A Study on Music Genre Classification Techniques

Samarth Gupta

Lovely Professional University, Phagwara, India

samarth.11911223@lpu.in

Abstract: One of the most challenging tasks in Music Information Retrieval is categorizing music files according to genre, which is very useful for music recommendation. In this study, we compare performance of various techniques used to classify music files and evaluate the performance. The techniques include KNN, Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, SVM and Artificial Neural Networks. These techniques use hand-crafted features both from time domain and frequency domain. Also, CNN is used in this work that uses spectrograms to classify music[1].

Keywords: Music Genre classification Techniques

1. Introduction

Music Industry is rising every day. The companies like Spotify and iTunes have contributed a lot to it. One of the biggest problems they face is examine and classify the music according to the genre. In order to perform music classification automatically based on the genre various machine learning and deep learning techniques can be used. This study explores the application of these algorithms to identify and classify the genre of a given audio file. Also, will put some flash on important features of Audio file. Music genre classification has various use cases. One of them is music recommendation.

2. Methodology

Figure 2.1 shows the steps followed in this work.

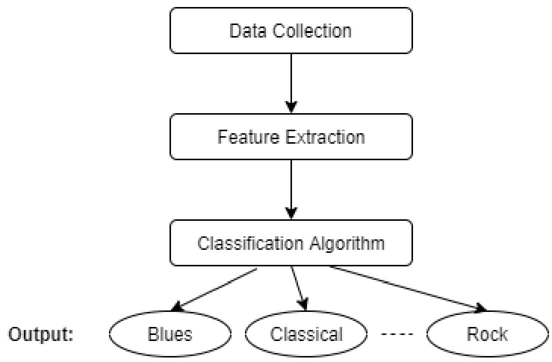

Figure 2.1 Methodology

2.1 Data Collection

Data collection is the first step in any classification task. In this work, we used GTZAN Dataset, which is a very popular Dataset for Genre classification. This dataset contains 10

[1] The code has been open sourced and is available at https://github.com/Sk-singla/Music-Genre-Classification.

Genre classes (blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock). Each class have 100 Music files in it.

2.2 Feature Extraction

Feature extraction gets the features from the data. To classify music, the main features to be extracted from audio file are:

- **Spectrogram**: A visual representation of the signal strength or loudness or pitch of a signal over time at different frequencies present in a waveform as shown in figure 2.2.1. By this we can see energy levels varying over time.

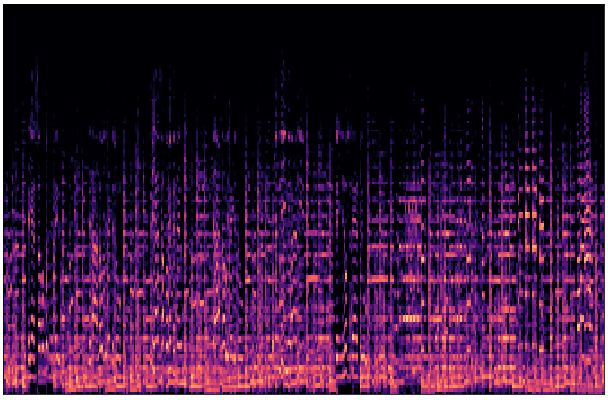

Figure 2.2.1 Spectrogram

- **Mel Spectrogram**: The Mel Scale, is the result of some non-linear transformation of the frequency scale. The Mel Spectrogram is a normal Spectrogram, but with a Mel Scale on y axis. This is what we are using for CNN.

- **Zero Crossing Rate**: The rate at which the signal changes from positive to negative or number of zero crossing withing a segment of a signal. It is used to measure smoothness of signal.

$$zcr = \frac{1}{T-1}\sum_{t=1}^{T-1} \mathbb{I}\{s_t s_{t-1} < 0\}$$

s_t is the signal of length t
$II\{X\}$ is the indicator function (=1 if X true, else =0)

- **Spectral Centroid**: It Indicates where the "centre of mass" for a sound is located and is calculated as the weighted mean of the frequencies present in the sound.

$$f_c = \frac{\sum_k S(k) f(k)}{\sum_k S(k)}$$

Where S(k) is the spectral magnitude at frequency bin k, f(k) is the frequency at bin k.

- **Spectral RollOff**: It is a measure of the shape of signal as show in Figure 2.2.2. It represents the frequency at which high frequencies decline to 0.

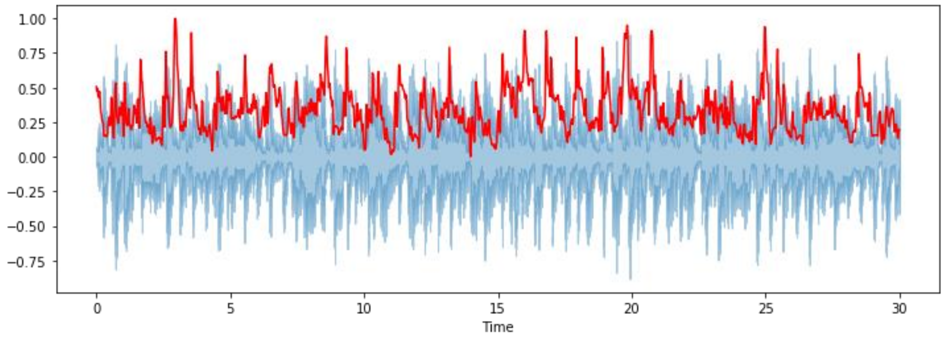

Figure 2.2.2 Spectral Rolloff

- **Mel-Frequency Cepstral Coefficients**: MFCCs of a signal are small set of features which concisely describe the overall shape of a spectral envelope. It models the characteristics of the human voice.

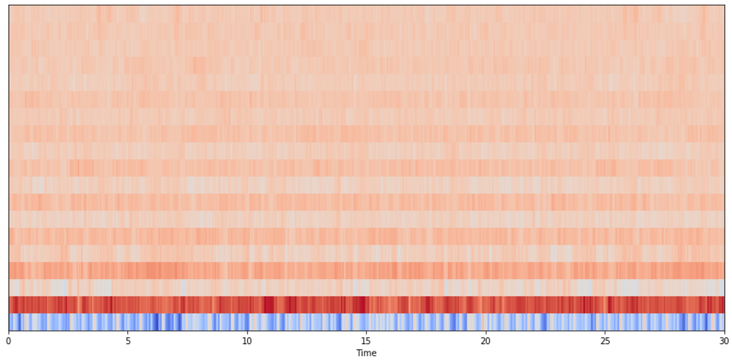

Figure 2.2.3 MFCC

- **Chroma Frequencies**: Chroma features are an interesting and powerful representation for music audio in which the entire spectrum is projected onto 12 bins representing the 12 distinct semitones of the musical octave.

- **Harmonics and Perceptrual:** Harmonics are characteristics that human ears can't distinguish It represents sound colour. Perceptrual understanding shock wave represents the sound rhythm and emotion

- **Tempo:** It represents how fast or slow a music piece is. It is expressed in Beats Per Minute (BMP).

2.3 Classification Algorithms

The extracted features are then used to classify music files according to their genre by using various classification algorithms. We are using various classifiers that are described below:

- **KNN:** A supervised machine learning algorithm, that takes labelled input data to process unlabelled data in future. When using in Music genre classification, it looks at familiar music and assumes they belongs to same category because they seems to be near to each other. When using k = 1, means the nearest neighbour, it gives maximum accuracy while classifying music genre.

- **Naïve Bayes:** Naïve Bayes is one of the highly practical Bayesian learning methods. This classifier is based on applying Bayes' theorem with strong independence assumptions between features. Bayes Optimal Classifier calculates the probability of a class from each group of existing attributes, and determines which class is the most optimal.

- **Decision trees:** Decision tree is a technique very useful in data mining to extract information of a data set, normally using a TDIDT (Top-Down Induction Decision Tree) algorithm. It is a tree shaped (if-else) classifier in which internal nodes represents features, branches represent decision rules and leaf nodes represent outcome.

- **Random Forest:** A forest contains multiple trees. Similarly random forest algorithms create given number of decision trees and get the best result by means of voting. It reduces the over-fitting by averaging the result. Larger the number of trees, more accurate is the result.

- **SVM:** A machine learning technique which is based on the principle of structure risk minimization is support vector machines. SVMs transform the original input data into a high dimensional space using a kernel trick. The transformed data is linearly separated using a hyperplane as shown in figure 2.3.1.

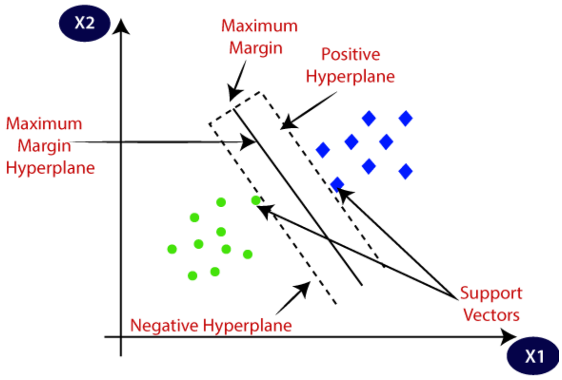

Figure 2.3.1 SVM

- **Logistic Regression:** This linear classifier is generally used for binary classification tasks. For this multi-class classification task, the LR is implemented as a one-vs-rest (OvR) method and uses the cross-entropy loss.

- **ANN:** Neural Networks are the ones that develop structure of our brain. Artificial Neural network is derived from that. Like the human brain that has neurons interconnected to one another, artificial neural networks also have neurons that are interconnected to one another in various layers of the networks.

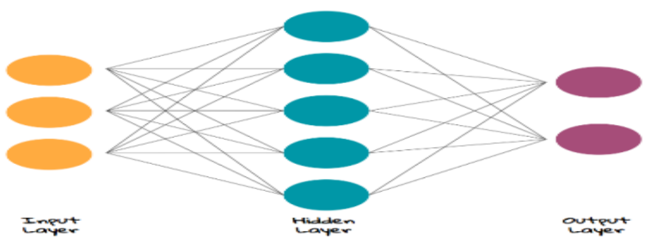

Figure 2.3.2 Neural Network Structure

- **CNN:** Convolutional Neural Networks is mainly used to process multidimensional vectors such as images. Understanding Images is done by CNN. So here we are using Spectrogram of audio file to classify genres. But it is found that it doesn't works so good in this case.

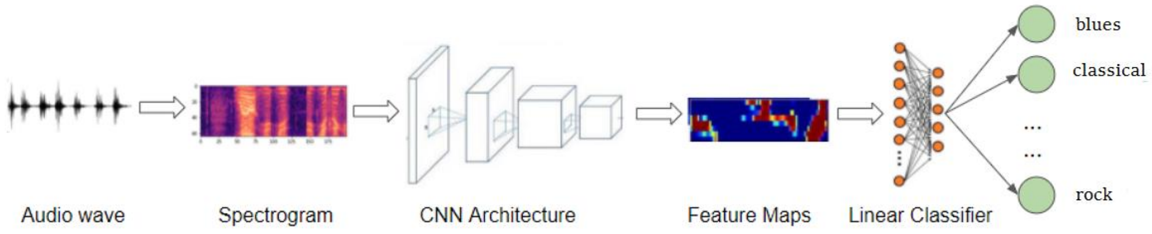

Figure 2.3.3 Architecture used for CNN

3. Evaluation

3.1 Model Performance Evaluation

To Evaluate the models described in section 2.3, the following metrics will be used.

- **Accuracy:** It refers to the percentage of correctly classified test samples.
- **F-score:** Based on the confusion matrix, it is possible to calculate the precision and re-call. F-score is then computed as the harmonic mean between precision and recall.

Table 3.1 Model Performance Evaluation

Classifiers	**Accuracy**	**F-Score**
KNN	0.68	0.7
SVM	0.66	0.67
Logistic Regression	**0.72**	**0.73**
Naïve Bayes	0.5	0.51
Decision Tree	0.45	0.45
Random forest	0.63	0.64
ANN	**0.68**	**0.72**
CNN	0.33	0.35

From table 3.1, It can be observed that Logistic Regression in Simple Machine Learning Algorithms and ANN in Deep Learning algorithms performed well for music genre classification task.

3.2 Feature Evaluation

Here we will find top 20 most Important features to classify music. From Figure 3.2, It can be observed that MFCC is most important feature of audio. Other than that, Zero Crossing Rate, Chroma Frequencies, Harmony and Perceptrual are also among top 20 features.

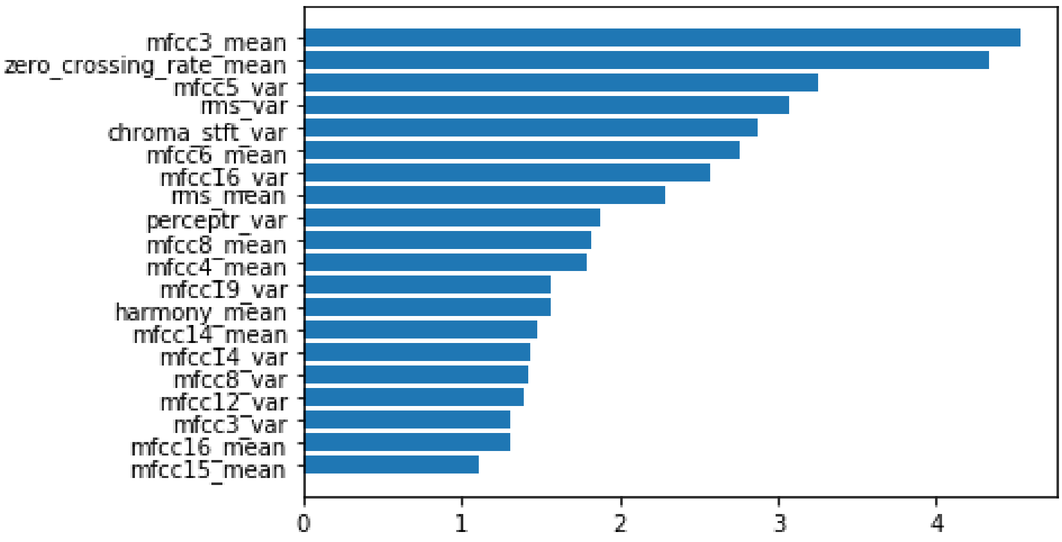

Figure 3.2 Top 20 Features

4. Conclusion

In this work, the task of music genre classification is proposed using 8 different machine learning and deep learning Algorithms. The techniques KNN, Logistic Regression, Naïve Bayes, Decision Tree, Random Forest, SVM and Artificial Neural Networks used hand-crafted audio features to predict music genre. But CNN used Spectrograms for classification. Logistic Regression is found to be the best performing model. But if we talk about Neural Networks, Simple ANN technique also performed well, which can even be further improved. Also, MFCC is found to be the Most Important feature for music classification. It has been found that the audio clips used in this study are bit noisy. Future studies can find a way to pre-process noisy data before feeding it into a machine learning model, in order to achieve better performance.

References

Olteanu, A. (2020, March 24). *GTZAN dataset - music genre classification*. Kaggle. Retrieved March 25, 2022, from https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification

Bahuleyan, H. (n.d.). *Music genre classification using Machine Learning Techniques*. Retrieved March 25, 2022, from https://www.researchgate.net/publication/324218667_Music_Genre_Classification_using_Machine_Learning_Techniques

Bressan, G. M., de Azevedo, B. C., & Lizzi, E. A. (2017). A decision tree approach for the musical genres classification. *Applied Mathematics & Information Sciences*, *11*(6), 1703–1713. https://doi.org/10.18576/amis/110617

Rawat, S. (n.d.). *Music genre classification using machine learning*. Analytics Steps. Retrieved March 25, 2022, from https://www.analyticssteps.com/blogs/music-genre-classification-using-machine-learning

Sugianto, S., & Suyanto, S. (2019). Voting-based music genre classification using Melspectogram and Convolutional Neural Network. *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. https://doi.org/10.1109/isriti48646.2019.9034644

Chauhan, N. (n.d.). *Audio data analysis using Deep Learning with python*. KDnuggets. Retrieved March 25, 2022, from https://www.kdnuggets.com/2020/02/audio-data-analysis-deep-learning-python-part-1.html

Anjum, F. (2020, February 4). *Music genre classification with python*. Medium. Retrieved March 25, 2022, from https://medium.com/analytics-vidhya/music-genre-classification-with-python-51bff77adfd6

Notes on music information retrieval¶. index. (n.d.). Retrieved March 28, 2022, from https://musicinformationretrieval.com/