Audio Classification

1 Problem Statement

Modeling audio classification where audio are of three category namely noise, speech, and music.

2 Explanation

In this problem we have data audio data where 3 different type of audio sample. These are noise, speech, and music. Here we have to make a model which will classify these three type of audio data to three respective category. First task will be making a model which will learn to classify these samples i.e. training model. After completion of training our model when we give any other audio record then it will classify whether it's a music or speech or noise.

3 Literature Review

Audio data are sequential data i.e. time dependent data. Here sequence w.r.t. time matters and because of this only machine learning algorithms are not preferred in any use case related to sequence data. There are other algorithms in deep learning like Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) where other dimensions are also taken care. In case of Audio (i.e. time dependent) RNN is most preferred algorithm as it takes time in consideration.

When we talk about audio so first thing is how to represent our audio data. For reading audio most common format to read audio is taking their amplitude as a sequence of time. There are many way to represent audio other then just amplitude and those are called audio features like spectogram, mfcc etc.

For this problem we are using zcr value as for these type of audio zcr value is very distinctive.

4 Data

Musan is a dataset for audio, which contains audio of three categories

1. music
2. speech
3. noise

Each type of audio is in different folder.

Link: http://www.openslr.org/17/

Sample audio :
Audio/Data/music/fma-western-art/music-fma-wa-0000.wav

etc.

file counts:
music: 43
speech: 13
noise: 101

5 Deliverable

Model should be able to classify other audio sample accurately.

6 Evaluation

Evaluation metric used: Accuracy

In first step of modeling we are apply rule based model so there we need to check how many are classified correctly And as data which we give to neural network model is balanced data so accuracy is good measure.

7 Data Ingestion

From Data folder we are reading these audio files using librosa library as array.

8 Data Analysis

As this data is recorded with good SNR value so considering data as good we are using actual data in our modeling without processing it.

9 Data Munging

Not Required

10 Data Exploration

Not Applicable

11 Feature Engineering

Here we are calculating zcr value for each audio file as for music, noise, and speech differentiation zcr is a good representation.

12 Modeling

This is done in two steps:

1. Identifying speech
For this zcr value are used. As for speech zcr is pretty high compared to music and noise.

2. Classifying music and noise

For this we are using mlp (multi layer perceptron) and rnn (recursive neural network) model. MLP is used as we have not many data points and multilayer will learn nonlinearity. RNN is used so that time dependency will be taken care and when we get some more data then it's performance will increase.

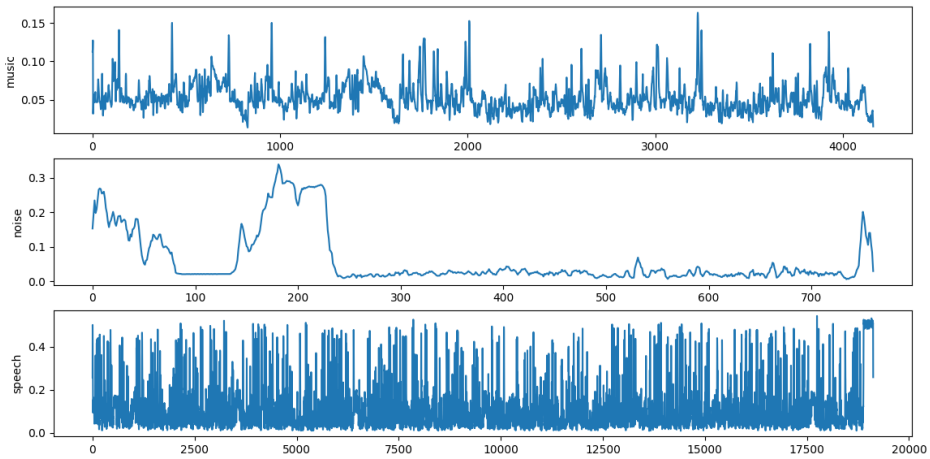
13 Optimization

Due to time taken by this model for now we are not apply hyper-parameter optimization as it will take lot of time.

14 Prediction

For the first model which is speech identification we are getting 100 % accuracy as this model is classifying all speech correctly.

15 Visual Analysis



Here we can see that for speech zcr plot is very high i.e more than 0.4 while in other case its less than 0..3 and 0.15.

16 Results

Speech separation was done with 100% accuracy as it classified all speech as speech. While for music we got 75% accuracy which was done on part of speech due to large training time.