INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.542

Systematic Survey on Review through Confusion Matrix

Swetha S, Dr.B. Prabhakara Reddy

PG Student, Dept. of C.S.E., Bheema Institute of Technology & Science, Adoni, India

Professor, Dept. of C.S.E., Bheema Institute of Technology & Science, Adoni, India

ABSTRACT: Nowadays, most of the products are being purchased through E-commerce websites. The reviews posted on these E-commerce websites help the customer or so-called end-user to choose the right product. But going through the vast set of reviews provided by other users is time-consuming. The five-star rating system followed by most of the E-commerce websites would also fail to help the user effectively as the ratings calculated won't separate the product-based reviews and the customer service-related reviews posted on these websites. This sometimes even damages the market value of the particular product when the negative ratings produced were obtained mostly because of considering in calculation, the negative reviews related to the bad customer services provided by the particular E-commerce website. I use Sentiment Analysis upon reviews posted on the E-commerce websites so as to provide to the customers, the feasibility of choosing the right product in the best E-commerce website in accordance with the reviews given to a particular feature of the product on that website. For example, a few customers tend to buy smartphones based on the camera quality and would only want to check the public reviews posted. This proposed system classifies the reviews of the product based on its features and quality. This system also classifies product-based reviews with the reviews given to the customer service. This system does opinion mining on reviews using Sentiment Analysis and uses Confusion Matrix for the classification of the reviews.Finally showing the efficiency of the proposed approach by comparing the proposed technique with two sentiment analysis techniques.

KEYWORDS: Sentiment Analysis, Confusion Matrix, Text Mining, Bag of Words

I. INTRODUCTION

Only after 22 years after the emergence of the internet there is no longer need to leave the house for shopping. From buying a book to booking a holiday, everything is possible with internet. A lot of changes are definitely caused by Internet and mobile internet.

Nowadays, we have iPad classrooms, Albert Heijn YouTube culinary lessons, and we send WhatsApp messages to customer care instead of calling. Physical purchasing, face-to-face suggestions, and even walking across town are slowly becoming obsolete. We are in the process of transitioning to a digital economy. The way individuals buy things has changed as a result of the current economy. Consumers are making more purchases online, widening the gap between online and offline stores. These customers have less time, expect convenience, want their problems addressed, and want to be surprised (Rabobank, 2016).One of the most significant communication channels is online review. Many studies have backed up the claims that it has a positive impact on sales. Online review management is becoming more and more incorporated into commercial communication strategies. According to studies, online customer reviews have a significant impact on customer behavior.

The purpose of this study is to find out what real users think about six major aspects of mobile products: design, battery, camera, screen, sound, and performance. Semantic analysis is a technique for obtaining user feedback or opinions on a given issue. To accomplish this, we first gather information from well-known online purchasing websites.Online reviews are a simple way to ensure people make the best decision possible in a short amount of time. As a result, onlinewe are having iPod schools, cooking with YouTube videos made by Albert Heijn and sending WhatsApp messages to customer services websites and get the reviews of the users related to mobile handsets. In our proposed work we use Naive Bayes algorithm for classification of reviews into positive and negative. Further, we determine the user's opinion upon the customer service provided by the E-commerce websites, so that the user can choose the certain product from better brand and the better E-commerce website while purchasing a particular product online users' reviews of mobile devices can be found on websites.. In our proposed work we use Naive Bayes algorithm for classification of reviews into positive and negative. Further, we determine the user's opinion upon the customer service

provided by the E-commerce websites, so that the user can choose the certain product from better brand and the better E-commerce website while purchasing a particular product online.

II. SENTIMENTAL ANALYSIS ON REVIEWS

As the commercial site of the world has fully undergone in the online platforms, people are trading products through different e-commerce websites. The objective of this project is to categorize the positive and negative feedbacks of the customers over different products and build a supervised learning model to polarize large amount of reviews. A study on amazon last 8 year revealed over 88% of online shoppers trust reviews as much as personal recommendationsAny online item with a big number of positive reviews serves as a persuasive endorsement of the item's validity. Books or any other online commodity, without reviews, on the other hand, creates a sense of distrust among potential customers. Simply said, more reviews appear to be more credible. People value other people's opinions and experiences, and reading a review on a product is the sole method to learn what others think about it. Opinions gleaned from users' experiences with certain items or topics have a direct impact on future customer buying decisions. Negative ratings, on the other hand, frequently result in sales losses.Understanding client input and polarising accordingly over a big amount of data is the goal for those who want to succeed. Similar work has been done using the Amazon dataset.

Disadvantages:
- It is complex to be applied on smallerdatasets.
- Customer service related reviews are not taken intoconsideration.
- Classification is not performed by using the features of theproducts.
- This project does not include the customer service related reviews.
- This does not provide a comparisons between E-commerce websites.
.

III. PROPOSED SYSTEM

The opinions expressed in product reviews provide valuable information to online retailers about their effective gain or loss, and standing worth in the market. For the project, a dataset of Amazon Electronic product reviews is taken. To retrieve the dataset, three information retrieval techniques are used.
1. Using the tool WebHarvy web scraper
2. Crawl data by developing a code
3.Collecting data sets available open-source online.

1.Sentiment Lexicon Analysis:
We are crawling Amazon reviews dataset wherein different electronic products are rated on a scale of 1 to 5. We then develop a dictionary of our own with words which might define sentiment in a review and have some weightage. Other previous studies just usedSentiWordNet. Our study however, uses an in-house developed Dictionary. We divide the review sentence into individual tokens.

I.Sentimental score:
The sentiment score shows a review's sentiment polarity. That is, the degree of how good or bad a review is. We calculated the sentiment score by incorporating the sentiment weightage of each word in the review. The score will be in the scale of -1 to
    1. Eg:-considerthereview–goodcameraandawesomebatterylife‖.Here,–good‖and
–awesome‖, both aretheadjectivesshowingsomesentiment. Wematch the weightage
of the sentiment defining words with the words in our Dictionary, and then, add all the weightages to calculate the total score. The sentiment score is calculated by using theequation.
W= weightage of the sentiment word in a review
L= square root of length of unique tokens
r = rth review

II. Negation of Feature Word Score:
Words such as adjectives and verbs are able to convey the opposite sentiment with the help of negative prefixes. So, there is a need to check the presence of any negative word before the sentiment word and then change its polarity to get the correct score. Else, it would consider it a good review and give it a positive score. For this problem, They have implemented an algorithm as
1.Check the index number of the sentiment word in the reviewstring,
2.Store two indexes before the index of the sentimentword,
3.Check the word store in two indexes with the negative wordsdictionary,
4.If the word is present, negate the polarity of the sentiment word,

5.Repeat this step for each sentiment word in the reviewstring.
 Inaninstancereviewlike‖Thebuilt-inspeakeralsohasitsusesbutsofarnothing
revolutionary‖,thetoken‖revolutionary‖signifiesapositivesentiment.However,the prefix ‖nothing‖negatesthissentiment, thus, making it negative. Therefore, identifying negative phrases is veryimportant.

III. Spam ReviewDetection
We normalize the product ratings (1 to 5) to a scale of -1 to 1. All the reviews have their respective ratings from -1 to 1. We calculate the sentiment score of each of them, and subtract the two. The reviews wherein the difference of the sentiment score and the rating is greater than 0.5, is considered to be spam. This is considered spam because a greater difference signifies a greater inconsistency between the two. For example,thereview‖Theinterfaceisreally uglyinmyopinion.And,androidisnota unifiedexperience,liketheiPhone‖hasoriginalrating4andnormalizedrating0.5.Itssentiment score is -0.375. The difference between the two is 0.875, which, being greater than 0.5, constitutes spam. All the spam reviews deduced a re deleted from the dataset

IV. Product Feature Analysis
In this project, the features of a product are considered, eg- for a Smartphone; the features like display, camera, battery life, speakers are con sidered. Then, the respective features in a series are subdivided into sub-features.
 Eg- Display may have sub-features like pixels, size, resolution, brightness. For each review, check for sub-features or features. And hence, the reviews remaining after spam filtering are categorized on the basis of features of the product.

2. Methodology
 The methodology in this project goes with three modules as shown in the below figure. The first module goes with the Data Collection and Pre-processing of data. A largesample of online reviews is collected from the e-commerce giant Amazon.com. Dataset collected is of the mobile phones. This consists of the Feature Selection and Sentiment Analysis. In this module the NRC dictionary is used to extract eight different emotions and their corresponding valence in the text including the entire reviews. This module performs the Classification and Cross Validation techniques. It calculates the polarity using the Positive score and Negative score obtained for each word in the dataset.

Polarity = Positive score − negative score

The polarity helps in determining the accuracy of the classification at the end of the results. So, first the positive score and the negative score are calculated separately and stored and finally the polarity is calculated.
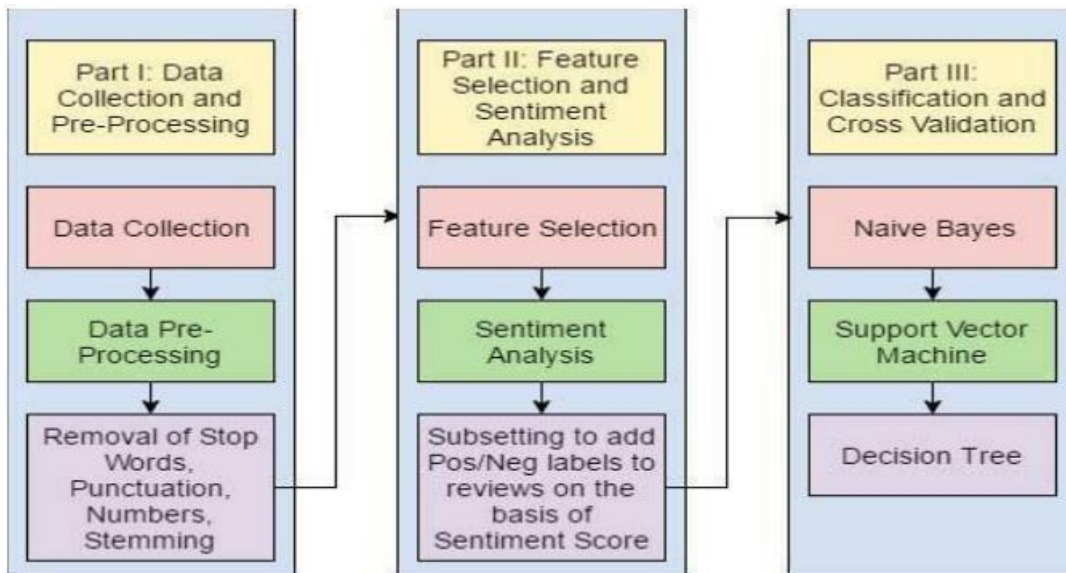

Fig 2 Framework of feature classification

3.Advantages:
- Huge data can be classified accurately using the samemethodology.
- Broad classifiation of reviews can bemade
- This project provides a methodology which produces a huge accuracy ofClassification

C. Experimental Evaluation

Filtering online informal text of product reviews may involve expletives and curse words. It requires more sophisticated methods to clean noise in raw text. We have created a _bag of words', containing around 350 words of expletives, and some other non-relevant words mentioned below. This is known as product feature extraction. It involves-

- Removal of special characters and punctuations, like (#, ^, *, etc.)
- Removal of irrelevant and malicious curse words
- Removal of repeating letters (stemming) Eg:(—happyyyy, hungryyyyl)
- Abbreviations extending As an instance, a review like —the battery life on this phone sucksl will be removed from the dataset in accordance to the abusive and vulgar nature of —sucksl.

1.Sample Reviews:

Fig 1.1 Positive review

Fig 1.2 Negative review

2. Calculation of Confusion Matrix

Fig 2.1 calculation for accuracy

Fig 2.2 Misclassfication error vs k

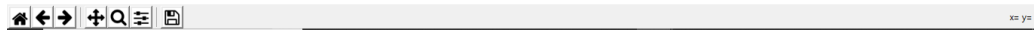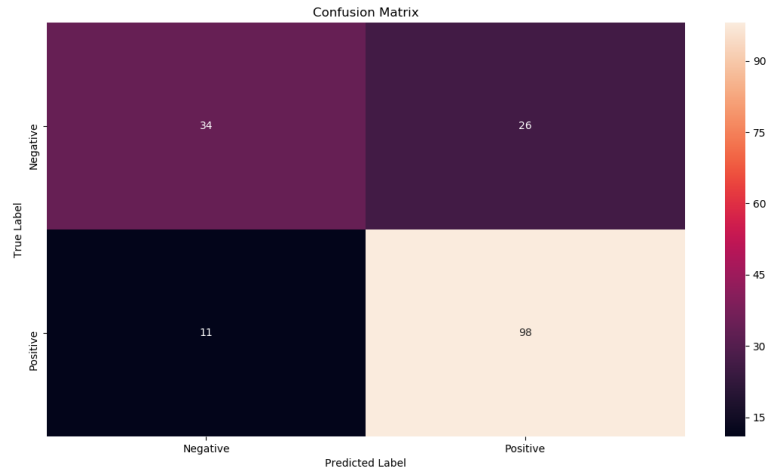Fig 2.3 caluclation for accury ,precision and recall
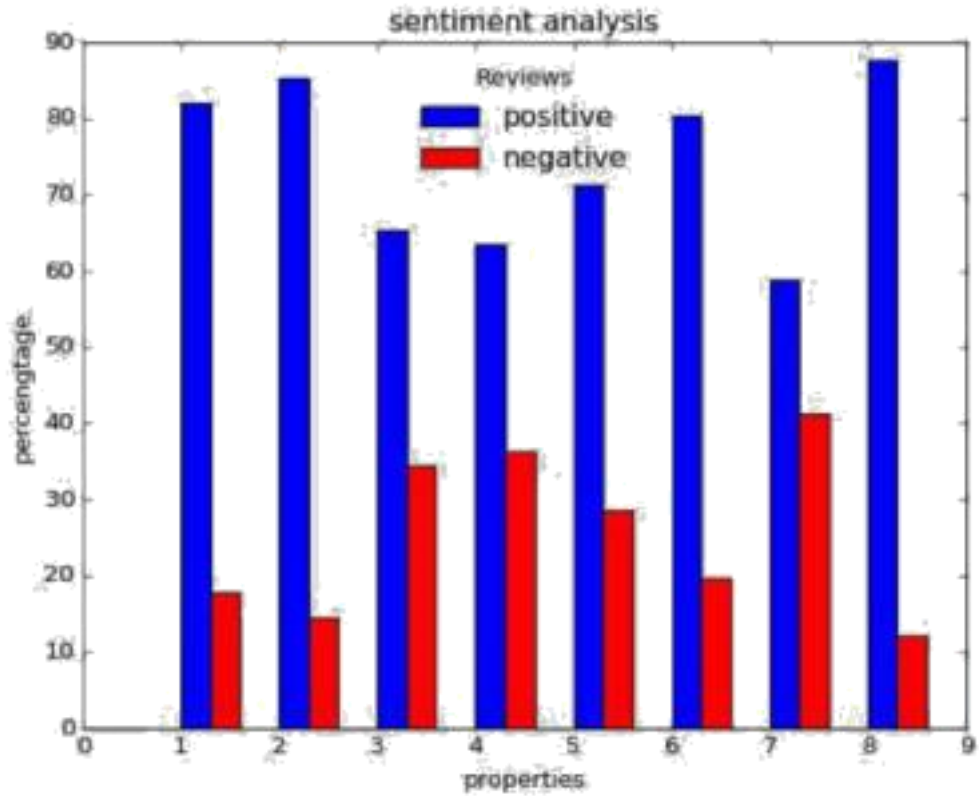
Fig 2.4 Confusion Matrix

Fig 2.6 Sentimental Analysis of positives(blues) and negatives(reds) reviews
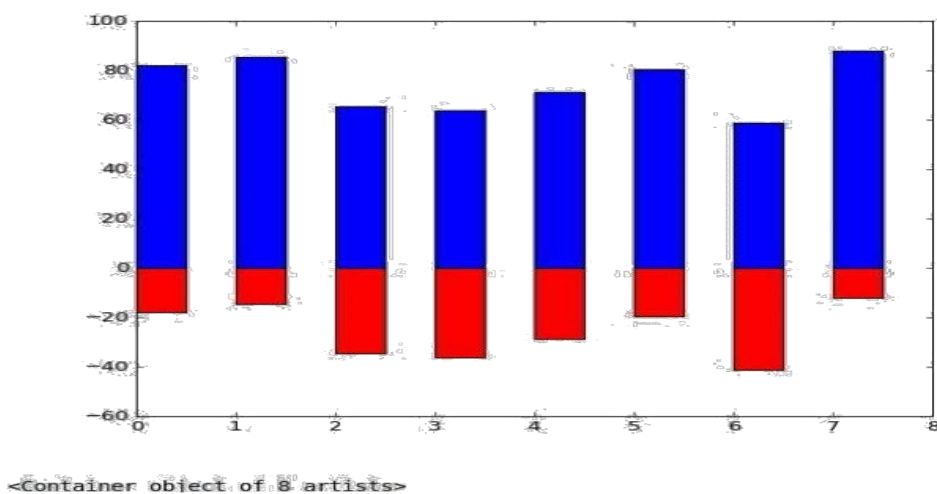
Fig 2.7 Output of positives(blues) and negatives(reds) of different mobiles

IV. CONCLUSION

As the customer reviews play an important role in the customer behavior, we have separated the product reviews according to its features so that it is made easy for the customer to choose better product brand in accordance with the reviews given to the different features of the product. Also the product quality can be analyzed. Further we propose to provide the ratings measure to such reviews for easier understanding of the customer. Also it helps to remove the certain scenario where the product quality and brand reputation is said to degenerate because of the low ratings calculated mostly due to negative reviews given to the customer service of the particular e-commerce website. Therefore, the customer service reviews should also be measured so as to produce to the customer the facility to choose the better e-commerce website

REFERENCES

1. **TanjimUlHaque, NudratNawal Saber, Faisal Muhammad Shah-** "Sentiment Analysis on Large Scale Product Reviews"-IEEE, International Conference on Research and Development,2018
2. **ZeeniaSingla, SukhchandanRandhawa, Sushma Jain** -"Sentiment Analysis of Customer Product Reviews Using Machine Learning" – IEEE , International Conference on Intelligent Computing and Control, 2017
3. **Shashank Kumar Chauhan, AnupamGoel, PrafullGoel, AvishkarChauhan and Mahendra K Gurve**– "Researchon Product Review Analysis andSpam Review Detection" -IEEE , 4th International Conference on Signal Processing and Integrated Networks , 2017
4. **Rekha, Dr. Williamjeet Singh**-"SentimentAnalysis of Online Mobile Reviews" -IEEE, International Conference on Inventive Communication and Computational Technologies , 2017, Roger S. Pressman, "Software Engineering: A Practitioners Approach", Edition - 7, 2014.
5. **GurneetKaur, AbhinashSingla**-"Sentimental Analysis of Flipkart reviewsusing Naïve Bayes and Decision Tree algorithm"- International Journal of Advanced Research in Computer Engineering & Technology (IJARCET),Volume 5 Issue1,January2016

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 **9940 572 462** 🟢 **6381 907 438** ✉️ **ijircce@gmail.com**