VIRTUAL EMOTION DETECTION BY SENTIMENT ANALYSIS

Rahul Kamdi

Assistant Professor, Yeshwantrao Chavan College of Engineering, Nagpur Maharashtra, (India).

Prasheel N. Thakre

Assistant Professor, Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, (India).

Ajinkya P. Nilawar

Assistant Professor, Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, (India).

J. D. Kene

Assistant Professor, Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, (India).

E-mail: jagdish.kene@gmail.com

https://doi.org/10.17993/3ctic.2022.112.175-181

ABSTRACT

As websites, social networks, blogs, and online portals proliferate on the internet, authors are producing reviews, opinions, ideas, ratings, and feedback. The emotional content of this writer may be about things like books, people, hotels, items, studies, events, and so on. These emotions have great value for businesses, for governments, and for people. The majority of the writer-generated material requires the usage of text mining algorithms and sentiment analysis, even if this information is meant to be instructive. Sentiment Analysis is a technique in Natural Language Processing (NLP) that tries to identify and extract assessments communicated within a given text. This paper intends to execute different content handling strategies in NLP and use of Valence Aware Dictionary for Sentiment Reasoning (VADER) Model that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion.

KEYWORDS

Natural Language Processing (NLP), Polarity Intensity, Sentiment Analysis, Virtual Emotion Detection, VADER.

1. INTRODUCTION

The emotions are of primarily of six types - love, happiness, anger, sadness, surprise and fear. Sentiment analysis is used to analyze the emotions present in a text. It is a method for determining if a given piece of writing is positive, negative, or neutral. The goal of sentiment analysis is to determine the writer's attitude, sentiments, and emotions in a written text using a computational treatment of subjectivity in the text. By application of sentiment analysis one can decipher the given sentence, paragraph or a document contains a positive or negative emotions or expressions in it. In sentiment analysis we classify the polarity of given text, it results by telling about opinion whether it's positive, negative or neutral.

2. LITERATURE SURVEY

Sentiment analysis has multiple ways and Vader is one of the best way (Mozetic et al., 2016) which is being used. Vader stands for Valence Aware Dictionary and Sentiment Reasoning. It works on a ruled based sentiment analysis and it contains list of lexical features which are labeled as per semantic orientation. By analyzing the intensity of wordings in the text, the sentiment score can be obtained of that text. Vader is smart enough to extract meaning of these words or texts as positive sentiments and words like Sad, bad, awful as negative sentiments. Vader only cares about the expressions in the text. Opinions like positive negative or neutral are the expression which is concerned for Vader.

Steven Bird & Loper (2009) express when a text is changed into its canonical or standard form, then it is called as Text normalization. A few processes have to be done to standardize the content and convert it into fitting structure which would then be given to the machine learning (ML) model. This helps in reducing unnecessary information that the computer does not require, thus subsequently improving efficiency. Library utilized for this is given in. Steps associated with this cycle are shown in Figure 1 and explained briefly in the following sections.

Fig 1. Text Normalization.

2.1 REMOVING STOPWORDS

A stopword is an ordinarily utilized word that can be disregarded, both when ordering sections for looking and while recovering them (Saif et al., 2014). Using pre-compiled stopword lists or more complex algorithms for dynamic stopword recognition, removing stopwords from textual data is a popular procedure for reducing noise. The Natural Language Toolkit (NLTK) in Python includes a list of stopwords for 16 different dialects.

2.2 TOKENIZATION

The method of breaking text into smaller units called tokens is commonly known as Tokenization (Stanford NLP Group, 2015). Tokens can be words, characters, or subwords in this case. As a result, tokenization can be divided into three categories: word, character, and sub-word tokenization.

2.3 DERIVING ROOT WORD

Nicolai & Kondrak (2016) explains in the zones of Natural Language Processing we come across circumstances where a word has many offshoots Stemming and lemmatization are the two main NLP

processes for generating root words. The root type of an inflected word is produced by both stemming and lemmatization (Samir & Lahbib, 2018). Stemming computation works by removing the word's postfix. Lemmatization thinks about morphological examination of the words. It restores the lemma which is the base type of all its inflectional structures.

2.4 FEATURE EXTRACTION

Machine learning calculations are unable to work on Crude content legitimately. The procedure of component extraction requires converting content into matrix or in vector form. The module can be used to extract features from database consisting of formats such as text and images and extract features in a format supported by machine learning algorithms the most popular strategies that includes feature extraction are Bag-of-Words and TF-IDF Vectorizer. Normalize with diminishing important tokens that appears in majority samples/documents (Mahajan et al., 2020).

2.5 POLARITY AND INTENSITY SCORE IN EMOTIONAL ANALYSIS

A key element of emotional analysis is to examine the body of a text to understand the concept it expresses. Emotional analysis is appropriate for positive or negative values, known as polarity. The perfect situation usually ends up being good, neutral or bad with the help of a polarity point calculation. In general, emotional analysis works better in a submissive text than in a single context text of purpose. Emotional analysis is widely used, as part of the analysis of social media in any domain, to understand the functioning of any system, to be controlled by people and that their response is based on their opinions.

Text textual analysis data can be calculated at most levels, either at the level of each sentence, at the paragraph level, or throughout the document. There are two major theories in emotional analysis. First is Learning Prescribed machine reading or in-depth learning: In this approach, traditional machine learning techniques with a TF-IDF model using the n-gram method. These divisions are the mindless Bayes of many lands, the orderliness of things, the closest neighbor k and the uninhabited forest. In all four classes, orderliness is achieved with minimal accuracy. Second is unsupported dictionary control: This method is to use a large learning process. The accuracy we get from reading a lot is much less than how to learn by machine. After obtaining excellent performance and fragmentation, the next step is to create a final model for back-to-work using certain advanced machine learning methods.

2.6 VALENCE AWARE DICTIONARY FOR SENTIMENT REASONING (VADER)

VADER is a model used for analysis of text sentiment from which it can detect both polarity (positive/negative) and emotion intensity or strength. VADER majorly relies on a dictionary that matches the lexical features to emotion intensities also known as sentiment scores for sentimental analysis (Beri, 2020). By summing up the intensity of each word in the text we can get the sentiment score of a text. Sentiment analysis statistically detects whether the polarity of a piece of text is negative or positive. Sentimental analysis is based on two approaches: polarity-based analysis, in which texts are classed as either negative or positive, and valence-based analysis, in which the intensity of the emotion or sentiment is considered.

2.7 WORKING OF VADER

VADER is a sentiment or emotion analysis method that uses lexicons of sentiment-related words. Each word in the lexicon is classed as positive or negative, and the strength of positivity or negativity is also examined using this method. Table 1 depicts the sentiment rating of an excerpt from VADER's lexicon, with higher positive ratings for more positive words and lower negative ratings for more negative terms.

Table I. Sentiment rating of the various words in a text.

Word	Sentiment rating
tragedy	-3.3
rejoiced	2.0
insane	-1.6
disaster	-3.2
great	3.3

According to Hutto & Gilbert in 2014, to determine if these terms are favorable or negative, a group of individuals must personally rate them, which is both time-consuming and costly (Hutto & Gilbert, 2014). The lexicon must contain a good coverage of the words of interest in your text; otherwise, accuracy will be poor. When the lexicon and the text are well-matched, this method is quite precise, and it even produces speedy results on vast volumes of text. VADER not only matches the words in the text with its lexicon, but it also takes into account certain aspects of the way the words are written as well as their context meaning.

3. EXPERIMENTATION AND PERFORMANCE ANALYSIS

In this paper, sentiment analysis by the utilization of VADER is performed. Firstly, utilize standardization methods in Natural Language Processing (NLP) for converting the test text into its vector form. The implementation of emotion analysis evaluated in Python using Natural Language Toolkit (NLTK) library. Figure 2 shows the virtual emotion analysis of the text where the emotional content is adjudged as 75% positive. Figure 3 shows the virtual emotion analysis of the text where the emotion is adjudged as 26% positive which is shown in red color depicting 26% negative. Figure 4 shows the sentiment analysis of long text where the sentiment is adjudged as 92% positive. By adding
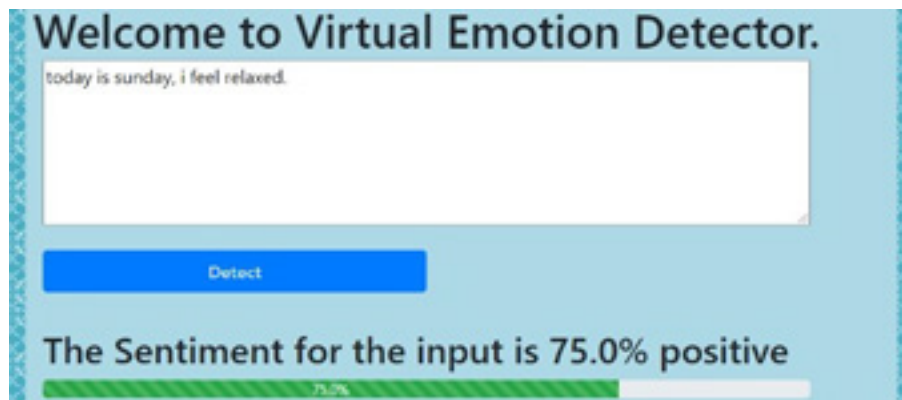
this feature of detecting gibberish, we are able to detect gibberish text where the algorithm returns 50% as can be seen in Figure 5.

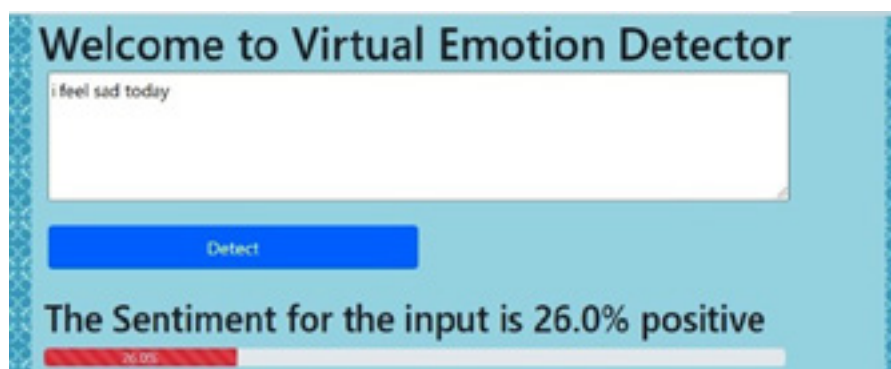Fig 2. Virtual emotion detection of text showing 75% positive polarity.

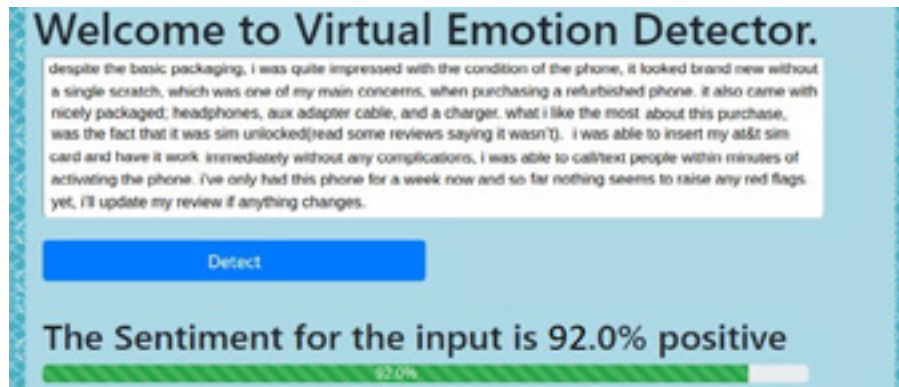Fig 3. Virtual emotion detection of text showing 26% negative polarity.

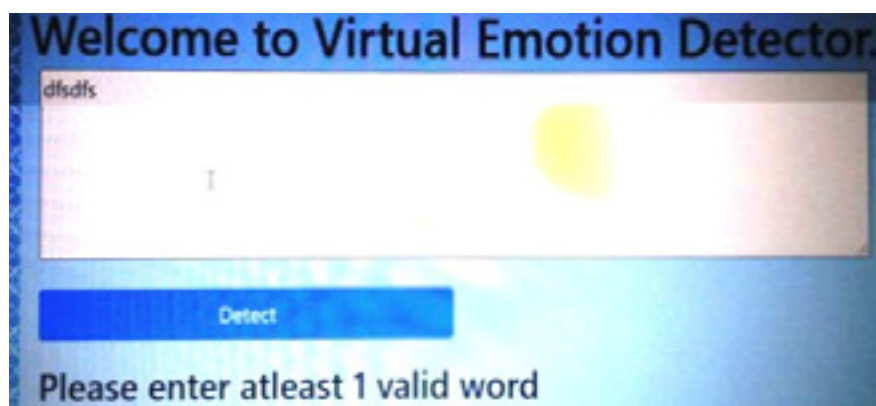Fig 4. Virtual emotion detection of long text showing 92% positive polarity.

Fig 5. Emotion detection of gibberish text.

4. CONCLUSION

The main focus of this paper is to calculate two scores of text: polarity and its intensity by using machine learning. The polarity range is between -1 to 1(negative to positive) and which help us to find whether the text is positive or negative. In this paper we have tested different emotions of a text by using natural language processing toolkit in python and quantified this emotion as positive or negative. Also we were able to identify a gibberish text i.e. words which does not make any sense or are rubbish.

REFERENCES

[1] Mozetic, I., Grcar, M. & Smailovic, J. (2016), 'Multilingual twitter sentiment classification: The role of human annotators', PLoS One 11(5), 1–26.

[2] Steven Bird, E. K. & Loper, E. (2009), Natural Language Processing with Python – Analyz- ing Text with the Natural Language Toolkit, O'Reilly Media.

[3] Saif, H., Fernandez, M., He, Y. & Alani, H. (2014), On stopwords, filtering and data sparsity for sentiment analysis of Twitter, in 'Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)'.

[4] Stanford NLP Group (2015), CoreNLP, Stanford University, Stanford USA. https://stanfordnlp.github.io/CoreNLP/index.html.

[5] Nicolai, G. & Kondrak, G. (2016), Leveraging inflection tables for stemming and lemmati- zation, in 'Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics', Vol. 1, Association for Computational Linguistics, Berlin, pp. 1138–1147.

[6] Samir, A. & Lahbib, Z. (2018), Stemming and lemmatization for information retrieval sys- tems in amazigh language, in M. A. A. N. E. Y. Tabii, M. Lazaar, ed., 'Big Data, Cloud and Applications. BDCA 2018', Vol. 872 of Communications in Computer and Informa- tion Science, Springer, Cham, Kenitra, Morocco, pp. 222–233.

[7] Mahajan, A., Ray, A., Verma, A., Kohad, S. & Thakare, P. N. (2020), 'Sentiment analy- sis using supervised machine learning', International Journal of Advance Research and Innovative Ideas in Education 6(6), 103–109.

[8] Beri, A. (2020), 'Sentimental analysis using VADER', Available online https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664.

[9] Hutto, C. & Gilbert, E. (2014), VADER: A parsimonious rule-based model for sentiment analysis of social media text, in 'Proceedings of the Eighth International AAAI Con- ference on Weblogs and Social Media', Association for the Advancement of Artificial Intelligence, Ann Arbor, MI, pp. 216–225.