Music Genre Classification

Group Project of ECE-GY-6123 Intro to Machine Learning

Group members:

Siqi Dong, sd4147

Peidi Tang, pt1348

Yu Fan, yf1399

Jiaxin Jiang, jj2919

***Abstract*—**We have made attempts to complete music genre classification work using neural network feeding with Mel-spectrogram during Lab 9 in this course. The main purpose of this group project is to complete a comparative study by implementing various common machine learning algorithms to classify the music dataset into its corresponding genres respectively to compare their performance this specific topic. These machine learning algorithms includes K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Convolutional Neural Network (CNN). This project concludes that CNN achieved the highest test accuracy on test set over other methods mentioned.

I. Introduction

Music Genre classification is a task that aims to predict music genre using the audio signal. Being able to automatize the task of detecting musical tags allow to create interesting content for the user like music discovery and playlist creations, and for the content provider like music labeling and ordering.

To solve the problem, it requires extracting acoustic features that are good estimators of the type of genres we are interested, followed by a single or multi-label classification or in some cases, regression stage. Conventionally, feature extraction relies on a signal processing front-end in order to compute relevant features from time or frequency domain audio representation. The features are then used as input to the machine learning stage. Instead of dealing with large variety of features, proposed scheme relies on MFCC and its variants which are introduced at the different stages to satisfy the need.

To compare the result across multiple architectures, we have took two approaches for this problem: One using the classic approach of extracting features and then using a classifier. The second approach, which is implemented on the code file here is a Deep Learning approach feeding a CNN.

II. Classification Methods

This project will use K-Nearest Neighbor, Support Vector Machine and Convolutional Network as our three machine learning methods because they are frequently used in classification problems.

A. K-Nearest Neighbor (KNN)

The k-nearest neighbor (KNN) classifier is a non-parametric classifier that polls the "K" nearest training points to a given test point and classifies that test point based on the majority vote of these "K" nearest neighbors. KNN relies on a local prominence of data with a certain label and works well with data that is described by a number of highly dense clusters, with each cluster representing a given label. While this algorithm is simple to implement, it is vulnerable to high dimensional inputs.

B. Support Vector Machine (SVM)

SVM is a supervised machine learning algorithm. It is a machine learning technique which is based on the principle of structure risk minimization is support vector machines. SVM constructs linear model based upon support vectors in order to estimate decision function. If the training data are linearly separable, then SVM finds the optimal hyper plane that separates the data without error.

C. Convolutional Neural Network (CNN)

CNN combine three architectural ideas to ensure some degree of shift, scale, and distortion invariance: local receptive fields, shared weights, and sub-sampling. These concepts can be modified and used in music classification based on convolution of spectrogram. First, we introduce the feature detector. Then we apply convolution operation on spectrogram using these filters. Once the feature map is obtained, we apply a sub-sample layer on each of the feature map.

III. Experiment

A. Dataset

In this project, we will use GTZAN as our dataset for all the machine learning methods. GTZAN is a dataset created by Tzanetakis et al. This dataset only contains raw audio wave file together with corresponding genre labels, and it is compounded of 1000 music excerpts of 30 seconds duration with 100 examples in each of 10 different music genres: Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Popular, Reggae, and Rock. GTZAN is of large-enough scale to support our music classification problem.

B. *Feature Extraction*

Feature extraction is of significant importance in our music classification task. It is difficult to determine features needed from scratch. The reason is that there are a great amount of features related to the tonal, timbre, pitch and rhythm of each music record, such as spectral centroid, tempo and so on. Thus, we would just extract features by using Librosa library in Python to get MFCC and its variants to simplify the feature extraction process.

C. *Performance*

We have measured the performance of listed machine learning methods by using test accuracy as metric. The test accuracy is defined as the ratio of the number of the results which were correctly classified to the total number of classified results.

As mentioned in III(A), our dataset contains 1000 music records in 10 genres. To evaluate the performance, we deployed and trained the KNN, SVM and CNN with the split training partition of our dataset and evaluated using the test partition. The results have been showed below:

- The genre prediction has been made using the Logistic Regression model with a total accuracy of 65.3%.
- KNN achieves test accuracy of 68.7%.
- Support Vector Machine achieve test accuracy of around 72.4%.
- And for the implementation using CNN, the average accuracy, computed as the mean of the diagonal elements, has been increased from 72.4% above using KNN to 83.2%.

IV. CONCLUSION

We explore the application of Logistic Regression, KNN，SVM，CNN for the task of music genre classification. The results have shown that this kind of networks need large quantities of data to be trained from scratch. What’s more CNN has better performance in music genre classification.