Machine Learning Engineer Nanodegree

Capstone Proposal

Music Genre Classification

Prakhar Kumar

13th April, 2019

Domain Background:

A Music Genre is a conventional category that identifies some pieces of music. Music can be divided into genres in many different ways. The artistic nature of music means that the classifications are subjective and some genres may overlap with one another. Some of the popular music genres are: hip-hop, rock, country etc.

It is very difficult to put a music into one category of music as it may have parts which belong to another category (genres) of music. There are differences between genres of music like hip-hop and heavy metal. Heavy metal music generally uses more bass and high-pitched sounds of musical instruments while hip-hop mainly focuses on the vocals.

Grouping the music into one genre is important because each individual is different as has a different taste in music. By grouping, we can allow users to listen to the music which they like.

Problem Statement:

The problem is to develop an ML application which can classify a music into different genres.

Datasets and Inputs:

The dataset used is GTZAN dataset. The dataset can be downloaded directly from this link - http://opihi.cs.uvic.ca/sound/genres.tar.gz

The dataset contains 10 classes/genres of music, these are:

- Metal.
- Disco.
- Classical.
- Hip-hop.
- Jazz.
- Country.
- Pop.
- Blues.
- Reggae.
- Rock.

Every genre listed above has 100 songs each of them are 30 seconds long.

The input to our application will be the songs listed above, we will have to convert them to appropriate format for our models to work.

Solution Statement:

The audio files obtained cannot be directly used with the Machine Learning techniques, we will first have to convert them. The techniques used to convert are FFT (Fast Fourier Transformations) and MFCC (Mel-Frequency Cepstral Coefficients). After having each song converted using the above methods, we will have to one-hot encode the target labels. There are 10 genres of music, each genre will be given a number from 0 through 9. After having the data in the required format, we will train 3 machine learning models namely – Logistic Regression, Random Forest and XGBClassifier on both type of features (FFT and MFCC) using the train dataset. We will then evaluate the model on test dataset using confusion matrices and roc auc score.

Benchmark Model:

The best model obtained from training on FFT Features will be taken as our benchmark model. To improve the performance, we will use the featurization techniques which have been developed by researchers for audio data – MFCC. We will compare the results of this featurization with the FFT results.

Evaluation Metrics:

The metrics used are – confusion matrix, roc auc score.

The metrics chosen are better than plain accuracy. The metrics are a best fit to the multi-class classification problem.

To read more on confusion matrices:
https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62

To read more on roc auc scores:

https://medium.com/greyatom/lets-learn-about-auc-roc-curve-4a94b4d88152

Project Design:

1. **Data Pre-processing:** The data obtained from the dataset is in the form of .au file, we will convert it to .wav file using "ffmpeg" software which is used to convert the audio file in format to another format. The software is easily available for download from - http://ffmpeg.org/download.html
2. **Data Visualization:** For audio data we will plot the spectrogram and power density for each of the classes/genres available.
3. **Model Training and Evaluation:** We will train 3 different Machine Learning Models on FFT Featurization and on MFCC featurization. The results will then be put up in table for comparison.

References:

- Featurization Techniques: https://www.analyticsvidhya.com/blog/2017/08/audio-voice-processing-deep-learning/
- Music genre Info: https://medium.com/giglue/top-10-genres-of-music-industry-7f19cdb177cb
- Metrics for Multi-class classification: https://medium.com/usf-msds/choosing-the-right-metric-for-evaluating-machine-learning-models-part-2-86d5649a5428