



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 Issue: III Month of publication: March 2018

DOI: <http://doi.org/10.22214/ijraset.2018.3537>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Sentiment Analysis of Application Reviews on Play Store

Kunal Varshney¹, Swapnil Tandel², Ajinkya Joshi³, Angelin Florence⁴.

^{1, 2, 3, 4} Department of Computer Engineering, St. John College of Engineering & Management, Palghar, Mumbai, Maharashtra, India.

Abstract: *As the web plays an increasingly significant role in people's social lives, it contains more and more information about opinions and emotion. Nowadays, as smartphones are evolving rapidly, mobile applications have become essential parts of lives. However, it is difficult for consumers to keep track and understand the applications speciality, because new applications are launching every day. Application developers have difficulties in finding out how to improve the application's usability based on overall ratings alone and would benefit from understanding the thousands of textual comments. This project proposes a methodology to automatically extract valuable informations (opinions) from the user comments to minimize the efforts needed in gathering user feedback for a specific application. The system is expected to provide the developer with more than one features like overall representation of comments, providing proper sentimental score of comments and the graphical representation and also the most frequently used words related to a particular application and its environment.*

Keywords: *Sentiment, Sentiscore, Classification, Feature Extraction, Data representation, Parts of Speech tagging.*

I. INTRODUCTION

Sentiment Analysis is nothing but the use of Natural Language Processing, text analysis and computational linguistics to recognize and extract meaningful information in source material. Generally, sentiment analysis aims to computationally recognising & compartmentalization of all opinions, in order to determine writer's attitude for a particular topic. Example positive, negative or neutral, expressed in the form of text. The attitude may be his or her judgement or evaluation, affective state or the emotional state of the author when writing, or the intended emotional communication that is to say, what kind of emotional effect author wants on reader's mind. Posting reviews online has become an increasingly popular way for people to express opinions and sentiments towards the products bought or services received. Analysing the large value of online reviews would produce useful actionable knowledge that could be of economic values to vendors and other interested parties.

II. REVIEW OF LITERATURE

A. Discovery of Ranking Fraud for Mobile Applications

Gathering data for reviews and ratings from application store, and other sources to determine the polarity of the sentences, based on aspects, large numbers of reviews are collected from the Web. For judging Applications on without Natural Language Processing^[1] basis ratings are downloaded side by side with reviews. Pre-processing data to remove any missing entries (using filtering technique) to determine the semantic orientation of the sentences a dictionary based technique of the unsupervised approach is adopted. To determine the opinion words and their synonyms and antonyms WordNet^[1] is used as a dictionary. Semantic matching for finding quality of review (Positive, Negative or Neutral) a large amount of reviews of users are collected on the web that needs to be explored, analyse and organized for better decision making. Opinion Mining^[1] or Sentiment Analysis comes under Natural Language Processing and Information Extraction task that identifies the user's point of view, explained in the form of positive, negative or neutral comments and quotes underlying the text. Primarily Aspect based opinion mining is a kind of Opinion Mining that identifies the user's thoughts of the given reviews and categorise the review for each feature.

B. Aspect Rating Analysis Based Product Ranking

This method is used to settle FMADM^[2] problem. It evaluates different ways for attributes or criteria, where each attribute is not interdependent with each other. According to Yoon Weighted average Product method using multiplication techniques to connect rating attributes, where each attribute rating should be raised to advance with the corresponding attribute average. This process is similar to the process of normalization.

C. Feature Extracted Sentiment Analysis of Customer Product Reviews

In the paper, main goal is to make it easy for customer in decision of purchasing their product instead of looking thousands of reviews before choosing the right product. The system uses JAVA. The system aims on sentiment based feature extraction algorithm^[3], by which goal of simplicity is achieved. In the proposed system, preprocessing is done, feature extraction (Part Of Speech tagging)^[3] with newly designed algorithm. The system also uses sentiment polarity (positive & negative). This system uses review matrix algorithm to extract features from a particular review, by using this method system can find easily if the product review is positive or negative. The paper suggests their own criteria to find polarity of product feature & feature extraction and classification can be improved using machine learning approach.

D. Design Approach for Accuracy in Movies Reviews Using Sentiment Analysis

In the paper, the objectives is to perform sentimental analysis on “Times of India” movie review database. System uses Random Forest Classification^[4] technique for achieving accuracy. This system follows lexical analysis^[4] approach using Senti Word Net for finding overall opinions of the movies reviews. This System uses Natural Language Processing for detecting part of speech^[4] like adjectives, verbs, nouns etc. The system calculates ratings as positive, negative and neutral. According to that accurate ratings are generated and that will helps to take a better conclusion while going to watch a movie.

III. METHODOLOGY

A. Classification

For the purpose of classification the system makes three copies of the comments. This is done so that it can take average opinion of people on the sentiment of the reviews and in this way the noise and inaccuracies can be minimized. The system then labels the comments in four classes according to sentiments expressed and observed in the comments: positive, negative, neutral, objective and ambiguous.

- 1) *Positive*: If the entire review contains positive, happy or cheerful attitude or if something is mentioned with positive feelings. Also if more than one sentiment is expressed in the comment but the positive sentiment is more dominant. Example: “Awesome application and works flawlessly on my tab!”.
- 2) *Negative*: If the entire comment contains negative, sad or displeased attitude or if something is mentioned with negative feelings. Also if more than one sentiment is expressed in the comment but the negative sentiment is more dominant. Example: “I hate this application as it opens too many tabs and lags too much on my tab”.
- 3) *Neutral or Objective*: If the reviewer expresses no personal opinion in the comment and gives less useful information. Advertisements of different products would be under this category. Example: “This application is downloaded on my phone recently”
- 4) *Ambiguous*: If some opinions are expressed by personal experience, but with no proper reference to it, then it is very difficult to understand the opinion of reviewer. Example: “I kind of like this application and also don’t like it at the same time”

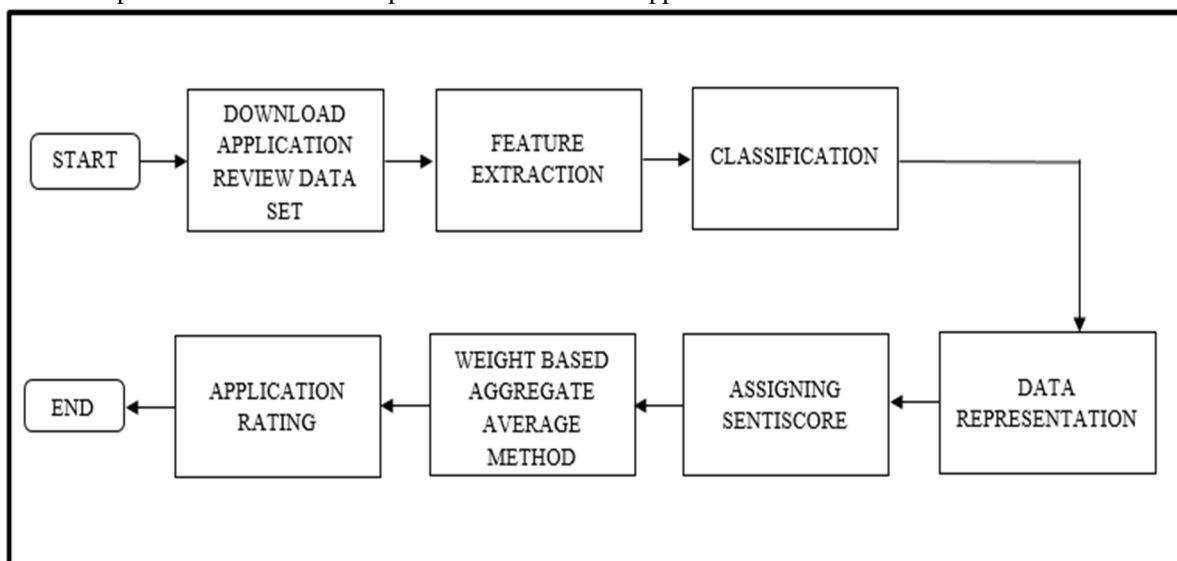


Fig. 1 Flow Diagram

B. Feature Extraction

As the system need to extract useful features from comments, Feature Extraction performed through Stanford core Natural Language Processing.

- 1) *Tokenization*: In Tokenization system breaks sentence into words, symbols called “tokens”, separated by white spaces or punctuation character. So that to identify individual word.
- 2) *Sentence splitting*: Here the system splits a sequence of tokens into sentences.
- 3) *Parts-of-Speech Tagging*: In Part of Speech tagging system assigns a tag to each word to identify grammatical part of the corresponding word. I.e. noun, verb, adjective, adverb.
- 4) *Named Entity Recognition*: Named Entity Recognition recognizes special names such as name of PERSON, DATE, TIME, LOCATION etc. With the default annotators, named entities are recognized. This Annotator provides a simple framework to allow a user to incorporate Natural Expression labels that are not annotated in traditional Natural Language. For example, a default list of regular expressions: ideologies (IDEOLOGY), nationalities (NATIONALITY), religions (RELIGION), and titles.
- 5) *Stop-words removal*: Stop words does not have any additional information, they are just common words. Examples include “a”, “an”, “the”, “he”, “she”, etc. These words can be repeatedly used multiple times by reviewer in the comments.

C. Data Representation

After the process of feature extraction the system performs classification on the extracted data and then classifies the data into two classes Positive, Negative. Then the system assigns Emotion score (For positive comment 1 will be added, for negative comment 1 will be subtracted).

D. SentiScore

The system calculates score of a word, sentence, and review of application, ranging from 0 to 5. Higher score suggests that the application has positive opinions and is popular on play store. Whereas score near to 0 suggests that opinions related to the application are mixed or highly negative. The system performs all these operation on the latest 40 comments for latest results, so that user develops confidence in the classifier. Number of positive classified comments or Number of negative classified comments. The ratio suggests if the number of positive comments is larger than negative comments on a particular keyword, the keyword contains overall popular opinion and vice versa. Finally a maximum of 40 comments are displayed for each class (positive, negative and neutral) so that the user develops confidence in our classifier.

E. Aggregated Average Method

The system calculates mean by outcome of sentiscore, to generate rating for a single application from bunch of reviews. Average ratings are calculated by addition of sentiscore divided by total number of sentiscore, which will generate accurate rating for the application.

IV. CONCLUSION

This system approaches for extracting application features in user review and their sentiments, which help developers to analyze opinions of users and use this information in future requirements. Also the problem faced by Natural Language Processing is overcome using lexicon approach. As the user reviews are of applications vary from category to category the proposed procedure is efficient. The results are sufficient for judging the android applications and developers are also able to predict the problem and the improvement needed in the applications for its popularity within less time.

REFERENCES

- [1] International Journal of Innovative Research in Computer and Communication Engineering “Discovery of Ranking Fraud for Mobile Applications” (An ISO 3297: 2007 Certified Organization) Vol. 5, Issue 5, May 2017.
- [2] International Conference on Global Trends in Signal Processing, Information Computing and Communication “ Aspect Rating Analysis Based Product Ranking” 2016
- [3] 2016 international conference on emerging technological trends (ICETT), “Feature Extracted Sentiment Analysis of Customer Product Reviews”.
- [4] International Conference on Electronics, Communication and Aerospace Technology “Design Approach for Accuracy in Movies Reviews Using Sentiment Analysis” (ICECA 2017)
- [5] IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING “Discovery of Ranking Fraud for Mobile Apps.”, VOL. 27, NO. 1, JANUARY 2015.
- [6] Ronen Feldman the text mining handbook-Advanced approaches in handling unstructured data Cambridge. Available: http://ijcai13.org/files/tutorial_slides/tf4.pdf



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)