Music Instrument Classifier

By Sudheer, Karthik & Pooja

TOC

Overview

Music instrument classification is a key part in music related products

It is the first step towards:

- Classifying music by genre
- Music search by instrument/genre

Preprocessing

Spectrum Cut	Filter the frequencies above/below certain Hz.
Frequency shift	All transformed samples shifted to be represented in a pitch of a certain Hz, so that instruments and not tones are learnt.
Partitioning Frequency Spectrum	Partitioning the frequency spectrum by taking average to reduce overfitting
Normalization	MinMax scaling so that all samples contribute.

Feature Types

Spectral (frequency)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Curabitur eleifend a diam quis suscipit. Class aptent taciti sociosqu ad litora et nec torquent per conubia nostra.

Temporal (time)

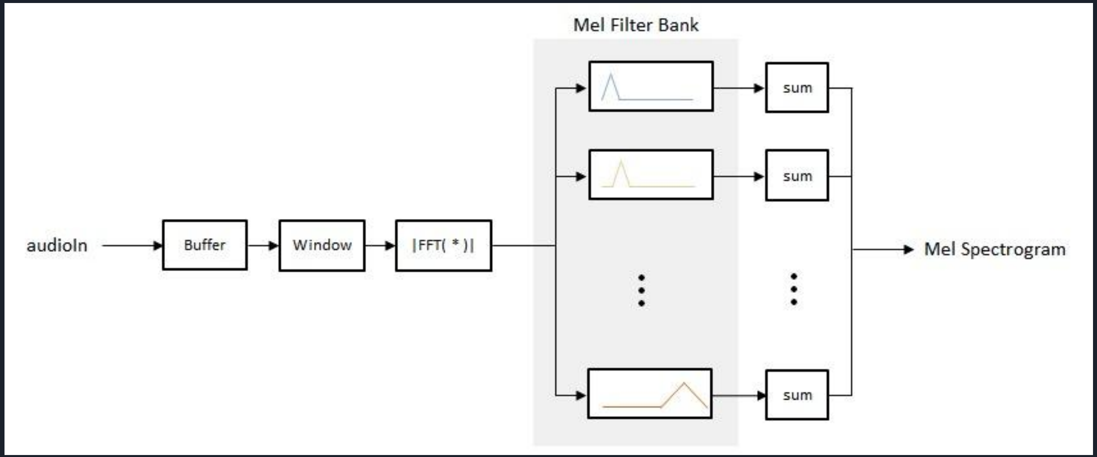
Amet, consectetur adipiscing elit. Curabitur eleifend a diam quis suscipit. Class aptent taciti sociosqu ad litora torquent per conubia nostra.

Feature of interest

- A spectrogram is a visual way of representing the signal strength, or “loudness”, of a signal over time at various frequencies present in a particular waveform.
- Mel scale is a perceptual scale that helps to simulate the way human ear works. It corresponds to better resolution at low frequencies and less at high.

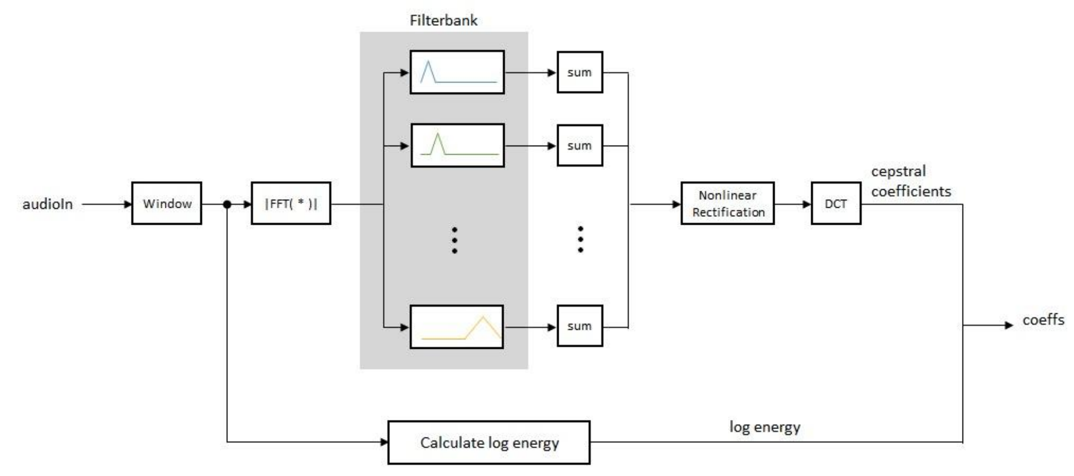
Feature of interest | Mel-Spectrogram

- Spectrogram represented in mel-scale is thus called **Mel-Spectrogram**
- Typically 32-64 bands in Mel spectrogram.



Feature of interest | MFCC

- MFCCs are an alternative representation of the Mel-frequency spectrogram often used in audio applications
- The MFCCs are calculated by applying the discrete cosine transform (DCT) to a mel-frequency spectrogram.
- MFCC is a very compressible representation, often using just 20 or 13 coefficients



Other Possible Features

- rms-energy envelope
- spectral centroid
- harmonic amplitude
- histogram differentials
- cross correlation

Models

Naive-Bayes	Baseline estimator against which improve the results.
SVM	Primary estimator trained.
Others	Gaussian Mixture, KNN, Hidden Markov, CNN, Kohonen self-organizing map

Dataset

Sourced from “Freesound General-Purpose Audio Tagging Challenge” on Kaggle.

Most research in this domain uses "clean" data with isolated sounds, consisting of a single instrument in each audio stream.

In real world data, each streams can have any kind of background noise and is not recorded under the same conditions, which adds complexity to the classification task.

Also real-world data has many instruments playing together.

Dataset

List of instruments

1. Acoustic Guitar,
2. Bass Drum
3. Cello
4. Saxophone
5. Double Bass
6. Flute
7. Violin or Fiddle
8. Snare Drum
9. Hi-hat
10. Clarinet

Hyper-parameter tuning

Grid search is performed separately for MFCC & Mel models to find the best value of PCA components

Summary

- SVM performs significantly better than GNB.
- Accuracy improves with more samples.
- To this end, given the limited dataset we sampled the data with replacement to provide sufficient data for training.

References

http://cs2.stanford.edu/projects_fall_2019/reports/26225883.pdf

https://www.kaggle.com/c/freesound-audio-tagging

https://www.mathworks.com/help/audio/ref/melspectrogram.html

https://in.mathworks.com/help/audio/ref/mfcc.html

[1] Eronen, A., Klapuri, A. (2000, June). Musical instrument recognition using cepstral coefficients and temporal features. In 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 00CH37100) (Vol. 2, pp. II753-II756). IEEE.

[2] Toghiani-Rizi, B., Windmark, M. (2017). Musical Instrument Recognition Using Their Distinctive Characteristics in Artificial Neural Networks. arXiv preprint arXiv:1705.04971.

Thank you!