PES
UNIVERSITY

A mini project report on

“Audio Classifier to Differentiate between Male and Female Voice”

Submitted in partial fulfilment of the requirements for the machine learning laboratory during 6th semester of

Bachelor of Technology
in
Computer Science & Engineering

Submitted by :

01FB16ECS262 : Prajwal S
01FB16ECS282 : R V Prithvi
01FB16ECS292 : Rakesh Reddy

Under the guidance of

Mr. Srinivas KS
Internal Guide
Assistant Professor

January – May 2018

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
FACULTY OF ENGINEERING
PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

PES UNIVERSITY
(Established under Karnataka Act No. 16 of 2013)
100ft Ring Road, Bengaluru – 560 085, Karnataka, India

FACULTY OF ENGINEERING

CERTIFICATE

This is to certify that the mini project entitled

‘Audio Classifier to Differentiate between Male and Female Voice’

is a bonafide work carried out by

01FB16ECS262 : Prajwal S
01FB16ECS282 : R V Prithvi
01FB16ECS262 : Rakesh Reddy

In partial fulfilment for the machine learning laboratory during sixth semester in the Program of Study Bachelor of Technology in Computer Science and Engineering under rules and regulations of PES University, Bengaluru during the period Jan. 2019 – May. 2019.

Signature

Signature
Dr. Shylaja S S
Chairperson

Name of the Examiners Signature with Date

1. ___________________

2. ___________________

TABLE OF CONTENTS

Definitions, Acronyms and Abbreviations

This section provides for definition of all terms, acronyms and abbreviations required for interpreting the High Level Design Document. Well known abbreviations need not be stated

References

This section describes the complete list of documents referred to prepare the High Level Design. This section shall describe the title, version number, dates, authors and publishers of the referenced documents whenever applicable.

If industry standard methodology is used for design, it will be clearly mentioned here. If however, other methodologies are used, the deviation from a standard methodology will be clearly described.

1.0 Introduction

1.1 Overview

Audio classification is a fundamental problem in the field of audio processing. The task is essentially to extract features from the audio, and then identify which class the audio belongs to. Our project is able to classify audio into two classes namely, Male Voice and Female Voice based on the various features that is been extracted from the audio files.

1.2 Scope

This can be employed in Personal Assistance to improve the interaction based on the gender of the person that is being identified in our project.

1.3 Objective

The efficiently classify audio file based on gender.

2.0 Literature Survey

CNN - https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148.

Audio Classification - https://ai.google/research/pubs/pub45611

Keras - https://arxiv.org/abs/1703.09179

3.0 Methodology

3.1 Proposed Approach

- Convolutional Neural Network is being used to classify Audio file based on the gender.
- The usage of CNN is motivated by the fact that they can capture / are able to learn relevant features from a speech at different levels similar to a human brain. This is feature learning.
- For a completely new task / problem CNNs are very good feature extractors. This means that you can extract useful attributes from an already trained CNN with its trained weights by feeding your data on each level and tune the CNN a bit for the specific task. Eg : Add a classifier after the last layer with labels specific to the task. This is also called pre-training and CNNs are very efficient in such tasks compared to NNs. Another advantage of this pre-training is we avoid training of CNN and save memory, time. The only thing you have to train is the classifier at the end for your labels.

3.2 High Level System Architecture

- CNN : In deep learning, a **convolutional neural network** (**CNN**, or **ConvNet**) is a class of deep neural networks, most commonly applied to analyzing visual imagery. CNNs are regularized versions of multilayer perceptrons. Multilayer perceptrons usually refer to fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer.

- 1D CNN - A 1D CNN is very effective when you expect to derive interesting features from shorter (fixed-length) segments of the overall data set and where the location of the feature within the segment is not of high relevance.

4.0 Environment Requirements

4.1 *Hardware Requirements*

PC with a minimum of 4GB RAM. Training can be improved using GPU's

4.2 *Software Requirements*

These are the python libraries that has been used

1. Numpy
2. Keras
3. Librosa
4. Glob
5. Pandas
6. Sklearn

The latest versions of the above mention libraries were sufficient for the completion of our project

4.3 *Data Requirements*

Audio recording of the word "STOP" from Google Speech Commands Dataset has been used in our project. 500 audio files were manually annotated based on gender for the creation of the used DataSet.

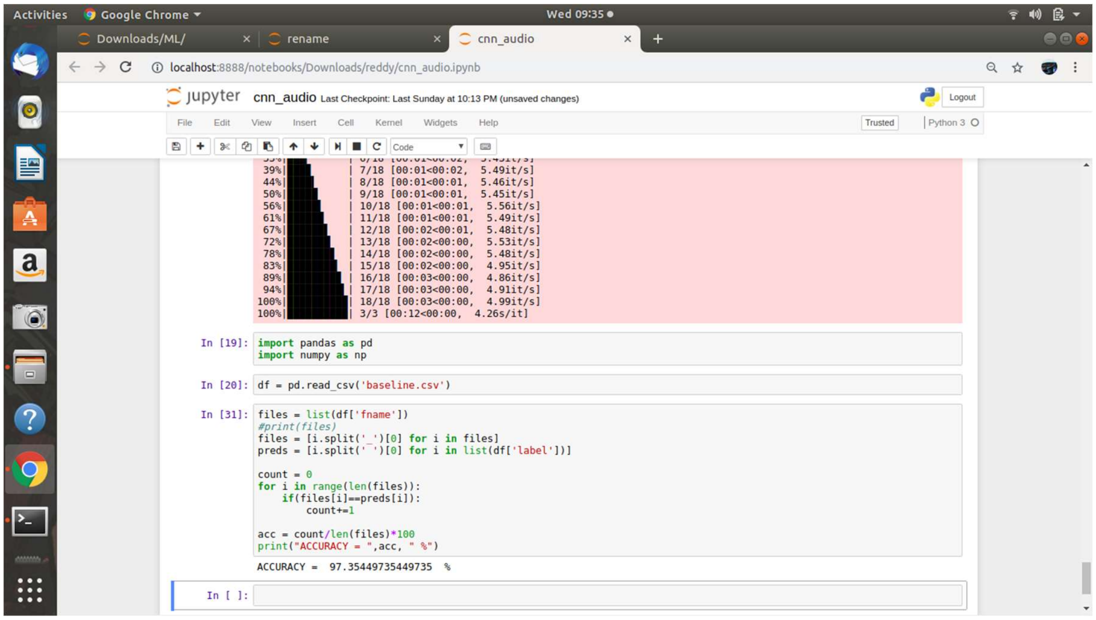
5.0 Proposed Approach

Our approach begins with extracting the audio file. An audio file is converted to a vector which captures the amplitudes of the audio signal at an interval specified by the sampling rate. This vector serves as the input to the neural network model.

The neural network model is a Convolutional Neural Network that is used to extract features over successive convolutional operations. The input is passed through 1D convolution layer twice and maxpool is applied to extract features and finally a dropout is applied. This phase is applied four times to extract maximum features possible. The features are then passed to a Dense layer to classify the gender of the audio file. A softmax activation function is used to return the probabilities of each class. The maximum probability class is predicted as the output.

6.0 Results

Our model is able to classify audio files with an accuracy of 97.354%.



```
 39%|█████     | 7/18 [00:01<00:02,  5.49it/s]
 44%|█████▌    | 8/18 [00:01<00:01,  5.46it/s]
 50%|██████    | 9/18 [00:01<00:01,  5.45it/s]
 56%|██████▌   | 10/18 [00:01<00:01,  5.56it/s]
 61%|███████   | 11/18 [00:01<00:01,  5.49it/s]
 67%|███████▌  | 12/18 [00:02<00:01,  5.48it/s]
 72%|████████  | 13/18 [00:02<00:00,  5.53it/s]
 78%|████████▌ | 14/18 [00:02<00:00,  5.48it/s]
 83%|█████████ | 15/18 [00:02<00:00,  4.95it/s]
 89%|█████████▌| 16/18 [00:03<00:00,  4.86it/s]
 94%|█████████▉| 17/18 [00:03<00:00,  4.91it/s]
100%|██████████| 18/18 [00:03<00:00,  4.99it/s]
100%|██████████| 3/3 [00:12<00:00,  4.26s/it]
```

```
In [19]: import pandas as pd
         import numpy as np

In [20]: df = pd.read_csv('baseline.csv')

In [31]: files = list(df['fname'])
         #print(files)
         files = [i.split('_')[0] for i in files]
         preds = [i.split(' ')[0] for i in list(df['label'])]

         count = 0
         for i in range(len(files)):
             if(files[i]==preds[i]):
                 count+=1

         acc = count/len(files)*100
         print("ACCURACY = ",acc, " %")

         ACCURACY =  97.35449735449735  %

In [ ]:
```

7.0 Conclusions

Convolutional Neural Network has worked well to classify audio files based on gender.

8.0 Future Work

Our model could be extended to classify more audio types.

9.0 References

CNN - https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148.

Audio Classification - https://ai.google/research/pubs/pub45611

Keras - https://arxiv.org/abs/1703.09179