CS 4641-B Machine Learning Final Presentation

Instrument Identification for Real-world Music Pieces

Group 6
Zhubo Zhou, Tianyi Zuo, Zhaoran Ma,
Junzhe Ruan, Yatong Bai

Table of content

Background

Identifying the instrument of a music piece has various applications:

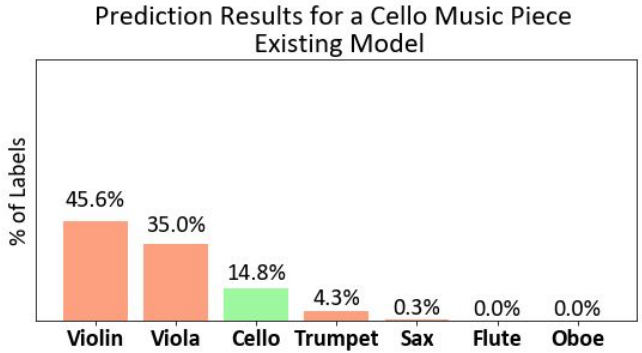
- Educational applications
- Entertainment applications

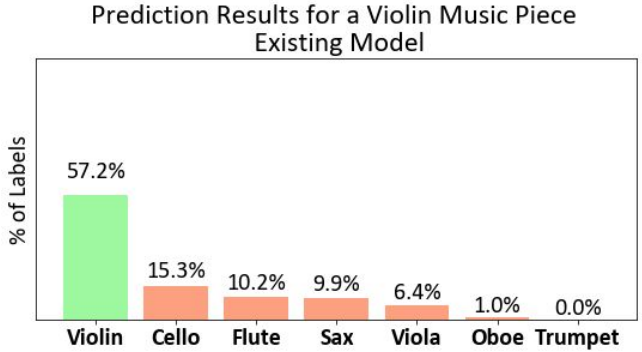
Goal

Apply supervised learning technique to train a model to identify instruments of music pieces.

Existing Model [1]

- Strength:
  - Works well for the Philharmonia dataset (75% training, 25% testing).
  - Fast to train
- Deficiency:
  - Performs badly when predicting a real-word music piece. (Example: Identified a cello solo piece as a viola solo)





Problem Statement

Improve the model so that it can correctly predict real-world music pieces.

[6] https://github.com/GuitarsAI/BasicsMusicalInstrumClassifi

Feature Engineering: Why MFCC?

Common feature extraction methods: MFCC, LPC, SO&CP

A previous paper [2] shows that instrument classification performs the best with MFCC.

Method	Accuracy
LPC	67%
MFCC	85%
SO&CP	53%

← **Best performance**

[2] https://arxiv.org/pdf/1909.08444.pdf

ML Model: Why SVM?

- Common classification methods:

Model	Accuracy
Logistic Regression	0.54
Decision Tree	0.52
LGBM	0.66
XG Boost	0.67
Random Forest	0.68
SVM	0.76

Best performance [3]

- CNN is also available, but is slow to train.
- SVM performs the best while being fast to train.

[3] https://arxiv.org/pdf/1912.02606.pdf

High-level Workflow

MFCC:

- The output feature vectors represent the tone
- How: take the DFT (discrete fourier transform) of the soundwave, and then take the DCT (similar to DFT) on the DFT result.
- A nice video: https://www.youtube.com/watch?v=OWc2slRAWcU&t=977s

Original Training Data: samples from Philharmonia [4]

- **20 types of instruments**
  - In this project, we picked 7 of them
    - Cello, violin, viola, trumpet, oboe, sax, flute
- **~ 1000 files for each instrument**

Instrument	Pitch	Duration (sec)	Loudness	File
Cello	E4	1	Fortissimo (Loud)	
Violin	As5	0.5	Fortissimo (Loud)	
Sax	F3	1	Pianissimo (Quiet)	

[4] https://philharmonia.co.uk/resources/sound-samples/

Improvement 1: More Training Data

- **Use the whole Philharmonia dataset**
  - The existing model used a small portion. We used the whole dataset to increase pitch range.
- **Data augmentation**
  - Overly pairs of music files with the same instrument.
- **Use real-world data for training (lower variance)**
  - Reason: all cello clips in the Phil dataset were recorded using one cello. Other cellos may have different tones.
  - We cropped real-world pieces into small clips (~0.3s), and added to the training data.

Improvement 2: Feature Engineering

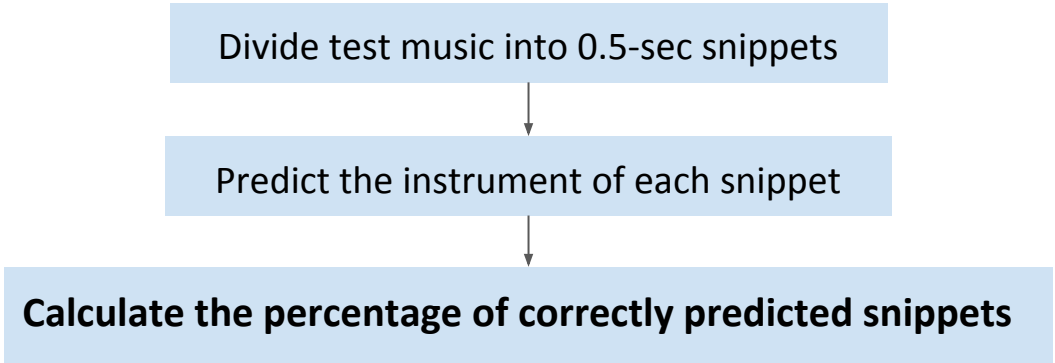
MFCC parameters were modified

	Old Parameters	New Parameters
FFT window length	2048	2048
# of mel bands	128	256 (⬆️)
Dimension of output	13	43 (⬆️)
	Transform files into 13D vectors	Transform files into 43D vectors

- New feature vectors preserve more information (lower bias)

Performance Evaluation

- Based on the percentage of correctly predicted snippets

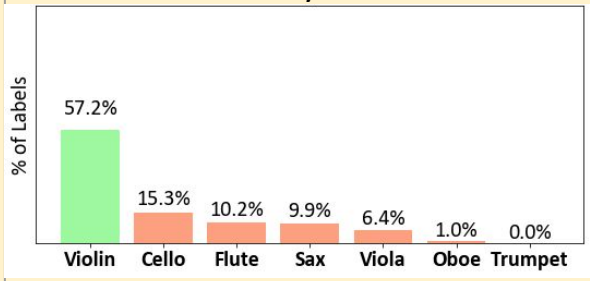
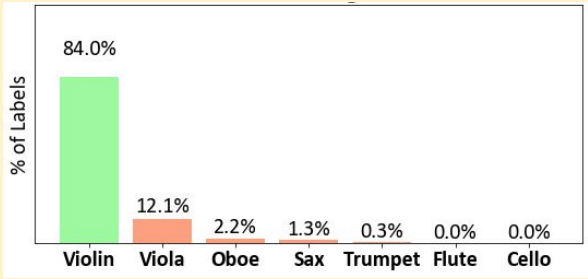
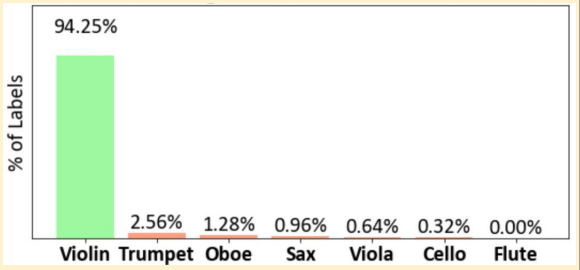
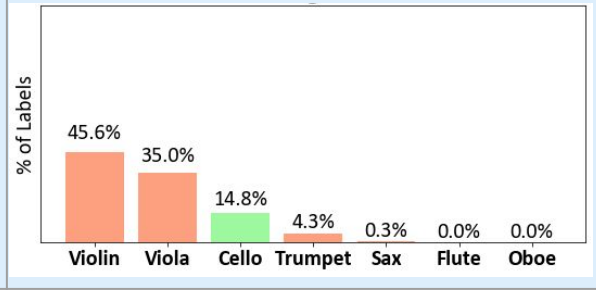
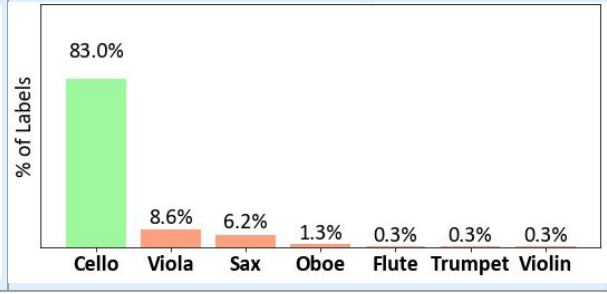
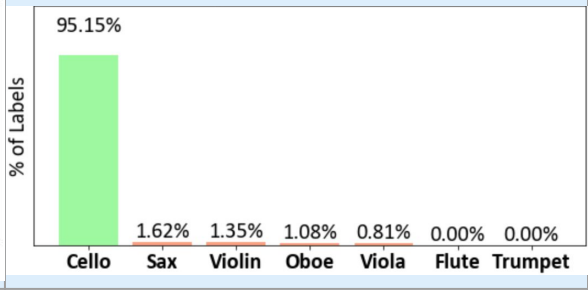


Why divide into snippets?

- Real-time prediction (can be used in embedded devices)
- Pieces in which instruments alternate to play

Results

Note: the music pieces for training are different from the pieces for evaluation.

	Existing model	More training data	More training data + new MFCC parameters
Violin piece	Prediction accuracy: **57.2%** 	Prediction accuracy: **84.0%** 	Prediction accuracy: **94.25%** 
Cello piece	Prediction accuracy: **14.8%** 	Prediction accuracy: **83.0%** 	Prediction accuracy: **95.15%** 

Discussion

- **Increase diversity of the training data:**
  - Layer music pieces together:

    Mimic multiple instruments being played at the same time
  - Introduce more real-world music pieces:

    Include timbres for different individuals of the same instrument
- **Tune MFCC parameters:**

Number of Mel Bands	Number of MFCC Features	Prediction Accuracy
⬆️	⬆️	⬆️
⬆️	--	⬇️
--	⬆️	--

Conclusion

- Our algorithm is able to correctly identify instruments (accuracy > 90%)

- By applying ML techniques to identify the instrument being played in a music piece, our work could potentially benefit the field by:
  - providing an alternative approach for musicians to acquire music scripts.
  - allowing producers to have higher freedom in post-processing of music recordings.

References

[1] Basic Music Instrument Classification (the model that we based on)
https://github.com/GuitarsAI/BasicsMusicalInstrumClassifi

[2] The data is from RWC music database:
https://github.com/wiku30/Musical_Instrument_Classification
https://github.com/wiku30/Musical_Instrument_Classification/blob/master/ppt/SST.pptx
https://arxiv.org/pdf/1909.08444.pdf

[3] Predominant Musical Instrument Classification based on Spectral Features
https://arxiv.org/pdf/1912.02606.pdf

[4] Philharmonia Dataset
https://philharmonia.co.uk/resources/sound-samples/

[5] Musical Instrument Identification with Supervised Learning (project inspiration)
http://cs229.stanford.edu/proj2019spr/report/6.pdf
http://cs229.stanford.edu/proj2019spr/poster/6.pdf

[6] Musical Instrument Timbres Classification with Spectral Features
https://link.springer.com/content/pdf/10.1155/S1110865703210118.pdf

Questions?

https://github.com/Bai-YT/CS4641-Music-Instrument-Recognition

baiyt98@gmail.com
jruan34@gatech.edu
zuotianyi0905@gmail.com
zzhou344@gatech.edu
zma98@gatech.edu