
Mohammad Shabaz, Urvashi Garg

Abstract: It has been discussed several times about the uses and applications of sentimental analysis which is the extractions of opinions, feelings or emotions of an individual or an organization through some of their descriptive data which is freely available. There are several approaches exists which are helpful in differentiating individual’s opinions by evaluating data which is present over social networking sites in the form of comments or tweets or posts, either through classification or clustering. Number of algorithms working over evaluation of sentiments or opinion which includes AS, naïve-base, k-means etc. In this paper we are going to introduce US_Frequency formula which provides the US_Frequency value that helps to find the most frequent used word from the data-set extracted from social network data which in result predict the common patterns found. The data-set which is used in this paper consists of live data extracted from social networking sites using API’s. The patterns are based on the similarity between different individual exists in the same group or profession.

Index Terms: Sentimental Analysis, Patterns Evaluations, Predictions, Opinion mining, Text Mining.

I. INTRODUCTION

Most of the data available over social networking site is in raw form, the major challenge is to transform this raw data into meaningful information. During this transformation large amount of data either scrapped or manipulated which creates biasedness. Thus we are unable to achieve the appropriate results. In order to tackle with these situation, it is necessary to create verified connectivity with the server of social networking sites. This connectivity can be done using API’s. Through this article we have showed the verified connectivity with the social networking sites for extracting the live data.

II. LITERATURE REVIEW

Devendra K. Tayal and Yadav Sumit Kumar:[1] The authors use bloom filter to analyse large set of data using Hadoop and formulate the sentiments to retrieve the results accurate and rapidly. This paper also shows the inverse proportionality of data size with its error probability.Min Zhao, Tao Zheng Zhang and et.al: [2]Using sentiment-oriented pointwise mutual information technique the authors find the words frequency and its intensity and thus evaluate the sentimental words for television programs.Beiming Sun and Vincent TY Ng: [3] Differentiating positive and negative words the authors compare different datasets that influences sentiments of an individual. Hase Sudeep Kisan, Hase Anand Kisan and et.al: [4]The authors use twitter data and performs sentimental analysis that works better over cloud. This work analyse the data using hashtag. The author evaluates the positive and negative sentiments of individual and checks collective intelligence.Jiao Huang, Haobo Sun and et.al: [5]Using semantic analysis the authors evaluate the comments of Chinese on the performance of mobile. The technique is also helpful to find the sentimental analysis. The authors focus on the neutral words rather than positive and negative words.Zhongyang Guo, Zengchang Qin and et.al: [6]Introduces mixed gram terms based model to classify sentiments. Static word vector are also used by the author for sentiments classification. The author initiates the learning process for a machine through word vector and mixed gram term using sentimental analysis. Feldman [7] introduces fundamental approach which is applicable on sentiment analysis. His main contributions are the collection of different research problems and acquire some most relevant ones into account. He divides sentiment analysis into document and sentence level analysis. lexicon based features are also covered under sentiment analysis.H. Du, B. Li and et.al: [8] The authors asked various questions based on sentimental analysis which includes about number of citations and papers on sentimental analysis and tries to find the different field associated with sentimental analysis.

III. PROPOSED METHODOLOGY

<table>
<thead>
<tr>
<th>NOMENCLATURE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>w</td>
<td>Frequency of word</td>
</tr>
<tr>
<td>T_s</td>
<td>Total no. of stemmed words</td>
</tr>
<tr>
<td>T_n</td>
<td>Not Stemmed words</td>
</tr>
<tr>
<td>T_w</td>
<td>Total number of words</td>
</tr>
</tbody>
</table>

The Proposed methodology depicts the actual functionality from start to end. The methodology is implemented using following points:
1. Connect to various Api’s of Social Networking Sites.
2. Extract data from comments and posts using keywords.
3. Perform random sampling of comments and posts of different individuals.
4. Calculate the frequency of particular word used in different posts/comments.
5. Calculating the frequency value lies between (0 to 1) of that particular word using US_Frequency value formula (Equation 1)

6. Based on the increasing order of the US_Frequency value arrange the words in data-warehouse repository.

7. Represent on bar-graph the statistics.

8. Evaluate common patterns found in comments and posts.

9. Further classify the words and predict the common patterns.

10. End the process.

US_Frequency formula provides the US-frequency value which helps to find the most frequent used word in the data-set. The formula is given by:

US_Frequency = Frequency of word (W) / Total no. of stemmed words (Ts)

OR

US_Frequency = W / (Ts + W) Equation (1)

Let, Total number of words be Tw

Stemmed Words be Ts

Not Stemmed words be Tn

Then, Tw = Ts + Tn Equation (2)

Also, Tw = n * m Equation (3)

where n is number of sentences and m is mean of words falls in sentences.

Now using Equation 1 we get,

US_Frequency = W / (n*m + W + Tn)

IV. DATASET

Starting by evaluating the common patterns which we found while connecting with different social networking sites through their API’s. We display all such findings using bar graph. We choose popular three of the following (See Section 1) and tried to find the common set of words using US_Frequency Formula.

Section 1

(a) Private Universities: We have chosen the three popular private universities namely Chandigarh University, Amity University and Lovely Professional University and get the most recent data about them from different social networking sites and display their most common used word as bar graph. The following data shows our findings. Figure 1, Figure 2 and Figure 3 respectively shows them.

(b) News Anchors: Three popular Indian News Anchor namely Abhisar Sharma, Anjana Om Kashap and Sudhir Choudhary are selected at random and we have tried to evaluate the common and frequent set of words they are often using on social media/sites. Figure 4, Figure 5 and Figure 6 shows this finding.
(c) **Entertainment Televisions:** Three popular Indian Entertainment Channels namely Sony TV, Zee TV and Star Plus are selected at random and we have tried to evaluate the common patterns. Figure 7, Figure 8 and Figure 9 shows this finding.

(d) **Politicians:** Three popular Indian Politicians namely Assadudin Owaisi, Jyotiraditya Scindia and Narendra Modi are selected at random and we have tried to evaluate the common and frequent set of words they are often using on social media/sites. Figure 10, Figure 11 and Figure 12 shows this finding.
Figure 11 Shows data of Jyotiraditiya Scindia

Figure 12 Shows data of Narendra Modi

Figure 13 Shows data of Shah Rukh Khan

Figure 14 Shows data of Ranveer Singh

Figure 15 Shows data of Priyanka Chopra.

(e) Bollywood Stars: Three popular Indian Bollywood Stars namely Shah Rukh Khan, Ranveer Singh and Priyanka Chopra are selected at random and we have tried to evaluate the common and frequent set of words they are often using on social media/sites. Figure 13, Figure 14 and Figure 15 shows this finding.

V. RESULTS AND DISCUSSION

From all the above shown findings that we generated using R-Programming we found the common and frequent words that every different professional uses on social media/sites. All such results are discussed later on in Result Section.

From all the above shown graphs in Fig 1-15 that we generated using R-Programming on R-Studio, we found the common and frequent words that every different professional uses on social media/sites. All Such Common words along with their US_Frequency Value which is obtained from the US_Frequency Formula (Eq. 1) can be shown in Table 1.

At, Value of $n=500$, $m=14$.

<table>
<thead>
<tr>
<th>Word</th>
<th>US_Frequency</th>
<th>Used By</th>
</tr>
</thead>
<tbody>
<tr>
<td>today</td>
<td>0.07</td>
<td>Politician</td>
</tr>
<tr>
<td>pluralism</td>
<td>0.12</td>
<td>Politician</td>
</tr>
<tr>
<td>congress</td>
<td>0.21</td>
<td>Politician</td>
</tr>
<tr>
<td>zero</td>
<td>0.21</td>
<td>Bollywood Actor</td>
</tr>
<tr>
<td>best</td>
<td>0.20</td>
<td>Bollywood Actor</td>
</tr>
<tr>
<td>said, just</td>
<td>0.125</td>
<td>News Anchors</td>
</tr>
</tbody>
</table>
TABLE 1 shows results obtained by common patterns.

<table>
<thead>
<tr>
<th>Word</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attend, admission</td>
<td>0.4</td>
</tr>
<tr>
<td>Suicidal</td>
<td>0.06</td>
</tr>
<tr>
<td>vote</td>
<td>0.071</td>
</tr>
<tr>
<td>zee</td>
<td>0.15</td>
</tr>
</tbody>
</table>

The Words which are described in Table 1 are used by only those professional which are described in the Introduction Section. These patterns may be changed based on their free available data. The sources of data are social networking sites such as twitter, Facebook etc.

VI. CONCLUSION

In this paper we have introduced US_Frequency formula which provides the US_Frequency value, helps to find the most frequent used word in the data-set extracted from social network which in result predict the common patterns found. The patterns are based on the similarity between different individual exists in the same group or profession.

ACKNOWLEDGMENT

We would like to acknowledge all the people who are either directly or indirectly supported us.

REFERENCES


AUTHORS PROFILE

Mohammad Shabaz is pursuing Ph.D. in Computer Science Engineering from Chandigarh University, Mohali, India.

Dr. Urvashi Garg is Ph.D. in Computer Science Engineering and working as Associate Professor at Chandigarh University, Mohali, India.