Rock or Not?

Defne Tuncer [1] **Kutay Barcin** [1]

Abstract

In the era of technology, millions of songs are brought to people everyday. The dramatic increase in the size of music collections has made the music genre recognition (MGR) an important task on machine learning. The goal of this paper is to give machines a chance to predict music genres given input features from music tracks. To do that, we apply various techniques based on machine learning on the dataset called FMA which consists of 161 sub-genres among 106,574 tracks, and published in 2017.

1. Introduction

When there is people, there is music. As people, living in today's world, music is always at our reach through technology. The ease of it has brought the demand of automatically generated playlists and customized music recommendations. The task in both those challenges is to be able to group songs in semantic categories. In this work we aim to model and classify music genres with the assumption of different music genres are also different at the bit level.

In this work we implemented and discussed baseline classification models to solve the problem of music genre recognition. These methods includes (i) Nearest Neighbor Classifier with/without dimensionality reduction and weighting hyperparameter, (ii) Logistic Regression through the one-vs-rest scheme, (iii) Support Vector Machine with linear kernel. To represent the audio tracks we planned to use Mel Frequency Cepstral Coefficients(MFCC) and Spectral Contrast features to begin with, which have been shown to be effective in the task of predicting genres.

[1]Department of Computer Engineering, Hacettepe University, Ankara, Turkey. Correspondence to: Defne Tuncer <defnetuncer@hacettepe.edu.tr>, Kutay Barcin <kutaybarcin@hacettepe.edu.tr>.

BBM406 Fundamentals of Machine Learning, Fall 2018.

2. Related Work

For the music genre recognition task, the most common datasets are GTZAN (Tzanetakis & Cook, 2002), Million Song Dataset (MSD) (Bertin-mahieux et al., 2011) and FMA: A Dataset For Music Analysis (Defferrard et al., 2017). While FMA is the most up-to-date dataset, it is especially suited for MGR as it features fine genre information.

In Music Information Retrieval (MIR), there have been various number of studies on building effective models to predict genre of music using audio features. An interesting study done by training multiple classifiers for the data and combining the results of the multiple classifiers into one single classification. (Sanden & Zhang, 2011) A paper analyzes into potent learning algorithms for genre classification based on audio waveforms. (Haggblade et al., 2011) Another study that considers the properties of the auditory human perception system proposed a music genre classifier. (Panagakis et al., 2009) Most of the studies done by exploring the timbre texture, the rhythmic content, the pitch content, or their combinations.

3. Dataset Exploration

The FMA dataset, a dump of the Free Music Archieve, includes 106,574 tracks with 161 sub-genres. In this task, we use 38,990 of the tracks with 15 top-genres sampled considering their metadata and popularity for computational efficiency. Our data includes clips of 30s and an unbalanced distribution among genres that differ from 24 to 14,182 clips per top-genre.

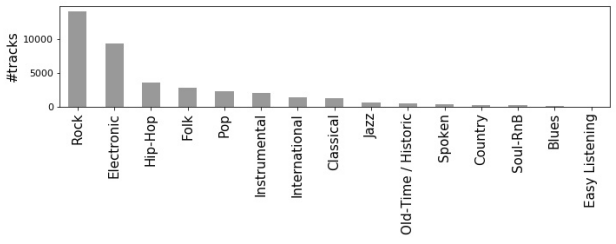

Figure 1. Top-Genre Distribution

We split our data preserving the percentage of tracks per genre as a reflection of population (stratified sampling) into training, validation and test by 80/10/10%. Thus, our training data turned into a matrix of 31,386 rows and 519 columns consists of 518 audio features and a genre label.

While extracting features, each clip is processed through librosa (McFee et al., 2018). Thus, each track contains 518 attributes categorized in 11 audio features; Mel Frequency Cepstral Coefficients (mfcc), Chroma Features (chroma cens, chroma cqt, chroma stft), Spectral Features (spectral bandwidth, spectral centroid, spectral contrast, spectral rolloff), RMS Energy (rmse), Tonal Centroids (tonnetz), Zero Crossing Rate(zcr). Each of these features are stored as statics, including kurtosis, max, mean, median, min, skew and std.

4. The Approach

Starting with the assumption that examples from the same genres are similar, they'll cluster closer to each other in the n-dimensional space, where n is the number of features. We discussed and applied several classifications methods in order to figure out which approaches best suited for our problem.

4.1. Classification Methods

4.1.1. Nearest Neighbors Classifier

In pattern recognition, k-Nearest Neighbors Classification (kNN) is a non-parametric method used for classification and regression. Although kNN is an easy implemented algorithm and performs well in a large number of classification problems, it suffers from the curse of dimensionality. Our model has a dimension space of 518 features which makes kNN vulnerable. In order to overcome this, we planned to apply Principal Component Analysis (PCA) to our matrix which reduces the input to a lower desired dimension. Figure 2. visualize scatter plot of two genres Rock and Classical after applying PCA to reduce the feature dimensions to three dimensions.

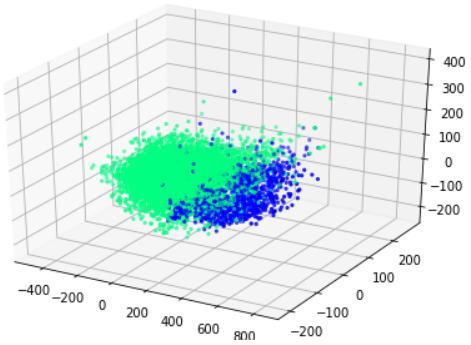

Figure 2. Scatter plot of Classical and Rock Genres

4.1.2. Logistic Regression

Logistic Regression is a technique from the statistics field and provides a probability score for observations. It is a go-to method for binary classification problems, however for our multiclass problem, we applied Logistic Regression through the one-vs-rest (OvR) scheme with a loss function.

4.1.3. Support Vector Machine

Support Vector Machine is a supervised learning method that can be used for classification. SVM works efficiently with high dimensional features even if the number of dimensions is greater than the number of samples. Various Kernel functions (SVC with linear kernel, SVC with radial basis kernel) can be implemented for the decision function. We preprocessed the data before applying Linear SVC using standard scaler to represent standard normally distributed data. Apart from Logistic Regression, SVM uses one-against-one approach for multi-class classification. This method is consistent, which is not true for one-vs-rest classification.

5. Experimental Results

For our baseline models, we approached with three classification methods: Nearest Neighbors(kNN)/Weighted Nearest Neighbors(wkNN), Logistic Regression and Support Vector Machine (SVM) with linear kernel. Each of the models was evaluated using the same training data of 31,386 clips, tuned on the validation data of 3,738 clips and tested on the 3,866 clips. The following tables show the accuracy performance obtained with all the features and Mel Frequency Cepstral Coefficients (MFCC) alone.

Table 1. Classification Methods for MFCC Only

Models	Train Acc.(%)	Test Acc.(%)
kNN	69.69	61.25
Weighted kNN	99.89	61.30
Logistic Reg.	65.04	62.55
SVM linear	64.37	61.77

Table 2. Classification Methods for All Features

Models	Train Acc.(%)	Test Acc.(%)
kNN	67.94	60.14
Weighted kNN	99.89	59.65
Logistic Reg.	73.20	65.13
SVM Linear	71.32	63.35

All the methods we used appear to have difficulties in capturing the non-linearities of the data, thus they achieve less accuracy than expected.

Both our kNN and Weighted kNN models are outperformed by Logistic Regression and Linear SVM regardless of the chosen features as expected. The reason behind of the issue is that Nearest Neighbors algorithm treats vectors as inputs which makes the method work poorly in high dimensions.

In order to improve kNN performances, we planned to apply Principal Component Analysis, short for PCA. For KNN with PCA, we found that first 3 principal component can only explain 26.17% variance, which is too low for PCA to have a good performance, and such that kNN-PCA only obtains 46.60% test accuracy. Due to the low variance on smaller dimensions, the accuracy rates of PCA algorithms weren't sufficient enough to outperform the actual kNN baseline results.

Another expected result was the 99.89% training accuracy in wkNN. Since we took k = 15 as for the number of neighbors, which controls the model flexibility, the distance parameter used as weight allowed a better learning in the training set. However, this situation led to overfit.

Support vector machine with linear kernel performed a classification with a linear decision boundary. When facing an input set of high dimensions, we came to a conclusion that we can improve our decision boundary by using a radial basis kernel in order to perform a well non-linear decision boundary. Thus, we might be able to increase both our training and test accuracy results.

As an addition to the MFCC feature set of 140 dimensions used in Table 1, combination of all the future sets created an input set of 518 dimensions. Despite the increase in dimension, all of our methods except for kNN, we achieved better results in both train and test accuracy. Therefore, with better feature and model selection, improved accuracy results can be achieved.

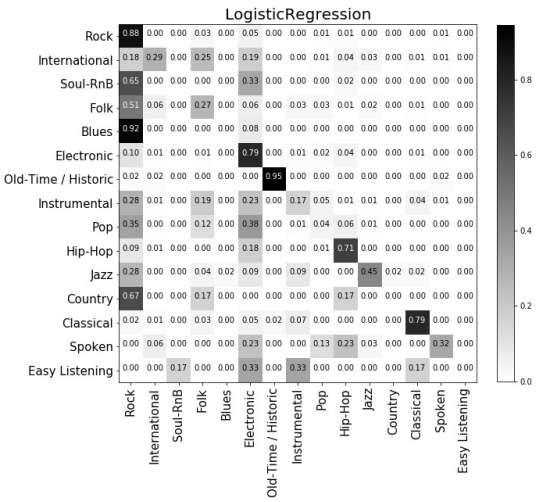

Figure 3. Logistic Regression: Confusion Matrix of All Features

Logistic Regression on the other hand outperformed all the methods we tested. Using Logistic Regression with one-vs-rest scheme is one of the reasons behind. One versus one approach, calculates the probability of each class assuming it to be positive using the logistic function. and normalize these values across all the classes.

Since we have an unbalanced data, we were expecting our normalized confusion matrix as appeared in Fig 3. As we observed the matrix, we figured out that as we are trying to preserve the percentage of the population we also miss-predicted most of the minority genres. As a future work we consider building genre-specific models for minority genres.

6. Conclusions and Future Work

In this work, we implemented and discussed various machine learning models including Nearest Neighbors Classifier, Support Vector Machines and Logistic Regression to recognize music genres using FMA dataset. Thus far, we have only selected one feature set, MFCC, among the eight others, and experimented our models with respect to the chosen feature as well as the combination of all features without further feature selection. Considering the fact that our input feature dimensions are significantly high, the baseline results are encouraging to build better models that can offer greater accuracies.

There is a number of extensions we plan to implement along the following: (i) Neural Network Implementation (ii) Balanced Data Experiment (iii) Genre specific models (iv) Baseline model updates including SVM kernels and multinomial logistic regression. We further consider applying a well model selection and feature extraction to get the optimal prediction values.

References

Bertin-mahieux, T., Ellis, D. P. W., Whitman, B., and Lamere, P. The million song dataset. In *In Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR*, 2011.

Defferrard, M., Benzi, K., Vandergheynst, P., and Bresson, X. Fma: A dataset for music analysis. In *18th International Society for Music Information Retrieval Conference*, 2017. URL https://arxiv.org/abs/1612.01840.

Haggblade, M., Hong, Y., and Kao, K. Music genre classification. 2011.

McFee, B., McVicar, M., Balke, S., Thom, C., Lostanlen, V., Raffel, C., Lee, D., Nieto, O., Battenberg, E., Ellis, D., Yamamoto, R., Moore, J., WZY, Bittner, R., Choi, K., Friesch, P., Stter, F.-R., Vollrath, M., Kumar,

S., nehz, Waloschek, S., Seth, Naktinis, R., Repetto, D., Hawthorne, C. F., Carr, C., Santos, J. F., JackieWu, Erik, and Holovaty, A. librosa/librosa: 0.6.2, August 2018. URL https://doi.org/10.5281/zenodo.1342708.

Panagakis, Y., Kotropoulos, C., and Arce, G. R. Music genre classification via sparse representations of auditory temporal modulations. *2009 17th European Signal Processing Conference*, pp. 1–5, 2009.

Sanden, C. and Zhang, J. Z. Enhancing multi-label music genre classification through ensemble techniques. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pp. 705–714. ACM, 2011.

Tzanetakis, G. and Cook, P. Musical genre classification of audio signals. volume 10, pp. 293–302, July 2002. doi: 10.1109/TSA.2002.800560.