Speech classification using MATLAB

Nasrun Sithara Ramees, Ahamad Rishard
School of Electrical Engineering and Computer Science(SEECS), NUST,
Islamabad, Pakistan

Abstract— Audio Classification is performed using Machine Learning that involves identifying various audio signals into different classes or categories. The goal of audio classification is to enable machines to automatically recognize and distinguish between different types of audios, including music, speech, and environmental sounds. The model we have created focuses on speech recognition of an audio signal that uses the Classification Learner application in MATLAB.

Keywords— Classification Learner, LPC, Linear SVM

I. INTRODUCTION

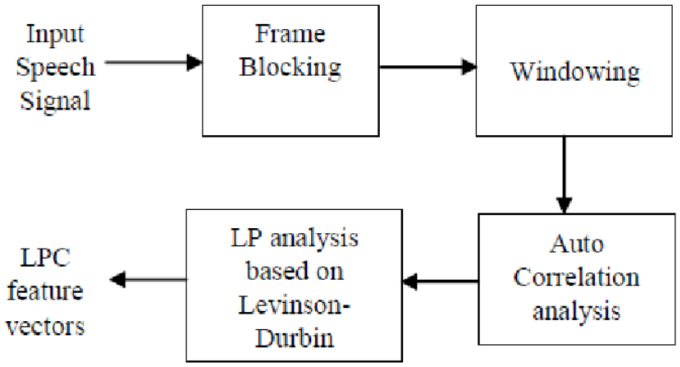
The speech recognition system we created in MATLAB using the Classification Learner application classifies spoken words. In real-life applications it is widely used in a variety of domains, including language learning applications, voice-controlled assistants, and transcription services.

II. PROPOSED METHOD

The basic objective of our project is training an ML model to detect two given audio samples. To achieve this, we started with a data collection approach. Here we are going to use two words TRUE and FALSE. Each of these words are recorded 30 times each and stored in our workspace.
Now 30 samples of TRUE audio and 30 samples of FALSE audio are passed to the feature extraction process. Here the features are extracted using the LPC function in MATLAB which will extract 10 LPC coefficients for each audio clip. LPC function basically represents any audio signal as a linear combination of its past samples. Each of the coefficients will be coefficient of a past sample and the formula is given by:

$$\hat{S}[n] = \sum_{k=1,p} a_k \, s[n-k]$$

Where S[n] is the audio sample. This is achieved by using active FIR filters which will break the signal into its past components values after passing the signal through a window. The below diagram taken from a research paper (reference given) shows the basic block diagram of the LPC function.



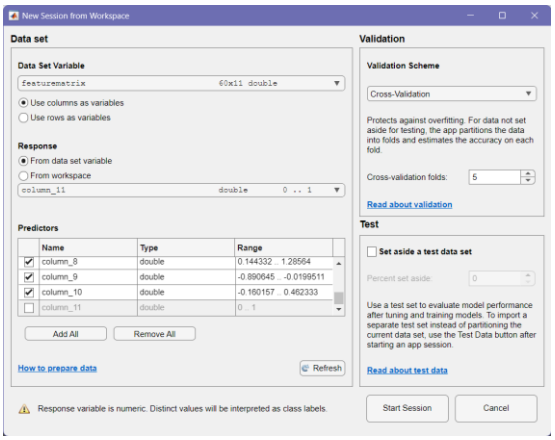
After extracting the features, we store the results in a matrix and pass that matrix for training purposes.

For training we use the classification learner inbuilt MATLAB app to feed in data to train the machine learning model. Here there are a variety of trained models, and we can choose the model which has the highest accuracy.
Coding comprises of three main parts.

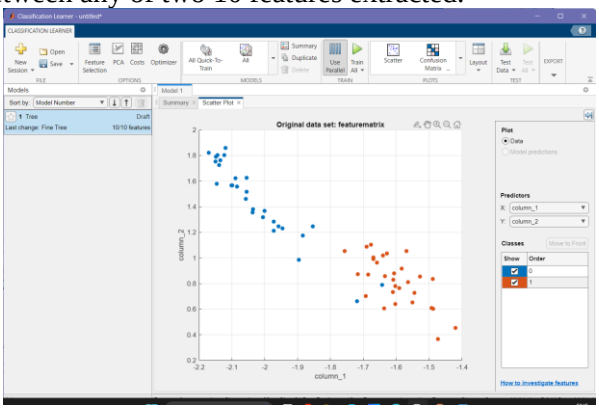
1. Data collection: Here the audio sample for 2 audio clips are collected and stored in the workspace of MATLAB. Each audio sample is taken for 3o samples thus giving a total of 60 audio files. The key approach here was to automatically name the audio files in a ordered sequence and store them in the workspace.
2. Feature Extraction: The audio samples were passed to the LPC function to extract the LPC coefficients and store them in a 60*10 matrix which will be used for training. Also, to name features of each audio file we created a 60*1 matrix which will be concatenated with the data matrix thus each row will be identified with a number to denote the audio samples.
3. Testing: After extracting the features training the model, we import the model to our work space and test a new sample of audio. The model will extract the features of the new audio sample and compare it with trained model and either predict it as TRUE or FALSE.

III. RESULTS AND DISCUSSIONS

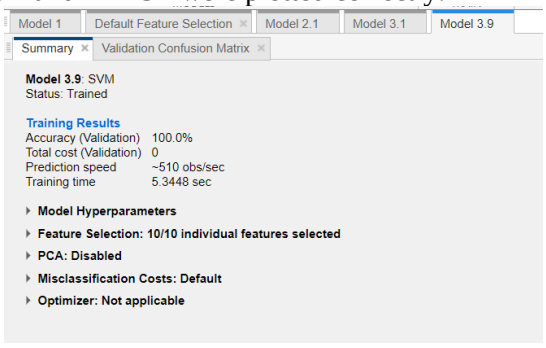
The LPC coefficient matrix is passed to the training classification learner.



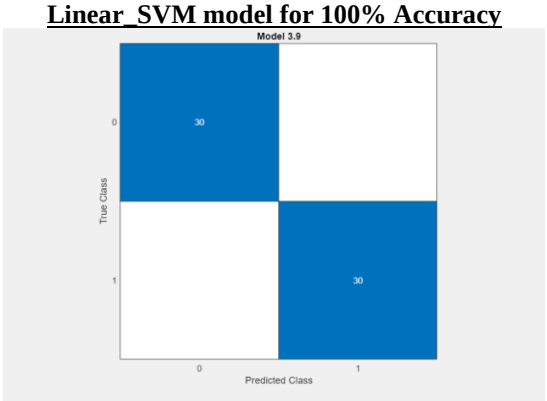
The feature is plotted in a scatter plot. Here we plot between any of two 10 features extracted.



After training the output of the samples were tested using a confusion matrix and the 30 samples of both TRUE and FALSE were plotted correctly.

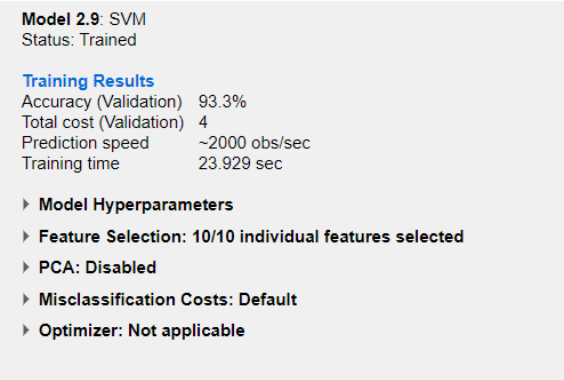


Linear_SVM model for 100% Accuracy

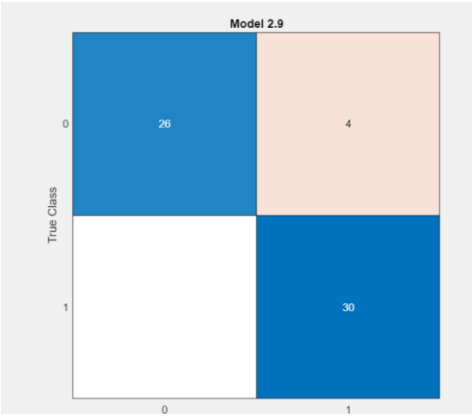


Confusion matrix for 100% accurate model

Since the data set, we used was a clean data set without any discrepancies we were able to achieve 100% accuracy in most of the models we trained. we selected the linear SVM model to test the data sets and verified the model by passing new audio signals of TRUE and FALSE and the model successfully detected the signals. The important feature here is whenever we pass any audio signals other than TRUE or FALSE the model will extract the features of that signal and compare it with the features of TRUE or FALSE and it will map that signal to either TRUE or FALSE depending on which is having the closest features.
After testing our results, we inserted 3 affected signals to train the TRUE data set and observe the results in the model. And after training we obtained a 93.3% accuracy for linear SVM. This can be modelled as an ideal case as always; we will not get pure data for training.



Linear_SVM model for 93.3% Accuracy



Confusion matrix for 93.3% accurate model

In these results, out 30 samples tested for TRUE we obtained 4 samples which were wrongly mapped as FALSE.

There are many applications that use audio classification.

- Music Genre Classification

Audio classification can be used for classifying music into various genres automatically, such as jazz, pop, electric, rock. This is a very useful application for arranging music libraries, analyzing music consumption patterns or recommending songs to users.

- Anomaly Detection

Audio classification can be used to detect unusual events in audio data. For instance, it can be used to identify unusual sounds in security systems such as alarms, gunshots or breaking glass.

- Forensic audio analysis

Audio classification can be used for forensic investigations to identify audio evidence, such as identifying voices, detecting tampering or determining the origin of audio recordings.

- Voice assists to detect emotions of a person.

We can train our model to detect variations in voice tones to detect if the person is feeling emotions such as sad, angry, or happy. Therefore, which enhances the AI model to respond accordingly.

IV. CONCLUSIONS

The above model was able predict the two words TRUE and FALSE signal with precise accuracy after training. The performance of the model is dependent on the quality and quantity of the training data; therefore, we can increase the training data set and train more than two words to detect. Also, to ensure that the training data does not have any defects we trained our own data set which is reliable. The model created should be able to work on unseen data. It is necessary to avoid data leakage and ensure the model is not learning specific characteristics of the training data that is not held in the real world.

REFERENCES

[1] Train models to classify data using supervised machine learning - MATLAB. (n.d.). https://www.mathworks.com/help/stats/classificationlearner-app.html

[2] ANALYSIS OF HEARTBEAT ANOMALIES FROM DIGITAL STETHOSCOPE AUDIO - UTP Electronic and Digital Intellectual Asset. (n.d.). http://utpedia.utp.edu.my/id/eprint/19171

[3] kavita, Dr & Saxena, Akash & Joshi, Jitendra. (2019). Speech Recognition Challenges by using Neural Network Approaches. 2348-2117.