Musical Instrument Chord Classification

Tianchen Yuan

Apr 02, 2022

Objective

Perform a binary classification to determine whether a chord is major or minor for piano and guitar.

Background Information

Definitions of some key words in music theory.

- **Note**: the most fundamental building block of music. There are 7 natural notes, represented as do–re–mi–fa–sol–la–si or A-B-C-D-E-F-G, which can be played with white keys on a piano. In between these natural notes are 5 half-step notes (e.g. A# between A and B) except B-C and E-F. They can be played with black keys on a piano. In total, there are 12 notes listed as follows and each pair of adjacent notes are half-step apart.

  A A# B C C# D D# E F F# G G#

- **Frequency**: each note has a corresponding frequency (Hz) that can be calculated as

$$f = 440 \cdot 2^{d/12} \tag{1}$$

  where d is the number of half-steps away from the reference note A.

- **Chord**: a harmonic set of notes/frequencies that are heard as if sounding simultaneously. A major chord consists of a root note, a major third (4 half-steps higher than the root), and a perfect fifth (7 half-steps higher than the root). A minor chord consists of a root note, a minor third (3 half-steps higher than the root), and a perfect fifth. For example, if the root is C, then the major chord is C-E-G and the minor chord is C-D#-G.

Dataset

The dataset is available at:
`https://www.kaggle.com/datasets/deepcontractor/musical-instrument-chord-classification`

The original data are stored in two folders named "major" and "minor". "major" folder contains 502 major chords in wav format. "minor" folder contains 357 minor chords in wav format.

Data Representation

- Time-domain representation of a single note:

$$p(t) = A\sin(2\pi f t) \tag{2}$$

  where A is the amplitude and f is the frequency.

- Frequency-domain representation of a single note: a spike at the corresponding frequency f.
- Time-domain representation of a pure chord:

$$c(t) = A_1 \sin(2\pi f_1 t) + A_2 \sin(2\pi f_2 t) + A_3 \sin(2\pi f_3 t) \tag{3}$$

- Frequency-domain representation of a pure chord: three spikes at the corresponding frequencies f_1, f_2, f_3.
- Time-domain representation of an instrumental chord: when a chord with 220Hz root frequency is played by an instrument, it also generates chords with root frequencies of 440Hz, 660Hz, 880Hz, 1100Hz and so on.

$$s(t) = c_1(t) + c_2(t) + c_3(t) + ... + \text{noise} \tag{4}$$

- Frequency-domain representation of an instrumental chord: multiple spikes including both the chords and the noise.

Suggested Steps

1. Listen to the audio files and understand the music theory behind these chords.
2. Explore different data processing methods to extract the frequency components that distinguish the two types of chords.
3. Create your own training/validation/testing data and labels.
4. Experiment with various models (both ML and DL) and compare the performance.
5. DO NOT copy the existing solutions online, at least try to understand and extend their works.

Reference

[1] https://www.kaggle.com/code/ahmetcelik158/mathematics-of-music-chord-classification
[2] https://en.wikipedia.org/wiki/Musical_note