

Sentiment Analysis on Reviews of E-Commerce Websites Using WordCloud and Naïve Bayes

Suyash Lokhande¹ Anil Kumar Gupta² Sarvesh Dhakunkar³ Suyash Chopda⁴ Vedant Sharma⁵
²Senior Member IEEE

^{1,3,4,5}Department of Computer Engineering
^{1,3,4,5}Dr. D.Y. Patil Institute of Technology, Pimpri, SPPU, Pune, India

Abstract— On E-commerce websites the sentiment analysis is very much essential and advantageous. The results stored in the web are present in an ever increasing amount. Also the amount of people searching items from the web is increasing. As a result, the reviews or posts by the user are increasing with each passing day. These reviews reveal the user’s attitude towards the product, which is very helpful for customers. This helps customers to know other people’s opinions on the product of their interest. Usually, E-commerce reviews involve many aspects of products, e.g. Appearance, quality, price, logistics, and many more. In this paper, we define the aspects of products, and present a sentiment analysis approach for E-commerce reviews which involves quantification of the sentiments of thousands of reviews. And finally, providing the customer summarized data about the expressed sentiments in way of intuitive and easy to understand charts and other visualization.

Keywords: Sentimental Analysis, E-Commerce, Word Cloud, Naïve Bayes

I. INTRODUCTION

Social media has become very important in day to day life of every individual. It allows its users to express what they think and feel about things on the web, such as products in E-commerce websites. These opinions of the users are called as reviews.

Today, the average consumer depends on user reviews and opinions expressed on the World Wide Web to a great extent to find out about any product before making a decision to buy, be it a T-shirt or an automobile. Review intents to understand the mood of the writer. It may either be positive or negative towards the product. These positive or negative emotions expressed by the people are known as sentiment.

So the data from the reviews by the customer is analyzed to make the data more dynamic. This is an essential field nowadays. In this age of machine learning based algorithms going through every review to understand a product is wastage of time where we can separate a review under a particular category to understand the demand of the product among all the users.

Different website allows different review structures to be followed. Some websites allow users to write the positives and negatives specifically while some are allowed to write a summary of the product. This paper mainly focuses on E-commerce product review extraction from the E-commerce websites and performing sentimental analysis on them.

II. IMPLEMENTATION AND METHODOLOGIES

This paper concentrates on mining reviews from the E-commerce websites like amazon.com, which allows the user

to freely write their views on various products. The reviews are extracted automatically from the website. It also uses Naïve Bayes and WordCloud for performing sentiment analysis and for displaying the results.. Following are the steps proposed for the data flow of the proposed system. The different processing elements of the system are as follows:

A. User Input and Availability-

The user enters the name of the product that he desires. The product name entered by the user has to be checked whether it is available in the E-Commerce site before proceeding with extraction.

B. Feature and Review Extraction-

We take the e-commerce site’s URL (e.g. amazon.com) as the input and extract all the text from the provided webpage using various tools like Scrapy and BeautifulSoup.

C. List of Product-

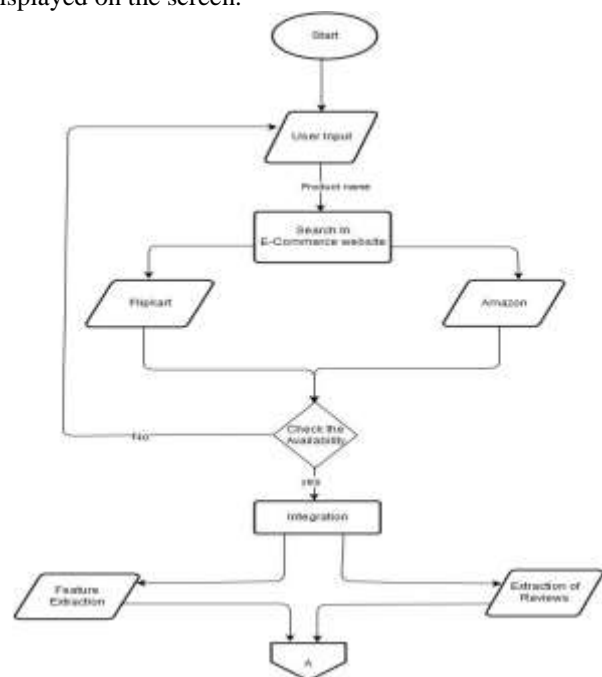
This will display a list of products from which we have to select a product of our choice to extract review.

D. Display Review List-

This generates the dynamic link and displays all the reviews of the selected product.

E. Analyzing Reviews-

Here we perform analysis on the reviews that are being extracted using tools like Wordcloud and are then results are displayed on the screen.



1) Word Cloud

A word cloud is used to visualize language or text data and is considered a weighted list, which gains expanding consideration and more application opportunities in the time of machine learning algorithms. Word Cloud has come out as a simple and outwardly engaging representation procedure for text. The Word Cloud narrows text down to those words that appear with higher frequency and are used in various contexts as a means to provide an overview of the text.

Word clouds have been contrasted with unweighted lists and other user interfaces in various studies. The outcomes demonstrate that users are on average increasingly viable in recognizing a particular term in an alphabetically ordered unweighted list than in an alphabetically ordered word cloud. However, the terms that are used frequently can be seen quickly and easily in word clouds due to their larger font sizes.

There are 3 major word cloud maps applied in social networks distinguished by their algorithm instead of appearance:

- 1) Frequency - In the frequency type, the size of font represents the number of keywords that appears in the collection. The frequency type is the most basic type used in mining text data.
- 2) Categorization-In the categorization type, the size of font indicates the number of subcategories of a collection. The categorization type is commonly used in geographical mappings. However, the categorization type can be transmitted to the frequency type with regular coding.
- 3) Mixed - In the mixed type, the data contain both frequency and categorization, which requires logically analyzing the complicated data before arranging the word cloud maps.

2) WordCloud on Reviews

The most frequently occurring words are found out in this analysis, which can give both the consumer and the designer an idea of what the users are feeling about the product or what are the key aspects of the product. The words represented in the word cloud with a set frequency can aid in highlighting the most commonly cited words in the reviews. The height of the word represents its frequency.

The word cloud of words having maximum is created by using 'WordCloud' packages for most common words used in reviews of high ratings and low ratings as shown below.

- 1) WordCloud for high rating reviews: In the figure, highlighted words like 'love', 'glad', 'fit', 'wanted', 'perfectly' paint a picture before the consumer about the product.
- 2) WordCloud for low rated reviews: In the figure, highlighted words like 'look', 'wear', 'wide', 'made' provides feedback of the consumers.



Fig. 1: WordCloud for the most frequent words used in review texts with high ratings.



Fig. 2: WordCloud for the most frequent words used in review texts with low ratings.

III. SENTIMENT ANALYSIS USING NAIVE BAYES

This concentrates on mining reviews from the websites like amazon.com, which allows user to freely write the view. It automatically extracts the reviews from the website. It also uses algorithm such as Naïve Bayes classifier, Logistic Regression and SentiWordNet algorithm to classify the review as a positive and negative review. The Fig.1, shows the data flow of the proposed system. The different processing components of the system are as follows:

A. Text Extraction

After the Login credentials, this module takes the amazon.com URL as the input and extracts all the text from the provide webpage.

B. Source Code Extractor

HTML source code of the webpage is extracted in this module.

C. List of Product

This module will display a list of products from which we have to select a product of our choice to extract review.

D. Display Review List

This module generates the dynamic link and displays all the reviews of the selected product.

E. Stop Word Dictionary

This function contains the stop word list which will be used to eliminate the stop words in the reviews.

F. Algorithm selection

This module allows the user to select any one algorithm among Naïve Bayes.

G. Calculate Performance

Once the algorithm is selected the training data is loaded and the performance of the algorithm is measured in terms of Recall, Precision and F-measure.

H. Display the Classification Result

This module displays two lists containing positive and negative review separately.

I. Positive and Negative Opinion Dictionary

This function contains the positive and negative word list which will be saved in the two separate text files and later it will be used for sentiment analysis.

J. Naïve Bayes Text Classification

The Bayesian arrangement is utilized similarly as a probabilistic strategy (Naive Bayes content classification). Utilizing suitable samples which reflect nice, terrible or impartial sentiments, same should recognize the middle of them. Basic feeling demonstrating combines a statistically based classifier with a dynamical model. Those credulous bayes classifier utilizes single expressions also saying pairs concerning illustration Characteristics. It allocates the input under nice or terrible. The unbiased classes, marks +1, -1 what's more 0 individually. This numerical yield drives a basic first-order dynamical system, whose state speaks to the mimicked enthusiastic state of the experiment's representation.

IV. CONCLUSIONS

This paper has presented the work of analyzing the product depending on its various specifications using sentimental analysis by WordCloud and Naïve Bayes. The motive behind this is to help the consumers in selecting the product of their choice. In this research, we have also presented supervised learning techniques to polarize a large amount of unlabeled product review dataset. We have found that emotion in user reviews can be seen as a meaningful additional meta-data.

REFERENCES

- [1] Florian Heimerl, Steffen Lohmann, Simon Lange, Thomas Ertl, Word Cloud Explorer: Text Analytics based on Word Clouds, 2014 47th Hawaii International Conference on System Science, 2014
- [2] Vamsee Krishna Kiran M, Vinodhini R E, Archanaa R, Vimalkumar K, User specific product recommendation and rating system by performing sentiment analysis on product reviews, 2017 International Conference on Advanced Computing and Communication Systems (ICACCS -2017), Jan. 06 – 07, 2017, Coimbatore, INDIA
- [3] Prashast Kumar Singh, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Amit Sharma, An approach towards feature specific opinion mining and sentimental analysis across e-commerce websites, 5th International Conference- Confluence The Next Generation Information Technology Summit (Confluenc), 2014
- [4] Santhosh Kumar K L, Jayanti Desai, Jharna Majumdar, Opinion Mining and Sentiment Analysis on Online Customer Review, IEEE International Conference on Computational Intelligence and Computing Research, 2016
- [5] Daniel Martens, Timo Johann, On the Emotion of Users in App Reviews, IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering (SEmotion), 2017