpful reviews based on customer's mentions about experiences." *Expert Systems with Applications* 39.15 (2012): 11830-11838.
- [6] Hu, Nan, et al. "Manipulation of online reviews: An analysis of ratings, readability, and sentiments." *Decision Support Systems* 52.3 (2012): 674-684.
- [7] Moreo, Alejandro, et al. "Lexicon-based comments-oriented news sentiment analyzer system." *Expert Systems with Applications* 39.10 (2012): 9166-9180.