Landon Buell

June 2022

Audio Classification of Musical Instruments with a Multimodal Neural Network

Abstract:

1. Introduction
   1. Digital signal classification is a broad task (many forms)
      1. Climate reports, finance, music, database organization
      2. Place each time-series signal into *category*
      3. Samples w/ similar properties belong to one category
   2. How to group samples efficiently?
      1. Too complex for traditional computing,
      2. Too large for humans to do
      3. Need to combine computer speed w/ human intelligence
   3. Enter the Neural network
      1. We wish to design a model and series of features for audio classification
      2. Design features to describe properties of the waveform
      3. Provide the features to a complementary model
      4. Learns a set of parameters to map features to musical instruments
   4. Data set
      1. ~16,000 audio samples from University of Iowa
      2. Each sample contains a single note (removed otherwise)
      3. Sampled at 44.1 kHz. Padded or cropped to be within sample range
      4. Predetermined set of classes, labeled according to the name of the file
      5. Collected features are preprocessed, scaled and passed into model
2. Feature Selections
   1. Overview of features
      1. Machine learning models require *features*
      2. Features are compact, low-dimensional representations of a larger data set
      3. Provided as input to a machine learning model
      4. Model evaluates
3. The Multimodal Architecture
   1. We are presented with multiple forms of features
      1. Some derived from time-space representation of waveform
      2. Some derived from frequency-space representation of waveform
      3. We can assemble these into a (1 x P) *feature vector*
      4. In the process, we have also built a spectrogram
      5. We can assemble this into a (M x N) *feature matrix*
   2. Now have two differing structures
      1. One is a 1 x P vector of
4. Performance Evaluation
   1. Standard metrics for classification
      1. Precision, Recall, F1, accuracy
      2. Descriptions of each
      3. Bonus: loss function
   2. Performance of each model
      1. Multilayer perception – Metric scores, rate of convergence
      2. Convolution Neural Network – Metrics scores, rate of convergence
      3. Hybrid Model – Metric scores, rate of convergence
      4. Comparison of all three models
5. Conclusions
6. References
7. Appendix