INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 6.078
(Volume 6, Issue 3)
Available online at: www.ijariit.com

Sentiment Analysis on Mobile Reviews using Naive Bayes Classification

Erukonda Harsha Praneetha
praneetha78792@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Chokkakula Pushpa
pushpa321ch@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Dundi Swathi
dundiswathi24@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Bushi Deepika
deepikadolly0707@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

Dr. B. Esther Sunanda
bonnie.sunanda@gmail.com
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

ABSTRACT

Now a day, due to increasing technology we can get the opinions of the users all over the world using web mining. People are always curious and show keen interest on what people perceive about various aspects of living and non-living things. In order to understand and analyse the various traits and varying personalities, opinion mining is needed. The main objective of Sentiment Analysis on Mobile Reviews using Naïve Bayes Classification is to develop a system that mines the positive as well as negative reviews of users. For this purpose, we first collect the data from the famous e-commerce websites and extract the reviews posted by users related to different mobile handsets. In this process, we define some common features like Camera, Battery Life, Overall Performance and Design and then find related reviews as a dataset for the analysis purpose. Negative and Positive words that are commonly used for expressing opinions are also collected. This system provides a sentimental analysis on various smartphone reviews diving them positive or negative or neutral using Naïve Bayes Classification. This is basically being obtained by studying and analysing the reviews posted by different users. Analysis of different words coupled in a sentence represents various sentiments and experiences of users and impacts the products available in the market. This analysis compiles a structural modeling approach and Bayesian Interface System to identify the polarity of the opinion which subsequently classifies positive and negative opinions.

Keywords— *Sentimental Analysis, Opinion Mining, Naïve Bayes Classification, Bayesian Interface System.*

1. INTRODUCTION

Sentiment Analysis is contextual mining of text which identifies and extracts subjective information in source material, and helps a business to understand the social sentiment of their brand, product or service while monitoring online conversations. It is the most common text classification tool that analyses an incoming message and tells whether the underlying sentiment is positive or negative or neutral. The applications of the Sentiment Analysis are in the area of 1) Business – Voice of Customer Experience Management, 2) Reputation Management – Pre-detection of fake publicity and 3) Competitor Monitoring. In recent years, the dramatic increase of smartphone and tablet applications has allowed users to comment on various service platforms at any time through mobile internet, social media, cloud computing etc. As the access to the information becomes easier, more and more consumers seek product information and performance reviews not only from the sellers but also from the other consumers. Reviews and ratings posted by people online are the examples and this has been an integral part of customer's buying-decision process. Sentiment polarity classification has gained much attraction from the fields of natural language processing and data mining.

2. EXISTING SYSTEM

In the existing system, sentimental analysis is done by taking a particular brand but not on overall based dataset. The rate of accuracy is low while the large datasets were analysed. It is not helpful to users to a large extent. It doesn't consider various aspects while analysing different brands of mobile handsets and also doesn't give much accurate results while dealing with large

datasets. It was done taking a small dataset containing 1000 reviews of two or three brands which may help considering accuracy factor but not in real-time. Also the classification is done in a binary model which considers a given review is either positive or negative.

3. PROPOSED SYSTEM

This Sentimental Analysis on Mobile Reviews using Naïve Bayes Classification system can perform sentiment analysis considering various models and brands in a large dataset containing up to 4,00,000 reviews posted by several users on an e-commerce website. The rate of accuracy of analysis will be more which can be achieved using efficient analysis tools like Bayesian Interface System and other approaches. This system can classify the brands individually and obtain the results accurately. It can generate both the data and pictorial representation of the results obtained about various mobile brands. With this system, there can be improved customer experience, online reputation can be built. Also manufacturers can get the feedback directly from users thereby having a chance for the improvement of their service or product and enhance the product features in a user-friendly way.

4. WORKING PROCEDURE

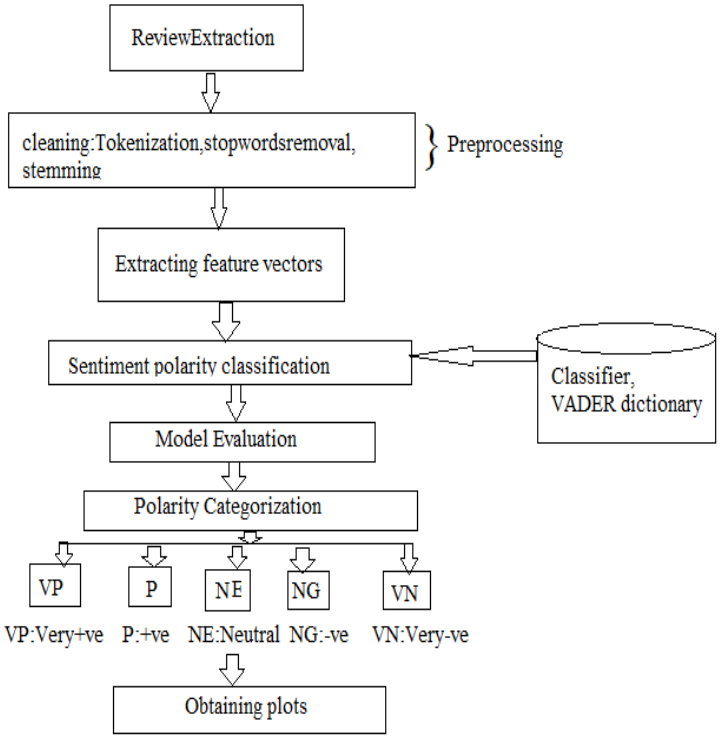

Fig. 1: Architecture of the System

4.1 Algorithm

Step 1: The dataset extracted from an e-commerce website undergoes pre-processing and cleaning which includes tokenization, stemming, removal of stop words and html tags etc.,

Step 2: Find the correlation between every two labels with respect to rating and observe the Positive as well as Negative correlation with respect to rating..

Step 3: To obtain features from the dataset, Vectorization and Naïve Bayes model training are performed on the dataset chosen.

Step 4: Model is evaluated and Classification report is obtained.

Step 5: Based on the polarity scores, reviews are categorized into five classes namely Very Positive, Positive, Neutral, Negative, Very Negative.

Step 6: Based on the category, sentiment value will be given in the range of 1-5 as Very Positive(5), Positive(4), Neutral(3), Negative(2), Very Negative(1).

Step 7: The top brands can be obtained based on the Ratings given by users and the Sentiment Value we got in the previous step.

Step 8: Finally, plot between Brands and their Sentiment Values are obtained.

4.2 Naïve Bayes Classification:

Naïve Bayes classifier is one of the supervised classification techniques which classify the text/sentence that belongs to particular class. This is a simple classification method based on Bayes rule. It relies on the simple representation of the document considered for classification. It is the probabilistic algorithm which calculates the probability of each word in the text/sentence and the word with highest probability is considered as output.

- Let us consider a document ‘d’
- A document a with a set of classes C = {c1,c2, … , cn}
- Consider a training set having m documents which is pre-determined that belongs to a particular class.

Now we train our classification algorithm using this training set and we get trained classifier. By using this trained classifier, we can classify the new document.

Bayesian Theorem Applied to Documents:

For a document d and a class C using Bayesian theorem,

$$P(C \mid d) = [p(d \mid C) * p(C)] / [p(d)]$$

The term p (d | c) is represented as Now representing the document a as a set of features (words or tokens) x1, x2, x3 …
We can then re-write P (d | C) as:

$$P(x1, x2, x3 \ldots xn \mid C)$$

P (C) is defined as total probability of a class which gives the frequency of class.

4.2.1 Bernoulli Naïve Bayes Model Evaluation: This model is popular for document classification tasks, where binary term occurrence features are used rather than term frequencies. This event model is especially popular for classifying short texts. It has the benefit of explicitly modelling the absence of terms. Note that a naive Bayes classifier with a Bernoulli event model is not the same as a multinomial NB classifier.

We obtain Classification Report and calculate accuracy percentage through this by finding precision, recall, f1 score. We also obtain confusion matrix.

Recall: The ability of a model to find all the relevant cases within a dataset. The precise definition of recall is the number of true positives divided by the number of true positives plus the number of false negatives.

$$recall = \frac{true\ positives}{true\ positives\ + false\ negatives}$$

Precision: Precision is defined as the number of true positives divided by the number of true positives plus the number of false positives.

$$precision = \frac{true\ positives}{true\ positives + false\ positives}$$

F1 Score: Combining Precision and Recall

In some situations, we might know that we want to maximize either recall or precision at the expense of the other metric. However, in cases where we want to find an optimal blend of precision and recall we can combine the two metrics using what is called the F1 Score. The F1 score is the harmonic mean of precision and recall taking both metrics into account in the following equation:

$$F_1 = 2 * \frac{precision * recall}{precision + recall}$$

4.2.2 Multinomial Naïve Bayes: With a multinomial event model, samples (feature vectors) represent the frequencies with which certain events have been generated by Multinomial Distribution (p1,....,pn) where pi is the probability that event *i* occurs (or *K* such multinomial in the multiclass case). A feature vector x=(x1,.., xn) is then a histogram, with xi counting the number of times event *i* was observed in a particular instance. This is the event model typically used for document classification, with events representing the occurrence of a word in a single document.

5. CONCLUSION

The main aim of this Sentiment Analysis on Mobile Reviews is to allow the customers to understand the audience and their perception on various brands of mobile phones. Sentiment Analysis models detect polarity within a text, whether it is a whole document, paragraph, clause or even a sentence. By automatically analysing customer feedback in many forms such as survey responses to social media conversations, brands are able to pay attention to the customers' needs and tailor the products or services such that they meet the customers' expectations and needs. The classifier we used here i.e., Naïve Bayes Classifier, is one of the best to perform Sentiment Analysis on a piece of text.

6. REFERENCES

[1] Shilpi Chawla, Gaurav Dubey, Ajay Rana "Product Opinion Mining Using Sentiment Analysis on Smartphone Reviews"
[2] Rekha, Dr. Williamjeet Singh Sentiment "Analysis of Online Mobile Reviews"
[3] Lopamudra Dey, Sanjay Chakraborty, Anuraag Biswas, Beepa Bose, Sweta Tiwari, "Sentiment Analysis of Review Datasets Using Naive Bayes and K-NN Classifier"
[4] Kavya Suppala, Narasinga Rao" Sentiment Analysis Using Naïve Bayes Classifier"
[5] Xing Fang & Justin Zhan" Sentiment analysis using product review data"
[6] Lin Zhangab, KunHuac, Honggang, Wangd, Guanqun Qiane Li Zhanga" Sentiment Analysis on Reviews of Mobile UsersSentiment Analysis on Reviews of Mobile Users"
[7] Shahid M. Awan, Nisar Hussain and Zaheer A. Gondal "Sentiment Analysis on Mobile Phone Reviews Using Supervised Learning Techniques"
[8] Upma Kumari; Arvind K Sharma; Dinesh Soni" Sentiment analysis of smart phone product review using SVM classification technique"
[9] Zeenia Singla,Sushma Jain,Sukhchandan Randhawa" Sentiment analysis of customer product reviews using machine learning"
[10] May Thanda Theint Aung, Aye kyaw "Sentiment analysis of the smartphone product reviews"

