

SENTIMENT ANALYSIS OF DIFFERENT LANGUAGES

Akshay Murali
Department of Computer Applications
Amal Jyothi College of Engineering
Kanjirapally,INDIA
akshaymurali2022a@mca.ajce.in

Shelly Shiju George
Department of Computer Applications
Amal Jyothi College of Engineering
Kanjirapally,INDIA
shellyshijugeorge@amaljyothi.ac.in

Abstract— Sentiment analysis is a fundamental interest in the field of Text mining research, which deals with the computational processing of sentiments, perspectives, and the subjective aspect of the text. This work focuses on mining and analyzing sentiments in five languages. The proposed research depicts a comprehensive analysis specifically in the fields of Hindi, Malayalam, Marathi, Kannada, and Gujrati. This study makes no use of datasets. It examines the significance of NLP and Speech Recognition from several perspectives, including their impact on the topic of Sentiment Analysis.

Keywords—NLP, Speech-Recognition, Sentiment analysis

I. INTRODUCTION (HEADING I)

Sentiment analysis using NLP is the use of machine learning techniques to identify whether a sentence has a positive polarity, negative polarity or neutral polarity. Nowadays, sentiment analysis has gained high importance in a wide spectrum of fields, ranging from marketing to forensics. For example, the sentiments of a person on any product or anything can be analysed through their speech or text. So, the task is to classify the data and gain knowledge. With the change in technology and development people started expressing their thoughts in English as well as in their mother tongue for e.g., Hindi, Marathi etc. So, the most spoken languages in India is focused that is assumed as the languages Hindi, Malayalam, Marathi, Kannada and Gujarati. So, a model to classify these 5 languages are anticipated. The model will recognize the speech from the user with help of Speech Recognition and then it will help you to derive polarity precision with help of sentiment Intensity analyzer.

II. BASIC TERMINOLOGIES USED

A. Google Speech Recognition API

Google has a great Speech Recognition API. This API converts spoken text (using microphone) into written text (Python strings), briefly Speech to Text. The API has excellent results for various languages.

B. Google Translation Library

Google translation is a free and unlimited python library that implemented Google Translate API. This uses the Google Translate Ajax API to make calls to such methods as

detect and translate. To install, either use things like pip with the package “google trans” or download the package and put the “google trans” directory into the python path. Features of this library is that it is Fast and reliable; it has auto language detection and can-do bulk translations.

C. Polarity Score

The polarity score is a float within the range [-1.0,1.0]. The subjectivity is a float within the range [0.0,1.0] where 0.0 is very objective and 1.0 is very subjective. Usage of this polarity score to get the sentiment polarity and labels for any text. Sentiment polarity for an element defines the orientation of the expressed sentiment, i.e., it describes if that text is expressed in positive, negative or neutral class given by the user about anything.

III. LITERATURE SURVEY

Rachna Bandana [3] suggested that using Naive Bayes and Linear Support Vector Machine (LSVM) to build the system model will be an advantage because different huge and small datasets are used for the preprocessing techniques and classification of the algorithm to observe various results. This work are basically using Movie Review Dataset to classify the reviews for analyzing the sentiments

Shalini K et.al [2] suggested that using the method CNN for sentiment analysis will be good because they are doing sentiment analysis of only Bengali and Telugu movie data. The technique is used for mining sentiments that generally measures the overview of opinions and analyses the information from the given content. The testing is done by dividing the training dataset of 2500 sentences as 2250 sentences for training and rest for Test data.

Saeed Mian Qaisar [4] suggested that using LSTM classifier based on the RNN algorithm will be much more efficient in analyzing the sentiments of the IMDb movie reviews. While handling the IMDb database section the performance accuracy of the LSTM classifier is evaluated by Confusion matrix and then provides a brief statement of Right and Wrong predictions.

Charu Nanda et.al [5] suggests that the model is implemented over Random Forest and SVM for Sentiment analysis. Basically, the work made an approach to classify the reviews in Hindi language. Algorithms with different

metrics are used to evaluate the performance which makes it much easier to recognize the intention of the text based on the dataset.

Reddy Naidu et.al [6] suggested that already existing Telugu senti-wordnet for the sentiment analysis for e-newspaper in Telugu language is used. It has achieved an accuracy of 74% for the subjectivity classification and 81% for the classification of the Sentiments.

IV. METHODOLOGY

A. Speech Recognition

Basically, speech recognition is a library for performing the task of recognizing the speech of various languages from the users, with help of various engines and different API's which helps computers to understand human language. Speech recognition can be installed by "pip install Speech Recognition". Basic requirements to access Speech Recognition you should have Python2.6, 2.7 or 3.3+.

B. Natural Language Processing(NLP)

Natural Language Processing is a specific area that focuses on making natural human language that can be used by the computer programs. NLTK or Natural Language Toolkit, it's a python package which is used for NLP.

C. VADER Sentiment

VADER (Valence Aware Dictionary and Sentiment Reasoner), it is branch of knowledge and rule-based sentiment analysis tool that is specifically aware of sentiments expressed by the person and works well on text from other domains. VADER does not require any type of training data. It can very well acknowledge the sentiment of a text.

D. Sentiment Polarity

The basic sentiment Polarity of an element gives a brief about the intention of the sentiment expressed. To calculate the polarity of a text, polarity score of words from the text is combined to get an overall polarity score. Polarity refers to power of an opinion.

E. Pause_Threshold

Pause threshold represents the minimum length of silence and the length will be in seconds that will concluded as the end of a sentence. Smaller values for pause threshold can be set then it will give the result of completing recognition more quickly.

F. Google Translation Library

Google trans is a free and unlimited python library that implemented Google Translate API. This uses the Google Translate Ajax API to make calls to such methods as detect and translate. To install, either use things like pip with the package "google trans" or download the package and put the "google trans" directory into your python path.

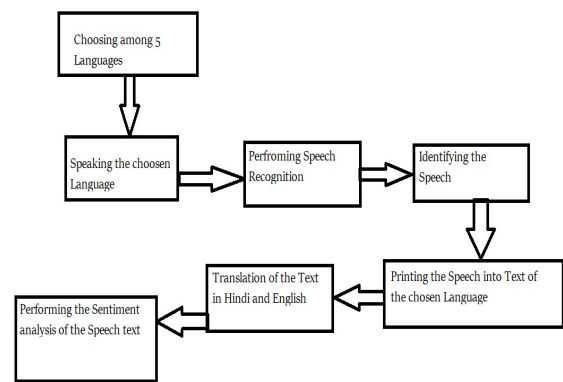


Fig 4.1: Working Model

V. PROPOSED SYSTEM

In the existing systems Sentiment analysis has been done in one or two languages, but in the proposed system sentiment analysis in the most spoken five languages of India are implemented. To perform sentiment analysis, previous works had used datasets like Movie Reviews, Twitter data or any other dataset but in the proposed system no datasets of any types are used, live input from the user is taken. User can choose among any of the 5 languages, given as options,

```

choose the language
1.Hindi
2.Malayalam
3.Marathi
4.Kannada
5.Gujarati
    
```

Fig 5.1: Choosing the language

then user have to speak in that chosen language, after reaching the pause threshold it will then recognise the speech from the user with the help of Speech recognition, it will be easy for the system to understand the speech given by the user. In existing system there is no use of translation, just the

sentiments are given as the output, but in the proposed system the speech from the user is converted into the chosen language and then the chosen language is translated into English and Hindi and then the sentiment analysis task is performed and shows the sentiment of the given speech. Speech Recognition plays an important role in this System.

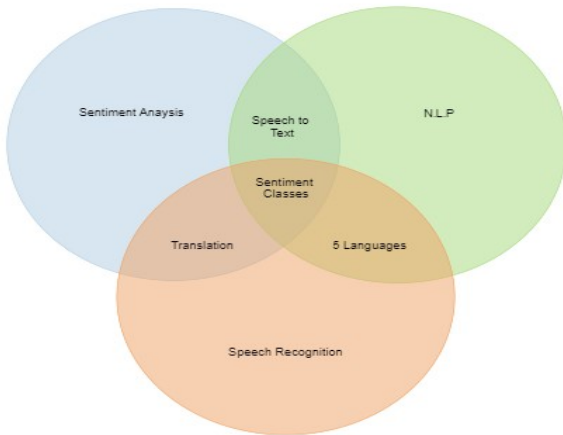


Fig 5.2: Sentiment analysis

VI. RESULT

In this proposed system there is no datasets used as the live input is taken from the user as their speech, before that user have to select any language among the five languages given, then the user can speak but after a silence of 0.7 seconds it will be considered as end of the sentence because of the pause threshold as small values in pause threshold will recognize speech faster, then the speech is converted to the chosen language text after the speech recognition. After the recognition and conversion of the speech into chosen language text then the Text is translated into English and then Hindi because these are most commonly used language in India. Then after the translation of chosen language text into English and Hindi, then the polarity score of the text is calculated if the float is within the range of [0.0,1.0] where 0.0 is very objective and 1.0 is very subjective, Subjective means an important positive or negative term while the Objective is considered as neither negative nor positive basically a neutral term. After calculating the polarity score the sentiment analysis of the sentence will be classified either it will be in positive class, negative class or in the neutral class.

```
choose the language
1.Hindi
2.Malayalam
3.Marathi
4.Kannada
5.Gujarati
```

Fig 6.1: Selecting the language

```
choose the language
1.Hindi
2.Malayalam
3.Marathi
4.Kannada
5.Gujarati
Listening
```

Fig 6.2: User giving Input

```
choose the language
1.Hindi
2.Malayalam
3.Marathi
4.Kannada
5.Gujarati
Listening
Recognizing
```

Fig 6.3: Recognizing User's speech

```
choose the language
1.Hindi
2.Malayalam
3.Marathi
4.Kannada
5.Gujarati
Listening
Recognizing
the query is printed=' ಸುಖವಾಗೋಣೋ '
```

Fig 6.4: Speech to text conversion

```
choose the language
1.Hindi
2.Malayalam
3.Marathi
4.Kannada
5.Gujarati
Listening
Recognizing
the query is printed=' ಸುಖವಾಗೋಣೋ '
Translated(src=en, dest=en, text=How are you, pronunciation=None, extra_data={'translat...'})
Translated(src=en, dest=hi, text=क्या हाल है, pronunciation=kya haal hai, extra_data={'translat...'})
```

Fig 6.5: Translation of chosen language text into English and Hindi

```
choose the language
1.Hindi
2.Malayalam
3.Marathi
4.Kannada
5.Gujarati
Listening
Recognizing
the query is printed=' ಸುಖವಾಗೋಣೋ '
Translated(src=en, dest=en, text=How are you, pronunciation=None, extra_data={'translat...'})
Translated(src=en, dest=hi, text=क्या हाल है, pronunciation=kya haal hai, extra_data={'translat...'})
Translated Sentence= Translated(src=en, dest=en, text=How are you, pronunciation=None, extra_data={'translat...'}) Transl
Dictionary= {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
```

Fig 6.6: Calculating Polarity score

```
choose the language
1.Hindi
2.Malayalam
3.Marathi
4.Kannada
5.Gujarati
Listening
Recognizing
the query is printed=' ಸುಖವಾಗೋಣೋ '
Translated(src=en, dest=en, text=How are you, pronunciation=None, extra_data={'translat...'})
Translated(src=en, dest=hi, text=क्या हाल है, pronunciation=kya haal hai, extra_data={'translat...'})
Translated Sentence= Translated(src=en, dest=en, text=How are you, pronunciation=None, extra_data={'translat...'}) Transl
Dictionary= {'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
It is a Neutral Sentence
Process finished with exit code 0
```

Fig 6.7: Sentiment analysis classifying into neutral sentence

Therefore, the User has chosen Malayalam language by selecting 2, then the user spoke □□□□ □□□ (Sugam aano) which means “How are you?” After the recognition of the speech and conversion of the speech to text of chosen language then it will translate the text to English and Hindi. i.e., How are you? and क्या हाल है (kya haal hai), then it will calculate the polarity score of the text and then the sentiment analysis will be done and the text will be classified into positive, negative or neutral statement. In the above Fig 6.7 the text was classified into the Neutral statement.

VII. CONCLUSION

The sentiment analysis of Indian languages using NLP has gained high importance in almost every field such as entertainment industry, business, forensics etc. Huge amount of user-generated content in a native language such as Hindi is available. This research work is created by collecting speeches from the users in five different languages by using speech recognition and NLP. Translating the speech of a particular language into English and Hindi and then sentimental score is analysed with the help of Sentiment Intensity analyser and classifying the sentence into positive, negative or neutral class by calculating the polarity score of the text.

VIII. REFERENCES

- [1] Kush Shrivastava and Shishir Kumar “A Sentiment Analysis System for the Hindi Language by Integrating Gated Recurrent Unit with Genetic Algorithm.” The International Arab Journal of Information Technology, Vol. 6, November 2020.
- [2] Shalini K, Aravind Ravikumar, Vineetha R C, Aravinda Reddy, Anand Kumar M and Soman K.P “Sentiment Analysis of Indian Languages using Convolutional Neural Networks.” 2018 International Conference on Computer Communication and Informatics (ICCCI -2018), July 11-12, 2019.
- [3] Rachana Bandana “Sentiment Analysis of Movie Reviews Using Heterogeneous Features.” 2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech), May 04-05, 2018.
- [4] Saeed Mian Qaisar “Sentiment Analysis of IMDb Movie Reviews Using Long Short-Term Memory.” 2020 2nd International Conference on Computer and Information Sciences (ICCIS), October 13-15, 2020.
- [5] Charu Nanda, Mohit Dua and Garima Nanda “Sentiment Analysis of Movie Reviews in Hindi Language using Machine Learning.” 2018 International Conference on Communication and Signal Processing (ICCSP), April 03-05, 2018.
- [6] Reddy Naidu, Santosh Kumar Bharti, Korra Sathya Babu and Ramesh Kumar Mohapatra “Sentiment analysis using Telugu SentiWordNet.” 2020 IEEE Conference on Computer Applications (ICCA), February 27-28, 2020.