# Instrument Classification

Yuqing Su

2022-07-30

**Abstract**
This work aims to compare different models on particular instruments (monophonic sound) recognition which is an important problem in the field of music information retrieval. Code (Jupyter Notebook) is included for easy reproducibility.

## 1 Introduction

Music is composed of different sounds which are determined by timbre. Humans can easily identify some instruments with characteristic timbre or instruments that they are familiar with. However, telling apart instruments with similar timbres and recognizing instruments that are not known to the audience can be big problems. Thus, artificial intelligence comes to help and the result is beneficial for musical analysis and music composition.
In 1999, 8 instruments are identified by GMM (Gaussian Mixture Models) and SVM (Support Vector Machine) with error rates 37% of GMM and 30% of SVM[2]. Further steps on instrument recognition focus on neural networks, especially CNNs (Convolutional neural networks). In 2019, Solanki et el achieved 92.80% accuracy on polyphonic musical instrument recognition using CNNs with eight layers[3]. In this research, SVM, GMM, ANNs, CNNs, and RNNs will be applied to the same instrument dataset and the accuracy will be compared.

# 2 Dataset and data-preprocessing

## 2.1 dataset

The dataset contains sound samples of 20 instruments[4]. 7 instrument sound samples are selected for convenience. The seven instruments are: guitar, flute, violin, cello, clarinet, trumpet and saxophone.

| guitar | flute | violin | clarinet | trumpet | cello | saxophone |
|---|---|---|---|---|---|---|
| 106 | 884 | 1502 | 846 | 485 | 889 | 732 |

Figure 1: Dataset overview

.

The rate of train dataset size and test dataset size is 7:3.

## 2.2 data-preprocessing

There are many ways of extracting sound features, such as Spectral Centroid, Time Domain Zero Crossings, Mel-Frequency Cepstral Coefficients(MFCC). Since "As a widely used feature in genre classification systems, MFCC is typically believed to encode timbral information, since it represents short-duration musical textures"[1] and MFCC behaves better than the other models in Liu, Jing, and Lingyun Xie's research[5], MFCC is chosen as the sound feature extracting method in this research.
MFCC is based on the short time Fourier transform. And "it combines characteristics of hearing perception and the mechanism of generating speech to get cepstral coefficients, which is widely used in various occasion of audio signal processing"[5].


Figure 2: MFCC process

.

13 features are extracted from MFCCs after processing the original .wav file. (Each row stands for a feature in Figure 3)

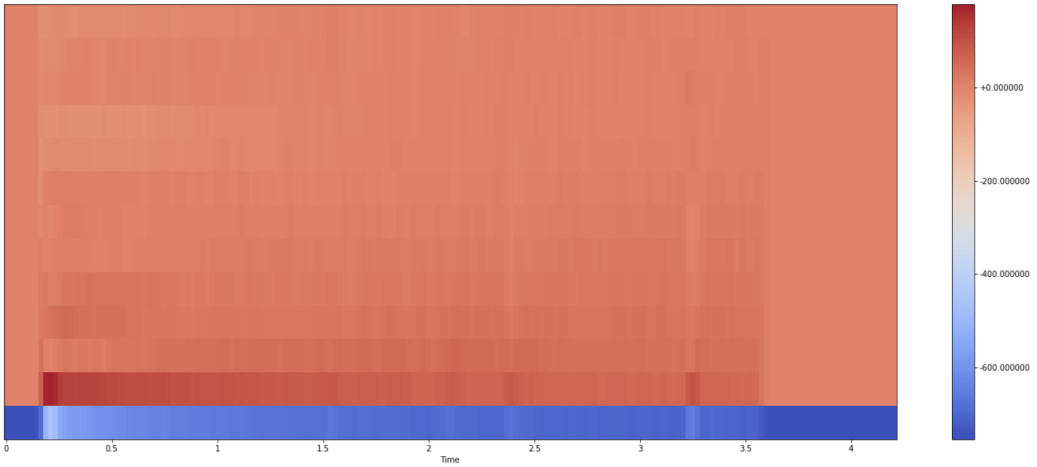


Figure 3: MFCC visualization

.

# 3 Classification Methods

This section will discuss SVM and logistic regression.

### 3.1 SVM(Support Vector Machines)

SVM basically means to find the optimal linear hyperplane among different dataset classes and meanwhile achieve minimum error and maximum margin. The decision_function_shape is set to 'ovo' though 'ovo' and 'ovr' generate the same result.

### 3.2 GMM(Gaussian Mixture Modeling)

GMM is is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. In this research, each instrument has its own GMM model. Although two GMM models for each instrument might improve the accuracy, but it fails to improve the accuracy significantly[2].

### 3.3 ANN(Artificial Neural Networks)

ANN uses layers of artificial neurons to transmit signals. Three hidden layers with 'relu' as the activation function are applied and the output layer applies 'softmax' as the activation function.

### 3.4 CNN(Convolutional Neural Networks)

CNN mainly uses small size matrices to extract features from a large dataset. 4 convolutional layers are applied with 'relu' as the activation function and the output layer applies 'softmax' as the activation function.

### 3.5 RNN(Recurrent Neural Networks)

RNN saves the output of processing nodes and feed the result back into the model (they did not pass the information in one direction only). 4 hidden layers are used with 'relu' as the activation function and the output layer applies 'softmax' as the activation function.

## 4 Results

SVM does a relatively good job and the time used is less than other models. Inaccuracy occurs when trying to distinguish cello and violin (with similar

timbre/music tone). GMM, which is often used in speech recognition, doesn't perform well especially for saxophone which only has an accuracy of 24.55%. Neural networks generally outperform GMM and SVM but requires longer running time. Overall, CNN might be a better choice on instrument classification. There are several ways to improve the accuracy. Two ways are practiced: 1. Add features from MFCCs results. The accuracy reaches up to 98.66% after extracting 40 features rather than 13. 2. Change layers inside the CNN model. The accuracy reaches to 94% 99% after changing 1D CNN to 2D CNN. However, the second improvement method is not recommended since the accuracy is unstable.

| | SVM | GMM | ANN | CNN | RNN |
|---|---|---|---|---|---|
| Accuracy | 73.35% | 62.41% | 88.33% | 96.82% | 94.83% |

Figure 4: Accuracy overview

.

# 5 Conclusion

CNN is the best model in this research and it has the capability of improvement since the layers are set artificially.

**References**

[1]. Li, Tom LH, and Antoni B. Chan. "Genre classification and the invariance of MFCC features to key and tempo." International Conference on MultiMedia Modeling. Springer, Berlin, Heidelberg, 2011.

[2]. Marques, Janet, and Pedro J. Moreno. "A study of musical instrument classification using gaussian mixture models and support vector machines." Cambridge Research Laboratory Technical Report Series CRL 4 (1999): 143.

[3]. Solanki, Arun, and Sachin Pandey. "Music instrument recognition using deep convolutional neural networks." International Journal of Information Technology (2019): 1-10.

[4]. “Sound samples,” Philharmonia. [Online]. Available: https://philharmonia.co.uk/resources/so
samples/.

[5]. Liu, Jing, and Lingyun Xie. ”SVM-based automatic classification of musical instruments.” 2010 International Conference on Intelligent Computation Technology and Automation. Vol. 3. IEEE, 2010.