Music Genre Classification

Mandeep Singh[1] **Parimal Gajbhiye**[1]

Abstract

Music genre classification aims to classify the music files in certain genre it belongs to. Automatic categorization of music into different genres is done using Machine learning/Deep learning tools. In this project, we are extracting features from the music data (GTZAN) and then using different learning models to perform the classification.

1. Technical Details

1.1. Dataset and feature extraction:

In this project, we have used the GTZAN dataset, which contains 1000 audio files of 10 different genres. This contains ten genres: Blues, Classical, Country, Disco, Hip-Hop, Jazz, Pop, Metal, Reggae, and Rock. There are 100 music clips of each genre. Each clip is 30 seconds long, with sample rate of 22050Hz.

We have extracted timbral texture features includes Spectral centroid, Spectral flux, Root Mean Square Energy (RMSE), Zero-crossing rate (zcr), Spectral contrast, Spectral bandwidth, Spectral flatness, Spectral roll-off and Mel-Frequency Cepstral Coefficients (MFCC). Chroma features used are chroma, tonnetz, also used tempo as a rhythmic content feature and pitch as a pitch content feature. We have used 20 MFCCCs. For timbral texture feautes which are frquency based features, we have used 1024 fft/window length and 512 hop length for Short-time Fourier transform(STFT). For each feature (except tempo) we have taken 6 statistics such as maximum, minimum, mean, standard deviation, kurtosis, and skewness. Hence total number of features for each sample becomes 31 (20 MFCCs + 11 other features) * 6(statistics) + 1(tempo) = 187. As we have total 1000 samples, the size of the data becomes 1000*187. Later taking 10 % of the samples for model testing.
For CNN & LSTM, we have divided each audio file into 10 segments (3 sec each) and then getting mel-spectogram (using 20 MFCCs) of each segment, to use in model.

[1]Department of Computational and Data Sciences, IndianInstitute of Science, Bengaluru, India.. Correspondence to: mandeepsingh, parimalgajbhiye <@iisc.ac.in>.

1.2. Machine Learning models used for Classification:

The following models are considered for the classification : XGBoost, Random Forrest, K-Nearest Neighbour (KNN), Support Vector Machine (SVM), Naïve Bayes, Deep Neural Network (DNN), Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM).

2. Results

The observed accuracy for different models are plotted in the graph below.

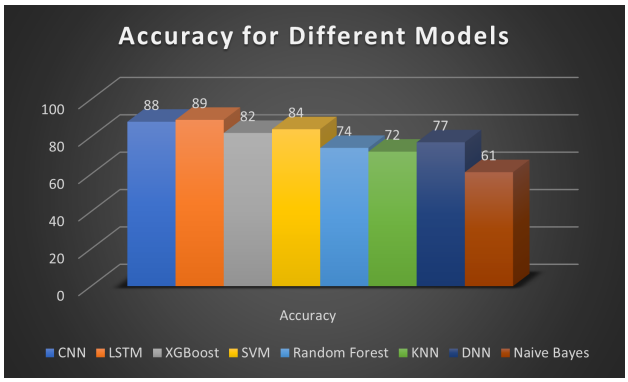

Figure 1. Accuracy graph

We can see that **LSTM** & **CNN** are giving best performance with **89** % & **88** % accuracy. Performance of XGBoost was improved by tuning the hyperparameters using RandomizedSearchCV. Similarly GridSearchCV was used to tune the hyperparameters for KNN, SVM and Random Forrest.

3. Contribution

For this project, we have referred papers (1),(2). From (1) we have studied the feature extraction from a audio sample for genre classification purpose. We have implemented paper (2), and used some additional audio features and implemented additional models such as CNN, LSTM, DNN, XGBoost.

We divided the task amongst ourselves as follows:

- Mandeep Singh : CNN, SVM, DNN, Random Forest, PPT
- Parimal Gajbhiye : Feature Extraction, XGBoost, LSTM, KNN, Naïve Bayes, Project report

4. Tools Used

The libraries that assisted us in obtaining the results are as follows:

- Numpy
- Matplotlib
- Pandas
- Seaborn
- Scikit-learn: Naïve Bayes, SVM, Random Forest, KNN
- Keras: CNN, LSTM

References

[1] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," in IEEE Transactions on Speech and Audio Processing, vol. 10, no. 5, pp. 293-302, July 2002, doi: 10.1109/TSA.2002.800560.

[2] A. Elbir, H. Bilal Çam, M. Emre Iyican, B. Öztürk and N. Aydin, "Music Genre Classification and Recommendation by Using Machine Learning Techniques," 2018 Innovations in Intelligent Systems and Applications Conference (ASYU), 2018, pp. 1-5, doi: 10.1109/ASYU.2018.8554016.