Sentiment Analysis/Emotion Recognition On Voice Using Feature Extraction

Ram Pardakhe[1], Nikki Jadhav[2], Mayur Nimbarte[3],
Kartik Sakhare[4]
Under the guidance of
Prof. Namrata Khade
[1, 2, 3, 4,]Stduents
Department of Computer Science and Engineering,
Priyadarshini Institute of Engineering and Technology, Nagpur, India.

Abstract: As The aim of this project is to create a more advanced software prototype that processes audio data and performs sentiment analysis on the audio file in order to track down the user's emotions. Computation of emotions is a very challenging task, much work has been done but much more improvements are also possible as the audio data needs to be distinguished by appropriate gender and to analyze the emotion out of it accordingly.

Keywords **:** feature extraction,mel-frequence cepstrum,speech emotion recognition.

I. INTRODUCTION

Understanding the emotions are greatly influenced through different patterns like voice, facial expressions, body language. The voice of a human out of this can also be broadly categorize in terms of emotions. Emotions in voice carries extra knowledge about human actions. Yet a few systems are exploring the broad field of emotional human interface. No established analytical methods, neither in the field of speech analysis nor in image processing, can reliably determine the intended or pure emotion. We have targeted only on the voice-emotion analysis based on the idea that humans are capable of detecting other human emotional state through voice input without any semantic understanding. Through further analysis, we can understand much better about the motives and intentions of people, whether they are happy, sad, angry or neutral and much more. Humans are easily able to determine the emotion of a speaker, but the field of emotion recognition through machine learning is an open research area. Sentiment analysis is the process of mining text and audio to identify and categorize the subjective opinion expressed by the user. Day by day publicly and privately available information over internet is constantly growing, a large number of texts and audios available in review sites, blogs and social media. What people are saying, how they are saying it, and what they mean ? Sentiment Analysis is the domain of understanding these emotions with software the help of software. It is used to determine whether the user's attitude towards a particular topic or product etc. is positive, negative or neutral and can also be used to determine the emotion of user so that a machine can be much more responsive and interactive to a user more efficiently It is also used in businesses to help user understand the social sentiment of their brand, product or services while monitoring online conversions.

We begin our study of emotion in speech by first detecting the gender of voice sample i.e. whether the sample audio is of male voice or female voice and then the emotion containing in it. Specifically, we investigate the various emotions in speech samples. In our analysis of emotions, we start by delineating the data used and through analysis, we investigated the best algorithms to select features that are relevant to predicting emotion. We also consider multiple machine learning models to classify emotion. Sentiment analysis in programming is implied as task can be generated, work can be ordered, and the program can be notified as data can be processed with appropriate algorithms.

II. METHODOLY

The project aims at designing a prototype which firstly checks whether an audio file is of male voice or of a female voice and then the emotion is either angry, sad, happy or neutral . At a high level, each audio file from the repository is processed and used to build a feature vector with corresponding labels that are either angry, sad, happy or neutral. After extracting features, multiple feature selection algorithm would be studied and the most suitable one with some advancements would be used for implementation and for checking for most relevant results. The next step would be apply various models to the features with the maximum scores. To determine the testing and training error, we use k-fold cross validation.

III. FEATURE EXTRACTION

The representation of time-domain of a voice can be very complex and in its original form, it does not provide that much good knowledge in the key characteristics of the sample and because of this. The most straightforward technique involves determining the average energy of the sample.. Duration also offers knowledge of emotion, as we need to get the statistics like the maximum, minimum, range, mean, and standard deviation of both the audio sample and spectrum. They also may indicate fluctuations in the volume or pitch that can be useful in determining the emotion. For both the audio and spectrum, we also derive skewness, the measure of departure of horizontal symmetry in the signal, and kurtosis, the measure of height and sharpness of central peak, relative to a standard bell curve.

IV. LITERATURE SURVEY

Sentiment analysis is gaining immense popularity from the past few years due to its advantages in business, social upliftment, for expressing and understanding human emotions,etc. Therefore,it is catching the attention of the researchers throughout the world and has become a hot topic of this decade. [1] In this research, we classify the feedback based on its category and sentiment. Several classification algorithms are used in opinion mining; two of them are NBC (Naive Bayes Classifier) and SVM (Support Vector Machine). This paper aims to classify feedback based on sentiments using NBC and SVM .Previous works are focused on eliciting results from unimodal systems. Machines are generally used to predict emotions by only facial expressions or only vocal sounds until now. [3]In prior studies, several modalities have been explored to recognize the emotional states such as facial expressions , speech , physiological signals, etc. Several inherent advantages make speech signals a good source for affective computing.[4] For example, compared to many other biological signals , speech signals usually can be acquired more readily and economically. This is the reason why the majority of researchers are interested in speech emotion recognition (SER).
[5] The study of sentiment analysis has resulted into applications which could be beneficial for police investigation i.e. by the vocal signals the police can detect whether the accused culprit is saying the truth or not.Other applications include interpreting the feeling of the person interacting with robots, audio surveillance, web-based E-learning, commercial applications, clinical studies, entertainment, banking, call centers, cardboard systems, computer games, etc. For classroom orchestration or E-learning, information about the emotional state of students can provide focus on the enhancement of teaching quality. Three key issues need to be addressed for successful SER system, namely, choice of a good emotional speech database, extracting effective features, and designing reliable classifiers using machine learning algorithms. In fact, the emotional feature extraction may be a main issue within the SER system. [7] Many researchers have planned important speech choices that contain feeling information, like energy, pitch, formant frequency, Linear Prediction Cepstrum Coefficients (LPCC), Mel-frequency cepstrum coefficients (MFCC), and modulation spectral.

Thus, most researchers like to favour the use of combined feature set that is composed of the many styles of options containing more emotional information . The last step of speech feeling recognition is classification. It involves categorizing the information within the kind of auditory communication or frame of the auditory communication into a selected class of feeling on the idea of options extracted from the information. In recent years in speech feeling recognition, researchers proposed many classification algorithms, such as [8] Gaussian mixture model (GMM) , hidden Markov model (HMM), support vector machine (SVM) , neural networks (NN) , and recurrent neural networks (RNN) . Some other types of classifiers are also proposed by some researchers such as a modified brain emotional learning model (BEL) in which the adaptive neuro-fuzzy inference system (ANFIS).

[9] Multilayer perceptron (MLP) are incorporate for speech feeling recognition. Another proposed strategy is a multiple kernel Gaussian process (GP) classification , in which two similar notions in the learning algorithm are presented by combining the linear kernel and radial basis function (RBF) kernel. [8] All previously publishled work generally used the Internation language database like berlin database , Spanish database .To our knowledge, the Indian local Emotion database never been used before. For this reason, we have chosen to compare them. In this chapter, we concentrate to improve accuracy as well.

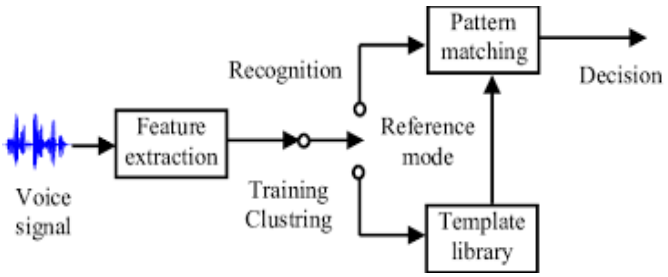
V. PROPOSED SYSTEM

Our project is divided into 4 stages, starting with data collection, data preprocessing, data classification and assessment. Data processing describes the process of preparing the data for audio analysis. Data classification gives the structure and pattern to the data, which is to be processed for sentiment analysis. The evaluation describes the classification results by comparing the entire pattern and then give results as follows.

i]. Data Collection

The data is collected from the audio file input and as a sample we have used around 200 audio files. The time-amplitude representation of sound can be represented in diverse types, it is in the raw form, and it does not provide any significant exact form, which will help for the audio analysis. Because of this characteristic of sound signals, we map this time-amplitude representation to extract more characteristics from the sound signals. Total energy in the signal, indicates the “volume” of the audio. Duration also offers insights into emotion, as do statistics like the max, min, average, range, mean, and standard deviation of both the signal and spectrum. These may indicate fluctuations in the volume or pitch that can be useful in determining emotion. The measure of height and sharpness of central peak, relative to a standard bell curve.

We also process the signal in the frequency domain through the Fourier Transform Pattern. We use audio pattern samples to get accurate representations of the frequency content of the signal at different points in time. In addition, we find the maximum and minimum frequencies with substantial power for each period, and use these values to determine the frequency range for each frame.



ii] Data Processing

After collecting the data, The Data Processing is the main part of our project with the pre-processing data stage to prepare the input data to be analysed afterwards and make into the pattern. The first process starts by removing all noise from the audio samples. Then it continues to merging all remain samples, and giving value for particular pattern. . After going through the data cleaning process and pre-processing data stage, both datasets have are compare.

iii]. Data Classification and assessment

After all above major parts it's time to merging and conclude the audio samples and give the basement by their value given by the processing method .Prior to the establishment of training data and test data, the collected data must be filtered first. The data is used for the establishment of training data as much as 70% and 30% as test data.and after the all process we have a core form of sample which is helpful for deciding the results.

Analyze and understand all the provided review comments thoroughly. Now make the required amendments in your paper. If you are not confident about any review comment, then don't forget to get clarity about that comment. And in some cases there could be chances where your paper receives number of critical remarks. In that cases don't get disheartened and try to improvise the maximum.

VI. CONCLUSION

Sentiment Analysis is an important area of Machine Learning .It plays a major role in data .mining, web mining and robotics. The proposed works are very less in dialectal languages like Indian regional languages even though many works already exists for universal languages like English,Spanish,etc. Although some of the works exist, it does not take into consideration fine-grained details like the aspect on which the user is commenting otherwise called as Aspect based sentiment Analysis. This work was an attempt to do aspect based sentiment analysis in regional languages. As a future work, emoticons can also be considered for machine learning and robotics, which directly impacts on human.

VII. REFRENCES

[1]"Analysing Employee Voice Using Real-Time Feedback" Andry Alamsyah1, Dessy Monica Ginting2School of Economics and Business Telkom University Bandung, Indonesia,(2018).

[2]"Analysing Vocal Patterns to Determine Emotion Maisy Wineman," Andy Sun,(2014).

[3] "Emotional Prosody Analysis on Human Voices" Natalie Ang, Doina Bein, David Dao, Luigi Sanchez, Justine Tran, Navie Vurdien Department of Computer Science California State University, Fullerton Fullerton, CA, USA,(2018).

[4] "Extracting Emotion from Voice" S. Fukuda* & V. Kostov, Department of Production, Information and Systems Engineering, Tokyo Metropolitan Institute of Technology Tokyo, Japan,(1999).

[5]"Alexa, My Love: Analysing Reviews of Amazon Echo" Yang Gao, Zhengyu Pan, Honghao Wang, and Guanling Chen Department of Computer Science University of Massachusetts Lowell, Lowell, MA 01852 ,USA,(2018).

[6]"Emotion Analysis: A Survey" International Conference on Computer, Communications and Electronics (CompTIA).Maniple University Jaipur, National Institute of Technology Jaipur & IRISWORLD,(2017).

[7]"Sentiment Analysis from Social Media in Crisis Situations" Ms. Harbinger Jeet Kaur Department of Computer Science & Engineering Sharda University Greater Noida, India,(2015).

[8]" Gender Identification by Voice" Kunyu Chen,Stanford university,(2014).

[9]"Hate speech classification in social media using emotional analysis",Ricardo Martins, Marco Gomes, JosJoao˜Almeida, Paulo Novais, Pedro HenriquesAlgoritmi Centre / Department of InformaticsUniversity of Minho, Braga – Portugal,(2018).

[10]"Interactive Voice Response using Sentiment Analysis in Automatic Speech Recognition Systems",amity university,uttar Pradesh,india,(2018).

[11]" Conversation-Based Natural Language Interface to Relational Databases" Majdi Owda, Zuhair Bandar, Keeley Crockett,(2007).

[12] ACOUSTICS, CONTENT AND GEO-INFORMATION BASED SENTIMENT PREDICTION FROM LARGE-SCALE NETWORKED VOICE DATA

[13]Acoustics,content and geo-information based sentiment prediction from large scale networked voice data,Zhu Ren1,2, Jia Jia1,2, Quan Guo 3, Kuo Zhang4, Lianhong Cai1,2Department of Computer Science and Technology, Tsinghua University, Beijing, China 2 TNList and Key Laboratory of Pervasive Computing, Ministry of Education,(2014).

[14] Aspect Polarity Recognition of Movie and Product Reviews in Malayalam, Thulasi P.K.PG Scholar Department of Computer ScienceNSS College of Engineering Palakkad,Kerala,(2016).

[15]"A Proposed Approach with Analysis of Speech Signals for Sentiment Detection"
16,Atul Tyagi,Nidhi Chandra,M.tech, Computer Science and Engineering,Amity University,Noida, India,(2015).