1 Abstract

Machine learning has been a major player in the role of audio processing and classification for decades. In this time, a great deal of work has been devoted to studying the performance of various model architectures and producing sets of features that can represent a waveform in a compact, efficient, and non-redundant way Citation?. In this work, we show that a spectrogram matrix and a feature-vector with predictors derived from both time-series and frequency-series representations of audio can be used to map that waveform to a potential source. This combination of features warrants a model design that combines a convolutional neural network (CNN) and a multilayer perceptron (MLP) to process the spectogram matrix and the feature-vector respectively. We detail the significance and behavior of branch of the network and explore how this hybridization architecture along with the chosen features produces improved classification performance while retaining computational practicality.

Perhaps we can say that we explore both time domain and frequency domain classifiers. Then we show that a hybrid network using the spectrogram and time series data gives improved results. To do so, it was necessary to link(fuse?) a CNN (on spectrogram) and MLP (on time series) at an internal layer in each architecture.

We might close by stating that the paper will also address the problem of creating good features, and will finish with a brief discussion of classifying instrument-like sounds from a chaotic music synthesizer and other sources.