Music Classification

Hasan Men

Gebze Technical University

1. INTRODUCTION

Analyzing music audio files based on genres is an active field of research in machine learning. There are a lot of feature extraction methods, datasets and example works [3][4]. In this project, I classified music genres with Mel Frequency Ceptral Coefficients features on GTZAN Marsyas dataset. Our difference from others is music sample time which is 6 seconds. Reducing sample/feature size is making difficult to classify similar genres. For classification, SVM, KNN, XGBoost, Random Forest Classifier and Decision Tree were used. Another difference and additional test is testing trained models on Youtube musics.

2. DATASET

There two dataset which used in this project. First one is called GTZAN dataset from Marsyas [1]. It's open source and open to any use. This dataset includes 1000 track example which each of them are about 30 seconds long. These musics are blong to 10 different genres. Some of them are, Jazz, Classic, Metal, Hip-Hop etc. This dataset was used only for train and dev test. The other dataset which is created from youtube mixed musics. For each class I cutted 5 different, 6 seconds samples like our train dataset.

3. FEATURE EXTRACTION

The accuracy of classifying a data highly depends on data and extracted features. Mel Frequency Cepstral Coefficients(MFCC) are mathematical coefficients for sound modelling and a representation of the short-term power spectrum of a sound [2]. MFCC is a way to represent time domain waveforms as just a few frequency domain coefficients(Figure 1).

I generate MFCCs from only 6 seconds of musics with librosa MFCC extractor[5]. Middle part of a music(12 and 18 seconds for GTZAN dataset) is a good for feature extraction because of getting actual part of music. MFCC Feature vector size is 420. For downloaded youtube musics, I listened the musics and selected the parts which I liked.

4. CLASSIFICATION AND RESULTS

On classification step, I shuffled all dataset and used K-Fold cross validation to get exact results. My K value is 5 and in each folding %80 train, %20 test data was used. After folding and classifying dataset, confusion matrix was used to calculate accuracy score.

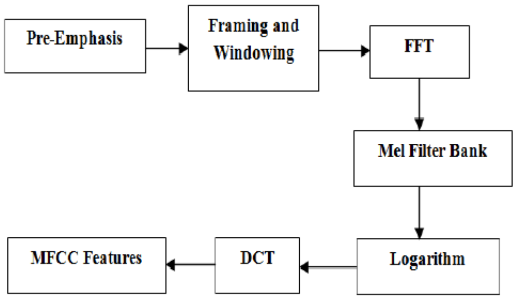
(1).png 

Figure 1: Block Diagram of MFCC

I prepared 2 section for classification. First one, classifies all data which has 4 classes which are blues, classical, jazz, pop and the other section classifies 10 class .

4.1. Classify for 4 Genre

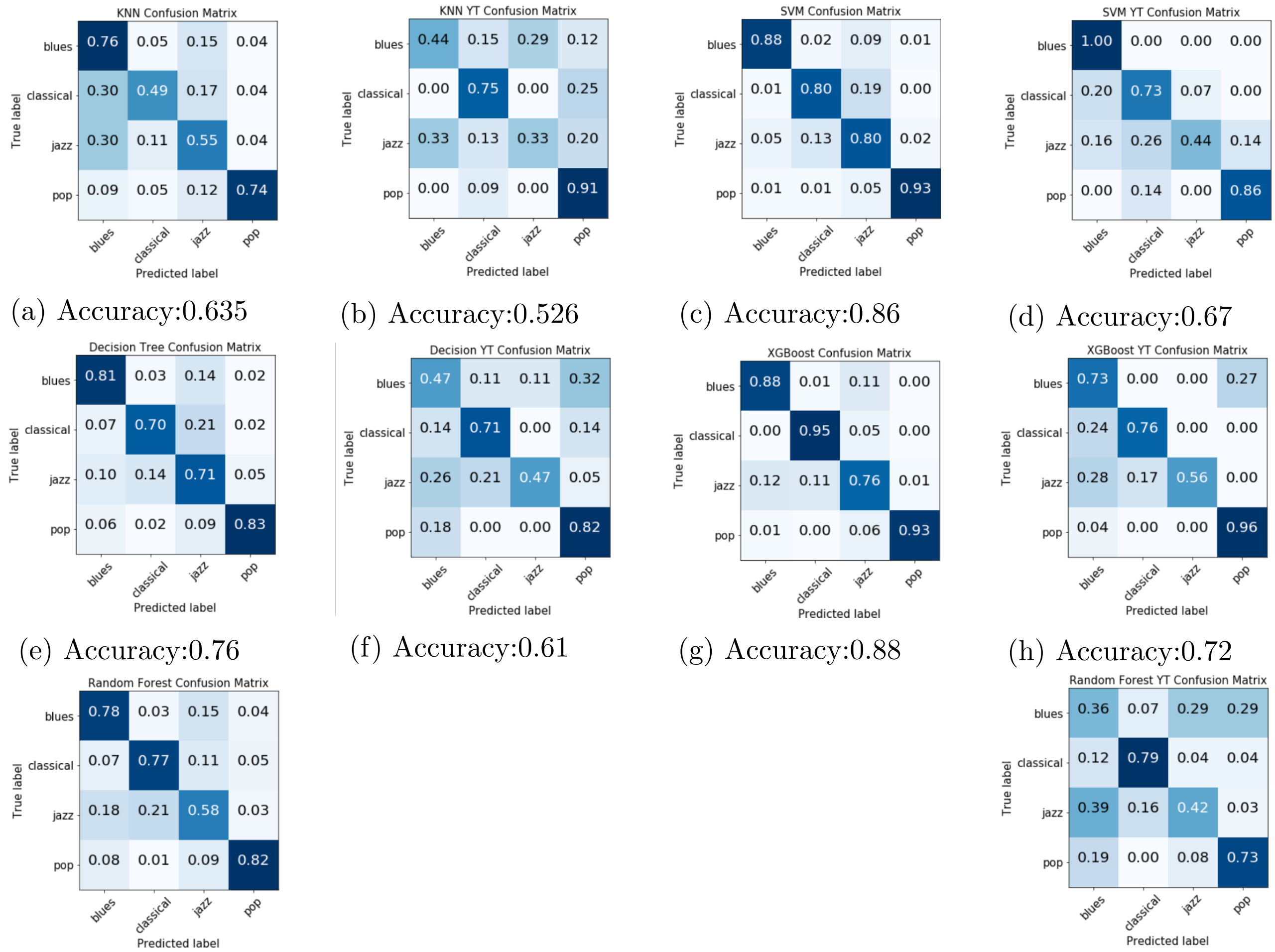
Old works about music classifications [3], [4] only used 4 genre type for classification and I also decided to test with blues, classical, jazz and pop genres. For additional tests, I used my youtube dateset and I tested them with trained models on each folding. The results are shown at below. The confusion matrices shows hat youtube test results are less than GTZAN result. I think that parts I choose from the musıc cause this difference. When I play with sample areas, results changes. Results of classifications are shown in figure 2.

4.2. Classify for 10 Genre

In this section, I trained selected algorithms with all dataset. All results are less than 4 class classification. When dataset is getting bigger, the number of similar music increases. Within this increase distinguishing MFCC features are getting hard. The most problematic part of big dataset is country musics. It actually contain common parts/rhythm with other species and we can see this result when we look Figure 3. Predictions for country and disco musics are too weak.

5. CONCLUSION

On both result, ensemble methods gave better accuracy. It was an expected scenario. When we look at the other classifiers, they works very fast but their accuracy are not acceptable on 10 class classification. If longer sample(15 seconds) are used in the feature, I am guessing the new results will be increased. The increasing of sample size can rebound feature vector but with using PCA combination acceptable results can be produced. The other opinion to improve accuracy is using Naive Bayes because there are a lot of correlation between genres like country and others.



(a) Accuracy:0.635 (b) Accuracy:0.526 (c) Accuracy:0.86 (d) Accuracy:0.67

(e) Accuracy:0.76 (f) Accuracy:0.61 (g) Accuracy:0.88 (h) Accuracy:0.72

(i) Accuracy:0.73 (j) Accuracy:0.58

Figure 2: Confusion Matrices for 4 Class Classification

References

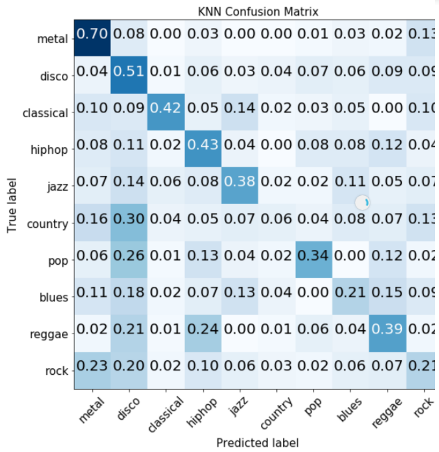
[1] Marsyas "GTZAN Genre Collection". http://marsyasweb.appspot.com/download/data_sets/, Last Visit: 23 May 2018.

[2] Mel-frequency cepstral coefficients. https://www.wikiwand.com/en/Mel-frequency_cepstrum, Last Visit: 23 May 2018.

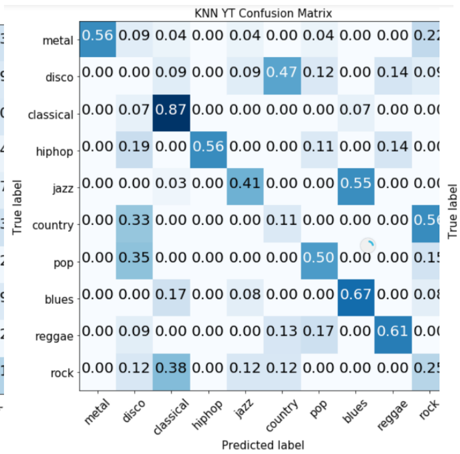
[3] Matthew Creme, Charles Burlin, Raphael Lenain. (Stanford University) Music Genre Classification, 2016.

[4] Madhura Dole, Teja Mukka. (The University of Texas at Dallas) Music Genre Classification, 2017.

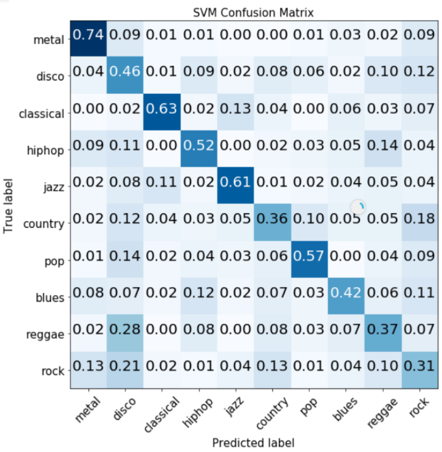
[5] Librosa MFCC Guide. https://librosa.github.io/librosa/generated/librosa.feature.mfcc.html , Last Visit: 7 Jun 2018



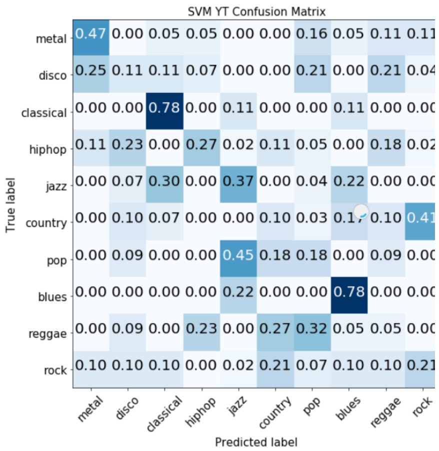
(a) Accuracy:0.36



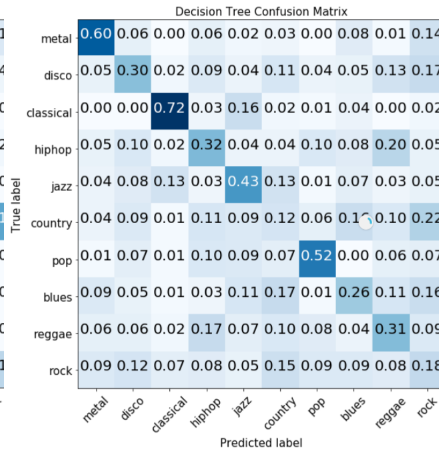
(b) Accuracy:0.43



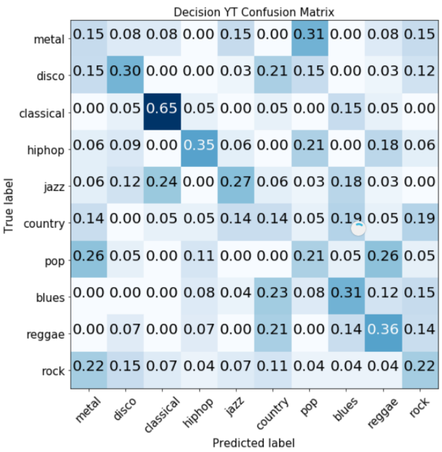
(c) Accuracy:0.49



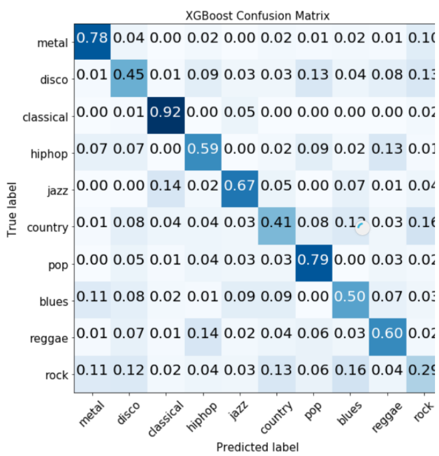
(d) Accuracy:0.26



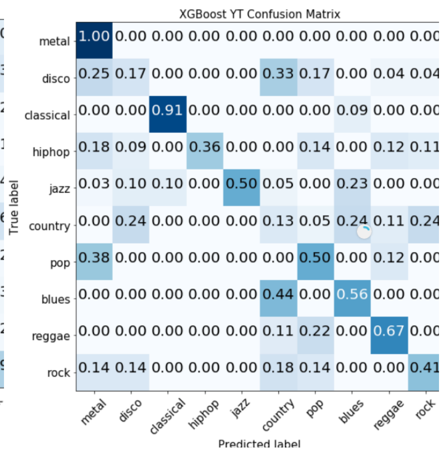
(e) Accuracy:0.37



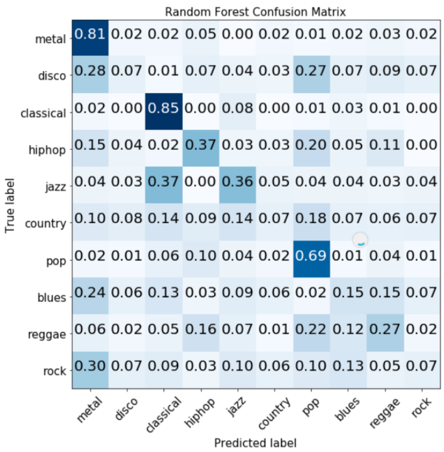
(f) Accuracy:0.30



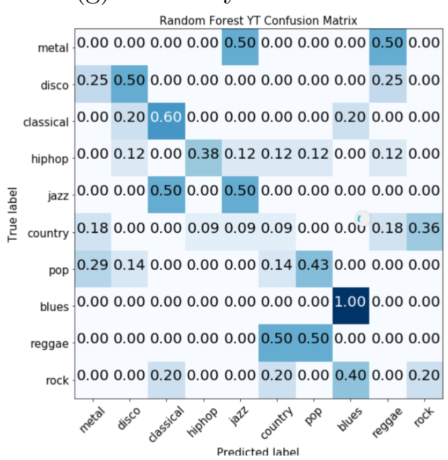
(g) Accuracy:0.60



(h) Accuracy:0.42



(i) Accuracy:0.37



(j) Accuracy:0.33

Figure 3: Confusion Matrices for 10 Class Classification