Verb Based Sentiment Research

A. Sharada, P. Preethi Krishna

Abstract— Sentiment Analysis is one of the leading research work. This paper proposes a model for the description of verbs that provide a structure for developing sentiment analysis. The verbs are very significant language elements and they receive the attention of linguistic researchers. The text is processed for parts-of-speech tagging (POS tagging). With the help of POS tagger, the verbs from each sentence are extracted to show the difference in sentiment analysis values. The work includes performing parts-of-speech tagging to obtain verb words and implement TextBlob and VADER to find the semantic orientation to mine the opinion from the movie review. We achieved interesting results, which were assessed effectively for accuracy by considering with and without verb form words. The findings show that concerning verb words accuracy increases along with emotion words. This introduces a new strategy to classify online reviews using components of algorithms for parts-of-speech..

Keywords: Parts-of-speech (POS) Tagging, verb words, TextBlob and VADER.

I. INTRODUCTION

The World Wide Web's conversion into a fresh, more participatory web has resulted in multiple rises in its value. Most of the customers share their comments, reviews or opinion about any product or their feature through social media sites on the Internet. Web users in today's web are no longer just customers but co-creators. So, this helps new customers to do analysis very quickly. For all this to happen we need to perform sentiment analysis.

Sentiment analysis has traditionally been a field of linguistics exploration. It is now a more appealing study area with significant progress in a brief period. Sentiment analysis helps to recognize the attitude of people which is expressed in terms of the e-text. The objective of sentiment analysis is to define the emotion of the writer by analyzing the opinion expressed. It focuses primarily on view or opinion that convey positive or negative polarity.

NLTK was established as part of a computer language course. NLTK is a major platform that works with human language information to build Python programs. It offers pleasant interfaces and tools, NLP library and active discussion forum.

This paper is organized by giving the study of an existing system in section 2. Section 3 explains about POS tagging and the type of verb word forms. Section 4 has the implementation of the proposed idea.

Dr. A. Sharada, , CSE Department, G. Narayanamma Institute of Technology & Science, Hyderabad, Telangana, India.
(Email: sharada@gnits.ac.in)
P. Preethi Krishna, CSE Department, G. Narayanamma Institute of Technology & Science, Hyderabad, Telangana, India.
(Email: Krishna.preethi95@gmail.com)

II. LITERATURE REVIEW

There are three levels to analyze the opinions or feelings: document-level, sentence-level, and word-level. This determines the polarity value by considering the entire document, sentence-wise, word-wise respectively in a specified text. The usual way to perform sentiment analysis is at document-level and sentence-level. These levels are categorized as positive polarity or negative polarity by considering the whole document or a sentence. These don't focus on the exact feature which is referred to. Another challenge is the way the situation can be expressed differently from individual to individual.

Aspect level sentiment analysis considers each word's polarity based on the context in the sentence which will lead to better accuracy score. To do word-level analysis POS tagging should be performed. Parts-of-speech tagging (POS Tagging) is a strategy to define and classify opinionated words as positive and negative based on semantic orientation. [5]

III. POS TAGGING

POS Tagger is parts of speech tagger which will assign a tag to a word in a sentence. POS is applied to words based on their syntactic context and role. Sentence when tokenized to perform POS it provides the following categories of tags: Noun, Verb, Adjective, Adverb, Determiner, etc. The tags are referred to as "Tagset". The Tagset may be coarse such as N for Noun, V for Verb, ADJ for Adjective, ADV for Adverb, PREP for Preposition, CONJ for Conjunction and so on. These different POS tagging categories contribute to metrics of sentiment in varying degree. The adjective, noun, adverb, verb, etc., played a crucial role in determining the precise polarity. It considers the relationship with adjacent and related words in a phrase or sentence.

It is very much important to perform POS tagging for word-level sentiment analysis as it helps in identifying and treating the different meaning of polysemous words. The main reason to perform POS is that we have polysemy words which mean there is a coexistence of many possible meanings for a word or phrase i.e. words in a different context can have different meanings. Example: the word "duck" can be treated as either a noun or a verb based on context. It is a noun when it refers to a bird and its verb when it refers to the motion or when to crouch down. So, it is very important to analyze words as they give the structure to a sentence.

POS taggers are used in sentiment analysis for two purposes. Firstly, words like nouns and pronouns can be filtered out by POS tagger because they have no feelings. Secondly, it helps to distinguish words that can be used as a

distinct vocabulary. [1] Parts-of-speech (POS) tagging can be implemented in distinct programming languages by using various tools.

A parts-of-speech tagger (POS Tagger) is defined as a multimedia part assigning parts of speech to every single word in a speech it reads. POS tagging can be torn into three classifications: Rule-based, Stochastic based and Hybrid tagging. [3]

Initially, a rule-based method was used to formulate language-specific guidelines to determine the label of each word. But this strategy failed to tag some unknown words, even with enough regulations.   In reality, these schemes fail when they find unknown words. Stochastic approach tag

text input using a statistical model. They are helpful in model construction based on earlier tagged information. These models are more promising to tag word recognizes as well as unknown but tagging correctly depends on the size of tagged training data. The hybrid approach incorporates the benefits of both the rule-based and stochastic-based. Words in this method are marked probabilistic first and then applied to tags as post-processing rules. [2]

3.1 Verb Sense Annotations

Penn's Treebank: It includes 36 POS tags and 12 additional tags for symbols of currency and punctuation. [4]

Table 3.1 Penn's treebank tag sets

Number	Tag	Description
1.	CC	Coordinating conjunction
2.	CD	Cardinal number
3.	DT	Determiner
4.	EX	Existential *there*
5.	FW	Foreign word
6.	IN	Preposition or subordinating conjunction
7.	JJ	Adjective
8.	JJR	Adjective, comparative
9.	JJS	Adjective, superlative
10.	LS	List item marker
11.	MD	Modal
12.	NN	Noun, singular or mass
13.	NNS	Noun, plural
14.	NNP	Proper noun, singular
15.	NNPS	Proper noun, plural
16.	PDT	Pre determiner
17.	POS	Possessive ending
18.	PRP	Personal pronoun
19.	PRP$	Possessive pronoun
20.	RB	Adverb
21.	RBR	Adverb, comparative
22.	RBS	Adverb, superlative
23.	RP	Particle
24.	SYM	Symbol
25.	TO	*To*
26.	UH	Interjection
27.	VB	Verb, base form
28.	VBD	Verb, past tense
29.	VBG	Verb, gerund or present participle
30.	VBN	Verb, past participle
31.	VBP	Verb, non-3rd person singular present
32.	VBZ	Verb, 3rd person singular present
33.	WDT	Wh-determiner
34.	WP	Wh-pronoun
35.	WP$	Possessive wh-pronoun
36.	WRB	Wh-adverb

The verb form of a word will represent an action or a state of being. This forms the main part of the predicate of a sentence. Verb word is an important opinion term for the opinions posted on social sites. Verb helps in identifying the intensity and diminishes the aspect of the sentiment in a sentence. From a linguistic perspective, verbs play an important role in the semantics of a sentence from a sentiment perspective, verbs give the sentiment and these verbs are called "opinion verbs". Hence, by considering verb words, it improves the sentiment classification score performance.

If the role of the verb is not considered in a sentence, then its sentiment may not be recognized as accurate. So, the verb form of the word is considered as most important because it asserts something about the subject of the sentence and expresses the action, event or state of being.

Table 3.2 Parts-of-speech tags for Verbs [1]

[1] Tag	[2] Definition
[3] VB	[4] Base form
[5] VBP	[6] Present tense not 3rd person singular
[7] VBZ	[8] Present tense 3rd person singular
[9] VBD	[10] Past tense
[11] VBG	[12] Present participle
[13] VBN	[14] Past participle

Table 3.2 lists all POS tagging tags for verbs.

(a) **VB:** is a simple base form of a verb.

(b) **VBZ:** is for 3rd person singular present. These refer to the verb words which are formed with simple affix the suffix −s to the end of the verb.

(c) **VBP:** is for non-3rd person singular present. These verb words express the habits and routines, general facts and truths, and thoughts and feelings.

(d) **VBN:** refers to the past participle. These are regular verb words with their past forms.

(e) **VBG:** is for gerund or present participle. This is done usually by adding −ing to the base word.

(f) **VBD:** is for normal past tense. These verb words are used to express activity, action, state of being in the past

IV. IMPLEMENTATION & RESULTS

In this work, the NLTK corpora, movie reviews dataset which contains 1000 positive and 1000 negative reviews is used. Next preprocessing is done that involves the removal of stop words, white spaces, etc. After preprocessing, by using NLTK it was POS tagged.
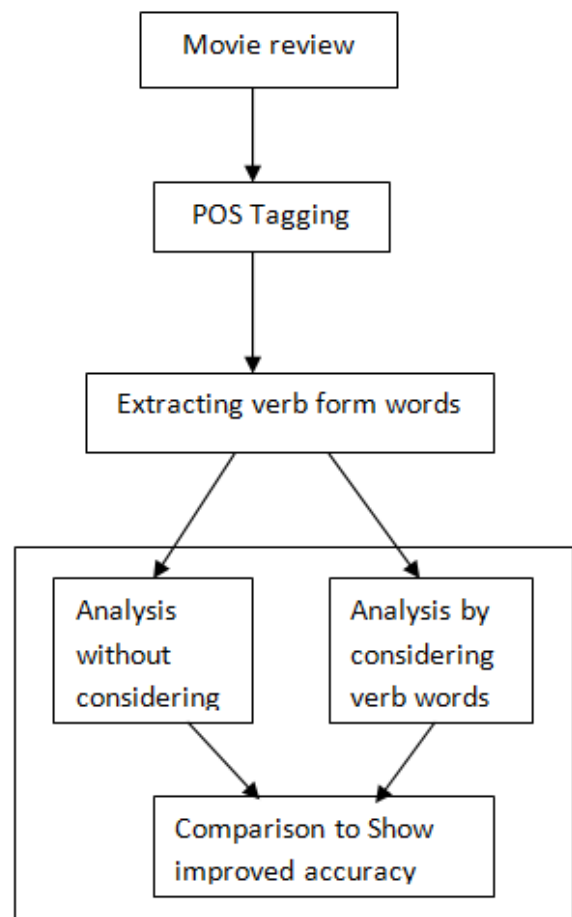

Fig: 4.1 Proposed methodology

A. POS Tagging

This is a process to tag individual words like noun, pronoun, verb, adjective, determiner, etc. This is performed by using TextBlob and VADER.

B. Extracting Verb form words

TextBlob: [7][9]

Textblob is a python library that provides a straightforward API for accessing its techniques and carrying out fundamental NLP functions. Based on its definition and context, parts-of-speech tagging, or grammatical tagging is a technique for marking words, present in a text. Simply put, it says if a word is a noun, or an adjective, or a verb, etc. TextBlob's sentiment feature returns two characteristics, polarity and subjectivity. Polarity is a float value within the range of [-1, 1] where 1 is positive and -1 is negative. A subjective sentence usually refers to opinion, emotions or judgments, while objective relates to factual data. Subjectivity is also a float value within the [0,1] range.

VADER:

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool. It is an open-source tool under the MIT license. Vader has the capability to perform with emojis, slang, and acronyms present in a sentence forming its best choice to perform sentiment analysis.

SentimentIntensityAnalyzer is the object for Vader which will help to perform all operations of sentiment analysis. "polarity_scores()" is the method used to determine the sentiment. Vader gives scores of positivity and negativity along with the value of how positive or negative a sentiment is.

Vader considers 5 heuristics: [6][8]

(a) *Punctuation*: Ex: "I like it" and "I like it!!!". In this second sentence give more intense emotion. Vader helps in recognizing these kinds of punctuation symbols to differentiate the intensity.

(b) *Capitalization*: Ex: "AMAZING Performance" and "amazing performance". In this capitalized word AMAZING has more intensity.

(c) *Degree modifiers*: Ex: "effing cute" and "sort of cute". Vader maintains a booster dictionary which contains a set of boosters and dampeners.

(d) *Shift in polarity due to "but"*: "but" connects two clauses with contrasting sentiments. Vader has a feature of "but" checker where it defines as 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.

(e) *Examining the tri-gram before a sentiment-laden lexical feature to catch polarity negation*: tri-gram means a set of three lexical features. Vader also has a feature of a negator list.

This model is the best choice to work with movie reviews and opinion articles. It has the advantage that it works exceedingly well with social media text and even works fast compared to other methods.

C. Evaluation

Following is the example to show the importance of verb words in determining sentiment analysis:

Example (i): We **played(VBD)** tennis in summer holidays. We **had(VBD)** a group of people who **were(VBD)** interested to **play(VB)** tennis match. But now the holidays **are(VBP)** over that no more playing of tennis.

Without Verb: Positive score: 0.298 Negative score: 0.086

With Verb: Positive score: 0.337 Negative score: 0.072

Example (ii): We like **playing(VBG)** tennis in summer holidays. We **have(VBP)** a group of people who **have(VBP)** interest to **play(VB)** tennis match. But now the holidays **are(VBP)** over that no more playing of tennis.

Without Verb: Positive score: 0.24 Negative score: 0.093

With Verb: Positive score: 0.343 Negative score: 0.071

Example (iii): We like to **play(VB)** tennis in summer holidays. We **are(VBP) having(VBG)** a group of people who **have(VBP)** interest in **playing(VBG)** tennis match. But now the holidays **are(VBP)** over that no more playing of tennis.

Without Verb: Positive score: 0.232 Negative score: 0.09

With Verb: Positive score: 0.326 Negative score: 0.068

By seeing at the above examples (i), (ii), (iii) can observe that different forms of the word "*play*" have given different scores.

IV. CONCLUSION

The task of sentiment analysis must go miles. Document-level and sentence-level sentiment analysis will result in inaccuracy, so word-level analysis should be carried out. Verb word plays a significant role in describing an action that enables us to enhance the precision of polarity of sentiment. POS tagging is conducted to obtain verb word and is contrasted against the emotional word to demonstrate the enhancement in polarity.

NLTK is the most widely used Python library in this study. The NLTK corpora itself collects the dataset. The precision of the outcomes for this dataset varies from 75% to 79%. A clear improvement is shown by the polarity score acquired by considering verb words.

To enhance accuracy, more techniques and algorithm for different tags need to be searched and applied. The work can also be expanded to other domains and languages.

REFERENCES

1. Akshaya R. Garje, K. V. Kale. Sentiment Analysis Text POS Tagging on Movie reviews using NLTK. International Journal of Advanced and Innovative Research. 2017.
2. Dharmishtha N. Mehta, Prof. Nikita P. Desai. A Survey on Part-Of-Speech Tagging of Indian Languages. 1st International Conference on Computing, Communication, Electrical, Electronics, Devices & Signal Processing (CCEEDS). 2015.
3. Kishorjit Nongmeikapam, Dilipkumar Khangembam, Wangkheimayum Hemkumar, Shinghajit Khuraijam, and Sivaji Badyopadhyay. Verb Based Manipuri Sentiment Analysis. International Journal on Natural Langugae Computing. Vol. 3, No.3, June 2014.
4. Mitchell P. Marcus, Beatrice Santorini, Mary Ann Marcinkiewicz. Building a Large Annotated Corpus of English: The Penn Treebank. Association for Computational Linguistics. 1993.
5. Vivek Kumar Singh, Mousumi Mukherjee, and Ghanshyam Kumar Mehta. Sentiment and Mood Analysis of Weblogs Using POS Tagging Based Approach. Communications in Computer and Information Science book series (CCIS, volume 168). 2011.
6. http://datameetsmedia.com/vader-sentiment-analysis-explained/
7. https://www.analyticsvidhya.com/blog/2018/02/natural-language-processing-for-beginners-using-textblob/
8. https://medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f
9. https://textblob.readthedocs.io/en/dev/