Music Genre Classification and Prediction - Proposal

William Ryan
wrryan@wisc.edu

Ryan Bali
rbali@wisc.edu

Danny Miller
djmiller23@wisc.edu

1. Introduction

As music has shifted towards digital platforms, the barriers to enter the industry as an artist have all but disappeared. Nearly any person with internet access can set up a Spotify account and start recording whatever song pops into their heads. In fact, as of February 2021, Spotify released reports stating that approximately 60,000 unique songs are uploaded onto their platform every day. Additionally, both music platforms and their artists benefit when a song becomes popular as both parties see increased revenue. Given that a song is much more likely to reach peak popularity if the track finds its optimal audience, it becomes clear how precise genre classification is an essential part of the music industry. Through the use of a quick and accurate classification model a company would be able to more easily sort through and distribute their ever growing library to the correct audience, thus leading to increased exposure for both their brand and their users.

In this study we hope to develop a supervised machine learning model that will be able to properly identify a song's genre. We have identified 1000 different 30 second audio clips that can be divided into 10 distinct musical categories. Using this data we hope to utilize k-nearest neighbors to identify common trends in the audio wavelengths for each style of music. This in turn will allow us to divide future test data songs into their correct genre.

A similar study conducted by Roberto Basili, Alfredo Serafini, Armando Stellato at the Unirversity of Rome Tor Vergata [1] found that certain aspects of music did influence which genre a song would be filtered into. Some of these attributes included speed, volume and rhythm. We will make sure to take into account similar variables as we conduct our own research.

2. Motivation

People who study music theory have spent a long time trying to understand sound and what inherently makes one piece of music different than another. What characteristics of the sound of a song differentiate it from another song? How does one visualize sound, and what tones and rhythms

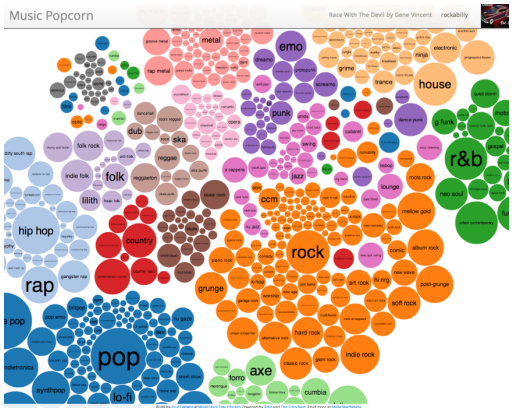

Figure 1. Example visualizing different music genres and how they connect via Music Popcorn. (Image source: https://musicmachinery.com/page/11/)

contribute to which specific genres of music? With music listeners always searching for the next best sound, it is becoming increasingly difficult today to categorize new music into different genres, and this problem is only accelerated with music that draws inspiration from multiple genres.

During the peak of the Covid-19 spread, people were forced to stay home and limit their interactions with others outside of their family. This of course limited activities that kids, teenagers, and young adults alike could do. There was a huge surge in television and movie watching, and of course in music streaming [2]. This pushed many people to explore different kinds of music and sound in the boredom of quarantine lockdown. This led many to listen to genres of music that they never had before and explore cross-related genres of music. This explosion of people widening their music taste due to social restrictions was a large motivation for this project. We want analyze whether we can quantify differences in sound and attribute these differences to specific genres.

3. Evaluation

The dataset that we will be using is the "Music Genre Classification" dataset provided by Andrada Olteanu on Kaggle [3]. It contains 100 different 30 second samples

of songs from 10 different genres in the Waveform Audio File format. Additionally, there is a dataset that provides 60 unique features such as tempo, Chroma STFT, and spectral centroid mean. The audio clips are 30 seconds each but we plan on decomposing them according to time segments obtained from the beginning, middle and end parts of the original music signal (time-decomposition). [4]

We plan on using many multinomial classification models and determining which models perform the best. Specifically, we plan on using support vector machines (SVMs), K-Nearest Neighbor (KNN), Logistic Regression (LR), and Gradient Boosting (XGB).

For our models we are using three specific evaluation metrics to determine the performance of our models.

- **Accuracy**: The accuracy simply refers to the percentage of correctly identified test samples. We will use the test/validation accuracy to get a basic idea of our models performance.

- **F-score**: We can use our confusion matrix to calculate the precision and recall and find the harmonic mean to compute the F-score evaluation metric.

$$F_1 = 2\frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} = \frac{\text{TP}}{\text{TP} + \frac{1}{2}(\text{FP} + \text{FN})}$$

- **AUC** In order to determine if our models can accurately distinguish between classes we will use AUC. We expect to achieve a AUC of higher than 0.5.

We are planning to build a baseline model with K-Nearest Neighbors without conducting any feature engineering. Then based on the performance of the baseline model, we will attempt to achieve better performance through feature engineering techniques.

Our goals for this project are to understand and communicate the underlying aspects of music that cause it to be classified into a specific genre. We would like to do this through feature importance techniques and visualizations. We also would like to create a model that is able to reliably classify music according to their respective genre. We will determine the success of our model according to the evaluation metrics listed above.

4. Resources

The Kaggle dataset [3] obtained for the musical genre analysis will consist of the following data fields and files:

- Genres Original: A collection of 10 genres with 100 audio files each, all having a length of 30 seconds

- Images Original: A visual representation for each audio file. One way to classify data is through Convolutional Neural Networks (CNN) because they usually take in some sort of image representation. If we choose to utilize this method, the audio files were converted to Mel Spectrograms to make this possible

- 2 CSV files: Containing features of the audio files. The first file has a mean and variance computed over multiple features that can be extracted from an audio file for each song (30 seconds long). The second file has the same structure, but the each song was split into 3-second audio files, thus increasing the amount of data we input into our classification models tenfold

Python will be the primary language used for modeling, feature engineering, and visualization. We will be utilizing the pandas and scikit-learn for our data handling and modeling procedures. We will also be using matplotlib and seaborn for data visualizations. Since many of us have experience in R we also may use the ggplot2 library for additional data visualizations.

Since we are dealing primarily with data in the Waveform Audio File format, we will be taking advantage of Python's librosa library to handle the music and audio analysis. Combined with matplotlib this will be particularly helpful for performing visualizations of audio data.

5. Contributions

As a group, it is decided that we are all taking responsibility to evenly divide computational and writing tasks. For this proposal, William worked on the evaluation portion and figures, Ryan worked on the motivation and resources portions, and Danny worked on the introduction and references portions of this proposal. All three of us worked on the formatting of the overall proposal and we are all responsible to proofread through all reports and proposals before final submission. For the computational tasks, we will schedule meetings in College Library to code and work together throughout the semester and set up group deadlines for each phase in the project to successfully generate a model that most accurately classifies and predicts the genre of different songs.

References

[1] R. Basili, A. Serafini, and A. Stellato. Classification of musical genre: a machine learning approach. In *ISMIR*, 2004.

[2] S. Hall. This is how covid-19 is affecting the music industry. `https://www.weforum.org/agenda/2020/05/this-is-how-covid-19-is-affecting-the-music-industry/` May 2020.

[3] A. Olteanu. Gtzan dataset - music genre classification. `https://www.kaggle.com/andradaolteanu/`

gtzan-dataset-music-genre-classification, Mar 2020.

[4] C. N. Silla, A. L. Koerich, and C. A. Kaestner. A machine learning approach to automatic music genre classification. *Journal of the Brazilian Computer Society*, 14(3):7–18, 2008.