AINA

AI and Analytics

Volume 1 • Edition 2019-20

Annual analytics magazine from the students of PGDBA

Post Graduate Diploma in Business Analytics: *Jointly offered by IIM Calcutta, ISI Kolkata, IIT Kharagpur*

Predominant Musical Instrument Classification based on Spectral Features

BY KARTHIKEYA RACHARLA, VINEET KUMAR, BHUSHAN CHAUDHARI , ANKIT KHAIRKAR, PATURU HARISH

With the aim to examine one of the cornerstone problems of Musical Instrument Retrieval (MIR), a spectral feature-based methodology for the classification of predominant instruments used in an audio sample is presented. For this purpose, the IRMAS dataset has been chosen. It includes clips of 3846 music samples with around 192 minutes run-time recorded from various sources in the last century, spanning multiple genres like country folk, classical, pop-rock, Latin-soul etc., making the data set diverse and better training.

Feature Extraction

While processing the audio dataset, it was found that despite having the same sound notes, the Spectrogram (the visual representation of the spectrum of frequencies of a signal over time) varies based on the instrument through which the musical note gets originated. This property of the spectrogram helps in capturing and predicting the instrument. For transformation from time domain to frequency domain, Mel Frequency Cepstral Coefficients (MFCC) were generated using FFT, then Mel Scale Filtering for scale transformation from frequency scale to Mel-scale. The audio spectrum is then analyzed by extracting MFCC's based on the default inputs of sampling rate (44.1 kHz) and hop size (hop length between the frames) is chosen as 512.

Additional Features

Based on the literature review, several additional features were used in this context such as:

1. Zero Crossing Rate (ZCR) – indicates the rate at which the audio signal crosses zero
2. Spectral Centroid (SC) – indicates the frequency at which the energy of a spectrum is centered upon, featuring the impression of the brightness of the audio sample
3. Spectral Bandwidth (SB) – represents the weighted average of the frequency signal by spectrum
4. Spectral Roll off (SR) – measures the frequency under which a defined proportion of the overall spectral energy belongs to.

Hann Windowing, a smoothing technique ideal for frequency resolution & reducing the spectral leakage is applied, while utilizing the Python inbuilt audio libraries - Essentia and Librosa. Several supervised algorithms were utilized for classification among which the SVM classifier has outperformed all the other model tryouts with an accuracy of 79%. Unsupervised techniques were also implemented among which the Hierarchical clustering was found to perform substantially well.

Having started as an academic project at ISI, this research paper was presented at the 7th International Conference on Signal Processing and Integrated Networks (SPIN) 2020, (having approximately 31% acceptance rate) organized at New Delhi during 27-28 February 2020, subsequently getting featured in the IEEE Xplore journal.

Paper Link: https://ieeexplore.ieee.org/abstract/document/9071125/