

A Novel Approach to Analyze the Sentiment with Conjunctive Words

Tanveer Khan¹, Yogesh Kumar²

¹Research Scholar, Dept. of CSE, BGIET, Sangrur, Punjab, India

²Assistant Professor, Dept. of CSE, BGIET, Sangrur, Punjab, India

Abstract – Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. It's also known as opinion mining, deriving the opinion or attitude of a speaker. Sentiment analysis is the process which helps in identifying people's attitudes and emotional states. In this paper we have discussed the existing work on sentiment analysis with different approach. We developed an algorithm that takes Punjabi paragraph as input and analyze the sentiment of paragraph by words, as well as a sentence on the basis of conjunctive word. The feature of this work is that it contains the dictionary of Punjabi words and conjunctive words which help to analyze the sentiment more accurately.

Key Words: Sentiment Analysis, Punjabi Language, Linguistic Resources, Opinion Mining, Subjective, Objective.

1. INTRODUCTION

Sentiment analysis [1] (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in the source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.

Sentiment analysis [2] aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document. The attitude may be his or her judgment or evaluation, affective state (that is to say, the emotional state of the author when writing), or the intended emotional communication (that is to say, the emotional effect the author wishes to have on the reader).

Sentiment analysis discovers the expressions in a text or record that contains some sentiment. Sentiment Analysis concentrates on sorting the content at the level of subjective and target nature. This undertaking is commonly characterized as grouping a given content (normally a sentence) into one of two classes: objective or subjective. There may be some objective facts or subjective sentiments in the text. It is mandatory to recognize between them. It also helps in deciding the elements and subject from text towards which viewpoint is sentiment is determined.

Subjectivity - Shows [3] that the content contains/bear's presumption opinion. Sentiment is ordered as objective (truths), positive (speaks of a condition of joy, bliss, delight or fulfillment) or negative (speaks of a condition of distress, misgiving, bitterness or dissatisfaction).

Objectivity - Shows that the content is without opinion, but contain facts.

On the basis of degree of polarity, a score can be given to the sentiment. In this manner, there are two research tips; first classifying the polarity of text to express the feeling in positive, negative or neutral. Second is the identification of subjectivity or objectivity. Due to its many aspects it is often referred to by different names such as opinion mining, sentiment classification, sentiment analysis, and sentiment extraction.

2. LITERATURE REVIEW

(Parul et.al, 2015) has discussed that Sentiment Analysis is to distinguish and group the assessments/feelings/opinions in composing content. Till date, English Language incorporates the majority of the examination work around there. In this paper, they talked about the different methodologies used to finish the opinion investigation and exploration work accomplished for Indian Languages like Hindi, Bengali and Telugu. An approach is proposed to determine the sentiment orientation, i.e. polarity of the Punjabi reviews by scoring method. Sentiment analysis is needed to be performed in Punjabi language because of the increase in Punjabi data on the web. Separate positive and negative condensed results are created which is useful for the client in choice making. They contrasted the outcomes and right now existing methodologies.

(Deepali et.al, 2014) discussed that Opinion mining and Sentiment analysis has emerged as a field of study since the widespread of World Wide Web and internet. The opinion refers to extraction of those lines or phrase in the raw and huge data which express an opinion. Sentiment analysis, on the other hand, identifies the polarity of the opinion being extracted. In this paper, they proposed the sentiment analysis in collaboration with opinion extraction, summarization, and tracking the records of the students. The paper modifies the existing algorithm in order to obtain the collaborated opinion about the students. The resultant opinion is represented as very high, high, moderate, low and very low. The paper is based on a case study where teachers give their remarks

about the students and by applying the proposed sentiment analysis algorithm the opinion is extracted and represented.

(Singh et.al, 2013) has stated an experimental work on a new kind of domain specific feature-based heuristic for aspect-level sentiment analysis of movie reviews. They have devised an aspect oriented scheme that analyses the textual reviews of a movie and assign it a sentiment label on each aspect. The scores on each aspect from multiple reviews are then aggregated and a net sentiment profile of the movie is generated on all parameters. They have used a Senti Word Net based scheme with two different linguistic feature selections comprising of adjectives, adverbs and verbs and n-gram feature extraction. They have also used our Senti Word Net scheme to compute the document-level sentiment for each movie reviewed and compared the results with results obtained using Alchemy API. The sentiment profile of a movie is also compared with the document-level sentiment result. The results obtained show that our scheme produces a more accurate and focused sentiment profile than the simple document-level sentiment analysis.

(Kaur et.al, 2014) has defined the meaning of sentiment Analysis is to recognize and characterize the conclusions/feelings/estimations in composed content. They discussed the various approaches used to accomplish the sentiment analysis and research work done for Indian Languages like Hindi, Bengali and Telugu. They proposed an algorithm by using subjective lexicon which is created by using Hindi Subjective Lexicon. Their approach proves good performance on the testing data. They compared the results with already existing approaches.

(Min Hye-Jin, 2012) have used NLP from a different perspective. They used NLP techniques to identify tense and time expressions along with mining techniques and a ranking algorithm. Their proposed metric has two parameters that capture time expressions related to the use of products and product entities over different purchasing time periods. They identified important linguistic clues for the parameters through an experiment with crawled review data, with the aid of NLP techniques. They worked on product reviews from amazon.com. Their results showed that their metric was helpful and free from undesirable biases.

(Deepali et.al, 2013) has explained on a movie domain reviews in Punjabi Language using N-Gram approach and machine learning technique/Naive Bayes. Fundamental issue while managing sentiment analysis on reviews is that surveys compass over numerous sentences. There are situations when a review contains various sentences and among them few sentences have inverse estimation. This survey is neither positive nor negative and falls flat while grouping.

3. OBJECTIVES AND METHODOLOGY

Sentiment analysis is a process of extracting information from user's opinions. The decisions of the people get affected by the opinions of other people. As there is a huge explosion of the users opinion on social media like Twitter, Facebook

and other user forums, then identification of sentiment becomes very difficult from this huge data manually. So, there is a need of automated sentiment analysis system. The main objective of this research work is to perform the sentiment analysis for Punjabi sentences.

3.1 Objectives

In order to perform this task, following objectives are proposed to be carried out.

- i. To study the existing approaches and techniques for sentiment analysis.
- ii. To study the existing algorithms of sentiment analysis in Indian languages.
- iii. To make an application that maintains the directory of positive, negative and neutral words.
- iv. To implement an algorithm that analyzes the sentiments of entered Punjabi words or sentences written in a paragraph.
- v. To extract the sentences/words according to sentiments.
- vi. To analyze the results of proposed algorithm with the help of positive, negative and neutral words database maintained in the application.
- vii. To evaluate the performance of the system.

3.2 Methodology

To achieve the above objectives, the methodology adopted has been discussed in this section

Step 1: Enter the Punjabi text as input.

Step 2: Parse the paragraph into token and store the words in an array list.

Step 3: Select the word from array list.

Step 4: Compare the selected word of paragraph with each word of database.

Step 5: If the result is true then

a) Find the status of word from database whether it is positive, negative or neutral.

b) Highlight the word according to their status.

c) Find the position of highlight word.

d) Find the sentence starting, ending position and position of conjunctive word.

e) Highlight the sentence according to the status.

f) Calculate the score of sentence display the result.

Else result is false then

a) Select next word from array until last element of array is traversed.

Go to step 4

Step 6: Display the results

The flowchart depicting the proposed algorithm is shown in Fig-1.

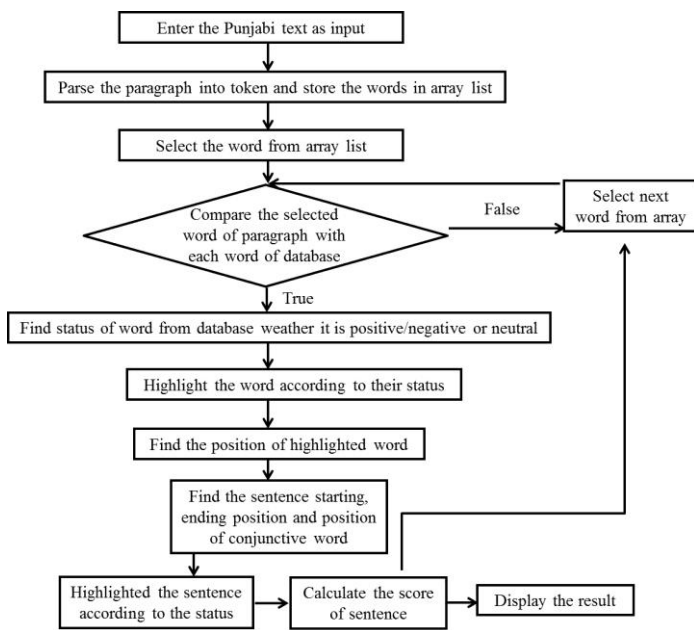


Fig 1. Flowchart showing the working of proposed algorithm.

This paper introduces sentence classification using the dictionary based approach for Punjabi language. We compared the various techniques and systems available. The developed system finds out the positive, negative and neutral words from selected paragraphs.

4. RESULTS

Dataset used for testing the system consists of 25 reviews retrieved from popular newspaper website and other Punjabi websites. On these 25 paragraph when test is applied it gives 90% of accurate results.

Test Case - 1

The following review contains 34 words comprising one positive word, one negative word, one neutral word and one conjunctive word.

ਮਨੁੱਖ ਇਕ ਸਮਾਜਿਕ ਜੀਵ ਹੈ | ਉਸਦਾ ਜੀਵਨ ਦੂਜਿਆਂ ਉੱਪਰ ਨਿਰਭਰ ਕਰਦਾ ਹੈ ਅਤੇ ਇਸ ਕਰਕੇ ਉਹ ਜਿੰਦਗੀ ਵਿਚ ਕੁੱਝ ਸੱਜਣਾਂ ਮਿੱਤਰਾਂ ਦੀ ਲੋੜ ਅਨੁਭਵ ਕਰਦਾ ਹੈ ਜਿਨ੍ਹਾਂ ਨਾਲ ਉਹ ਆਪਣੇ ਦੁੱਖ-ਸੁੱਖ ਵੰਡ ਸਕੇ।

Positive word: ਵੰਡ

Negative word: ਨਿਰਭਰ

Neutral word: ਸਮਾਜਿਕ

Conjunctive word: ਅਤੇ

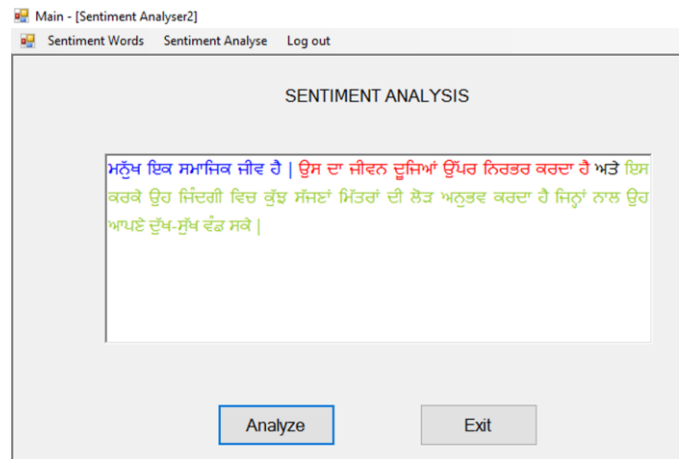


Fig-2 Highlighting the sentence on detecting the positive, negative and neutral word.

In Fig-2, the designed algorithm identifies the positive word ਵੰਡ, negative word ਨਿਰਭਰ and neutral word ਸਮਾਜਿਕ and highlight the sentence in green color for positive sentence, red color for the negative sentiment and blue color for neutral sentence. In this test case ਅਤੇ is a conjunctive word.

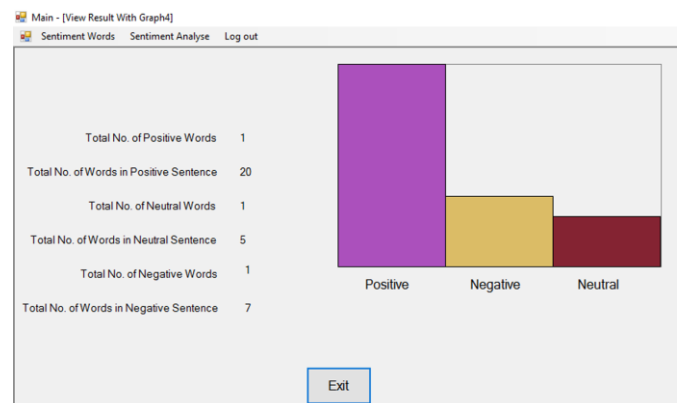


Fig-3 Snapshot showing the graphical representation of the result.

In the above Fig-3, the detail of the result comprising the total no. of words and total no. of words in sentence is shown and the result is represented in the graphical form for test case 1.

Test Case - 2

The following review contains 34 words comprising one positive word and one negative word.

ਭਗਵਾਨ ਤਾਂ ਹਜ਼ਾਰਾਂ ਦੇ ਰੂਪ ਵਿੱਚ ਹੋ ਸਕਦਾ ਜਾਂ ਸਾਰੇ ਜੀਵ ਖੁਦਾ ਦੇ ਵਰਗੇ ਹੋ ਸਕਦੇ ਨੇ | ਇਸ ਲਈ ਇਨਸਾਨ ਆਪਸ ਵਿੱਚ ਜੰਗ ਲੜ ਕੇ ਜਾਂ ਧਰਮ ਦੇ ਨਾਂ ਹੇਠ ਦਾ ਸੰਘਾਰ ਕਰਦੇ ਨੇ |

Positive word: ਰੂਪ

Negative word: ਸੰਘਾਰ

Conjunctive word: ਜਾਂ

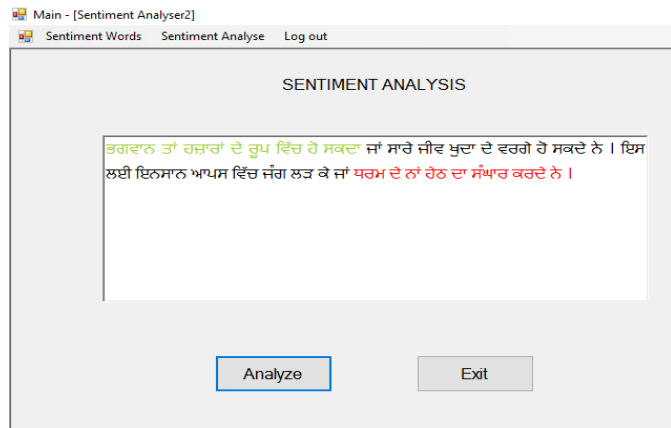


Fig-4 Highlighting the sentence by detecting the positive and negative word.

In Fig-4, the designed algorithm identifies the positive word ਰੂਪ and the negative word ਸੰਘਾਰ and highlight the sentence green, according to its positive sentiment and red according to its negative sentiment. In this test case ਜਾਂ is a conjunctive word.

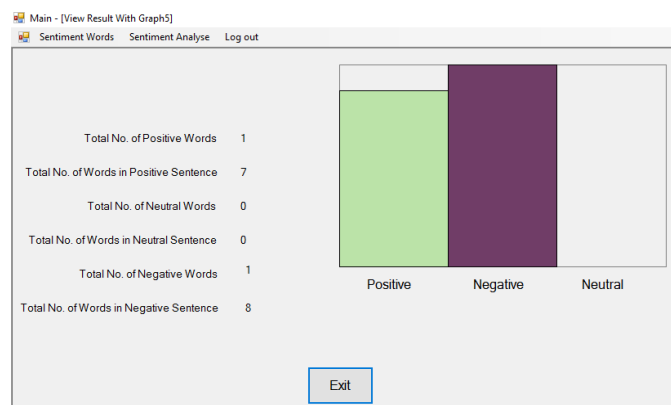


Fig-5 Snapshot showing the graphical representation of the result.

In the above Fig-5, the detail of the result comprising the total no. of words and total no. of words in sentence is shown and the result is represented in the graphical form for test case 2.

Test Case – 3

The following review contains 25 words comprising one positive word, one negative word and one conjunctive word. ਜਦੋਂ ਅਸੀਂ ਮਿੱਤਰ ਨਾਲ ਖੁਸ਼ੀ ਵੰਡਦੇ ਹਾਂ ਤਾਂ ਉਹ ਦੁਗਣੀ ਹੋ ਜਾਂਦੀ ਹੈ ਪਰ ਜਦੋਂ ਦੁੱਖ ਨੂੰ ਵੰਡਦੇ ਹਾਂ ਤਾਂ ਉਹ ਔਧਾ ਰਹਿ ਜਾਂਦਾ ਹੈ |

Positive word: ਖੁਸ਼ੀ

Negative word: ਦੁੱਖ

Conjunctive word: ਪਰ

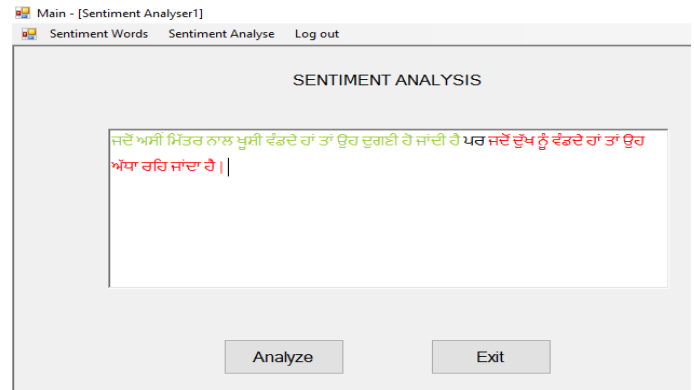


Fig-6 Highlighting the sentence by detecting the positive and negative word.

In Fig-6, the designed algorithm identifies the positive word ਖੁਸ਼ੀ and the negative word ਦੁੱਖ and highlight the sentence green, according to its positive sentiment and red according to its negative sentiment. In this test case ਪਰ is a conjunctive word.

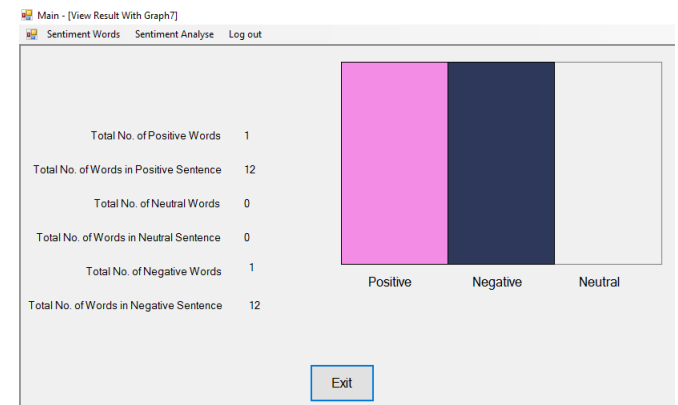


Fig-7 Snapshot showing the graphical representation of the result.

In the above Fig-7, the detail of the result comprising the total no. of words and total no. of words in sentence is shown and the result is represented in the graphical form for test case 3.

4. CONCLUSIONS

Sentiment analysis is the process which helps in identifying people’s attitudes and emotional states. The feelings of the people can be expressed in positive or negative ways. Mostly, parts of speech are used as

feature to extract the sentiment of the text. Sentiment analysis is an evolving field with a variety of user applications.

Further, we evaluated the accuracy of the system, from which we identify the conjunctive words in a Punjabi sentence to increase the accuracy of sentiment analysis. With the help of conjunctive words, we can calculate the sentiment of the sentence up to the conjunctive word instead of highlighting and calculating the result for the whole sentence.

For e.g. ਫਿਲਮ ਤਾਂ ਵਧੀਆ ਸੀ ਪਰ ਗਾਣੇ ਬੇਕਾਰ ਸੀ |

Positive word: ਵਧੀਆ

Negative word: ਬੇਕਾਰ

Conjunctive word: ਪਰ

The developed algorithm divides the sentence in two sentiments, positive and negative to increase the accuracy of sentiment analyzing.

5. FUTURE WORK

A system is developed so as to increase the precision and lower the recall rate from the sentiment analysis technique by applying the data is collected from different newspaper site and Punjabi essay websites. The system built can be enhanced by adding provision for translating the Punjabi string into English or another dialect.

Furthermore the current version of this algorithm does not facilitate the function of synonyms and antonyms but can be done in the future research processes.

ACKNOWLEDGEMENT

I am thankful to my guide Mr. Yogesh Kumar, Assistant Professor in Department of Computer Science & Engineering for their valuable support. Also, I am thankful to my Department for the technical support.

REFERENCES

- [1] Sentiment Analysis, Available at https://en.wikipedia.org/wiki/Sentiment_analysis.
- [2] Arora P., & Kaur B., "Sentiment Analysis of Political Reviews in Punjabi Language." International Journal of Computer Application, Vol.-126, No.-14, Sept 2015.
- [3] Kaur A., & Gupta, V. , "A survey on Sentiment Analysis and Opinion Mining Techniques." Journal of Emerging Technologies in Web Intelligence, Vol.-5, No.-4, Nov. 2013.
- [4] Deepali Virmani, Vikrant Malhotra and Ridhi Tyagi, "Sentiment Analysis Using Collaborated Opinion Mining". International Journal of Soft Computing & Engineering, Vol.-4, Issue-CCIN-2K14, March-2014.

- [5] Singh, V., & Piryani, R., "Sentiment Analysis of Movie Reviews a new feature-based heuristic for aspect-level sentiment classification." DOI: 10.1109/ iMac4s.2013. 6526500.
- [6] Kaur A., & Gupta, V., "Proposed Algorithm of Sentiment Analysis for Punjabi Text." Journal of Emerging Technologies in Web Intelligence, Vol.-6, No.-2, May 2014.
- [7] Hye-Jin Min, Jone C. Park, "Identifying helpful reviews based on customer's mentions about experiences." Expert system with Applications, Vol.-39, Issue-16, Nov-12.
- [8] Deepali, Navneet Garg. "Movie Review Mining in Punjabi." International Journal of Application or Innovation in Engineering and Management, Volume-2, Issue-12, Dec-2013.