International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 12 (2015) pp. 31807-31816

http://www.ripublication.com

Sentimental Analysis for Social Networking Sites Using Different Techniques

[1]Balrajpreet Kaur and [2]Roshan Srivastava

[1]DepartmentofComputerScience and Engineering,
Lovely Professional University, Jalandhar, India
[2]Asst. Professor, Department of Computer Science and Engineering,
Lovely Professional University, Jalandhar, India

Abstract

Sentiment gives an idea about feeling of a person about something. Sentimental analysis is a type of opinion mining which is used to check the positivity and negativity of data. Sentimental analysis gives us feedback of people through which we can update the product. Sentimental analysis is needed for making updation in business brochure or product. In this paper we calculate the positive and negative score of words. These score help in improving the accuracy of sentiment analysis. In this paper we also compare the results. We collect the data from social networking sites for analyzing opinion of people on different product.

Keywords: Polarity, Opinion, Sentiwordnet

I. INTRODUCTION

There is a big influence of internet on our daily life. The internet provides us social media to connect to the world. We use Facebook, twitter, LinkedIn and more sites to get connected to people. We also give opinions and review about a process or product. Sentiment analysis is made up of two different keywords, first is sentiment which defines emotion, opinion and behaviour about particular thing. Second one is analysis which defines as examination of structure i. e. breaking the sentences for getting result by which we can further improve the performance of a product or process. Both of these keywords give idea about any product. The main aim of sentiment analysis to get the reaction or behaviour of a person about product or specific subject.

There are four types of sentiment analysis level [1]:-

- **Document level of sentiment analysis**:-In document level the opinion is classified either in positive or negative. In document level we check single entity for result.
- **Sentence level of sentiment analysis**:-Sentence level refine the review of document level analysis. It also filters the sentence containing nothing.
- **Aspect based sentiment analysis**:-Aspect based sentiment analysis focus on all sentiment expression. It recognizes the aspects which contain the opinion. Aspect based sentiment analysis is basically used for getting the opinion which hide due to categorization of words i. e. the sentence which contain both positive and negative opinion.
- **Comparative sentiment analysis**:-In comparative sentiment analysis we get opinions by comparing it with similar product. In this we use sentiment lexicon acquisition which contains three approaches manual approach, Dictionary based approach and corpus based approach.

II. SENTIMENT ANALYSIS BASIC WORKING

The basic working of sentiment analysis includes collection of data from blogs, websites and reviews. The collection of data from different sources acts as a corpus. After that we process the documents by excluding the statements which are not having any sense. Next step is to apply different techniques for analysis for getting result.

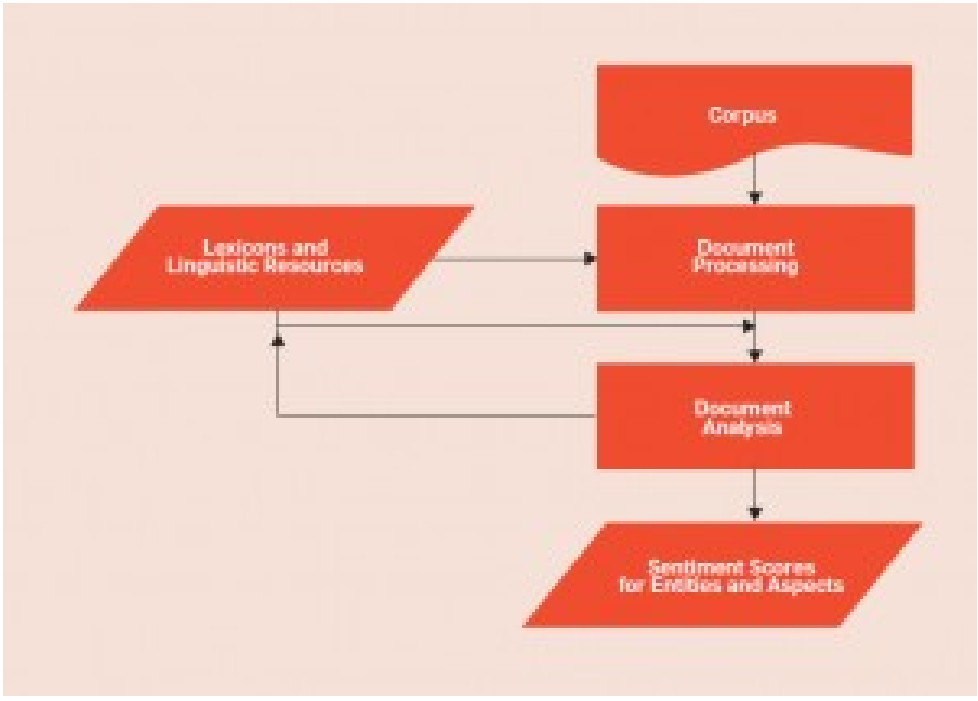

Fig. 1. Architecture of sentiment analysis

III. LITERATURE REVIEW

Marc Cheong et al[1]detected the hidden patterns from twitter messages. They used two techniques (a) the Cheong and Lee's content analysis framework and (b) SOM algorithm for visualization. For data gathering they used Twitter message corpus. They searched messages which helped them in getting useful data for clustering. They also used massage attributes and user attributes. In message attributes they determine the common message indicator, pictures and the type of device used by twitter user. In user attribute they determine the habits of user by checking their post, account age and friends following.

For clustering and visualization they use SOM algorithm. SOM is a technique based on artificial neural network. In SOM clusters are made in the form of Map.

In this paper the authors consider three topics:-

1) Iran election issue in 2009.
2) The iPhone OS 3. 0 software launch
3) US President Obama's foreign policy.

In SOM different colors like red, blue, green and yellow are used for visualizing result.

Tetsuya et al[2] used sentiment analysis approach to check the positive and negative opinion for specific topic from a document. In this paper the positive indicate favourable opinion and negative indicate unfavourable opinion.

NLP is used for capturing favorability. There are three identifications used:-

(a) Sentiment expression
(b) Polarity and strength of the expression
(c) Their relationship to the subject

The authors used notation which consist of the following information:-

- Polarity:-positive (good), negative (bad), or neutral is denoted by g, b, or n, respectively, and sentiment transfer verbs are denoted by t.
- Part of speech (POS):-Currently, adjective (JJ), adverb (RB), noun (NN), and verb (VB) are registered in our lexicon.
- Sentiment term in canonical form.
- Arguments such as subject (sub) and object (obj) that receive sentiment from a sentiment verb or arguments that provide sentiment to and receive sentiment from a sentiment transfer verb.

In this paper the algorithm was used which contain upper limit of 50 words before and 50 words after. For identifying sentiment expression POS tagging was used. For POS tagging, author used Markov-model based tagger. In this paper the techniques fails to identify positive result if there is any negative word in the sentence. For example, "it is difficult to take a bad picture from this camera ". The result of this sentence polarity was negative according to above algorithm because the sentence containing word 'bad'.

Alena et al [3] proposed Sentiful is a lexicon based system for assigning scores to word for analysis of sentiment. In this paper the structure relations are purposed. The positive score and negative score tell about the polarity of a sentence. The author uses sentiword net for giving value to words on the basis of polarity.

For generating the sentiment lexicon the first step is to collect the content of word. There are four formulas are used to calculate positivity and negativity:-

$$Pos\_score = \left[\frac{\sum_{i=1}^{pos} Intensity(i)}{pos}\right], \quad (1)$$

$$Neg\_score = \left[\frac{\sum_{i=1}^{neg} Intensity(i)}{neg}\right], \quad (2)$$

$$Pos\_weight = \left[\frac{pos}{pos + neg}\right], \quad (3)$$

$$Neg\_weight = \left[\frac{neg}{pos + neg}\right], \quad (4)$$

There are three scores object score, positive score and negative score. The score range from 0. 0 to 1. 0 and the sum up to 1. 0. If the positive score is greater than negative score the word is positive and if negative score is greater than positive score then the word is negative else if object score is greater and negative positive score are equal then the word is neutral.

There are four methods for Expanding Sentiful:-

(a**)** **Finding new lexical units through synonymy relation**:-In this method we derive new words from synonymy relation. In this authors consider the words which are having same meaning and assign sentiment score to them.

(b) **Examining direct antonym relation**:-In this method authors consider the words which having opposite meaning. For example, carelessly and carefully.

The authors calculate the positive and negative score of word through this relationship. By this method new words are added to WordNet.

(c) **Examining Hyponymy relation**:-In this method the relations based on hierarchy between words. For example attainments => success => winning. By this method we get a list of hyponyms from word net and it also remove the duplicate words from word net.

(d) **Method to derive and score morphologically modified words**:- Morphologically words are derived through suffixes and prefixes. The meaning of many suffixes can change the meaning of word especially in term of sentiment.

In this paper the algorithm of derivation and scoring of new words give high accurate result. Through sentiful method we can expand sentiment lexicon and improve the result of polarity. The different tables in this paper help us in examine the words polarity.

Neethu Mohandas et al[4] focused on mood extraction from Malayalam text. The authors used NLP for part of speech tagging and PMI formula for SO calculation. The semantic orientation (SO) gives the result in numerical form through which we can check the polarity of sentence.

This paper gives idea to implement sentiment analysis on different languages such as Hindi, Punjabi, and Marathi etc. By this method we can get result on applications contain different languages. We can calculate sentiment analysis of different movies which are based on area language and their comments are also in that language.

Peter D. Turney et al [5], used semantic orientation concept to calculate the positive and negative review. The author uses average semantic orientation to get result of review in positive and negative. In this paper the author explain point wise mutual information (PMI) and latent semantic analysis technique.

The author done experiment by using these technique and calculate the result on the basis of words which they included in the database. The authors also compare the result from both techniques on different data and explain about technique with experimental result. They show result in the form of diagram also which is in graphical form.

Julia Kreutzer et al[6] give information about sentiwordnet. In this paper the author explains the applications and usage of wordnet. The authors also explain the structure of sentiwordnet. They also give examples of scores and scoring review of words. The authors also explain about the analysis and scoring technique. They also include the things which are recently updated in the sentiwordnet and also show the scores of positive, negative and objective word. The authors also explain the score of various word which doesn't provide any sentiments such as grandmother, brother etc. The author also explains the problems in sentiwordnet.

OM P. Damini et al[7] explained the concept of PMI. PMI is basically used to measure the co-occurrence between two words. For calculating PMI by various methods they used different formulas. They evaluate their result by correlating their

result with gold_standard dataset. The author used two types of co-occurrence i. e. document level and corpus level.

IV. OBJECTIVE

We introduce two main techniques for sentiment analysis. The first technique is PMI which is used to check the co-occurrence of two words. By this technique we can analyze which part of sentences have more positive result and which words are co-occur more. After that we use sentiwordnet formula for calculating the positive and negative score by using values which are present on sentiwordnet. By calculating the positive and negative values we can easily calculate semantic orientation. The techniques which we are using make result more accurate and can describe the polarity of data. The tagging of data is also useful which provide small-small parts of sentence in different categories which help in analyzing the sentence.

V. RESEARCH METHODOLOGY

In this methodology we first collect the sentences from social networking sites such as Facebook and twitter. After that collocation function is used this function is already predefined in the Ipython. Through this function we get the words which occur together in the text. Now we use PMI method for counting the co-occurrence of words many times in the text manually.

After this we use tagging of text by using some lines of code. After executing the code we get the adjectives and adverbs from the text. After that we apply sentiwordnet code for finding the value of words then we manually write the values and find the average of positive and negative value of words. After that we apply semantic orientation by subtracting negative values average from positive value average. If we get the positive value as result this means the text is positive either if negative result then the result is negative.

There are some points which is considered during this methodology:-

1. If there are two negative values together then it became the positive value.
2. If there is any word value which we don’t get in the sentiwordnet through python then we can check the value from website.
3. If there is any positive word with not then its value is considered as negative value.

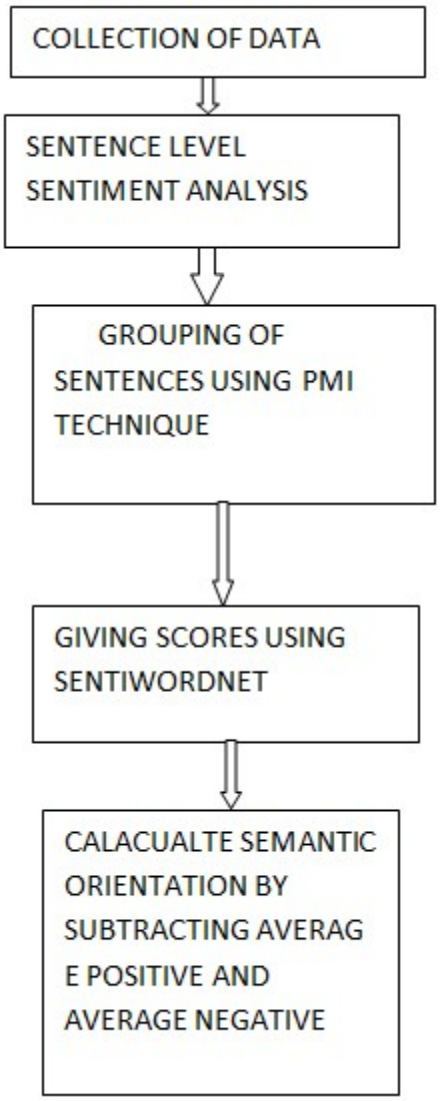

Fig. 2. Research Design.

VI. IMPLEMENTATION AND RESULT

In this section the proposed method implemented result is shown. We implemented the methodology in Ipython which is included in Anaconda tool. We take comments of people on Samsung mobile through Facebook page of Samsung. We take 200 comments line. These comments included simple as well as complex sentences. We first remove those sentences which don't include any adjective and adverb.

After that we count the words which are repeated according to our categories in the text file and we also check the words through collocation method. Along with this we check the sentiwordnet value of word available in Ipython.

The screenshots of sentiwordnet value:.

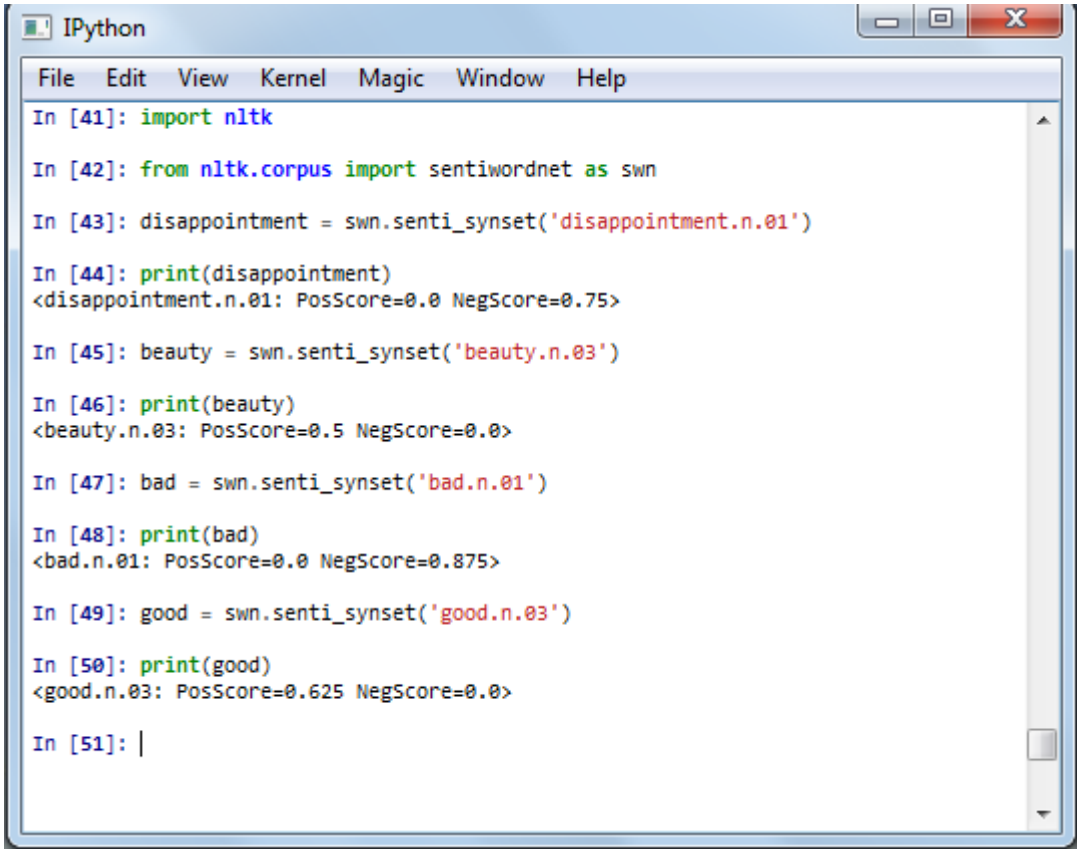

Fig. 3. Sentiwordnet value of words.

For polarity result i. e. semantic orientation we subtract value of average positive from average negative.

Avg positive=0. 625+0. 5/2=0. 5625

Avg negative=0. 875+0. 75/2 =0. 8125

Semantic orientation =-0. 25

The result is negative so the opinion of people to the product is negative.

TABLE OF WORDS WITH POSITIVE AND NEGATIVE VALUE

S. NO	WORDS	PVALUE	NVALUE
1	BEAUTIFUL	**0. 5**	**0. 0**
2	GOOD	**0. 625**	**0. 0**
3	BAD	**0. 0**	**0. 875**
4	DISAPPOINTMENT	**0. 0**	**0. 75**

PVALUE=POSITIVE VALUE, NVALUE=NEGATIVE VALUE

Tab. 2. Results comparison of different methods.

S. NO	TECHNIQUE	ACCURACY
1	**SENTIFUL**	**86. 3%**
2	**SVM**	**80. 09%**
3	**NAÏVE BAYES**	**85. 8%**
4	**PMI**	**82. 8%**
5	**PROPOSED METHODOLOGY**	**87. 2%**

VII. CONCLUSION

Sentiment analysis provides us the feelings of people about any product or subject. Companies needed sentiment analysis for decision making and updating in product. Sentiment analysis provides favourable and unfavourable opinions about any product or any particular subject.

The techniques such as NLP, PMI and sentiful analysis provide us useful results. These techniques give us result in the numerical form or values. The result by these values helps us to analyse result and compare it with other techniques. The result by semantic orientation tells about the polarity of data. The more is positive value in semantic orientation the more positive is the opinion of people. The analysed result from sentiment analysis provides various advantages in decision making in business.

VIII. REFRENCES

Journals

[1] Marc Cheong, Vincent Lee in, 2010, “A Study on detecting patterns in twitter Intra-topic User and message Clustering”, IEEE.

[2] Tetsuya Nasukawa, Jeonghee Yi in, 2003 “Sentiment Analysis: Capturing favorability using natural language processing”.

[3] AlenaNeviarouskaya, Helmut Prendinger, Mitsuru Ishizuka in, 2011, ” Sentiful: A lexicon for sentiment analysis”, IEEE.

[4] Neethu Mohandas, Janardhanan PS Nair, Govindaru V in, 2012, ” Domain Specific Sentence Level Mood Extraction from Malayalam Text”, IEEE.

[5] Peter D. Turney, MichaelL. Littman, ”Measuring praise and criticism :Inference of Semantic Orientation from association”.

[6] Julia Kreutzer and Neelewitte, 2013-14., Opinion Mining using Sentiwordnet".

[7] OM P. Damani, October, 2013, "Improving pointwise mutual information(PMI) by incorporating significant co-occurrence".

[8] B. Huberman, D. Romero, and F. Wu, 2008, "Social networks that matter Twitter under microscope".