
REAL TIME SPEECH EMOTION RECOGNITION USING MACHINE LEARNING

Dr. Nirmaladevi J¹, Aarthi K V², Vasundhara B³, Diwaan Chandar C S⁴, Abinaya G⁵

¹Associate Professor, Department of Information Science & Engineering

^{2,3,4} Second Year, Department of Information Science & Engineering

⁵ Second Year, Department of Information Technology

Bannari Amman Institute of Technology, Sathyamangalam, Erode, Tamilnadu - 638401

Abstract

A speech recognition system, which can detect emotions contained in the dataset such as sad, happy, neutral, angry, disgust, surprised, fearful and calm expressions. In real time we can use this application in various decisions. Although we are in a pandemic situation all processes are taking place only through online like job interviews, doctor appointments etc. In these cases this application is very useful whether in what state they are and to detect their emotions through speech. Here we are using a library called Librosa. Librosa is a python package for music and audio analysis. It provides the building blocks necessary to concoct music information retrieval systems. It was developed by Brian McFee, assistant professor of music technology and data science at NYU, and creator of Librosa, a python package for music and audio analysis. Librosa upholds a few elements connected with sound records handling and extraction like burden sound from a circle, register of different spectrogram portrayals, symphonious percussive source detachment, conventional spectrogram decay, stacks and translates the sound, Time-space sound handling, successive demonstrating, coordinating consonant percussive partition, beat-simultaneous and some more.

Keywords: *Speech Emotion Recognition; Librosa; CNN; Machine Learning; NLP*

1. INTRODUCTION

Speech emotion recognition is an indispensable tool to improve man-machine annexation. It is also used to judge a person's psychological, physiological state. Emotions play an important role in a person's day to day life. Emotions are mental states escorted by neurophysiological changes kindred with thoughts, feelings, behavioral responses and a degree of contentment and discontentment. Emotions are often entwined with mood, temperament, personality, disposition and creativity. Almost there 27 types of emotions put in place by humans. Emotions are reactions that human beings experience in response to events or situations. Speech emotions recognition(SER) is mostly beneficial for applications, which need human-computer interaction such as speech synthesis, customer service, education, forensics and medical analysis. Speech being a primary medium to pass information, we humans can also understand the intensity and mood of the speaker by the speech data generated. Recognizing emotional conditions in speech signals is a challengeable area for several reasons. First issue of all speech emotion methods is to select the best features, which will be powerful enough to distinguish between different emotions. The idea of creating this project was to build a machine learning model that could detect emotions from speech we have with us all the time.

2. LITERATURE REVIEW

[1] **Title:** "Speech Emotion Recognition"

Author: Ashish B. Ingale, D. S. Chaudhari

The database for the speech emotion recognition identifies emotional speech samples and the features extracted from these speech samples are the energy, pitch, linear prediction cepstral coefficient (LPCC),

Mel frequency cepstrum coefficient (MFCC). The average accuracy of most of the classifiers for speaker independent systems is less than that for the speaker dependent. This achieved an accuracy of 70% for seven emotional states.

[2] Title: “Emotion Recognition from Audio Using Librosa MLP Classifier”

Author: Prof. Guruprasad G, Sarthik Poojary, Simran Banu, Azmiya Alam, Harshith K R

In these they detect a person’s emotion just by their voice Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset. The primary aim of the project is to contribute to the theoretical debate currently occupying music and emotion research by systematically comparing evaluations of perceived emotions using two different theoretical frameworks: the discrete emotion model, and dimensional model of affect.

[3] Title: “Speech Emotion Recognition System using Librosa for Better Customer Experience”

Author: Subhadarshini Mohanty, Subasish Mohapatra, Amlan Sahoo

An exact implementation of the speaking speed can be investigated to see whether there is any inaccuracy, thereby resolving some of the model's flaws. Pitch has a more effective way to generate more leads and engagement to generate higher revenue by analyzing the speech. Trying to figure out how to remove the audio clip's aimless stillness. Exploration of various acoustic features of sound data is being investigated to see if they may be used in the realm of speech emotion identification.

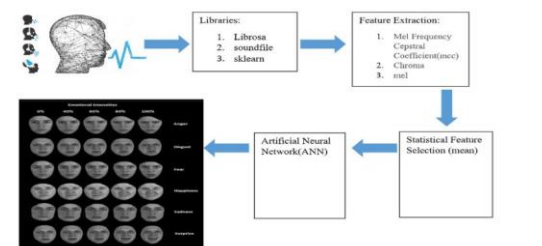
3. LIBROSA VS CNN

Librosa is a python package used for analyzing music and audio through which we can detect emotions. Librosa is primarily utilized for working with audio data by visualizing the audio signals performing automatic speech recognition. The model is trained using several user speech inputs made by different audios. The audio data is analyzed and then its features are extracted by librosa library. Convolutional Neural Networks (CNNs) have shown great potential for audio classification and have excelled in classifying images. Previously we have tried it with CNN. But the accuracy is lower than accepted. Also it takes more time to train the model.

4. METHODOLOGY

- Stockpile the dataset.
- Enthroned librosa soundfile.
- Extricate the features for data processing.
- Using MLP classifiers the internal artificial neural network is used for the purpose of classification.
- Through these we can detect the emotions.
- Preparing the dataset , here we can download and convert the dataset.
- Loading the dataset process is about loading the dataset in python which involves extracting audio features.
- By pedagogy the model after we prepare and load the dataset, train it on a suited sklearn model.
- By examining the model we can measure how good our model is doing.

5. WORKFLOW



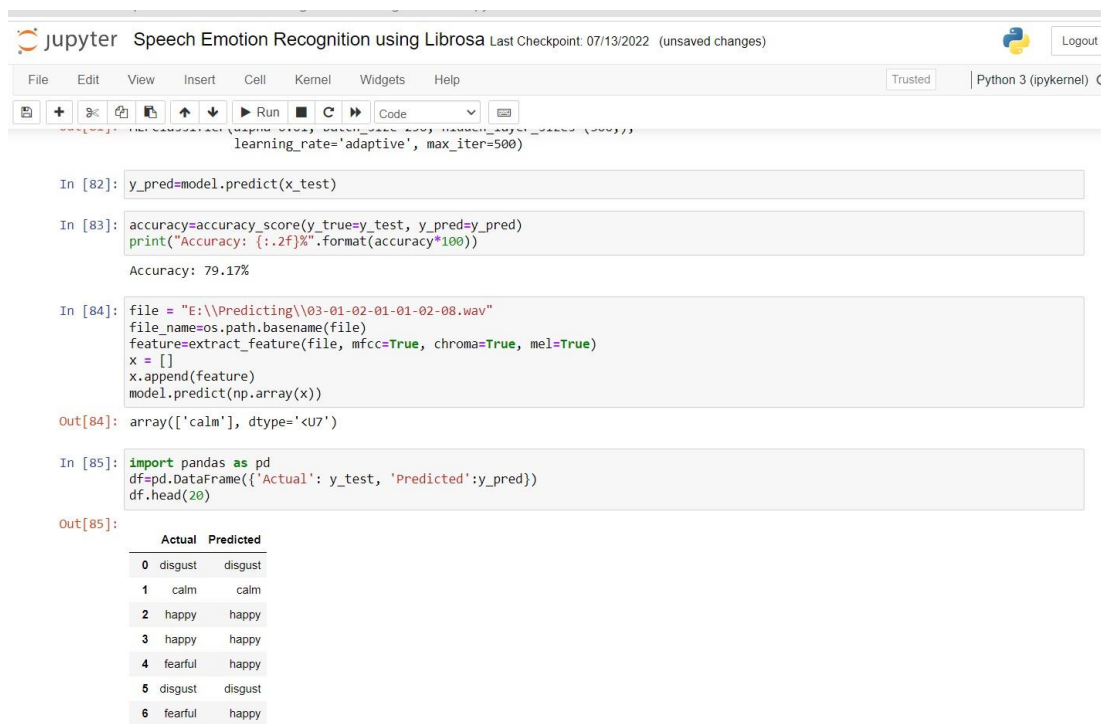
6. RESULT ANALYSIS

Previously we have tried it with CNN(Conventional Neural Network) but the accuracy is lower than the expected level. Also it takes more time to train the model. Later we used Librosa library which is more accurate and takes less time compared with CNN. Librosa is basically used when we work with audio data generated by automatic speech recognition. It helps to visualize audio signals and also do the feature extractions in different signal processing techniques. This model is trained using several user speech inputs made by different actors. The audio data is analyzed and then its features are extracted using the Librosa library.

7. CONCLUSION

The use of several classifiers in speech emotion recognition systems is illustrated. The signal processing unit in which appropriate characteristics are extracted from the available speech signal and another is a classifier that classifies emotions from the speech signal are the key components of a speech emotion detection system. In real time we can use this application in various decisions. Although we are in a pandemic situation all processes are taking place only through online like job interviews, doctors appointments etc. In these cases this application is very useful whether in what state they are and to detect their emotions through speech. Also by extracting more effective features of speech, accuracy of the speech emotion recognition system can be enhanced.

In previous research they have achieved an accuracy of 70% for seven emotional states. In another study Support Vector Machine for speech motion recognition of the four different emotions with an accuracy of 73% was obtained, We have obtained the accuracy of 79.17% by using real time circumstances.



```
learning_rate='adaptive', max_iter=500)

In [82]: y_pred=model.predict(x_test)

In [83]: accuracy=accuracy_score(y_true=y_test, y_pred=y_pred)
print("Accuracy: {:.2f}%".format(accuracy*100))
Accuracy: 79.17%

In [84]: file = "E:\\Predicting\\03-01-02-01-01-02-08.wav"
file_name=os.path.basename(file)
feature=extract_feature(file, mfcc=True, chroma=True, mel=True)
x = []
x.append(feature)
model.predict(np.array(x))

Out[84]: array(['calm'], dtype='<U7')

In [85]: import pandas as pd
df=pd.DataFrame({'Actual': y_test, 'Predicted':y_pred})
df.head(20)

Out[85]:
```

	Actual	Predicted
0	disgust	disgust
1	calm	calm
2	happy	happy
3	happy	happy
4	fearful	happy
5	disgust	disgust
6	fearful	happy

8. REFERENCES

- [1]Speech Emotion Recognition [Ashish B. Ingale, D. S. Chaudhari]
- [2]Emotion Recognition from Audio Using Librosa MLP Classifier [Prof. Guruprasad G, Mr. Sarthik Poojary, Ms. Simran Banu, Ms. Azmiya Alam, Mr. Harshith K R]



- [3]Speech Emotion Recognition System using Librosa for Better Customer Experience [Subhadarshini Mohanty, Subasish Mohapatra, Amlan Sahoo]
- [4]Puri, Tanvi, et al. "Detection of Emotion of Speech for RAVDESS Audio Using Hybrid Convolution Neural Network." *Journal of Healthcare Engineering* 2022 (2022).
- [5]Xu, Mingke, Fan Zhang, and Wei Zhang. "Head fusion: improving the accuracy and robustness of speech emotion recognition on the IEMOCAP and RAV.
- [6]"Using Deep and Convolutional Neural Networks for Accurate Emotion Classification on DEAP Dataset," *Proceedings of the Twenty-Ninth AAAI Conference on Innovative Applications (IAAI-17)*.
- [7]B. Yang, M. Lugger, "Psychological motivated multistage emotion classification exploiting voice quality feature." F. Mihelic, J. Zibert, *Speech Recognition, InTech*, 2008, chapter 22.
- [8]Steven R. Livingstone, Frank A. Russo,"The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English",*journal.pone.0196391*,May 16, 2018 .
- [9]S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis using physiological signals," *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 18–30, 2012.
- [10] Tarunika, K., R. B. Pradeeba, and P. Aruna. "Applying machine learning techniques for speech emotion recognition." *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*. IEEE, 2018.