Sound Event Detection Using a Neural Network

Aditya Agarwal, Syed Munawwar Quadri, Savitha Murthy, Dinkar Sitaram
Cloud Computing and Big Data, PES University, Bangalore

September 21-24, 2016, Jaipur, India

Poster Size : 100 cm long and 90 cm wide.

Abstract

This paper proposes a sound event detection system that is trained using a minimally annotated data set of single sounds to identify and separate the components of polyphonic sounds. Single sounds are represented as MFCC vectors and the system is trained using a Feed Forward Neural Network. Sounds are preprocessed using Principal Component Analysis and Non-Negative Matrix Factorization.

Introduction

An auditory scene could have either a general context that consists of multiple / polyphonic sounds, or a characteristic context containing singular / monophonic sounds. In an auditory scene, when a recognizable event is present; this event is known as a sound event.

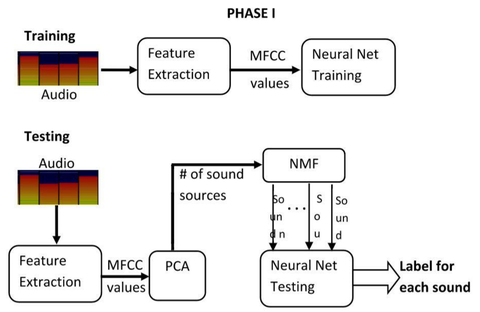
In real life auditory scenes, sound events tend to occur as overlapping events. Despite this, humans are easily able to recognize the sound events present.

The primary aim of this paper is to explore accuracy in sound event detection for overlapping noises with minimal use of training data, contextual information and annotations. The paper proposes a system that does not use context to identify and separate sound events in an auditory scene. This is an unsupervised method for detecting sound events.

Implementation

I) Dataset: Utilizes an initial seed corpus of 7 single sounds present in the publicly available Urban Dataset (Urb_data).

II) Feature Extraction: MFCCs from 25 milli-seconds audio frames sampled at a rate of 10ms (Hamming) with no frame overlap were extracted.



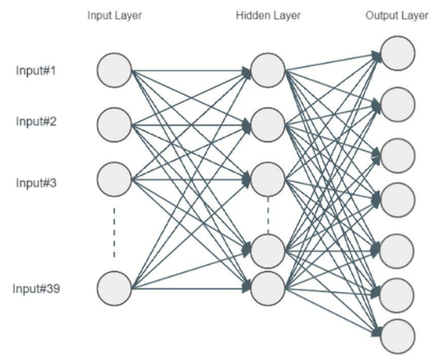
III) Training the Network: A Feed-Forward NN is used for sound event detection using labeled data. It is a 2 layered system.

IV) Source estimation and separation: Principal Component Analysis (PCA) estimates the number of different sound components computationally. The FASST toolbox implements the Non-negative Matrix Factorization Algorithm to separate the overlapping sounds into a number of individual sounds specified by the PCA.

V) Testing: After obtaining the source separated sounds from NMF, they are given to the NN for classification.

Hardware and Software Configuration

A neural network is composed of an input layer, multiple layers of hidden units and an output layer. In the proposed framework, we have implemented a neural network with one input layer, 1 hidden layer and 1 output layer essentially making it a 2 layered network. There are 39 neurons in the input layer (corresponding to 39 MFCCs) and 30 neurons in the hidden layer. The output layer consists of 7 neurons which classifies 7 different known classes, corresponding to 7 single sounds from the Urb_data.



The NN hyper-parameters such as number of hidden units, mini batch size, learning rate and initial weight and bias etc. are selected by a grid search over the parameter values.

The neural network was trained for 100 epochs with a mini batch size of 10, a learning rate of 3 and with an accuracy of average 70% (for correctly classifying sounds from the Urb_data).

Conclusions

The paper proposed methods to detect sound events in overlapping noises, with minimal use of training data, contextual information and annotations. The system does not use context to identify and separate sound events in an auditory scene.

The system gives a reasonable accuracy of classification for source separated sounds obtained from the synthetically overlapped sounds created from the Urban Sound Dataset and naturally occurring overlapping sounds.

The promising results from the Neural Networks are obtained because of its ability to learn new data thus making it an ideal candidate for our framework.

Acknowledgements

We have greatly benefited from the contributions of many a individuals and the following organizations and are indebted to their valuable guidance and assistance without which the research project would not have been a success:

Faculty and Researchers associated with the organizations:

- Cloud Computing and Big Data.
- Department of Computer Science and Engineering, PES University.

We would also like to thank TEQIP II program for enabling us to register and present in this conference.