International Journal of Engineering & Technology, *7 (3.4) (2018)* 163-166

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Classification of Reviews on Mobile Phones Using Text Mining Techniques

Livin Davis[1]*, V. Vaidhehi [2]

[1] PG Scholar, Computer Science Department, Christ University, Bengaluru, Karnataka, India
[2] Associate Professor, Computer Science Department, Christ University, Bengaluru, Karnataka, India
**Corresponding author E-mail:livindavis48@gmail.com*

Abstract

People register their opinion or feedback regarding the products in different forum. This research work is based on the classification of reviews regarding the different mobile phones. Dataset from Amazon pertaining to the opinions for mobile phones is used in this work. Opinion which is expressed as text is classified as positive opinion or a negative opinion using text mining techniques. Opinion mining helps to understand the customers in a better way. This work shows the visual representation of words by using word cloud and to classify the reviews on a two point scale. From the dataset, randomly 197 reviews are taken out of which 148 reviews are classified as positive, 49 reviews are classified as negative.

Keywords: *Classification, Opinion mining, text mining*

1. Introduction

In recent years, the use of smart phones has increased rapidly. It plays an important role in day to day life. The growth of mobile technology has increased widely in such a way that users can give their feedback on different products. User opinions are widely used in analyzing the business growth in different domain like communication, online shopping, photography etc. Opinions are expressed as text, reviews ratings, emoticons.

Opinion mining is a field of study that analyzes people's opinions, sentiments, interest, attitudes, and emotions towards entities such as products, services, organizations, individuals, events, topics and their attributes. It is also known as sentimental analysis and refers to the use of natural language processing for tracking the interest of the customer about a particular product. This comes under the text mining category. In opinion mining, the sentences are analyzed as text and the decision is made accordingly Opinion mining has different techniques which identifies the user's type of interest towards the product through reviews and make a decision whether the review is a positive or a negative review. User reviews are feedbacks given by public or the customer to the organization about their experience in using the products.

This research work is based on the analysis of reviews on mobile products. Analysis of opinions which is expressed as text is classified as positive opinion or a negative opinion. The main objective of this research is to classify the review using text mining techniques. The dataset used for this research work is collected from Amazon that contains 1, 94,336 reviews given by different customers on different mobile products and mobile accessories.Dataset obtained is in the unstructured format and by applying appropriate preprocessing techniques, data is made ready for mining. Reviews are stored as separate documents. Each document is classified as positive opinion or negative opinion by calculating the sentence polarity.

This paper is organized in different sections. Section 2 describes the various research work related to the mobile reviews based on text processing and opinion mining. Section 3 elaborates the methodology followed in this approach. Results are discussed in Section 4.

2. Existing Related Work

In Mongkol [1] et al discussed feature based opining mining on smart phone reviews given in Thai language. Polarity of words is used to classify the reviews. Classification of opinion mining techniques is elaborated in [2]. In Lin et al [3] sentimental analysis on reviews of mobile users based on the length of the review. Richa et al [4] they have worked in opinions based on document level on movie reviews. Document is classified as positive, negative and neutral based on the polarity of sentences. In Phong et al [5] research is based on "Phrase-Based Extraction of User Opinions in Mobile App Reviews" (PUMA).Puma is based on extracting phrase from the reviews using part-of-speech, and to cluster phrases having similar meanings.

In S. Vijayarani et al [6] proposed knowledge discovery in text which is used in further text analysis. In Eivind et al [7] discussed opinion mining in hotel reviews and visual analysis based on reviews using Google Maps is also implemented. In Pravesh Kumar Singh [8] et al provide a detail review on the various techniques of opinion mining and sentimental. In Gaurav Dubey et al [9] is based on reviews analysis of web based data using POS tagging and Rule- Mining. In Anil Kumar [10] et al discus the user Sentiments from Kannada Web Documents using semantic learning approach.

In Minara et al [11] is based on product rating using sentimental analysis. They have analyzed the reviews of mobile phones. In Yuefeng et al [12] proposed a new architecture for opinion mining

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

by integrating opinion and customer data. Po-Wei Liang [13] et al is based on automatic analysis of sentiments on twitter reviews. Sentiment is identified as positive or negative. Hai Son Le [14] et al has worked on classification of opinion using supervised learning approach. In Venkata Rajeev [15] et al proposed a system for comparing and recommending different online products through different websites.

From the literature, it is inferred that the reviews on mobile products is not explored much. Also analysis of reviews from Amazon on mobile products is not carried out. Building opinion database for mobile products is not yet initiated. Most of the literature is focused on binary polarities. This research work is focused on classifying the review as positive and negative for a review on mobile products collected from Amazon [16].

3. Methodology

Opinion mining is proposed to understand customer's reaction, interest, attitude, emotion, sentiment towards their interested area. In the recent years, opinion mining has attracted many researchers. In this research work, text processing steps are used to classify the documents using python and R. The various steps involved in this research work is shown in Fig1.

3.1. Data Collection

Dataset of mobile user's reviews of Amazon is used in this research work. The dataset contains 1, 94,336 reviews given by different customers on different mobile products and mobile accessories starting from 2004 to 2014. The dataset contains the following attributes namely user_id, user_name, ratings, summary, review_time, date, review_text. The dataset from Amazon is in JSON file format. CSV format of dataset is shown in Figure2.

3.2. Preprocessing Methods

Preprocessing method is one of the important tasks in the text mining techniques. The detail of preprocessing steps is shown in Fig3.

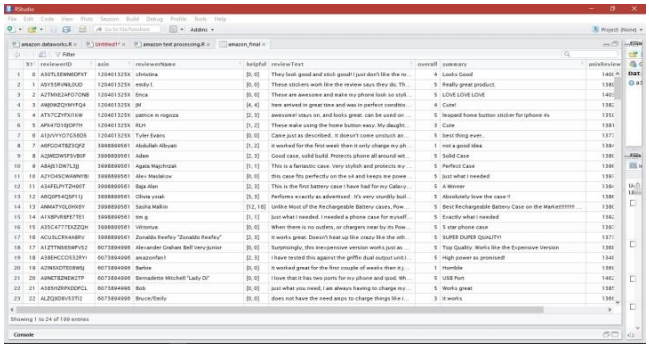
3.2.1 Extraction:

The dataset that is collected is in the form of JSON format. This has to be converted to CSV. Each review is extracted and stored as separate documents. After extraction it is found that the reviews contain Minimum one word to a maximum of ten sentences. Size of the review document ranges from 1KB to 2KB

3.2.2 Stops Words Removal:

The dataset contains many stop words which makes the text heavier. Examples for stops words: the, in, a, an, with, etc. these words are removed from the documents because those words cannot be considered as keywords

Figure2: Amazon dataset



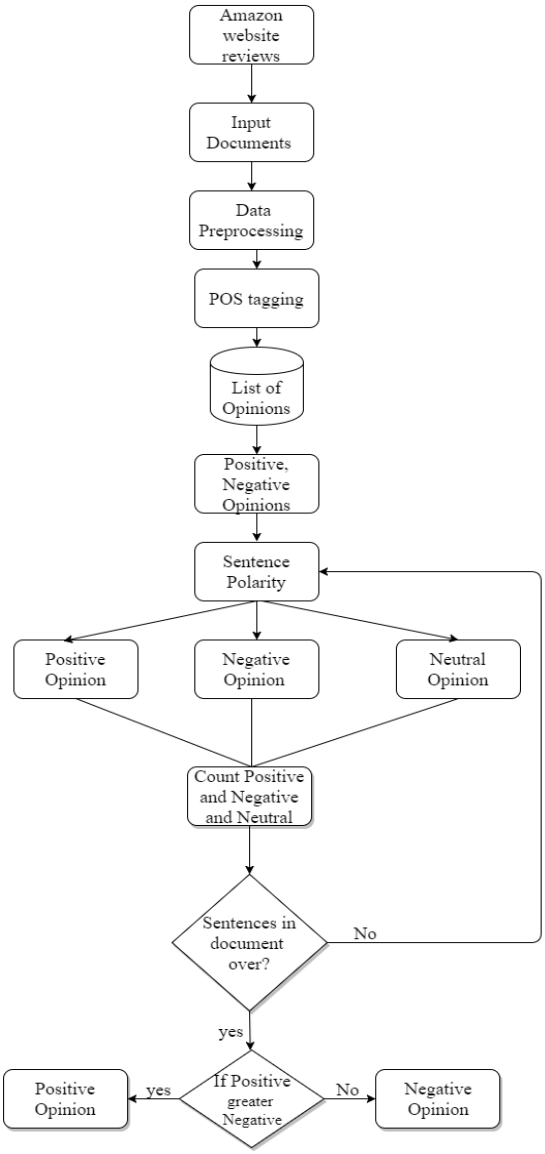


Figure 1: Proposed framework

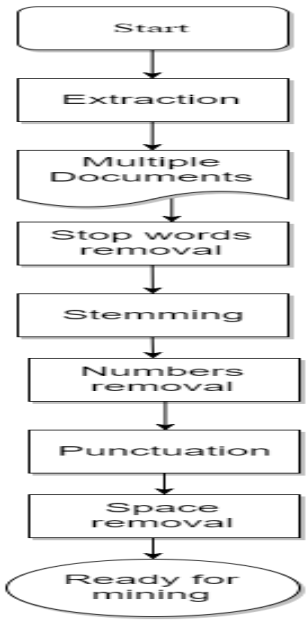


Figure 3: Preprocessing steps

3.2.3 Stemming: This method is used to identify the root of the word. The main aim of this step is to remove the suffixes, to reduce the number of words, to save time and space. For example consider the words search, searched, searching, and searches all these words can be stemmed to the word "search".

3.2.4 Numbers Removal: This method is used to identify the numbers in the text and remove all the numbers in the documents, removes all the ID, mobile numbers, and other integer etc

3.2.5 Punctuation: This method removes all the punctuation in the text to make it lighter and save space

3.2.6 Space Removal: This method is used removes all the white spaces which makes the text h

3.3. POS Tagging

Preprocessed are sent to the POS tagger to tag the words present in the sentence with appropriate POS tag. This technique is necessary to determine the opinion of the words. This can be implemented using RPOSTagger Package in R.

3.4. Opinion Words Extraction

After the POS tagging, the next step is to extract the pinion words. Opinion words are stored in database. The extracted opinion words are matched with the list of words from database. If the words are matched with the database, then the opinion is identified. If the words are not matched, then the words are compared with the synonyms with the help of Word Net to identify the opinion. This process is repeated for every sentence in each document.

3.5. Polarity Detection

For every sentence the polarity is calculated by using opinion words. The polarity of the document is determined by counting the positive, negative and neutral values for all the sentences present in the document. By comparing the polarity values of sentences present in the document, polarity values for the document are determined. Once the document is identified as positive or negative, then, it is moved to the corresponding folder

4. Results and Discussion

From the dataset of 1, 94,336 reviews, randomly 197 reviews are considered for analysis. All the above mentioned steps like preprocessing, POS tagging, opinion word extraction and polarity detection is applied. Data visualization package in R studio is used to display the opinion words as word cloud. The word cloud for positive words is shown in figure 4. The frequency of words is shown in figure 5.



Figure 4: Word cloud

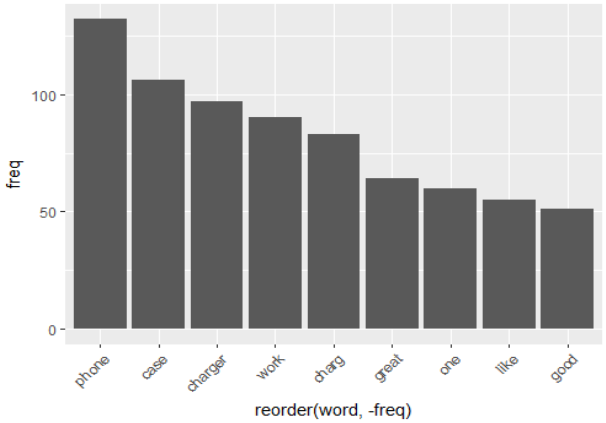


Figure 5: frequency of words

In the word cloud, the most frequently used words are shown in large font and the less frequently words in small font. Out of 197 reviews, 148 are classified as positive and 49 reviews are classified as negative. These classified reviews are stored in separate folders.

5. Conclusion

This work shows the visual representation of words by using word cloud and to classify the reviews on a two point scale. This analysis will help to identify the customer requirements and to know their opinion about the products in a better way. This classification will help to improve the strategic planning of an organization. The current research work can be further extended to obtain the comparison of opinions collected from different sources like Flip kart, eBay etc. Once the classification is done, machine learning models can be built to improve the accuracy of the system. Also, opinions given through emoticon and also on emojis are too analyzed to identify the user's reaction towards products. Analysis of products based on brands and other criteria can also be considered. The size of the dataset can be increased for better efficiency.

References

[1] Mongkol Saensuk, "Feature-Based Opinion Mining on Smart-Phone Reviews" IIAE International Conference on Intelligent Systems and Image Processing 2015

[2] Nidhi Mishra, "Classification of Opinion Mining Techniques", International Journal of Computer Applications (0975 – 8887) Volume 56 No.13, October 2012

[3] Lin Zhanga, "Sentimental analysis on review of mobile users", International Conference on Mobile Systems and Pervasive Computing"2014

[4] Richa Sharma, "Opinion mining of movie reviews at document level" International Journal on Information Theory (IJIT), Vol.3, No.3, July 2014

[5] Phong Minh Vu, "phrase-based extraction of user opinions in Mobile app reviews" ACM, 2016

[6] Dr. S. Vijayarani, "Preprocessing Techniques for Text Mining-An Overview" International Journal of Computer Science & Communication Networks,Vol 5(1),October 2016

[7] Eivind Bjørkelund, "A Study of Opinion Mining and Visualization of Hotel Reviews",ACM December,2012

[8] Pravesh Kumar Singh, "Methodology study of opinion mining and sentimental analysis techniques" International Journal on Soft Computing (IJSC) Vol. 5, No. 1, February 2014

[9] Gaurav Dubey, "User Reviews Data Analysis using Opinion Mining on Web", International Conference on Futuristic trend in Computational Analysis and Knowledge Management, 2015

[10] K. M. Anil Kumar, "Analysis of Users' Sentiments from Kannada Web Documents", International Multi-Conference on Information Processing, 2015

[11] Minara P Anto, "Product rating using Sentimental analysis", International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) , 2016

[12] Mohd Ridzwan Yaakub, "Integration of Opinion into Customer Analysis Model",IEEE, International Conference on e-Business Engineering,2011
[13] Po-Wei Liang, "Opinion mining on social media data", IEEE, International Conference on Mobile Data Management,2013
[14] [14] Hai Son Le, "Aspect Analysis for Opinion Mining of Vietnamese Text", International Conference on Advanced Computing and Applications, 2015
[15] Vekata Rajeev P "Recommending products to customers using opinion mining using online products reviews and features. IEEE 2013.
[16] R. He, J. McAuley Modeling the visual evolution of fashion trends with one-class collaborative filtering. WWW, 2016 J.McAuley, C.Targett, J.Shi, Van den Hengel. Image-based recommendations on styles and substitutes. SIGIR, 2015