The 12th International Conference on
Computer Science & Education (ICCSE 2017)
August 22-25, 2017. University of Houston, USA

Musical Instrument Classification Utilizing a Neural Network

Therrick-Ari Anderson
Dept. of Computer Science, Technology & Mathematics
Lincoln University
Jefferson City, Missouri, USA
therrick-ari.anderson448@my.lincolnu.edu

Abstract— **This paper discusses a method for classifying musical instrument audio signals utilizing a neural network. This research will identify the most salient features to evaluate within a neural network that will quickly detect an instrument from another. Feature extraction and selection are crucial steps in helping distinguish musical signals. Feature extraction is the process of obtaining specific characteristics from a data sample. Feature selection is the process that follows extraction in which the most relevant features are chosen to represent each sample. Once relevant features are selected they are applied to the neural network as possible inputs. In this study, the neural network distinguishes between two classes of instruments (e.g., trumpet or tuba). Various features are evaluated to identify which elements worked best.**

Index Terms— **Neural networks, classification, musical instruments, machine learning**

I. Introduction

The essential idea behind musical instrument classification is creating a computer system (neural network) that distinguishes between classes of instruments based on descriptive features. As human beings, we have the ability to classify things based on certain characteristics. In research, we have the ability to classify items using other means such a neural network. A neural network is a computational model comprised of a highly connected system of neurons that relate specific inputs to desired outputs. A neural network works by training data iteratively, adjusting the strengths of the connections to obtain the correct responses [6]. The classification of musical instruments also involves analysis of audio signals to obtain descriptive information. Obtaining and selecting these descriptive features are crucial steps in the overall development of the neural network. Musical instrument classification is a critical component in designing automatic indexing and database retrieval systems. Additionally, musical instrument classification is used in a variety of applications including music transcription, genre classification, song identification, music indexing, etc. This study strictly focuses on classifying musical notes. In this research study, a music information retrieval system (Mirtoolbox 1.6.1) is used to analyze and extract features. In this paper, we discuss the process of classifying musical instrument using a neural network.

II. Machine Learning

This paper addresses machine learning as a branch of artificial intelligence. In our context, machine learning is the process of teaching computers to learn in a manner similar to how we as human learn from experience [7]. The purpose of this research study is to allow the neural network to learn from the data and establish connections between inputs and outputs. Machine learning algorithms utilize computational strategies to gain knowledge from data. Additionally, the algorithms adaptively enhance their execution as the amount of learning samples increases [7]. Machine learning is broken into two main learning techniques: supervised and unsupervised. Supervised learning uses known input and output data and trains the model to generate feasible predictions when new data is applied. Unsupervised learning finds hidden patterns or internal structures within the data. It is often used to develop inferences from the datasets including inputs and unknown outputs [7]. Supervised learning techniques are applied in this study.

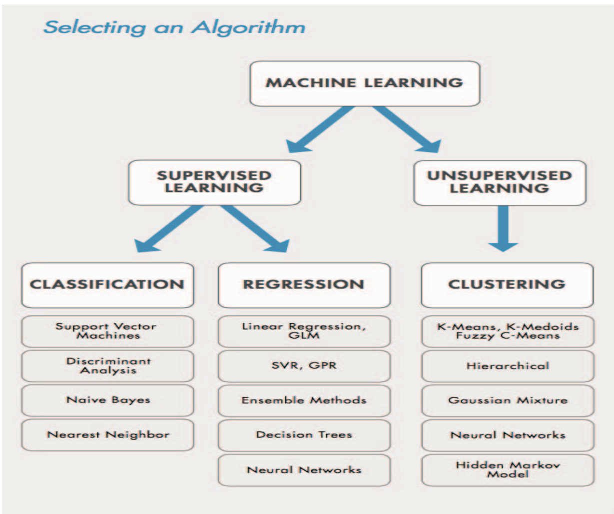

Fig 1. Different types of learning algorithms and processes.

III. Classification

Supervised learning methods consist of classification and regression [7]. This research study addresses a binary classification problem. In a binary classification problem, a single training set is limited to two classes (for example, determining whether the audio signal is a trumpet or tuba). Classification is the process of assigning individual objects into defined groups based on certain characteristics. Understanding and identifying these characteristics are essential to defining each class.

IV. Background

There have been a number of studies done to address musical instrument classification. Bormane and Dusane created a new classification technique for musical instruments utilizing various techniques in wavelet transform [1]. Park and Lee proposed a method utilizing convolutional neural networks [8]. Furthermore, they have explored a new technique in classifying musical instruments using learned features from convolutional neural networks. Livshin and Rodet presented a process for recognition of multiple instrument solo recordings that utilizes a gradual descriptor elimination feature selection algorithm [5]. These sources were used to gain insight in the process of classifying musical instruments. The main method in this study involved using Mirtoolbox 1.6.1 [4] to extract features which lead to the development of the neural network.

V. Basic Methods

A. Gathering Data Samples

In this research study, the musical instrument audio samples were gathered from the University of Iowa Musical Instrument Samples database [2]. The University of Iowa Musical Instrument Samples is a notable data source and has been used in over 270 publications [2]. Furthermore, various note samples were collected for B-flat trumpet and tuba. There was a total of 12 note samples, six for each instrument, used in the feature extraction process. Additionally, there were 12 samples used as test data within the neural network. In this study, I wanted to limit the data samples to focus on two distinct instruments.

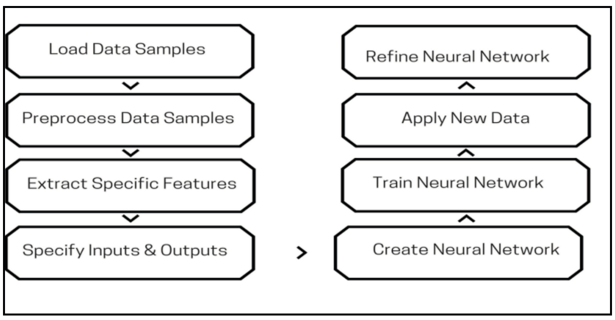

Fig 2. Basic procedure for the study.

B. Preprocessing

After obtaining these samples, some preprocessing was administered. This process involved removing redundant information such as silence before the note attack and decay. The purpose of this process to make sure the analysis of the audio signal would include relevant information only. After preprocessing these audio files, they sounded approximately four seconds each.

C. Feature Extraction & Selection

The most important factors in this research study are feature extraction and feature selection. Feature extraction is the process of gathering characteristics from a data sample. This information is often represented numerically and/or graphically. Feature selection is the process of determining the most important features or variables that give the best predictive potential in modeling your data. In this study, 19 features were chosen (see Table 1) from the various fields in Fig 3. They were composed of statistical information from the audio features including standard deviation and mean values.

Audio Features Extracted
tonal.keyclarity.Mean
tonal.keyclarity.Std
spectral.centroid.Mean
spectral.centroid.Std
spectral.brightness.Mean
spectral.brightness.Std
spectral.skewness.Mean
spectral.skewness.Std
spectral.flatness.Mean
spectral.flatness.Std
timbre.spectralflux.Mean
timbre.spectralflux.Std
timbre.zerocross.Mean
timbre.zerocross.Std
dynamics.rootmeansquare.Mean
dynamics.rms.Std
rhythm.attack.time.Mean
rhythm.attack.time.Std
fluctuation.peak.PeakPosMean

Table 1. The 19 audio features extracted, including the operation performed on each field in Fig. 3 (e.g. standard deviation and mean).

D. Audio Features

According to Bormane and Dusane, a musical instrument sound includes four perceptual characteristics: pitch, loudness, duration and timbre [1]. Perceptual characteristics are features related to perception and sensory experience. Audio features are features that are generated from raw signal data. In this study, there were approximately six audio features used. They include dynamics, fluctuation, rhythm, spectral, timbre, and tonal. Dynamics are features based on the loudness or quietness of an audio signal. Fluctuation is the variation in the audio signal including the overall rising and falling (e.g. attack, decay, sustain, and release). Rhythm is the pattern of movement within the audio signal. Spectral are features based on frequency, they are gathered by converting the time-based signal into the frequency domain. Timbre is the quality of a sound by which a listener can determine that two sounds of equal loudness and pitch are dissimilar. Tonal is the colorfulness or tone of the audio signal. These audio features were selected based on the musical information retrieval system capabilities.

VI. Musical Information Retrieval Toolbox

Content-based musical information retrieval systems are used to obtain musical data from audio signals. Mirtoolbox 1.6.1 was used to gather numerical and graphical information from the audio samples, and it also served as the primary tool for feature extraction [4]. Mirtoolbox is a tool that runs within MATLAB. This toolbox extracts multiple statistical features from audio signals. Some features include dynamics, fluctuation, timbre, tonal, spectral, and rhythm. The feature extraction process within the toolbox involved the using a function called "mirfeatures". This function along with creating a for loop allowed each audio sample to be processed and produce statistical data for each sample.

```
1×12 struct array with fields:

  dynamics
  fluctuation
  rhythm
  spectral
  timbre
  tonal
  FileNames
```

Fig 3. Return value of running the "mirfeature" function within MATLAB, producing a 1 by 12 structure with 7 fields. There are 12 total samples in this research study.

VII. Neural Network Architecture

The neural network utilized conjugated back propagation. There was a total of 19 input values of statistical data from each of the 6 audio features fields (dynamics, fluctuation, timbre, tonal, spectral, and rhythm). There were a total 2 outputs (i.e., classification as trumpet, tuba). Lastly, there were a total of 20 hidden neurons used.

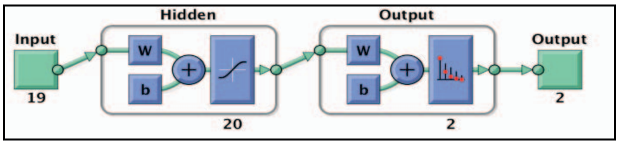

Fig. 4 the neural network basic architecture.

VIII. Results

After successfully implementing the neural network, it provided decent return values. There was a total of 12 samples used for testing. 83% percent of these samples were fairly close to their predicted desired outputs. The neural network correctly classified 5 of 6 trumpet samples. With regard to the tuba samples, it successfully classified 4 of 6 tuba samples. Given the number of samples used for testing, the neural network predicted 75% correctly. Some individual results were very accurate with respect to the desired outputs, and others were not. The neural network is currently in the development process, and it will be refined in the future to produce better results.

IX. EXPERIMENTAL COMPARISON

Bormane and Dusane's classification technique utilizing wavelet transform produced results for 14 of 16 instruments [1]. These instruments consisted of strings, brass, keyboard, and woodwinds. In their experiment, they had 10 testing notes. More so, for the classification of the cornet and tuba, they received 100% accuracy for tuba and 70% for cornet. Park and Lee's method utilizing convolutional neural networks produced high performances in their confusion matrix [8]. The experiment used 20 instruments including woodwinds, strings, brass, and piano. In the confusion matrix, trumpet and tuba produced high percentages ranging between 90% and 100%. Lastly, Livshin and Rodet's technique utilized a gradual descriptor elimination feature selection algorithm [5]. Their experiment used woodwinds, piano, and strings. The average recognition rate of 1-second solo recordings resulted in 88.13%.

X. Conclusion

Overall, the neural network performs fairly well, given the test samples used (see Table 2). The neural network was able to generate correct outputs for the majority of the trumpet samples resulting in 83% of samples producing the desired output. The tuba test samples, on the other hand, did okay when applied to the neural network they resulted in 66% of the test samples producing values close to zero. Furthermore, some outputs were completely off for both instrument classes; this may be due to the features used as well as unknown factors within the neural network. All in all, this research study is in the trial and error phase, and more development and enhancements will be applied.

Test Samples Results				
Samples	Actual Output (Trumpet)	Actual Output (Tuba)	Desired Output (Trumpet)	Desired Output (Tuba)
Trumpet Sample 1	1	0	1	0
Trumpet Sample 2	1	0	1	0
Trumpet Sample 3	1	0	1	0
Trumpet Sample 4	1	0	1	0
Trumpet Sample 5	0.9996	0.0004	1	0
Trumpet Sample 6	0.4767	0.5233	1	0
Tuba Sample 1	0.9972	0.0028	0	1
Tuba Sample 2	0.4900	0.5100	0	1
Tuba Sample 3	0.3537	0.6463	0	1
Tuba Sample 4	0.7035	0.2965	0	1
Tuba Sample 5	0.0732	0.9268	0	1
Tuba Sample 6	0.0001	0.9999	0	1

Table 2. Results of the test samples applied to the neural network. The output values in green represent outputs that were exact or close to their desired outputs. The results in red indicate incorrect outputs compared to the desired outputs.

ACKNOWLEDGMENT

Funding was provided by the National Science Foundation, (HBCU-UP) Award # HRD-1410586 to David Heise, Lincoln University, Jefferson City, MO 65101. Faculty Advisor/Mentor: Dr. David Heise, heised@lincolnu.edu.

REFERENCES

[1] Bormane, D. S. and Dusane, M. "A Novel Techniques for Classification of Musical Instruments," vol. 3, pp. 1–7 (2013).

[2] Fritts, L. (n.d.). Musical Instrument Samples. Retrieved February 23, 2017

[3] Kostek, B. & Czyzewski, A. "Representing Musical Instrument Sounds for their Automatic Classification," pp. 1–7

[4] Lartillot, O., Toiviainen, P., & Eerola, T. (n.d.). MIRtoolbox. Retrieved February 23, 741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[5] Livshin, A. & Rodet, X. "Instrument Recognition Beyond Separate Notes - Indexing Continues Recordings," pp. 1–4

[6] MathWorks. (2016). Applying Supervised Learning. Retrieved February 23, 2017

[7] MathWorks. (2016). Introducing Machine Learning. Retrieved February 23, 2017

[8] Park, T. and Lee, T. "Musical instrument sound classification with deep Convolutional Neural Network using Feature Fusion approach". pp. 1–11