

Sentimental Analysis on Product using Statistical Measures

P Krishna Raj¹, Rathan Nayak N², Sanjana Kini M³, Smitha Pai J⁴

^{1,2,3,4}Student, Dept. of Computer Science and Engineering, CEC, Karnataka, India

Abstract – As online shopping becomes increasingly more popular, many shopping websites encourage existing customers to add reviews of the products purchased. These reviews given by the users will make an impact on the purchasing decisions of potential customers. At amazon.com for instance, some products receive hundreds of reviews. It is overwhelming and time restrictive for most customers to read, comprehend and make decisions based on all of these reviews. Customers most likely end up reading only a small fraction of the reviews usually in the order which they are presented on the product page. In the proposed method, opinions are classified using various statistical measures to provide ratings to help the sentimental analysis of big data. Experimental results demonstrate the efficiency of the proposed method to help in analysis of quality of product, marketer's evaluation of success of a new product launched, determine which versions of a product or service are popular and identify demographics like or dislike of product features, etc. Analysis of plenty of words coupled in a sentence represent various sentiments of users and the various experiences and impact that product has given them. This analysis compiles all structural modelling approach and Bayesian Interface system to identify the polarity of the opinion which subsequently classifies positive and negative opinions. In this paper, we present a product ranking model that applies weights to product review factors to calculate a product ranking score.

of review is classified with respect to the sentiment classes that is the polarity of reviews are deciding which are positive and negative reviews.

In recent years customer review has a big influence on users across the internet. Users tend to decide over the reviews. Because there are amount of reviews from other users, it is difficult for the user to come to a conclusion. Most of the research for sentiment analysis has focused on many commercial applications such as hotel review, movie review etc. In this project we focus on product reviews. Merchants started providing metadata for the products sold online. For customers it was difficult to make a decision about the product only with the details provided. As a solution, online merchant enabled forums, which allows customers to express their opinions and get reviews. Consumer reviews are more trusted than the description comes from manufacturers. This is because a consumer review serves to explain what the product is about and how it works. Both the good and badly written reviews matter. Consulting reviews is now a logical step in purchasing cycle of all types. There are two main problems hinder customer from fully utilizing the reviews. Explosive growth of reviews, i.e., it becomes hard for the customers to read all the reviews and he may get a biased view, if he reads only few of those reviews. Secondly, customers mainly look for the features that serve them specifically. But from the thousands of reviews it is practically impossible for the customers to identify the reviews which speak about the specific product feature.

We take the unstructured data into consideration which will be filtered to remove the noisy data and pre-processed to evaluate the sentiment of the mobile phone reviews. The proposed work will help the future buyers to make better decisions on the basis of analysis of feedback received by a particular smart phone brand. It will also allow manufacturers to meet consumer expectations better on basis of feedback received.

1. INTRODUCTION

Sentimental analysis can be seen as Natural Language Processing (NLP) task that aims to analyze opinions, sentiments and emotions expressed in unstructured data. Today sentiment analysis and opinion mining place a vital role for decision support system. The main objective of sentiment analysis is to analyze the polarity of comments by extracting the features and components of the objects that have been commented on document or text. Firstly, the data is the collected from the e-commerce store. It consists of customer reviews about products such as phone, television, camera etc. However, for our project the focus is on electronic product. The data is collected dynamically through web scrapping. Then, the polarity of the opinions must be analyzed. For instance, the reviews including "good" and "bad" in it will hold a positive and negative opinion. Secondly, for accurate result to be obtained the opinion strength should be determined. The strength of the words of the user's opinions are taken into account. For example, "good" and "excellent" indicate different levels of positive sentiment. The addition of "very" before any word can also be used to determine the strength. Finally, the classification

2. LITERATURE SURVEY

Abinaya R "Automatic Sentiment Analysis of User Reviews": proposed a dictionary based classification for accurately classifying the reviews as positive, negative and neutral. To enhance the accuracy in the classification of neutral reviews, Support Vector Machine algorithm has been implemented. Both the product owner as well as the user can identify the quality of the product based on the sentiment graph that is generated on basis of reviews for each of the product video. A comparative study of the sentiment graphs are performed in order to improve the efficiency of visual representation[1].

ZeeniaSingla "Statical and Sentiment Analysis of Consumer Product Reviews": proposed sentimental analysis to extract subjective information from the text. In this research, data analysis of a large set of online reviews for mobile phones is conducted. They have not only classified the text into positive and negative sentiment but have also included sentiments of anger, anticipation, disgust, fear, joy, sadness, surprise and trust. This description about the classification of reviews are helpful to evaluate the product as a whole, enabling better-decision making for consumers. The classification of data is very efficient this way as the accuracy of SVM after cross validation is equal to 84.87%. This approach can be extended to mine customer requirements keeping in mind the designer's concerns. So, helpfulness of reviews can be computed to provide the designer with maximum useful information enabling him to improvise the product or roll a new product in the market by meeting maximum customer requirements[2].

Shivaprasad T K "Sentiment Analysis of Product Reviews": proposed a method to extract the opinion from the given review and the analysis process including the classifications like natural language processing (NLP), computational linguistics, text analytics and classifying the polarity of the opinion. In the field of sentiment analysis there are many algorithms that exist to tackle NLP problems. Each algorithm is used by several applications. In this research work, they have shown the taxonomy of various sentiment analysis methods and also shows that Support Vector Machine (SVM) which gives high accuracy in comparison with Naïve Bayes method[3].

Rekha "Sentiment Analysis of Online Mobile Reviews": proposed a work to analyze the opinions of the user's positive and negative reviews. For this purpose they first collect the data from the famous online shopping websites and retrieve all the reviews of the users that are related to mobile handsets. In this process, they defined six common features, they are Camera, Battery, Screen, Sounds, Design and hardware/software performance, of three handsets and then find and collect related reviews as dataset for proposed algorithm. The proposed algorithm has successfully filtered the positive and negative reviews counting. Then this counting is used to get the confusion matrix results, which shows the overall result performance of the handset[4].

M. Lovelin Ponn Felciah "A Study on Sentiment Analysis of Social Media Reviews": proposed a study on huge volume of data present in the web for internet users. This survey focuses mainly on various techniques and methods that are used for classifying the opinion from social media datasets and its related future aspects. It also provides the applications and budding challenges that arise due to the sheer volume of data in the Internet. Finally, Sentiment analysis is an emerging field in research for Decision Support System[5].

Neen Devasia "Feature Extracted Sentiment Analysis of Customer Product Reviews": proposed a semantic based approach to extract product features. An algorithm, which employs typed dependencies, is introduced especially for this purpose. Recursive Deep model is mainly introduced and used to identify sentiment orientation of review sentences. Each product feature is shown under a classification which is constructed and shown in review matrix which gives importance and polarity of the product. The results of this experiment shows that it is effective and has achieved the desired objective. This can be further improved by adding implicit feature extraction. Some review sentences don't give the exact associated feature term, they only include sentiment words. The product feature that do not appear in the review sentences but are actually referred to are called implicit features. In this work, sentiment from the review sentences computed solely using Recursive Deep Analyser. This can be improvised using machine learning techniques, so that the Deep Analysers can be taken as one of the feature of classification.[6].

Chhaya Chauhan "Sentiment Analysis on Product Reviews": had proposed Sentiment analysis which is a tool that can be used to analyse unstructured data that generates objective results. These techniques basically allows a computer to understand and determine the sentiment of humans. It analyses the text or speech to determine the sentiment. It can be used in product reviews to understand human sentiment towards a particular topic. The main focus of this research work is to review algorithms and techniques to form a summary of a product from an extract feature, analyse and form an authentic review. Using higher level natural language processing tasks, further work will include product reviews on websites when extracted the Naive Bayes classifier can give very good results when good features were found the results vary for all the n-grams and it has been observed that unigrams and bigrams give the best results. It was also found that feature presence was a better measure for sentiment analysis than feature frequency. The results vary for different type of techniques and depend on the kind of dataset used. Also, POS tagging and use of negation improves accuracy. Some techniques of unsupervised learning can also be taken into account for aspect level sentiment analysis. Using POS tagging we can include findings of the features and opinions and classifying the opinion based on its semantic orientation as determined by Word Net[7].

Yu Mon Aye "Sentiment Analysis For Reviews of Restaurant in Myanmar Text": proposed the creation of Myanmar sentiment lexicon for food and restaurant domain and analyses the Myanmar text reviews of customers using lexicon based sentiment analysis for the recommendation. To our knowledge, this is the first work for sentiment analysis of Myanmar text comments. The first challenge is the absence of annotated data and sentiment lexicons. In this work, they addressed the approach to sentiment analysis for Myanmar Language and generate the context-independent sentiment rules for Myanmar Language[8].

R.Nithya -“Sentiment Analysis on Unstructured Review”: proposed a classification method S based on subjectivity and polarity detection. Today when a consumer wants to buy a product the first thing, they do is check the rating and review of the product online which is easily available in online shopping site such as Flipkart, Jabong, pepperfry etc. Some of the website like reevoo, buzzillions, bizarte, Amazon and many more which discuss about positive and negative facts related to product which companies launch in the market. These types of analysis are important as well as a basic requirement for the sellers to go through market analysis, market segmentation, product penetration, branding, quality analysis and so on. Here, the proposed system classifies the most identified features using supervised learning method Naive Bayes and determined their neutral, positive and negative polarity distributions[9].

Yuwa Sawakoshi “An Investigation Of Effectiveness Of Opinion And Fact Sentences For Sentimental Analysis Of Customer Reviews”: proposed a customer review regarding the travel statistics, which provides even the smallest piece of information which makes a major influence on users in accommodation by the spread of internet. In this research, in order to fasten the picking out the information from the reviews, they proposed a simple method to classify them based on the sentences into “Opinion sentences” and “Fact sentences” using SVM. The effectiveness of “opinion sentences” is to estimate the values by classifying reviews into Negative and Positive using SVM[10].

3. PROPOSED WORK

In this project we are using machine learning along with the Support Vector Machine Classifier. In machine learning, the support vector classifier are supervised learning models with associated algorithms that analyze the data used for classification and regression analysis. In SVM model the points in space are the example for the representation and are mapped so that the examples of the separate categories are divided by a velar gap that is wide as possible. Further, new examples are mapped into that same space and predicted to belong to a category based on which side of the gap they fall. We basically, take the comma separated value (CSV) file of the particular product selected from the list given and it undergoes Corpus Generation. Corpus generation mainly consists of removing of the stop words and cleaning the corpus. SVM classifier classifies this dataset along with the testing data provided. There will be two separate datasets called the training and the testing data. The training data is the one which includes all the reviews irrespective of any products selected. Whereas, the testing data includes the reviews of only the selected product. Based on the classification the user will be able to obtain the output in the form of either pie chart of the bar graph.

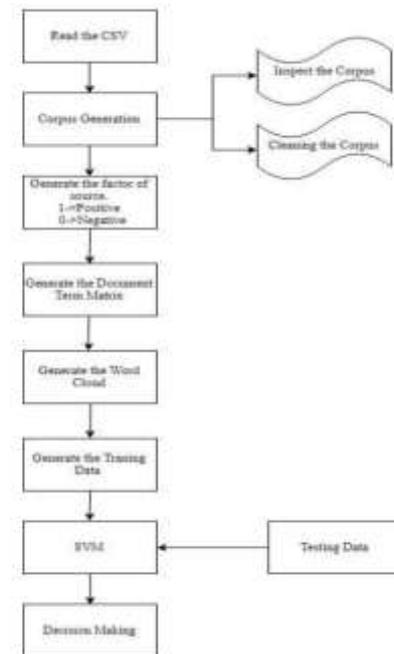


Fig 1- Architecture diagram of the system

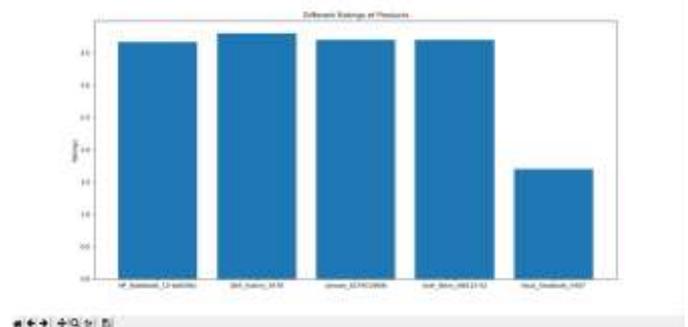


Fig 2- The bar graph which shows the overall statistics of the products used.

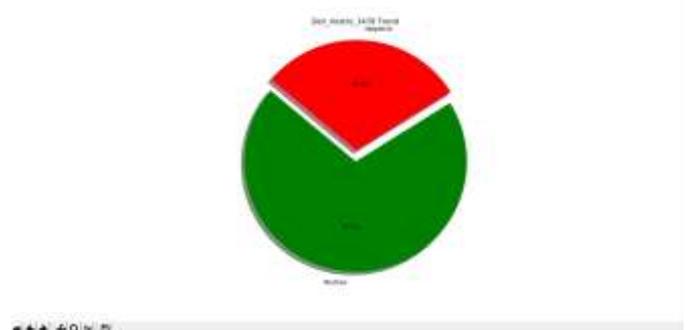


Fig 3- The pie chart that shows the classification of positive and negative review of the individual project

4. CONCLUSIONS

In this proposed work, we analyze the reviews of the users. We have downloaded the dataset from Amazon Web site of the required products. These reviews are tested using train data and are classified into positive and negative reviews using Support Vector Classifier. In this we have considered products like mobile phones, laptops and televisions. Decision is made as per the training set. The result is then represented using pie chart. Also the best model among each products is also shown using bar graphs.

There is lot of research work that is needed to be done in future so as to improve the performance. Sentimental analysis should be applied for any new applications so as to help many customers. This extensibility will be of great benefit to the industry and customers.

REFERENCES

- [1] Abhinaya R, Aishwarya P, Bhavana S, Thamarai Selvi ND - "Automatic Sentiment Analysis of User Reviews". 2016 IEEE International Conference on Technological Innovation in ICT for Agriculture and Rural Development. 978-1-5090-0615-1/16/\$31.00 © 2016 IEEE.
- [2] ZeeniaSingla, Sukhchandan Randhawa and Sushma Jain- "Statical and Sentiment Analysis of Consumer Product Reviews". 8th International Conference on Computing, Communication and Network Technologies 2017. IEEE-40222.
- [3] Shivaprasad T K, Jyothi Shetty-"Sentiment Analysis of Product Reviews".2017 InternationalConference on Inventive Communication and Computational Technologies. 978-1-5090-52974/17/\$31.00 ©2017 IEEE.
- [4] Rekha, Dr. Williamjeet Singh-"Sentiment Analysis of Online Mobile Reviews".2017 International Conference on Inventive Communication and Computational Technologies. 978-1-509052974/17/\$31.00 ©2017 IEEE.
- [5] M. Lovelin Ponn Felciah, R. Anbuselvi-"A Studyon Sentiment Analysisof Social Media Reviews". IEEE Sponsored 2nd International Conference on Innovation in Information Embedded and Communication Systems. 978-1-4799-6818-3/15/\$31.00 ©2015 IEEE.
- [6] NeenaDevasia, Reshma sheik-"Feature Extracted Sentiment Analysisof Customer Product Reviews". 2016 International Conference on Emerging Technological Trends. 978-1-5090-3751-3/16/\$31.00 ©2016 IEEE.
- [7] Chhayya Chauhan, Smriti Sehgal-"Sentiment Analysison Product Reviews". International Conference on Computing, Communication and Automation. ISBN:978-1-5090-6471-7/17/\$31.00 ©2017 IEEE.
- [8] You mon Aye, SintSint Aung-"Sentiment Analysis For Reviews of Restaurant in Myanmar Text". SNPD 2017. 978-1-5090-5504-3/17/\$31.00 ©2017 IEEE
- [9] R.Nithya, Dr.D.Maheshwari- "Sentiment Analysis on Unstructured Review".2014 International Conference on Intelligent Computing Applications. 978-1-4799-3966-4/14/\$31.00© 2017 IEEE
- [10] Yuwa Sawakoshi, Makoto Okada, Kiyota Hashimoto- "An Investigation Of Effectiveness Of Opinion And Fact Sentences For Sentimental Analysis Of Customer Reviews". 2015 International Conference on Computer Application technologies. 978-1-4673-8211-3/15\$31.00 © IEEE DOI 10.1109/CCATS.2015. 33.