Knowledge Based Algorithm for Analysing Customer Opinion

Bhavna Bhardwaj[1], Avinash Navlani[2]
[1, 2] School Of Future Studies and Planning, Devi Ahilya Vishwavidyalaya, Indore

Abstract- *This article is to propose a new algorithm for opinion mining purpose using Knowledge based approach. Opinion mining is basically a text mining utility in which tends towards finding the review or feedback for some particular product or topic by using text mining tools and techniques. Their exist some approaches for this like apriori algorithm, naïve bayes classifier etc. In this paper a new algorithm is introduced which may try to do the same in an easier way. In knowledge based approach we make the cluster for positive and also negative catchphrases to contrast them and the archives, states and as per the words scores are allotted to every report, expression and in light of the scores we can figure out if the assessment is positive, negative or nonpartisan. The parameters that we discovered here are - OpinionScore to retrieve the scoring of each phrase, MeanOpinionScore to find the average opinion, PercPosScore denotes the percentage of positive feedback, and PercNegScore denotes the percentage of negative feedback.*

Keywords- Opinion mining, Knowledge based approach, text mining, natural language processing.

I. INTRODUCTION

With the explosive growth of social media on the web, individuals and organizations are increasingly using the content in this media for their decision making [1]. In present era if you want to buy a product, then you are no longer bound to ask only family or friend's view because there are numerous customer reviews available on the web. Similarly if you want to know about the vision on anything (consumer product, any political issue, any global issue etc.) on global basis, you can find that by using some techniques. These are known as Opinion Mining or Sentiment Analysis.

For example if you want to know about any product's feedback you can check out numerous website which will provide direct customer review and can take decision accordingly. Here it is okay if the reviews are in small number but if there exist some thousand reviews than definitely nobody is going to consider all. At this point text mining can really help you out i.e. by using some text mining techniques the summarized view of whole data can be driven easily i.e. the reviews are positive or negative or neutral. This is known as sentiment analysis or opinion mining.

II. LITERATURE REVIEW

Social Media like Twitter, Facebook, and Blogsnowadays are really very helpful to know about the reviews about anything like any product, movie, some political issue or anything else, people continuously post their thoughts over topics on such websites and these comments and post are really helpful for us to make a decision. Even many of companies can use opinion mining tools and techniques to determine the actual reputation of their product.

Many companies are attempting to automate the process of filtering. This will leave them with genuine blogs and filter out the noise. Thus, even at a commercial level, it becomes important to understand conversations and look at the relevant content and absorb its functioning properly [2].

Presently social networks have secured several millions active and passive web users around the planet. The fast and continuous growth of these informal communities has proven undeniable, facilitating interconnection between users and high rate of data trade. According to the Nielsen report in March 2009, Social Networking has been the global consumer phenomenon of 2008. Two-thirds of the world's internet population visits a social network or blogging site and the sector now accounts for almost 10% of all internet time [3].

Opinion mining is one of numerous zones of computational studies that arrangement with supposition situated natural language processing. Such feeling focused studies incorporate, among others, type qualifications, feeling and inclination acknowledgment, positioning, importance calculation, points of view in content, content source identification and opinion-oriented summarization [4].

Opinion Mining can be redefined as classification of a textual entity into one of the elements or another of the predefined set described above. Sentiment analysis can be done at different levels - document, section, paragraph, sentences, or phrase levels. Most sentiment analysis work is done at a document level, but there are works on sentiment analysis at phrase and at sentence levels Studies have shown that sentiment analysis at phrase level proved to be more difficult than at higher levels [5].

III. DATA PREPROCESSING

Before applying the algorithm to the data set it needs a lot of processing for which we will use natural language processing techniques. Data pre-processing is the routine of creating a new data set in a way that it can be directly used by any mining technique and can give efficient results. Firstly we will remove stop words and white spaces from the data set, after that punctuations and numbers will be removed, then if there exist some hyperlinks in dataset those has to be removed. At last the whole data set will be converted into lower case because lower and upper case letters has different ASCII codes. And when this data set is processed completely after this it is ready to be an input to our algorithm.

IV. PROPOSED METHODOLOGY

Here for opinion mining purpose we are trying to find out a new way to get efficient results. In this we are going to use positive and negative dictionary. These lexicons will be used to score the particular phrases according to their word usage. First step will be data extraction which can be taken from any social networking website or any other website which provide the reviews of the products. Then data pre-processing techniques are applied to the data set. Next step is of comparing it with the lexicons and after that by using our formulas scores will be given to each phrase, and using those score other parameters will be determined. The whole procedure can be understood by following diagram.

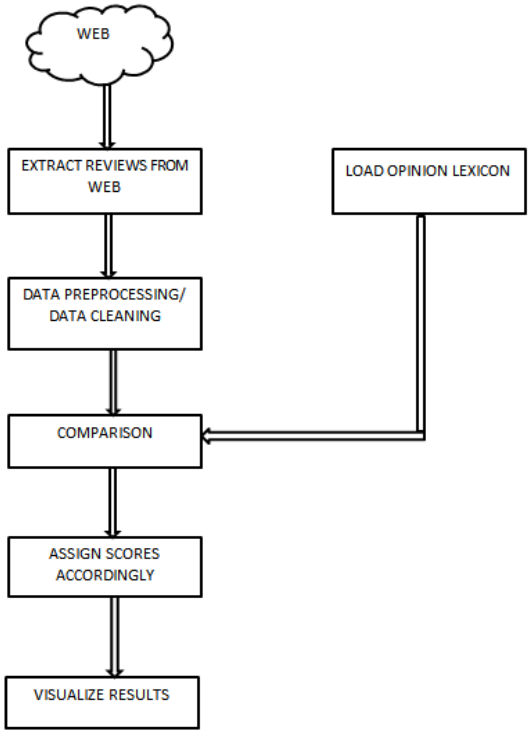

Figure 1: Generalize Process for Opinion mining through Knowledge Based Approach

A. MATHEMATICAL EXPRESSION:

The mathematical expression that will be applied for the algorithm is-

$$\text{OpinionScore} = \sum(\text{pos}) - \sum(\text{neg}) \quad \ldots(1)$$

$$\text{MeanOpinionScore} = \frac{OpinionScore}{N} \ldots(2)$$

$$\text{PercPosScore} = \frac{Count(PositiveOpinionScore)}{N} \ldots(3)$$

$$\text{PercNegScore} = \frac{\text{Count(NegativeOpinionScore)}}{\text{N}} \ldots(4)$$

Here , OpinionScore is the final score for each phrase/ document, Sum(pos) is total positive words in phrase/ document, Sum(neg) is total negative words in phrase/ document, MeanOpinionScore is average opinion of the whole data-set, N is the total no of phrases/ document, PercPosScore is the percentage of positive feedbacks and PercNegScore is the percentage of negative feedbacks. By using such formula we retrieve our result.

V. ALGORITHM

Algorithm for analysis of opinion on the basis of textual data

Input:

Review/ dataset, sum(pos)=0 and sum(neg)=0

Output:

OpinionScore, MeanOpinionScore, PercPosScore, PercNegScore.

Procedure:

1: Extract the reviews from social networking sites, blogs etc.

2: Clean the data.

3: Split the phrases into words.

4: Compare each word with the positive and negative lexicon.

5: if word== Positive

Sum(pos)+1;

if word== Negative

Sum(neg)+1;

6: OpinionScore=Sum(pos) – Sum(neg);

7: if OpinionScore> 0

Feedback is positive;

Else if Opinionscore< 0

Feedback is negative;

Else

Feedback is neutral;

8: count total positive, negative and neutral feedbacks and find out overall results.

9: Visualize the results through graphs.

VI. RESULT AND ANALYSIS

For analysis of the above algorithm we used statistical programming in R for the implementation purpose. We took some of the reviews of a product from web and the results are-

```
> result$score
[1]  2  3 -1 -1  2 -1 -2
```

Figure 2: OpinionScore for each review

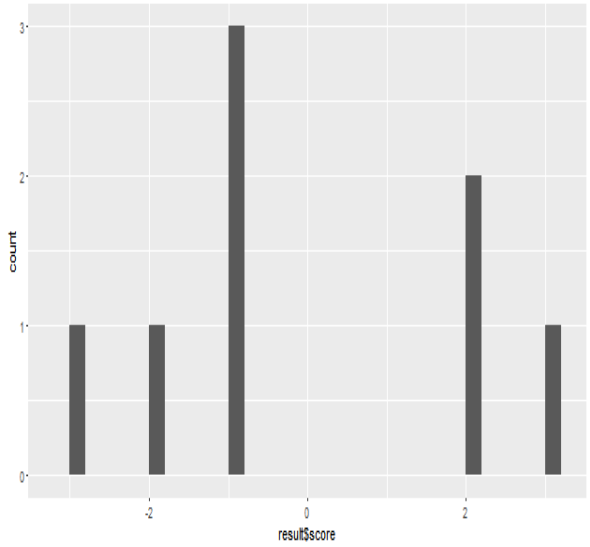

Figure 3: Visualization of OpinionScore

The Mean OpinionScore here is 0.24 which shows that the overall feedback is positive for the product. The data-set here used is very small that's why the result is not that much clear but larger dataset will give more efficient results.

VII. CONCLUSION AND FUTURE WORK

The above approach discussed also known as knowledge based approach in this we already have an array of positive and negative words with us and by comparing and using some other natural language techniques the results are determined. This method is really helpful to find out feedbacks from web, although it has some disadvantages also to be work on.

Sentences like "she is not a good student." Actually gives negative feedback but in this method it may give a neutral feedback, because when it compares the sentence with positive and negative array it will give +1 to "not" and also +1 to "good" and the final result will become zero hence feedback is neutral. Such problem has to be resolve.

Another problem with this method is, it do not deal with Sarcasm, which nowadays is a big part of informal communication. So there is a need to find out a way handle such emotions.

REFERENCES

[1] **Liu,** Bing,Web Data Mining, Chapter 9 Opinion Mining and Sentiment Analysis, Department of Computer Science, University of Illinois, Chicago, 851 S. Morgan St., Chicago, IL, 60607-7053, USA.

[2] Alex W., (2009), "Mining the Web for Feelings, Not Facts", New York Times, 2009-08-23, Retrieved on 2009-10-01, "http://www.nytimes.com/2009/08/24/technology/internet/24emotion.html?_r =2&", (Date: November 05 2015).

[3] Global Faces and Networked Places, A Neilsen report on Social Networking's New Global Footprint, March 2009. Neilsencompany.

[4] Pang and L. Lee. Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1-2):1-135, 2008. ISSN 1554-0669.

[5] Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pages 347-354. Association for Computational Linguistics, 2005.