**PROJECT REPORT**

**Problem Description and Motivation:** The project is divided into two tasks that involve audio event detection and classification.

* **Task1 -** The aim of the first task is to classify each audio file corresponding to an event class, e.g., children-playing, dog-barking, drilling, etc. It is a multiclass classification problem. The evaluation metric is accuracy.
* **Task2 –** The test audio file will contain a sequence of events occurring one after the other. The aim of this task is to find out that sequence of events. A sequence can contain at least 1 and at most 5 events. Thus, it is a multilabel classification problem along with relative ordering. The evaluation metric is edit distance.

**Feature Extraction:** To train the models, we extract some features from the audio files. We find the absolute short-time Fourier transform (DFT over short overlapping windows) of the audio files using Nfft = 1024. The number of time frames is not the same for all the files. To make it uniform, we used zero padding. For task 1, it was made as 401 and for task 2, it was made as 2005(5\*401).

**TASK 1:**

**Data Set:** The data set used in the project comprises of 1761 audio files in the .wav format along with the corresponding labels. There are 10 classes into which the labels are divided.

The number of data samples for every class are varied and hence due to data imbalance, shuffling and randomly selecting 20% of the data points did not provide a good training and validation set. Also, the dataset has a high level of correlation as almost all the audio corresponding to one class have been taken from the same source which might lead to overfitting. Also, since the dataset is quite small, the model may lack robustness. Thus, to overcome this, based on the source, we separated the given data into 10 lists and added approx. 20 percent of data points from each list to the validation data. The remaining 80 percent of each list was kept as the training data and shuffled.

**Models Implemented and Observations:** To improve the prediction accuracy, we have implemented the following 3 models:

* **CNN Model with 4 Convolutional 2D Layers**
* **CNN Model with 6 Convolutional 2D Layers**
* **DNN Model with 4 Dense Layers**

Hyperparameters/ parameters for all the models are:

|  |  |
| --- | --- |
| Epochs | 150 |
| Loss | Categorical Cross entropy |
| Optimizer | Adam |
| Metrics | Accuracy |

**Note:** The model overfits very easily and after approx. 100 epochs, all the models reach a training accuracy of 98% percent and do not get further updated.

**Result:** We aggregated the findings of each model and passed them into a Voting Classifier and predicted the output class based on most votes (hard voting). This increased the prediction by an:

Accuracy on validation data set – 70.5%

Accuracy on the sample\_test\_task1 – 100%

**TASK 2:**

**Data Set:** The data provided for this task was not provided separately. To create the dataset, we chose an arbitrary integer between 1 to 5 and concatenated that many randomly chosen files end to end. A total of 600 data samples with dimensions (513, 2005, 1) were created to train the model, out of which 120 were randomly chosen to be the validation data. The one-hot encoded train and validation labels are (5,10) in dimension where the extra files have a zero vector.

**Model Implemented:**

We have implemented Convolutional recurrent Neural Network (CRNN) for single channel sound event detection. We have used GRU for point change detection, i.e., where one audio ends and the next one starts. Also, a Time Distributed layer was used which allows us to evaluate on every temporal slice of the input.

The flow chart explains the model used:

Hyperparameters/ parameters for all the models are:

|  |  |
| --- | --- |
| Epochs | 150 |
| Loss | Categorical Cross entropy |
| Optimizer | Adam |
| Metrics | Edit Distance |
| CNN Pool Size | [25,4,4] |
| Number of Filters | 16 |

Accuracy on validation data set – 35%, Accuracy on the sample\_test\_task2 – 33%

Submitted by: Shivam Goel(180714), Vartika Gupta(180849)