Analyzing Vocal Patterns to Determine Emotion

PROJECT SYNOPSIS

OF Internship

BACHELOR OF TECHNOLOGY

In

Computer Science

SUBMITTED BY

Team Members

Kshitiz- R970218030

Animesh Anand- R970218005

Aryaman Singh- R100218007

Kartikey Semwal - R142218074

GUIDED BY

Dr. Prakash G.L

(Professor SOCS)

Internship Report

Table of Contents

Executive Summary

1. Background

1.1 Aim :

Analyzing Vocal Patterns to Determine Emotion

1.2 Technologies:

- Python
- Visual studio Code
- Jupyter Notebook

1.3 Hardware:

- GPU Extensive System
- CPU Extensive System
- Minimum 4 gb RAM
- Dedicated Graphic Card

1.4 Software

- Programming Language: Python
- Libraries: numpy, matplotlib.pyplot, tensorflow, librosa, pandas, keras, sklearn, model.json, os, sys

2. System

2.1Requirements

2.1.1 Functional requirements

- Simulates the functioning of an automatic car.

2.1.2 User requirements

- Working System with jupyter notebook and python installed in system.

2.1.3 Environmental requirements -

- Jupyter Lab using Python

2.2 Design and Architecture

Analyzing Vocal Patterns to Determine Emotion, this project make use of SAVEE & RAVDESS dataset to train model for classification and test it for audio data provided by the user to classify feature present in the audio by MFCC feature extraction and use it to predict the nature of an individual with certain range of emotion.

Algorithm:

- Mel Frequency Cepstral Coefficients (MFCC)
- Convolutional Neural Network (CNN)

2.3 Implementation-

2.1 Data Input:

User need to provide audio data sample to be used in trained model for classification and prediction.

2.2 Data Pre-processing:

The MFCC features will be extracted from the audio data sample provided by user and fed into the trained model for classification and prediction.

2.4 Testing

2.4.1 Test Plan Objectives

- To check the model accuracy for labeled testing data and random testing data.
- To determine the accuracy for gender and emotion prediction.

2.4.2 Data Entry

- Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) Database
- Surrey Audio Visual Expressed Emotion (SAVEE) Database

2.4.3 Test Strategy

- Check the model against provided tested data to determine whether the model works the way we want it to.
- Check the model with user testing data to check the model prediction performance up to the mark.
- Check it for noisy data to test its working limit.

2.4.4 System Test - Successful

2.4.5 Performance Test - Successful

2.4.6 Basic Test- Successful

2.4.7 User Acceptance Test -Successful

2.4.8 System - Successful

2.7 Evaluation

2.7.1 Table

1: Performance

- Gender at **97%** absolute accuracy
- Emotion at **47%** absolute accuracy
- Gender and Emotion at **46%** absolute accuracy

2.7.2 STATIC CODE ANALYSIS - Successful

2.7.3 WIRESHARK -Not used in our Project

3. Conclusions

Building the model was a challenging task as it involved a lot of trial and error methods, tuning etc. The model is very well trained to distinguish between male and female voices and it distinguishes with 90% accuracy. The model was tuned to detect emotions with more than 70% accuracy. Accuracy can be increased by including more audio files for training.

4. Further development or research

- In Future, to implement it in an application as well as make an online website in which the user would be able to perform live classification on his/her voice.
- To increase Efficiency for better classification and understandability of features present inside an audio.
- Perform Classification not only on small phrases but on proper sentences as well to grasp a wider range of emotions a human being feels through the conversations.
- Including Hindi Language data set and performing classification to increase the versatility of the model and make it more robust.
- Increasing model accuracy by doing data augmentation and training it with noisy data to increase its working accuracy.

5. References

•(PDF) MFCC-Based Feature Extraction Model for Long Time Period Emotion Speech Using CNN (researchgate.net)

•Audio Classification Using CNN — An Experiment | by The Experimental Writer | AI Graduate | Medium

•Some Commonly Used Speech Feature Extraction Algorithms | IntechOpen

•https://link.springer.com/chapter/10.1007/978-981-15-9293-5_31

•Ellis, Daniel P. W. "PLP and RASTA (and MFCC, and Inversion) in Matlab using Melfcc.m and Invmelfcc.m." PLP and RASTA (and MFCC, and Inversion) in Matlab Using Melfcc.m and Invmelfcc.m. 1 Jan. 2005. Web
M. Liberman, et al., "Emotional Prosody Speech and Transcripts," Philadelphia; Linguistic Data Consortium. 2002. - References

•Emotional Prosody Speech and Transcripts - Linguistic Data Consortium