International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 4 960 - 963

Product Review Analysis Tool

Prof. Dipti Pawade
Assistant Professor, Department of IT
K.J.Somaiya College of Engineering
Vidyavihar, Mumbai

Khushaboo Rathi
Department of Information Technology
K. J. Somaiya College of Engineering
Vidyavihar, Mumbai

Shruti Sethia
Department of Information Technology
K.J.Somaiya College of Engineering
Vidyavihar, Mumbai

Kushal Dedhia
Department of Information Technology
K. J. Somaiya College of Engineering
Vidyavihar, Mumbai

Harshada Sonkamble
Assistant Professor, Department of CS
Vishwatmak Om Gurudev College of Engineering Mohili, Maharashtra

***Abstract*—**Nowadays, people give their opinion regarding a particular product bought by them or the service provided by the service provider. Analyzing large number of online reviews will help in producing the useful knowledge which could be of economic values to the customers for selecting particular product and for vendor to improve the quality and sales. In our system, customer/users comments are fetched from the facebook and through sentiment analysis percentage of positive, negative and neutral comments are calculated. We have planned to compare results for five similarity measures, viz, Jaccard Similarity Measure, Cosine Similarity Measure, Dice Similarity Measure, Overlap Similarity Measure, Simple Matching Similarity Measure.

Keywords*-Sentiment Analysis, Jaccard Similarity Measure, Cosine Similarity Measure, Dice Similarity Measure, Overlap Similarity Measure, Simple Matching Similarity Measure*

*****

I. INTRODUCTION

Generally individuals and companies are always interested in others opinions. If someone wants to purchase a new product, then firstly, he/she tries to know its reviews. Similarly, companies also excavate deep for consumer reviews. Digital ecosystem has a plethora for same in the form of blogs, reviews etc. But there is huge data on Internet. Going through each and every review is very tedious and time consuming. To resolve this tiresome task we developed a product review system, as a JAVA application. The system automatically fetched the revives from the Facebook page related to that product and through sentiment analysis calculate the percentage of positive, negative and neutral comments. Sentiment analysis is considered to be a good tool to analyze and classify the human sentiment, emotions or opinion about particular thing [2]. If percentage of positive comments predict that product is good and user has affirmative outlook for the same.

Vikas Thada et al. [1] have done comparative analysis for Jaccard, cosine and dice similarity coefficient to give most pertinent document for the specified collection of keywords. From the experimental results they conducted best fitness values were obtained using the Cosine similarity coefficients followed by Dice and Jaccard. Shraddha et. al [3] have proposed product review analysis tool based Jaccard and Cosine similarity measure to overcome the shortcomings of State Vector Machines (SVM).

II. METHODOLOGY

We have used five similarity measures in our tool, the comments are processed and result is given by the Jaccard, Cosine, Dice, Simple Matching and Overlap similarity. As we know there is some difference in accuracy of every algorithm. So while processing we use these five similarity measures and then afterword give final result as average of these five.

Jaccard similarity coefficient (Jaccard index) states likeness between the finite sample sets as the intersection divided by size of the union of the sample sets (Formula 1) [4].

$$J(X,Y) = \frac{|X \cap Y|}{|X \cup Y|} \quad \ldots\ldots\ldots(1)$$

Where X, Y are the finite sets.

Cosine Similarity measures the similarity between two vectors using Euclidean dot product formula 2 [3].

$$x.y = ||x||||y||cos\theta \ldots\ldots\ldots(2)$$

Dice's similarity coefficient measures calculate the similarity index by comparing the letters in dataset string and keyword.

Simple Matching Similarity(SMC) is a statistic used for comparing the similarity and diversity of sample sets.Given two objects, A and B, each with n binary attributes, SMC is defined as formula 3 [4]

$$SMC=\frac{\text{Number of matching attributes}}{\text{Number of Attribute}} = \frac{M_{00}+M_{11}}{M_{00}+M_{01}+M_{10}+M_{11}} \ldots(3)$$

 ISSN: 2321-8169
960 - 963

Where:

M_{00} is the total number of attributes where *A* and *B* both have a value of 1.

M_{01} is the total number of attributes where the attribute of *A* is 0 and the attribute of *B* is 1. M_{00}

M_{10} is the total number of attributes where the attribute of *A* is 1 and the attribute of *B* is 0.

M_{11} is the total number of attributes where *A* and *B* both have a value of 0.

Overlap Similarity coefficient is also known as Szymkiewicz-Simpson coefficient. It is defined as the size of the intersection divided by the smaller of the size of the two sets (formula 4).[4]

$$overlap\ (X,Y) = \frac{|X \cap Y|}{\min(|X|,|Y|)} \ldots\ldots\ldots (4)$$

Where, *X* and *Y* are two sets.

III. System Overview

Figure 1 demonstrates the system architecture.

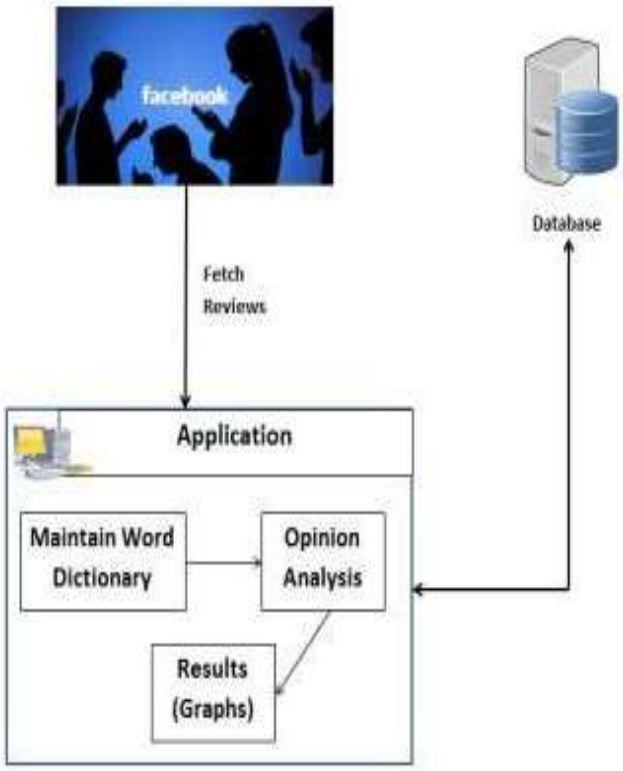

Figure 1: System Architecture

In our application we are fetching the live comments from Facebook. Firstly we need to specify the name of that product for which the reviews are to be fetched. Then the comments or reviews from the Facebook pages are dynamically fetched and responded to the system. Then to classify those reviews we already need to maintain a dictionary of words (positive, negative). And that words dictionary is stored and maintained in the database. So as soon as we get any words as positive or negative that we need to add in the dictionary specifying its proper category. Then after reviews are fetched this dictionary is compared to the same. Before comparing, the fetched reviews are first broken down into the tokens. Then algorithms are applied to sort comments into positive, negative and neutral section. This will be calculated to generate the result. Then the results are displayed in the forms of bar graph specifying the number of positive, negative and neutral reviews. The whole sentiment analysis flow is shown in figure 2.The overall process has following steps:

Step 1:
DocumentCorpus- Itisacomment/review page. Thefetched reviewsareintheformofthedocuments whichare takendynamicallyfromthe website.

Step 2:
Parsing- It means breaking paragraph or data block into small parts like sentences. Here, the whole paragraph containing comments is broken down to sentences.

Step 3:
Tokenization- The sentences which are broken down from paragraph are further separately tokened into or broken into Keywords.

Step 4:
Stemming- It is the technique to find base or root form of words which we have tokenized from sentences for example “helping and sharing” is converted into “help share”. Stemming is very important in this process.

Step 5:
TF-IDF- It stands for Term Frequency-Inverse Document Frequency is use to find the frequency of a word in a document this will help in improvement of the result.

Step 6:
Sorting- According to the frequency of the words they are rated and then sorting is applied.

Step 7:
Filtering- Here, important keywords are taken and unnecessary words are filtered or rejected.

Step 8:
Final Keyword- List of final keywords that are tokenized, stemmed, sorted and filtered are ready to be compared with the dictionary which is maintained.

Step 9:
Labelling- After comparing the keywords with the dictionary, it is labelled as positive, negative or neutral.

Step 10:
Transform- In this step the string is converted to integer. So the words are converted into integer as “1” and “0” .

Step 11:
Final Data set- All the above processing results in generation of final data set using which percentage of positive or negative or neutral comments is displayed in terms of text as well as bar graph.

 ISSN: 2321-8169
960 - 963

IV. RESULTS

Figure 3 shows the interface to manage dictionary. Here the facility to add, delete or import text file containing wordis provided. The advantage of this system is that one can add words of any language or slang words.

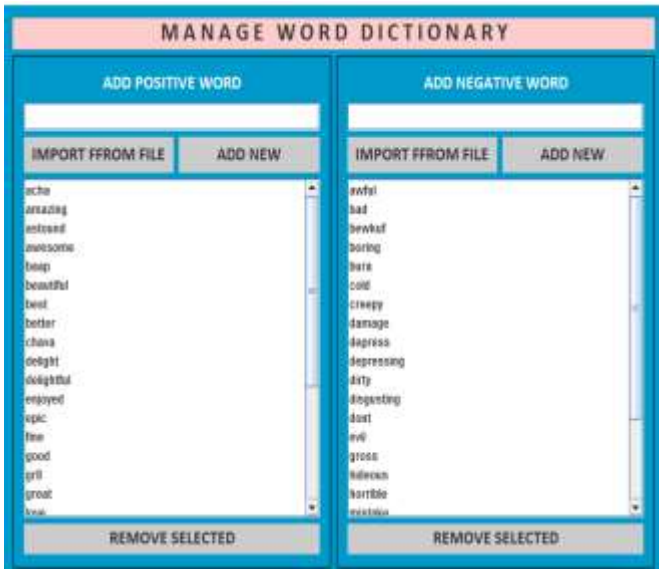

Figure 3 Manage Directory

Figure 4 Fetching Comment

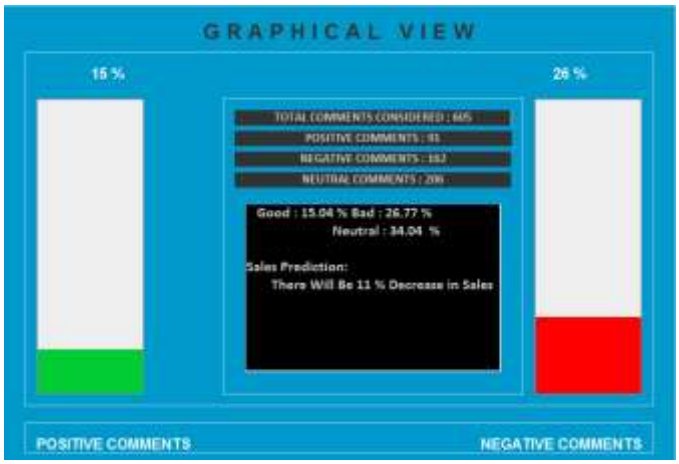

Figure 5Review Analysis

Figure 4 shows the comments fetched from the Facebook foe a particular product. Figure 5, represents the review analysis for the fetched Facebook page

V. CONCLUSION

The product review system proposed in this paper considers different similarity measure to give best results to the user. This helps vendors and management people to take strategic business decisions and also the customers to choose the best product.

As of now we have taken reviews from Social Networking website like Facebook. In future, the reviews can be also collected from Twitter, YouTube or other similar kind of Websites. This will help in getting more kind of reviews on a particular product or many. This will move towards the accuracy of the result of a product. A product can be reviewed from many sites at a time as well. Languages can also be updated as required. Many languages can be introduced as well which the comments will be stored and processed accordingly.

REFERENCES

[1] Vikas Thada, Dr Vivek Jaglan, "Comparison of Jaccard, Dice, Cosine Similarity Coefficient To Find Best Fitness Value for Web Retrieved Documents Using Genetic Algorithm," International Journal of Innovations in Engineering and Technology (IJIET), Vol. 2 Issue 4 August 2013.

[2] [2] Fabon Dzogang, Marie-Jeanne Lesot and Maria Rifqi, Expressions of Graduality for Sentiments Analysis - A Survey, Fuzzy Systems (FUZZ), IEEE International Conference, july 2010.

[3] [3] Shraddha Deshpande, Mrunmayee Shinde, Jonika Rathi, Shanu Gandhi, Vaishali Deshmukh, "Sentiment Analysis Tool using Cosine and Jaccard Implementation", International Journal of Computer Applications, Volume 115 – No. 12, April 2015.

[4] [4] Seung-Seok Choi, Sung-Hyuk Cha, Charles C. Tappert , "A Survey of Binary Similarity and Distance Measures", Systemics, Cybernetics And Informatics Volume 8 - Number 1 - Year 2010.

[5] [6] Malmaz Roshanaei, Shivakant Mishra,"An Analysis of Positivity and Negativity Attributes of Users in Twitter,"IEEE /ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014) page 365-370, July 2014.

 ISSN: 2321-8169
960 - 963

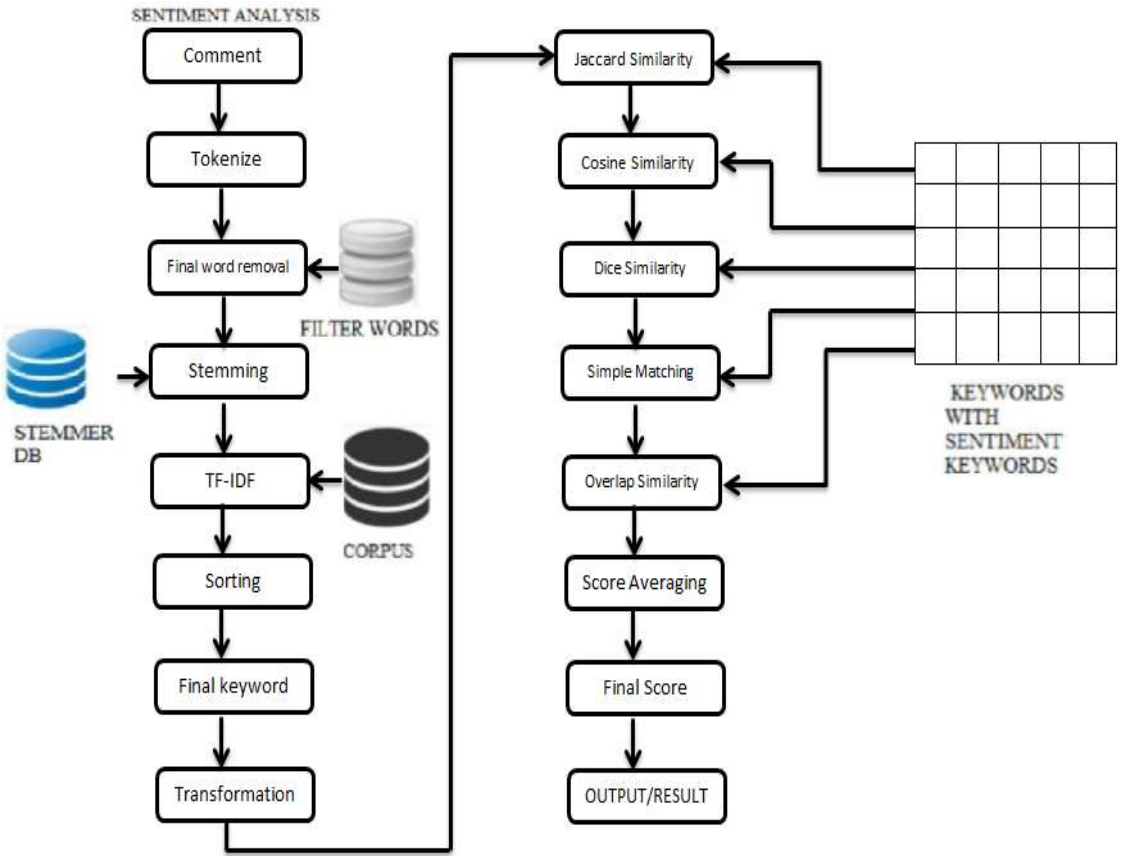

Figure 2 Flow of Sentiment Analysis