

Customer Feedback Analysis Using Machine Learning

Prof.Sonali J. Mane¹, Abhishek Parve², Bhavesh Patil³, Akash Kamble⁴

¹Assistant Professor of Department of Information Technology, Bharati Vidyapeeth College of Engineering Kharghar, Navi Mumbai

^{2,3,4}Department of Information Technology, Bharati Vidyapeeth College of Engineering, Kharghar, Navi Mumbai

Abstract—Internet users are growing exponentially for the last one decade. With rapid expansion of e-commerce, almost all the products are sold on the web. Customer gathers complete information about the products of their interest before making a purchase decision. Opinions and reviews placed by customers help the business as well as other customers in right decision making and improving shopping experience. As the number of people writing reviews is growing rapidly, it becomes difficult to understand and track every customer needs. Some reviews are usefully too long, ambiguous, incomplete and short which makes it more time consuming and difficult to analyse for business. So Customer Feedback Analysis System is required to provide a solution to such problem which will keep track of all those reviews and present statistical report for enhancing the quality and performance of the product.

Index Terms— Text mining, Natural Language Processing (NLP), crawling, Support Vector Machine (SVM), Vectorization, Stemming.

1. Introduction

In recent studies it has been found that 70% of users read and consult an online review before making a purchase. 90% of consumers read online reviews before visiting a business. And 88% of consumers trust online reviews as much as personal recommendations. Reviews have become a validator for users to actually consider to buy the product. Online reviews have an incredible power to shape consumer behaviour. Depending upon the size of the business and the volume of reviews it receive, the businesses need to optimize, prioritize, and be very selective in evaluating which reviews to respond to, and how. For large organizations with numerous locations, and a high volume of reviews, the objective may seem impossible. Having a clear and defined strategy is a must for tracking and managing of those reviews and gains a strong sense of trust with the customers.

In the proposed system, the idea is to mine and summarize all the reviews. We have classified all those reviews based on product quality, service related issues and comment. The system provides a statistical report of the product. We have used Support Vector Machine for modelling and prediction of the review dataset. This would enable business to identify frequently reported issues and the trend of the issues. This knowledge would

ensure enhanced product and support quality. In a process to derive the target outcome from the unstructured raw text, the first step is to identify suitable data source. 'www.amazon.com' an e-commerce website is considered as a data source for the system.

We have followed the below steps in deriving the insights:

- Crawling product reviews from e-commerce website (www.amazon.com) with the help of WebHarvy (Extraction tool).
- Applying text Pre-processing steps to mine the raw text.
- Building the Text data classifier and then applying the classifier on the comments to predict the target label.
- Providing the statistical report.

2. Review of Literature

Chandrasekhar Rangu, Shuvojit Chatterjee and Srinivasa Rao Valluru, HCL Technologies Ltd Hyderabad, India, 2017, proposed a model that enhances Product Quality using Text Mining Approach. The paper showcased different text mining techniques to identify issues of a product. Related to Text pre-processing, they have used different Natural language processing techniques to reduce the vocabulary size and to build robust classifier. Along with SVM classifier they tried other classification technique like Bayes, Random Forest and chosen SVM as best classifier for this kind of data. The limitations of this approach is that the classification model expects minimum 10–15 words after removing non-usable characters during pre-processing. Hence, smaller size review comments are not taken into consideration.[1]

Nikhita Mangaonkar and Sudarshan Sirsat, 2017, proposed a Neuro Linguistic Programming Approach to evaluate customer product experience. They proposed a set of technique required to summarize and filter product reviews, in a specific manner so as to understand the overall Customer experience based on text mining and neuro linguistic programming. The objective is to provide a language a medium through which a customer can be profiled and be retained for a life time, benefitting the Organization business. While reading customer reviews / feedback on a product is

finding words that fall in categories of Visual description, Auditory description and Kinesthetic description.[2]

Arun Manicka Raja, Godfrey Winster , Swamynathan S, 2016, has been proposed a review analyzer system based on performing the sentimental words' analysis for sentiment classification. From the experimental results on the digital camera review datasets, we can confirm that the accuracy of existing classifiers at sentence level outperform Bag-Of -Words (BOW) method. With these results, this method may be extended by checking the patterns of adjective or adverb and noun phrases combination with which the individual reviews' weight can be computed. Then the computed weight can be used to evaluate the product models with the help of weight based ranking.[3]

Kevin Bratawisnu, Refi Rifaldi Windya Giri, Rudi Rinaldi, 2017, proposed Text Network Analysis, a research technique that focuses on identifying and comparing the network relationships between words, sentences and system to model interactions that generate new knowledge or information modelling the perception of consumers well. The advantages of network text analysis than any other method likely Multi-Dimensional Scaling (MDS) to see consumer perception is lower cost, more real time and reach a larger area. So that business enterprises can use to Enhance customer relationship management (CRM) Because It Allows a business to follow public opinion about its brand and respond quickly intervening to customer issues.[4]

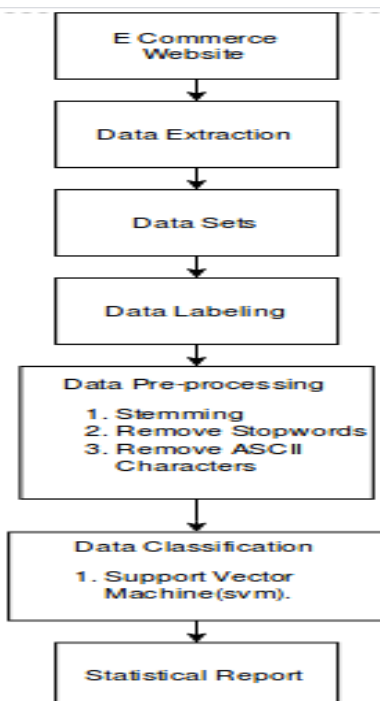


Fig1-System Flow

3. METHODOLOGY

Data Extraction

In general data extraction or crawling from different web pages or community sites is done with the help of an Extraction Tool . 'Webharvy' is used for data extraction of reviews from amazon sites. For this analysis various electronics brand are chosen along with their sub-products i.e. Washing Machine, Television, Refrigerator etc. And then we have crawled all pages review comments with all the selected fields information for each comment. Each comment typically contains "Review Title", "Username", "Rating", "Usefulness", "Comment" and "Date" fields. Review Title gives the information in brief about the comment, Username is the name of the user who commented, Usefulness gives how many have actually liked the comment, Comment is the actual user review comment available in text and Date is the timestamp value of each comment.

Text Pre-Processing

Text pre-processing is the most important phase for any text mining task. It plays an important role in deriving the right patterns from the data. The typical text pre-processing involves Stop words removing, reducing the size of the vocabulary, handling special and unwanted characters. As user comment is a free flowing text, there is very high possibility to have so many special characters, punctuations, symbols, numeric values and white spaces. We have applied stemming to reduce the size of the vocabulary to ensure minimal corpus size. The processed text will be converted into structured format by creating Term Document Matrix (TDM). The term frequencies and inverse document frequencies for each processed review will then be calculated.

Customer Feedback Analysis:

The data obtained after the Pre-processing, the Customer Feedback Analysis System will perform following activities:

- i) Problem Identification
- ii) Text Classification
- iii) Statistical Report generation

i) Problem Identification

The phrases are identified manually after reading various review comments for main categories namely Product Related and Support Related and few of them are selected. For each of the phrases, the associated words to identify the sub-issue under the main category are chosen. For Example, under Product Related Issues we have identified various features/Component/functionality of the product along

with phrases like *Not-Working /Only Problem /Issue /error /main problem/ don't work.*

There is a very high possibility of getting junk words along with these phrases. It is reduced by doing proper pre-processing in the initial phase. Below are the few review comments with identified phrases and corresponding sub-issues

Example : “My Product-X is not working most of the time, unexpectedly the Feature-Y goes off and it gets recovered only after restarting the device.. “

ii) Text Classification

In any classification task, it is important to have labels for the data in order to train the classifier. In our case, the pre-annotated reviews covering both Product and Support related comments are considered. The goal of text classification is to identify the right label for each review comment in the test dataset. It is ensured that the training data representing all levels/categories of the target in order to attain better classification accuracy.

‘Support Vector Machine’ classification methodology is tested with the data. The Stratified random sampling will be used to divide the trained and test datasets with different SVM kernels (radial/polynomial/linear/sigmoid) with optimal model parameters (gamma, cost, and degree) for achieving higher classification accuracy.

Following is the description of the classification methodology

- Support Vector Machine
In machine learning, support vector machines (SVM) are supervised learning models with associated learning algorithms that analyses data used for classification and regression analysis.
In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

iii) Statistical Report Generation

For each identified feature/component under the two main categories Product and Support, the percentage of the customer sentiment is calculated.

This information captured under Product and Support related are then grouped. The graphical representation in the form of Bar plot and Pie charts is provided as the output.

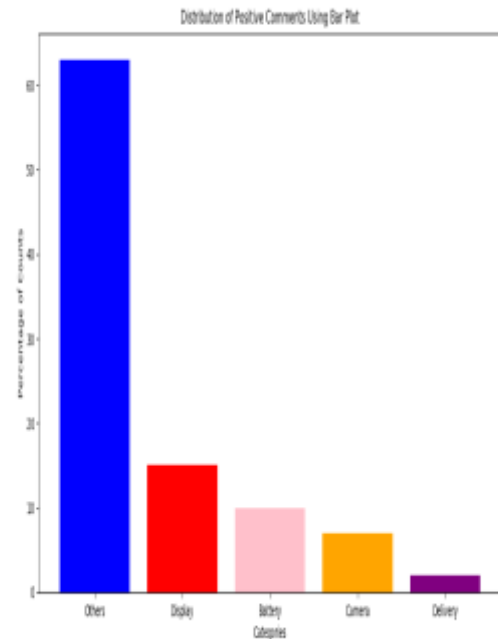


Fig2-Bar Plot of Customer Reviews

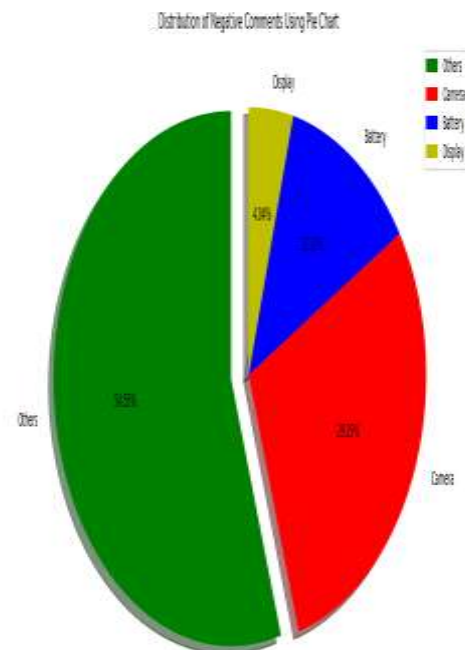


Fig 3- Pie chart of Customer Reviews

4. Conclusions

The reviews received from the customers help business to make correct decision and to increase their profit. With the expansion of internet large amount of feedback/reviews can be received. "Customer Feedback Analysis" will summarize all the reviews which will help product manufacturers to know customer opinion on different product features and help in improving the market strategy as per customer needs.

It will help the support teams in identifying most frequent issues faced by customer, so that some customizations can be made to resolve them in quick time. The system will help business to maintain their market position and attract more customers.

4. References

1. Chandrasekhar Rangu, Shuvojit Chatterjee and Srinivasa Rao Valluru, HCL Technologies Ltd Hyderabad, India, 2017
2. Nikhita Mangaonkar and Sudarshan Sirsat, 2017. Neuro Linguistic Approach to evaluate Customer Product Experience Based On Neuro Linguistic Programming.
3. Arun Manicka Raja, Godfrey Winster, Swamynathan S, 2016. Review Analyzer System Based On Sentiment Analysis
4. Bruce, R., and Wiebe, J. 2017. Recognizing Subjectivity: A Case Study of Manual Tagging. Natural Language Engineering
5. Daille, B. 2017. Study and Implementation of Combined Techniques for Automatic Extraction of Terminology. The Balancing Act: Combining Symbolic and Statistical Approaches to Language. MIT Press, Cambridge
6. Dave, K., Lawrence, S., and Pennock, D., 2017. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. WWW'03.
7. Tait, J. 2017. Automatic Summarizing of English Texts. Ph.D. Dissertation, University of Cambridge.
8. Bhumika, Prof Sukhjit Singh Sehra and Prof Anand Nayyar. A Review Paper On Algorithms Used For Text Classification- (Ijaiem)-2017
9. Mita K. Dalal, Mukesh A. Zaveri- Automatic Text Classification: A Technical Review, International Journal of Computer Applications (0975 – 8887)-Aug-2016
10. Ahmed Faraz- An Elaboration Of Text Categorization And Automatic Text Classification Through Mathematical And Graphical Modelling- An International Journal (CSEIJ), Vol.5, No.2/3, June 2016