Musical Genre Classification of Audio Signals

Aditya Sharma[1] **Mahesh Kumar**[1]

Abstract

AI based Classifier techniques are an essential component of Music Recommender systems based on the preferred genre of the user. In this project, we apply several Machine Learning and Deep Learning techniques to Design and Analyze the performance in terms of accuracy of various Multi-Class Classifiers for the classic GTZAN dataset [2]. We start by replicating the results predicted by the models used in Reference [1] i.e. k-NN as well as the Generative model : K-means initialization and subsequent application of the GMM model. On top of that, we've also applied SVM, Random Forest Classifier, Deep Neural Network on the audio features and a CNN on the Mel-Spectrogram.

1. Technical Details

1.1. Feature Extraction

From the Audio files (1000 samples), we extract the Timbral (19), Rhytmic (6) and Pitch Content Features (5). We take the Mean and Variances of the Timbral Features and the top 5 MFCC features to replicate the procedure given in [1]. We use these features for the application of the GMM and k-NN to validate the results given in [1]. Further, SVM, Random Forest Classifier (RFC), Deep Neural Network are applied on the extracted audio features (30 dim) with a test-train split of 30% . We also extract the Mel-Spectrogram Images (128 x 640 dim) and apply a CNN model on it .

1.2. Model Details

k-NN – 4 Neighbours . GMM, intialized with K-means with 3 component GMMs per class . SVM with C (Regularization Constant) = 0.1, tolerance = 0.001 . RF classifier with 100 estimators , DNN–2 Hidden Layers (H1=256, H2=128), with ReLu activations , Dropout =0.35 and L2 Regularization=0 for each layer. For CNN – 3 Conv. Layers with Max Pooling and Batch Normalization each with Filter Size 3x3 and Pool size 2 x 4 and 2 Flattened hidden layers with 128, 64 neurons respectively and the activation Elu with Dropout of 25% and 0.1% of L2 Regularization.

[1]Indian Institute of Science, Bengaluru, India.

2. Results

Table 1. Classification accuracies for various Models.

MODEL	TRAIN ACC.	TEST ACC.	BETTER?
GMM	54.33± 0.2	56.70± 4	
K-NN	72.30± 1	59.0± 4	
SVM(RBF)	87.84± 1	62.72± 1	√
RFC	99.00± 0.1	65.00± 2	√
DNN	90.00± 2	72.00± 2	√
CNN	99.20± 0.6	73.50± 2	√
DNN- 3 SEC SPLIT	97.00± 0.5	91.00± 0.5	√

GMM generative model gives about 55-60 % and k-NN gives us about 60% accuracy as predicted by [1] . This is in close agreement with our results as shown in Table 1. SVM shows a marginal increase to 62.72% on the testing data and Random Forest proves itself more accurate (65%) due to it being a combination of multiple trees. DNN shows an apprecibly good result with a robust 72 % accuracy. CNN shows a good 73% only with taking mel-spectrograms which is very appreciable. Finally and Most Importantly, we have implemented the 3s bifurcation of audio signals and it starkly improves our accuracy taking it upto 92 % .

3. Novel Contributions

We have splitted 30 seconds audio file which have too much information for the model to take at once, so we splitted a single audio file into 10 audio files each of 3 seconds.We implemeneted CNN and also implemented The DNN 3s classifier which showed a very good result due to Data Division and Augmentation.

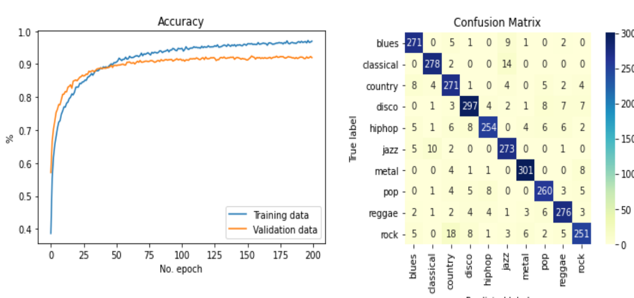

Figure 1. Accuarcy & Confusion Matrix of DNN: 3 Sec Features

3.1. Summary

In the project, we compared the performances of various pre-processing techniques and concluded that the MFCC feature extraction is the best amongst them.

Feature Extraction is collaborated between both the teammates. Individual contributions are as follows:

Aditya Sharma: GMM, SVM, RFC, DNN.

Mahesh Kumar: K-NN, DNN, CNN.

4. Tools Used

We used Python and following Libraries are used to perform the experiments:

1) Librosa- For Extracting Timbral Features.

2) Sklearn- For Preprocessing.

3) Keras- For DNN and CNN.

4) Google Colab and Jupyter Notebook- IDE

5. References

1. George Tzanetakis and Perry Cook. ”Musical Genre Classification of Audio Signals”

2. Data Source. http://opihi.cs.uvic.ca/sound/genres.tar.gz