

VOICE EMOTION DETECTION USING ML CLASSIFIER

Course: Machine Learning LAB(CSE-4212)

Saharab Hossain|Tarekul Islam|Sudipta Roy Chowdhury

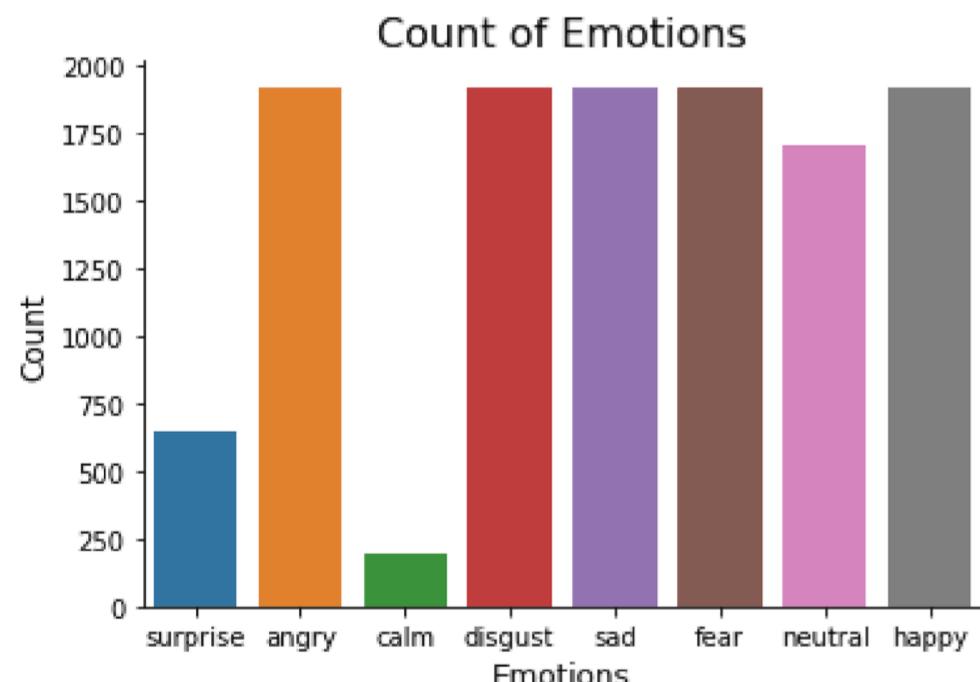
INTRODUCTION

Detecting emotions is one of the most important marketing strategy in today's world. You could personalize different things for an individual specifically to suit their interest. For this reason, we decided to do a project where we could detect a person's emotions just by their voice which will let us manage many AI related applications.

DATASET DESCRIPTION

We use four kinds of dataset.

- Crowd-sourced Emotional Multimodal Actors Dataset (Crema-D)
- Ryerson Audio-Visual Database of Emotional Speech and Song (Ravdess)
- Surrey Audio-Visual Expressed Emotion (Savee)
- Toronto emotional speech set (Tess)



METHODOLOGY

Step1: Upload Data

Step2: Data Preprocessing

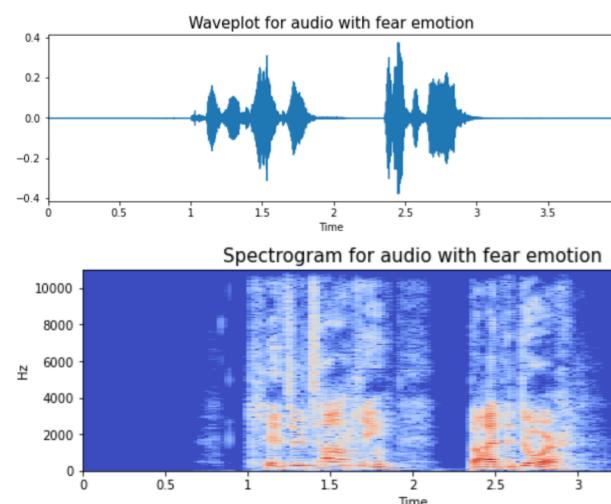
Step3: Data Augmentation

- Simple Audio
- Noise Injection
- Stretching
- Shifting
- Pitch

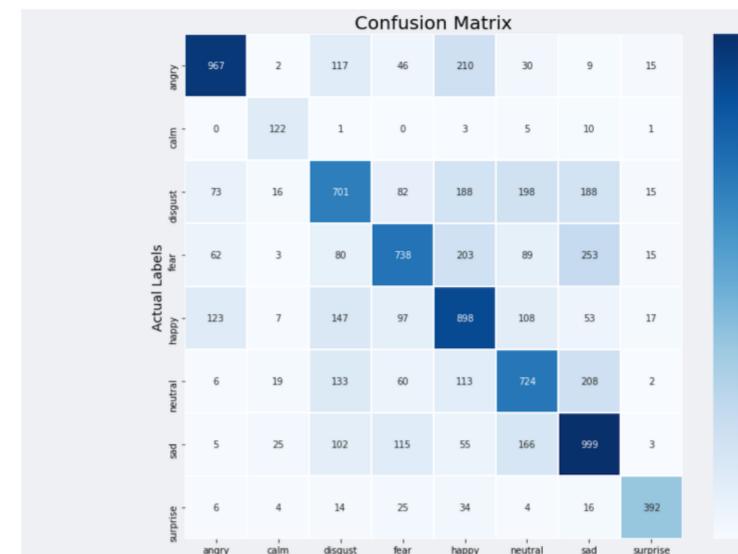
Step4: Feature Extraction

Step5: Data Preparation

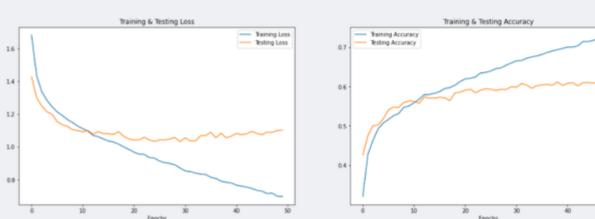
Step6: Modelling



This two figure shows the wave and spectrum view of an audio. Shows the sequence of the audio with respect to time and hertz.



This confusion matrix is an 8 X 8 matrix that is used to evaluate the performance of the Sequential model classification model, where 8 is the number of target classes.



This figure shows the Sequential model performance with train and test accuracy and train and test loss.

RESULT

We can see our model is more accurate in predicting surprise, angry emotions and it makes sense also because audio files of these emotions differ to other audio files in a lot of ways like pitch, speed etc.. We overall achieved 61% accuracy on our test data.

MODEL	ACCURACY	RECALL	Precision	SUPPORT
Sequential model	0.61	0.65	0.63	9122

CONCLUSION

In this project we discussed the basic process for emotional speech detection and the most important part of emotion detection of speech is feature extraction. Above review states that the prosody features such as pitch, formant, intensity, energy and spectral features such as LPC and MFCC are most commonly used feature.

FUTURE WORK

We will build a mobile application by using this project which will helpful for detection emotion by any kind of voice and can collaborate with them.

REFERENCES

- [1] Md. Touseef Sumer, "Salient Feature Extraction For Emotion Detection Using Modified Kullback Leibler Divergence" in International Journal of Research in Engineering and Applied Science(IJREAS), Jan 2014
- [2] Akshay S. Utane Dr. S.L.Nalbalwar, " Emotion Recognition Through Speech Using Gaussian Mixture Model And Hidden Markov Model" in International Journal of Advanced Research in Computer Science and Software Engineering, April 2013
- [3] Mohammed E. Hoque, Mohammed Yeasin, Max M. Louwerse, "Robust Recognition of Emotion from Speech"