



SENTIMENT ANALYSIS ON PRODUCT REVIEWS

¹Prof. Swati Powar, ²Mr. Sonu Parab, ³Mr. Sahil Panchal, ⁴Mr. Siddesh Parkar,
Department of Information Technology Finloex Academy of Management and
Technology Ratnagiri, Maharashtra, India^{1,2,3,4}

ABSTRACT

In the digitalized world recommendation of any products to attract customers that meet their requirements is very crucial for the vendors to survive in the global market. Sentiment analysis is an intellectual way of identifying user's sentiments towards certain entities. Natural Language Processing (NLP) has the process of analyzing sentiments as one of its prominent fields. In today's e-commerce world sentiment analysis is crucial part because it capture the opinion of any product. The growth in e-commerce has led to increasing customer reviews about various products which are available on the internet. Reviews are not only for product but also the service given to the customer. The objective is that identifying set potential feature in the review and extracting review. The approach proposed in this paper is novel and serves as a better alternative to rate a product based on its technical specification by analyzing large number of user reviews which are the subjective information in source materials i.e. e-commerce website by applying Natural Language processing, . Several methods have been developed in recent years in order to accomplish this task. In this paper, we discuss various levels of sentiment analysis followed by comparison among different approaches to sentiment analysis.

Keywords—Sentiment analysis, product reviews, naive bayes.

INTRODUCTION

“We live LESS in real and MORE in virtual, this statement is turning out to be a reality with the rapid development of the Internet and Social media platforms and the ultimate increase in the number of social media users. With the expansion of e-commerce, the numbers of people that buy the products and review them increased. Different websites allows different review structure to be followed. Some websites allows user to explicitly write the advantageous and disadvantageous, in some cases along with the summary, in other cases there will not be any restrictions on user to write review, so that they can write however they want and express their feelings. Sometimes fake review is given on particular product.it has become possible to use sophisticated machine learning algorithms for this purpose. In this paper, we are providing solution to classify a particular review based on the data. Sentiment analysis is also called as opinion mining, which makes system to gather opinion about product.by obtaining important information from review it helps user while buying any product, and also helps best product between two. Many different techniques or approaches have been presented for analysis of sentiments in product reviews. Some of the methods are discussed in this paper. These approaches are basically categorized into machine learning based and lexicon based approaches. Approaches based on machine learning include some supervised and unsupervised classification algorithms. Lexicon based methodologies consist of dictionary based and corpus-based approaches. Sentiment analysis has a remarkable

importance in the competitive environment of the market and the dynamic needs of people. It provides meaningful information at both customer level as well as enterprise level [3]. It has many applications in business intelligence, recommender systems, managing brand reputation, politics and government intelligence. [4]

CLASSIFICATION OF SENTIMENT ANALYSIS

DOCUMENT LEVEL SENTIMENT ANALYSIS

In this classification, the sentiment is being analyzed in the whole document and is categorized as positive, negative or neutral. In this level of classification, extract that subjective text which could be inferred in the complete sentiment of the overall document. Thus for polarity classification it is supposed that the document should be focused on a single object and should contain opinion from a single holder. In this Sentiment analysis particular document like particular product when we take that review it goes through all reviews so it will find that whether it is positive or negative.

SENTENCE LEVEL SENTIMENT ANALYSIS

At this sentiment analysis the sentiments are being analyzed goes through text appearing at the sentence level. It is a considered to fined grained classification as compared to the classification at document level in this the polarity was categories only in three forms either positive, negative or neutral . This type of classification can be done in the two ways- Grammatical Syntactic Approach or semantic approach. The grammatical approach considers the grammatical hierarchy or structure of the various sentences occurring in the text by taken into consideration various parts of speech. And the latter one counts the occurrences of positive as well as

negative words in order to propose the polarity of the analyzed sentence. For example words like amazing, wonderful, worst.

1.3 Entity or Aspect Level Sentiment Classification Instead of looking at language construct document, paragraph, sentence, classes or phrases aspect level directly looks at the opinion itself. This type of classification also known as Polarity Classification. It is based on the idea that on opinion consist of a sentiment (positive or negative) and a target. It is used to analyze in sample amount of text in which every sample is being labeled as either positive, negative or neutral sample depending on the overall response received that is being expressed in that particular text.

1.3FEATURE LEVEL SENTIMENT CLASSIFICATION

Product features or attributes referred to as components are being analyzed under this classification covering the sentence level as well as document level. Opinion is being derived from the already extracted features.

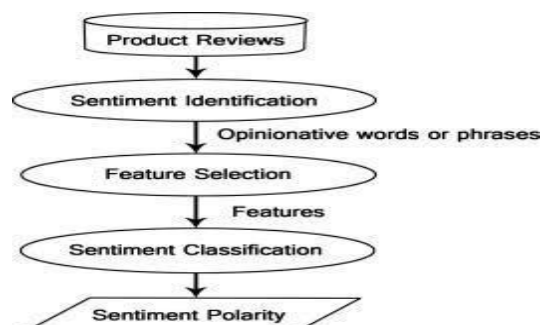


Figure 1: Sentiment Analysis process on product reviews.

RELATED WORK

In the E-commerce the classification of reviews are more important. The ongoing research work related to the Opinion mining and Sentiment Analysis are given in this section.

In [1], presented a study on classifying the document using similarity metric from balanced review sets. Several investigations on the review states that it is possible to classify the reviews based on the nature. Based on the analysis, we could conclude that reviews even though were mixed in nature; it is possible to group them into clusters as needed. It is assumed that the reviews were trust worthy and aren't fake ones. We are in the process of designing a algorithm to detect fake reviews.

In [2], Sentiment analysis and opinion mining is a field of study that analyses people's sentiment, attitude's or emotion towards certain entities. This project on sentiment analysis and opinion mining tackles from E-commerce website flip cart are selected as the input to this project. This will enable the system to analyses and then generate result (for reviews) which have been sorted according to various geographical regions.

THE PROPOSED METHOD

We propose the below methodology for solving the problem. First we perform data pre-processing for noise removal which includes discarding language stop words URLs or links, mentions, hash tags, special characters and word stemming. Once we are done with pre-processing, then we perform feature extraction using bag of words (BOW) and Term frequency- inverse definition frequency (TFIDF).

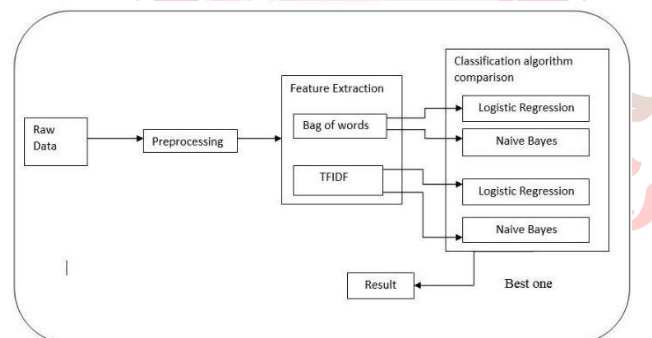


Figure 2: Proposed architecture of sentiment analysis

METHODOLOGY

We propose the above methodology for solving the problem. First we perform data pre-processing for noise removal which includes discarding language stop words URLs or links, mentions, hash tags, special characters and word stemming. Once we are done with pre-processing, then we perform feature extraction using bag of words (BOW) and Term frequency- inverse definition frequency (TFIDF). Then we use logistic regression classifier and naive bayes classification algorithms on both of the extracted features i.e. BOW and TFIDF to classify the reviews.

A. Logistic Regression

Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function. The recommended sample size for each category of independent variable is at least 10 observations per estimated parameter. Logistic regression uses large sample size which decreases the chances of errors

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k - E \quad (1)$$

Y= Dependent variables

b= coefficient of variable X X= independent variable

E= error term

B. Precision

The Precision defines as the ratio of correctly classified over number of all experimental classifications

Precision = correctly classified / (correctly classified + Errorly classified)

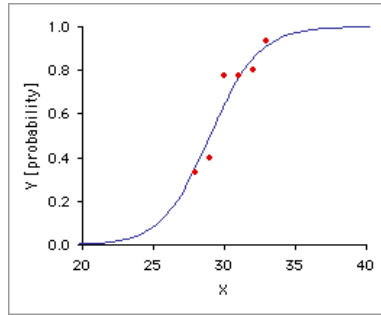


Fig. 3 Logistic Regressions

F1-Score

. F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account.

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

B. Naïve Bayes

Naive Bayes classifier

```
In [21]: from sklearn.metrics import classification_report
print(classification_report(y_test, predictions))
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.66 | 0.90 | 0.76 | 138 |
| 1 | 0.87 | 0.60 | 0.71 | 162 |
| accuracy | | | 0.74 | 300 |
| macro avg | 0.76 | 0.75 | 0.73 | 300 |
| weighted avg | 0.77 | 0.74 | 0.73 | 300 |

E-ISSN NO:2349-0721

RESULTS

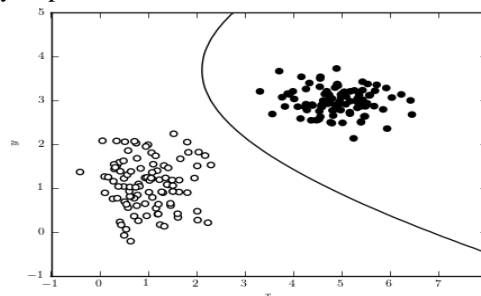
It is a classification technique based on Bayes Theorem. Naïve Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. It is easy to predict class of test data set.

$$P(c | x) = P(x | c) P(c) / P(x) \quad (2)$$

P(c | x) = posterior probability of the target class P(c) = prior probability of class

P(x | c) = likelihood which is the probability of predictor class

P(x) = prior probability of predictor



Bernoulli Naive Bayes Classifier

```
In [30]: from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred2))
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.65 | 0.89 | 0.75 | 138 |
| 1 | 0.86 | 0.59 | 0.70 | 162 |
| accuracy | | | 0.73 | 300 |
| macro avg | 0.76 | 0.74 | 0.73 | 300 |
| weighted avg | 0.77 | 0.73 | 0.73 | 300 |

Logistic Regression Classifier

```
In [27]: from sklearn.metrics import classification_report
print(classification_report(y_test, y_pred))
```

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0 | 0.71 | 0.84 | 0.77 | 138 |
| 1 | 0.84 | 0.70 | 0.77 | 162 |
| accuracy | | | 0.77 | 300 |
| macro avg | 0.77 | 0.77 | 0.77 | 300 |
| weighted avg | 0.78 | 0.77 | 0.77 | 300 |

RESULTS AND ANALYSIS

The performance of the classifications methods can be found out by using some of the following parameters:

A. Recall

The Recall is known as true positive function and defines as the ratio of correct instances classified as given class over the number of actual total in that class.

Recall = correctly classified / (correctly classified + Missed classified)

CONCLUSION

The aim of this paper is to compare and study multiple classification algorithms in on different size of datasets.

The aim is achieved by performing a comparative study of classification algorithms. After this study, we can say that all the classification algorithms performance is same when size of dataset is less (Training dataset) but, when we increase size of dataset then logistic regression classifier's performance is better than other two and it gives better accuracy

REFERENCES

- [1] Ms. Swati Powar, Dr. Subhash Shinde "Named Entity Recognition and Tweet Sentiment Derived From Tweet Segmentation using Hadoop", IEEE & CSI sponsored 1st International Conference on Intelligent Systems & Information Management Oct.2017.
- [2] Sayali Pednekar, Komal Patil, Rutweek Sawant "SENTIMENT ANALYSIS ON ONLINE PRODUCT REVIEWS" International Journal of Research Education.
- [3] Prashant Kumar Singh, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, Amit Sharma "An approach towards feature specific opinion mining and sentimental analysis across ecommerce websites", 5th International Conference- Confluence the Next Generation Information Technology Summit, (Confluence) 2014.
- [4] Anil Singh Parihar, Bhagyanidhi. "A Study on Sentiment Analysis of Product Reviews" 2018 International Conference on Soft-computing and Network Security (ICSNS).
- [5] A.Nisha Jebaseeli,, E.Kirubakarm." A Survey on Sentiment Analysis of (Product) Reviews"International Journal of computer

Applications (0975 – 888) Volume 47– No.11, June 2012.

- [6] Raheesa Safrin, K.R.Sharmila, T.S.Shri Subangi, E.A.Vimal “SENTIMENT ANALYSIS ON ONLINE PRODUCT REVIEW” in International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 04 | Apr -2017
- [7] Periakaruppan Sudhakaran, S hanmugasundaram Hariharan “Classifying product reviews from balanced datasets for Sentiment Analysis and Opinion Mining”, Joan Lu3 in 2014 6th International Conference on Multimedia, Computer Graphics and Broadcasting
- [8] Santhosh Kumar K L, Jayanti Desai, Jhama Majumdar, “Opinion Mining and Sentiment Analysis on Online Customer Review” 2016 IEEE International Conference on Computational Intelligence and Computing Research

