Analyzing social media remark using sentimental analysis

[1]T.Devishree, [2] K.S.Jeen Marseline

[1]Research Scholar, [2]Head
Department of Information Technology
Sri Krishna Arts and Science College, Coimbatore, TamilNadu, India

***Abstract*: Sentiment analysis is the process of determining whether a piece of writing is positive, negative, neutral. It is also known as opinion mining, deriving the opinion or attitude of speaker. Sentiment Analysis is contextual mining of texts which identifies and extracts subjective information in source material and helping a business to understand the social sentiment of their brand, product or service while monitoring online concentration. Sentiment Analysis is the most common text classification tool that analyses an incoming message and tells whether the underlying sentiment is positive, negative or neutral. This paper deals with two forms which is *polarity-based*, where pieces of texts are classified as either positive or negative, and *valence-based*, where the intensity of the sentiment is considered.**

Index Term: Sentimental analysis, score

I. INTRODUCTION

Sentiment Analysis is simply the process of working out (statistically) whether a piece of text is positive, negative or neutral. The majority of sentiment analysis approaches take one of two forms: **polarity-based**, where pieces of texts are classified as either positive or negative, or **valence-based**, where the intensity of the sentiment is taken into account. For example, the words 'good' and 'excellent' would be treated the same in a polarity-based approach, whereas 'excellent' would be treated as more positive than 'good' in a valence-based approach. Sentiment analysis has applications across a range of industries - it's great for anything where one can get unstructured opinion data about a service or product. One application of sentiment analysis is for companies that have Twitter or other social media accounts to receive feedback. Obviously it's bad business for these companies to leave negative feedback unanswered too long, and sentiment analysis can give them a quick way to find and priorities these unhappy customers.

II. PROPOSED SYSTEM

In the broad field of Sentiment Analysis, here performed a piece of work by analyzing the tweets of people on the trend topic. The people's feel and words differs from each of individual towards the topic, their unique feel is important and also it's very challenging for the analysis process. First part of this paper, say about the technique tool used for the sentiment classification. VADER (Valence Aware Dictionary for sEntiment Reasoning) is a model used for text sentiment analysis that is sensitive to both polarity (positive, negative) and intensity (strength) of emotion. This belongs to a type of sentiment analysis that is based on lexicons of sentiment-related words. In this approach, each of the words in the lexicon is rated as to whether it is positive or negative, and in many cases, how positive or negative. In Table.1 below it's an example from VADER's lexicon, where more positive words have higher positive ratings and more negative words have lower negative ratings.

Table.1

Sentiment metric	**Value**
Positive	0.45
Neutral	0.25
Negative	0
Compound	1

Primarily, VADER sentiment analysis relies on a dictionary which maps lexical features to emotion intensities called sentiment scores. The sentiment score of a text can be obtained by summing up the intensity of each word in the text. When it analyses a piece of text it checks to see if any of the words in the text are present in the lexicon. For example, the sentence "The life is good and the atmosphere is nice" has two words in the lexicon (good and nice) with ratings of 1.9 and 1.8 respectively. That produces four sentiment metrics from these word ratings, which you can see below. The first three, positive, neutral and

negative, represent the proportion of the text that falls into those categories. Table.2 As in the example sentence was rated as 45% positive, 55% neutral and 0% negative. The final metric, the compound score, is the sum of all of the lexicon ratings which have been standardized to range between -1 and 1.

Table. 2

Word	Sentiment Rating
Tragedy	-3.5
Happy	2.0
Danger	-3.1
Great	3.0
Insane	-1.1

Mechanism of VADER:

VADER sentiment analysis takes this into account by amplifying the sentiment score of the sentence proportional to the number of exclamation points and question marks ending the sentence. It first computes the sentiment score of the sentence. If the score is positive, it adds a certain empirically-obtained quantity for every exclamation point (0.252) and question mark (0.19). If the score is negative, it subtracts. There are other contextual elements, like punctuation, capitalization, and modifiers which also impart emotion. It takes these into account by considering five simple heuristics.

The first heuristic is **punctuation**. Compare "I like it." and "I like it!!!" It's not really hard to argue that the second sentence has more intense emotion than the first, and therefore must have a higher VADER sentiment score.

The second heuristic is **capitalization**. "SPLENDID." is definitely more intense than "splendid performance." And so it takes this into account by incrementing or decrementing the sentiment score of the word by 0.83, depending on whether the word is positive or negative, respectively.

The third heuristic is the use of **degree modifiers**. The effect of the modifier in the sentence is to increase the intensity, while in the second sentence, if it is to decrease the intensity. It maintains a booster dictionary which contains a set of boosters and dampeners.

The fourth heuristic is the **shift in polarity due to "but"**. Oftentimes, "but" connects two clauses with contrasting sentiments. The dominant sentiment, however, is the latter one. For example, "I love ice-cream but I don't want to eat it anymore." The first clause "I love ice-cream" is positive, but the second one "I don't want to eat it anymore." is negative and obviously more dominant sentiment-wise.

It implements a "but" checker. Basically, all sentiment-bearing words before the "but" have their valence reduced to 50% of their values, while those after the "but" increase to 150% of their values.

The fifth heuristic is **examining the tri-gram before a sentiment-laden lexical feature to catch polarity negation**. Here, a tri-gram refers to a set of three lexical features. It maintains a list of negation words. Negation is captured by multiplying the sentiment score of the sentiment-laden lexical feature by an empirically-determined value -0.74.

III. IMPLEMENTATION

As said already that sentiment analysis is such a broad field and it grows day by day, this work have implemented in Python language which runs on emerging platform named Anaconda. Python is an interpreted high-level language for general-purpose programming. Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural and has a large comprehensive standard library. Here performed a piece of work by analyzing the tweets of people on the current trend topic on twitter. VADER outputs sentiment scores for a piece of text. We need to load the SentimentIntensityAnalyser object in from the VADER package, we'll assign it to another name *Sid*, to make it a bit easier to use. The implement of VADER package coded as:

```
from nltk.sentiment.vader import SentimentIntensityAnalyzer

sid = SentimentIntensityAnalyzer()
```

Finally, we'll use the polarity_scores () method to get the sentiment metrics for a piece of text.

```
sid.polarity_scores(sentence)
```

VADER doesn't just do simple matching between the words in the text and in its lexicon. It also considers certain things about the way the words are written as well as their context. Let's see few sentences and scores from the work.

(i) It all begins today! I will see you at 11:00 A.M. for the swearing-in. THE MOVEMENT CONTINUES - THE WORK BEGINS!

neg: 0.0, neu: 1.0, pos: 0.0, compound: 0.0,

Here, the capitalizing increases the intensity of positive. One of the things that VADER recognizes is capitalization, which increases the intensity of both positive and negative words.

(ii) Wow, television ratings just out: 31 million people watched the Inauguration, 11 million more than the very good ratings from 4 years ago!

neg: 0.0, neu: 0.733, pos: 0.267, compound: 0.8066,

(iii) Thank you, @Samsung! We would love to have you..!!

neg: 0.0, neu: 0.49, pos: 0.51, compound: 0.8065,

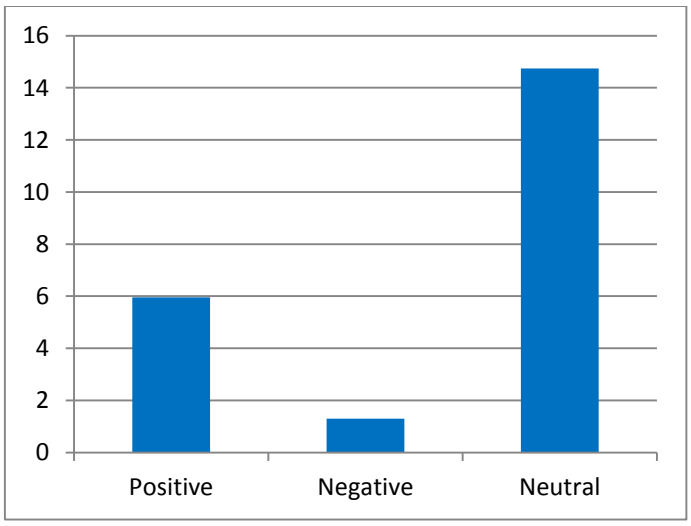
Here, rather the exclamation symbol increases the scores than a simple text.

(iv) Despite what you hear in the press, healthcare is coming along great. We are talking to many groups and it will end in a beautiful picture!

neg: 0.0, neu: 0.735, pos: 0.265, compound: 0.8516,

The tweet comments which have been scored individually and the scores have plotted as graph and segregated into positive, negative and neutral. This is implemented to know overall sentiment of the tweet comments of the people. As for the sample, have analyzed few comments and the plot (fig.1) will be like:

Fig.1 Plot for the sentimental analysis



According to the plot, from the whole tweet comments using the sentiment scores which conclude the overall sentiment of the peoples' are neutral which we could identify by the scores calculation that attain from the sentiment scores and sentiment intensity analyzer.

IV. CONCLUSION

The complex information analyzed and the voluminous views of people classified accordingly to the sentiment. The tweets comments of the people on the trend topic, the datasets acquired from the data site and does the NLTK VADER sentiment intensity analyzer. That classifies the sentiment from the sentences as positive, negative and neutral by the method of sentiment scores. After attain the scores and then perform plot to know overall sentiment from the people's comments.

REFERENCES

[1] Feldman, R., 2013. Techniques and applications for sentiment analysis. Communications of the ACM 56, 82–89.

[2] Liu, B., Zhang, L., 2012. A Suvery of opinion mining and sentiment analysis, in: Aggarwal, C.C., Zhai, C. (Eds.), Mining Text Data. Springer US, Boston, MA. chapter 13, pp. 415–464.

[3] RuiXia,FengXu,JianfeiYu," Polarity shift detection, elimination and ensemble: A three stage model for document-level sentiment analysis" Information Processing and Management 52 (2016) 36– 45.

[4] A. Hassan, A. Abbasi, and D. Zeng, "Twitter sentiment analysis: A bootstrap ensemble framework," in Social Computing (SocialCom), 2013 international Conference on. IEEE, 2013, pp. 357-364.

[5] K. Mouthami, K. N. Devi, and V. M. Bhaskaran, "Sentiment analysis and classification based on textual reviews," in Information Communication and Embedded Systems (ICICES), 2013 International Conference on. IEEE, 2013, pp. 271–276.