Speech Emotional Recognition

Group Members :
Nayan Kumar (20BAI10386) ,
Shubham Tejani(20BAI10152) ,
Parv Paliwal (20BAI10228) ,
Bhavyanagana Kanthed (20BAI10380).

Guide : Dr. Md Shah Fahad

Project Group ID:- 26

ABSTRACT

Speech Emotion Recognition, abbreviated as SER, is the act of attempting to recognize human emotion and the associated affective states from speech. This is capitalizing on the fact that voice often reflects underlying emotion through tone and pitch.

Emotion recognition is a rapidly growing research domain in recent years. Unlike humans, machines lack the abilities to perceive and show emotions. But human-computer interaction can be improved by implementing automated emotion recognition, thereby reducing the need of human intervention.

In this project, basic emotions like calm, happy, fearful, disgust etc. are analyzed from emotional speech signals.

INTRODUCTION

Speech emotion processing and recognition system is generally composed of three parts, the first being speech signal acquisition, then comes the feature extraction followed by emotion recognition. The conventional neural networks of Multi- Layer Perceptron (MLP) type have been increasingly in use for speech recognition and also for various other speech processing applications.

Speech recognition is the process of converting an acoustic signal, captured by microphone or a telephone, to a set of characters. They can also serve as the input to further linguistic processing to achieve speech understanding, a subject covered in section. As we know, speech recognition performs tasks that similar with human brain.

MODULES AND METHODS

We use machine learning techniques like **Multilayer perceptron Classifier (MLP Classifier)** which is used to categorize the given data into respective groups which are non linearly separated.

Mel-frequency cepstrum coefficients (MFCC), chroma and mel features are extracted from the speech signals and used to train the MLP classifier.

For achieving this objective, we use python libraries like Librosa, sklearn, pyaudio, NumPy and soundfile to analyze the speech modulations and recognize the emotion.

RESULTS

The results obtained from the evaluation phase show the effectiveness of the model compared to the baselines and the state of the art on the **RAVDESS dataset.** In particular, Fig. 1 shows Confusion Matrix and Table I. shows the values of precision, recall and F1-score obtained for the emotional classes like angry, happy, neutral and sad. The accuracy of our SER system comes out to be **65%** accurate.

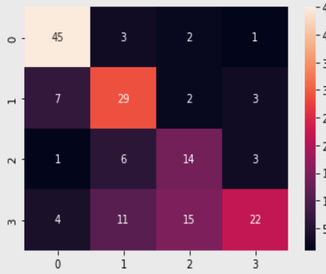

Fig 1. Confusion Matrix

	precision	recall	f1-score	support
angry	0.79	0.88	0.83	51
happy	0.59	0.71	0.64	41
neutral	0.42	0.58	0.49	24
sad	0.76	0.42	0.54	52
accuracy			0.65	168
macro avg	0.64	0.65	0.63	168
weighted avg	0.68	0.65	0.65	168

Table I. Classification Report

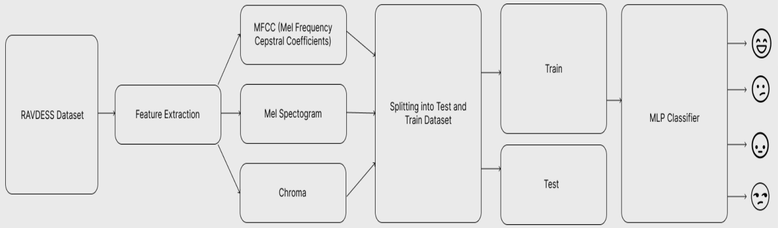

Fig 2. Speech Emotion Recognition System

DISCUSSION

This paper shows that MLPs are very powerful in classifying speech signals. Even with simplified models, a limited set of characters can be easily identified. We have obtained higher accuracies as compared to other approaches for individual emotions. .The results obtained in this study demonstrate that speech recognition is feasible, and that MLPs can be used for any task concerning recognizing of speech and demonstrating the accuracy of each emotion present in the speech.

CONCLUSIONS

The performance of a module is highly dependent on the quality of pre-processing. Mel Frequency Cepstrum Coefficients are very dependable. Every human emotion has been thoroughly studied, analyzed and the accuracy has been checked. The results obtained in this study demonstrate that speech recognition is feasible, and that MLPs can be used for any task concerning recognizing of speech and demonstrating the accuracy of each emotion present in the speech.

REFERENCES

- https://www.stat.purdue.edu/~vishy/papers/VisMur02b
- https://www.sciencedirect.com/science/article/pii/S1110 8_6
- https://pdfs.semanticscholar.org/d945/29ae612b76287c 7_dc
- https://bit.ly/3ECPt6O