Music Genre Classification

Ruochen Kong
ruochen.kong@emory.edu
Emory University
Atlanta, Georgia, USA

ABSTRACT

Searching music by genre is a function implemented in almost all music apps, so an accurate algorithm to classify the music is required. The existing research papers generally classified music into nearly five genres, but in the real world, much more detailed genres exist. This rough classification may further cause an unsatisfaction of the user experience. In this project, I aim to investigate the signal of the music in 19 genres, extract distinguishable features, and create a classification model. The result may form a more detailed automatic classification model for music.

KEYWORDS

Signal Processing, Data Mining, Machine Learning

1 INTRODUCTION

Major music applications, such as Spotify, Apple Music, or Gnoosic, have all allowed users to search by genre. The categorization quality, however, is somewhat unsatisfactory. In order to reduce the number of incorrectly classified music, algorithms in use tend to reduce the number of available genres. As stated by Tamatjita (2016), the more genres to be classified, the lower the accuracy [2]. This drawback is understandable because even humans will feel harder to distinguish between more genres, but pursuing a more accurate automatic classification method is still necessary. The goal of this project is to establish a method to classify musics into more genres. Several previous reaserch paper has already developed methods to classify music with a high accuracy, but few of them achieved with more than 10 different genres. In the real world, however, users may choose series of genres in order to find their favorite songs. Then, in this condition, these limited generes can not satisfy. In this project, I will develop the model on a dataset posted on Kaggle [1]. This dataset, labeled with 19 different genres, has 19.9k individuals in the training dataset and 5k in the testing dataset. To develop the model, I will first visualize several music signals in different genres to determine possible features, such as the beat, the volume, or the frequency. I will then base on common classification models, such as CNN or Random Forest, to either improve on them or form an embedded model. Finally, I will evaluate the result with commonly used metrics, including ROC, AUC, F1-score, precision, recall, and accuracy.

2 PROBLEM DEFINITION

The music data are stored as *.ogg* files which can be converted into a series of frequencies with existed packages. Because the music is binaural, when converting into arrays, they will have a dimension as $(n, 2)$ with n represents the number of recorded signals and 2 represents the left and right channel.

Figure 1 and 2 represents an example of electronic music and instrumental music. As can be seen from the figures, these two kinds of music tend to diverge significantly on the view of signals. For example, electronic music may have shorter beats, which will be extracted as a feature to be considered. Other possible features will be determinded in further analysis.

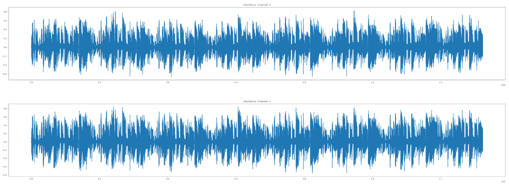

Figure 1: Genre: Electronic

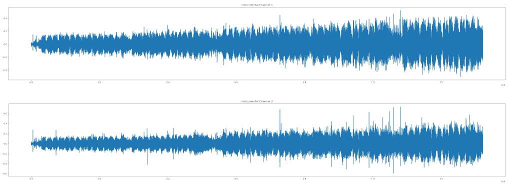

Figure 2: Genre: Instrumental

Assume the extracted features are stored as matrix $X \in \mathbb{R}^{n \times p}$ with n represents number of music and p represent number of features, and also denote $y = \{y_i : y_i \in \{0, 1, \ldots, 18\}\}$, then the goal is to create a model $\mathcal{M} : X \rightarrow \hat{y}$ with minimal distance between y and $\hat{y}$. The distance would be a combination of L1-norm and L2-norm.

3 MATERIALS AND METHODS

3.1 Dataset

The dataset is separated as $19,922$ training data and $5,078$ testing data. These music files are then transformed into arrays by the *soundfile* package in python. The labels are listed in order as electronic, rock, punk, experimental, Hip-Hop, folk, chiptune/glitch, instrumental, pop, international, ambient electronic, classical, old-time/historic, Jazz, country, Soul-RnB, spoken, Blues, and easy listening.

3.2 Evaluation Metrics

The most common ones are:

$$ACC = \frac{TP + TN}{P + N}, \qquad F_1 = \frac{2TP}{2TP + FP + FN}$$

$$precision = \frac{TP}{TP + FP}, \qquad recall = \frac{TP}{P}$$

Since this is not a binary classification task, an average value would be used.

REFERENCES

[1] Rob. 2022. Music classification. https://www.kaggle.com/c/kaggle-pog-series-s01e02/overview

[2] Elizabeth Nurmiyati Tamatjita and Aditya Wikan Mahastama. 2016. Comparison of music genre classification using Nearest Centroid Classifier and k-Nearest Neighbours. In *2016 International Conference on Information Management and Technology (ICIMTech)*. 118–123. https://doi.org/10.1109/ICIMTech.2016.7930314