Study of different Machine Learning models for Music Genre Classification

Abhishek Kumar [*1] **Khadatkar Sameer Raju** [*1]

Abstract

Music apps nowadays have millions of songs in their database. For the proper organization of these songs, their classification according to the genre is essential. In this project, we are trying to use different Machine Learning models for doing this classification.

1. Technical Details

1.1. Dataset and its Features Extraction

In this project, we used the GTZAN dataset, which contains 1000 audio files of 10 different genres. This dataset has ten classes, Blues, Classical, Country, Disco, Hip-Hop, Jazz, Pop, Metal, Reggae, and Rock. Each class is containing 100 audio clips. Each clip is 30 seconds long in .wav format and with sample rate of 22050Hz.

We use timbral texture features as our features, which includes frequency domain features such as Spectral centroid, Spectral flux, Root Mean Square Energy (RMSE), Zero-crossing rate (zcr), Spectral contrast, Spectral bandwidth, Spectral flatness, Spectral roll-off, and the Mel-Frequency Cepstral Coefficients (MFCC). The number of MFCCs used is 20. And for each frequency domain feature, we get their statistics such as maximum, minimum, mean, standard deviation, kurtosis, and skewness from the ravel form of their data. So our final length of the feature vector for each sample becomes 28 (frequency domain features) × 6 (statistics) + 1 (the tempo of the music) = 169. If there are N samples in the training set, the size of training data becomes N×169. To extract these features, we used the floating point time series form of the music file. Some features require constructing the spectrogram for which the FFT length used was 1024, and the hop length used for Short-time Fourier transform (STFT) was 512.

To find the features for CNN, we first converted the 1D shape of time-series data of a sample to a 2D form using a window size = 33000 and window offset = 16500. Due to which shape of each data sample became (39,33000) for a 30 seconds .wav file. Then the 128 point mel-spectrogram was calculated for each 33000 size data for hop length of 256, which change the size 33000 to (128, 129), so the final size of each feature became (39,128,129). This data was given as input to CNN.

1.2. Machine Learning models used for Classification

Models that we are using: K-Nearest Neighbour (KNN), XGBoost, Support Vector Machine (SVM), Deep Neural Network (DNN) and Convolutional Neural Network (CNN). Except XGBoost all models have discussed in our MLSP class. For XGboost, it refers to eXtreme Gradient Boosting decisions tree which enhances both speed and performance. In this, we predict the errors of previous models and add them to arrive at the final prediction. For adding new models, XGboost uses gradient descent algorithm to minimise loss.

Models	KNN	XGBoost	**SVM**	DNN	CNN
Accuracy	59%	65%	**76%**	72%	69.8%

Table 1. Accuracy on test data for different models

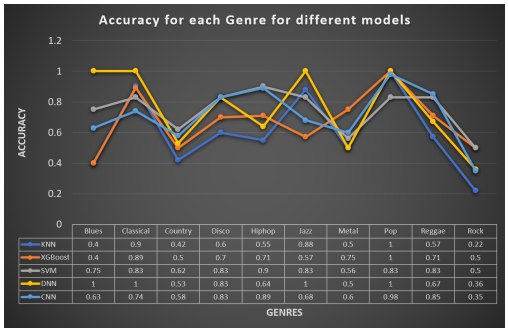

Figure 1. Speedups Graph

2. Results

The accuracy on test data for different models are shown in Table 1. As can be observed, SVM is the best performing model for us. In Figure 1, we can observe that pop and the classical genre were easy to classify, whereas rock was

[1]Department of Computational and Data Sciences, Indian Institute of Science, Bengaluru, India. Correspondence to: abhishekkum2,sameerraju <@iisc.ac.in>.

comparatively difficult to classify. We were expecting good results from CNN, but somehow we were not getting the desired results, as we couldn't make enough trials for the model's training because it took a lot of time on our system.

3. Tools used and Contributions

For this project, we read some research papers (Tzanetakis & Cook, 2002),(Ghildiyal et al., 2020), through which we got some insight about the new approaches to extract features of the audio (.wav format) file, which are more appropriate for the music genre classification.

Python Libraries that helped us to get the results are :

- Numpy
- Pandas
- Librosa: Extracting features
- Scikit-learn: KNN, SVM
- xgboost
- Keras: DNN,CNN
- Matplotlib

We distributed our work among us in the following manner:

- Abhishek Kumar : DNN, KNN, XGBoost and report writing.
- Khadatkar Sameer Raju : Feature Extraction, SVM, CNN and PPT.

References

Ghildiyal, A., Singh, K., and Sharma, S. Music genre classification using machine learning. In *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 1368–1372. IEEE, 2020.

Tzanetakis, G. and Cook, P. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing*, 10(5):293–302, 2002.