Single Event Detection - Sound Classification

Swathi Pratapa

October 12, 2022

Abstract

Audio Detection plays a very important role in detection and responding to certain events automatically. It can be useful in having triggers that mark events, which later can be used to analyse the events and take appropriate measure. For example, in case of a construction site near residential area can be analysed for higher than appropriate noise levels to can be detected and be used to help better the community space. Classification of bird species based on their sounds is also a very interesting example of Sound Classification, which can be helpful in identifying birds and their presence in a certain location. The report is based on the similar lines of Single Event Detection, where an event is being detected based on the sound produced at the event. The input files are spectrograms which are classified into ten separate classes like Bark,Crying and Sobbing, Doorbell, Knock, Meow, Microwave oven, Shatter, Siren, Honking and Footsteps. The spectrograms are pre-processed and then later fed to Logistic Regression Model and also to an ANN model. It is observed that ANN provided better accuracy than compared to the Logistic Regression model.

1 Introduction

Spectrograms are features extracted from a sound signal. It is a feature of the Time-Frequency domain. It is a visual representation of the spectrum of the signal with respect to time. The spectrograms are created by implementing Short Time Fourier Transform(STFT) on the time continuous sound signal and later on converting the frequency and time to log-scale. These spectrograms are images where amplitude or loudness is represented by color. The X-axis is time and Y-axis the frequency. Spectrograms can play a role in detection and classification of sound signals. These can be used as markers or triggers to identify events.

In this report the use of pre-processing the specrogram, the accuracy of prediction when used in a Logistic Regression Model and then an ANN model is explored.

2 Literature Survey

Audio Classification is an important task in various real world problems. Successful classification of audio, or identification of sound can be used in solving various problems like sentiment analysis, commands to assistive devices, classification of bird species based on their chirps, identification of rare bird species.

https://towardsdatascience.com/audio-deep-learning-made-simple-sound-classification-step-by-step-cebc936bbe5 : - This article explores the use of CNN for audio classification.

As spectrograms can be treated as images and the problem could be converted to an image classification problem than a sound classification proble. However, the sample size of the training data is 1000. This makes use of CNN less desirable as the number of samples is very less than the trainable parameters.

3 Methods used to Pre-Process the Spectrogram

3.1 Max frequency value over the time range of the spectrogram

The spectrograms received are a constant size in the frequency(Y-axis) and the length of the image varied for different spectrograms. One method explored is to take the maximum value across each

frequency. This created a 128x1 vector. These values are later binned together in a bin of 8 each and then the mean of each bin was stored in a 16x1 vector. This vector is the input data that is fed to the logistic regression and ANN models.The intuition behind this approach is that each of the sounds would have different amplitudes or loudness.

3.2 Other Methods: Summing of all the values across each frequency

Other Method that was initially used for pre-processing the spectrogram was to sum all the values across each frequency in the spectrogram and create a 128x1 array whose values are later binned together in a bin of 8 each and then the mean of each bin was stored in a 16x1 vector. This vector is the input data that is fed to the logistic regression and ANN models.

4 Methods used for Classification

4.1 ANN Approach with pre-processing mentioned in section 2.2

The ANN model is fed with a 128x1 vector with the vector values being the sum of all the values for a particular frequency value. This was later binned and reduced to a 16x1 vector.

This method was providing an accuracy of 43 percentage.

So, the model was changed to a Logistic regression model where the input was 16X1 vector and the output is the argmaxclass$_o$*nehot*$_e$*ncoded*$_v$*ector*

4.2 Logistic Regression Approach with max value across each frequency in the spectrogram

The maximum value across each frequency is taken and all the given 1000 spectrograms are pre-processed using the same logic. This created an Xtrain with size [1000,16] where each spectrogram had 16 distinct features used for classification. The output was the index given to each class. i.e; Bark =0, Meow = 4, etc

This method was giving a score of 48 percentage.

For the given 1000 spectrograms, the Training and Test data was split as 0.9 and 0.1 respectively.

The following are the metrics for the method for the 100 samples that were used as Testing Data.

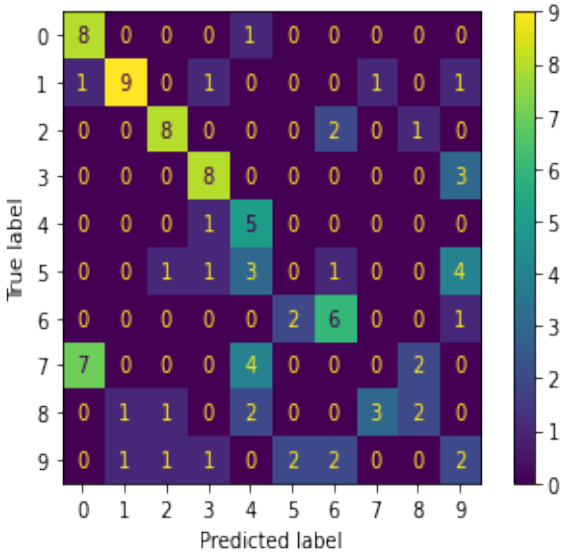

Figure 1: Confusion Matrix.

Table 1: **Metrics - Logistic Regression :With given 1000 training data split into 100 samples for testing**

Metric	Value
Accuracy	0.505
Precision	0.533
Recall	0.505
F1 score	0.493

4.3 ANN Approach with max value across each frequency in the spectrogram

The same input as given in the above Logistic regression model is given to ANN model. The input is 1000x16 data and output is 1000x1 where each input is mapped to one of the 10 classes.

For the 1000 spectrograms provided as the training data, following is the description.

1. 1000 spectrograms are preprocessed and used for both Training and Validation.
2. Data is split as 900 training and 100 test samples.
3. The model is created and compiled. The best model depending on the validation loss is saved to the working directory. The model is fit and the best model is saved
4. The best model is used to predict the testing data
5. The approach uses Adam Optimizer with the learning rate = 0.01. Batch size and epochs are iterated intuitively to get the best model

Following are the metrics for the 1000 spectrograms split into validation and testing:

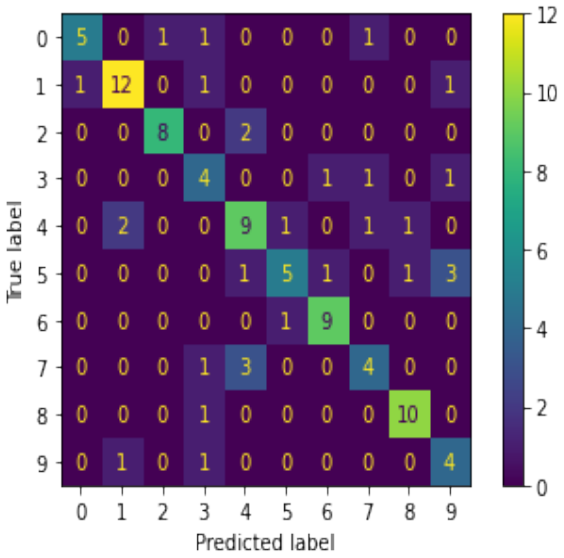

Figure 2: Confusion Matrix.

The model is first fit and saved. The learning rate is changed accordingly for each compilation to fit the model better to the training data. The best model is saved based on the validation loss metric. The model is saved when the validation loss at epoch x is less than the validation loss at epoch y (y¡x). The model is replaced every time the validation loss is improved.

Table 2: **Metrics - ANN Approach :With given 1000 training data split into 100 samples for testing**

Metric	Value
Accuracy	0.7
Precision	0.715
Recall	0.7
F1 score	0.70

5 Testing of ANN model

5.1 ANN Approach testing

With the evaluation metrics, ANN is considered as the better approach to solve the classification problem.

Hence the best model is utilised in testing the data.

However, It can be seen that for the new test samples the ANN approach gave a considerable less accuracy when compared to the validation set used while training the model.

Pre-processing and feature extraction from the spectrogram plays a huge role in accurate prediction of the test data.

Following are the metrics for the ANN:

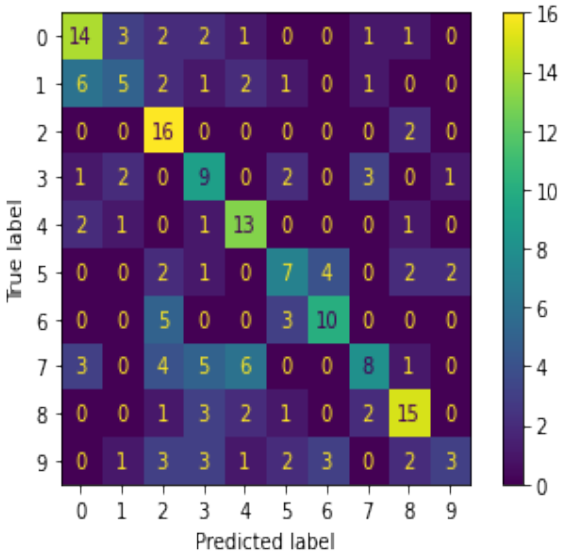

Figure 3: Confusion Matrix.

Table 3: **Metrics - ANN Approach :With given 100 samples for testing**

Metric	Value
Accuracy	0.497
Precision	0.504
Recall	0.497
F1 score	0.478