Music Genre Classification Using Machine Learning

McKinley Morris, Alex Pruitt, Jason Liu, Vaishak Menon

October 4, 2024

Project Overview

This project aims to investigate the problem of automatically classifying music tracks into specific genres using machine learning techniques. This is an interesting and relevant problem because music recommendation systems, streaming platforms, and music categorization tools all benefit from accurate genre classification. Improving the performance of genre classification can enhance user experience and allow for more precise music suggestions based on listening habits.

Challenges

We anticipate several key challenges for this project:

- **Audio Features:** Extracting relevant and meaningful audio features (e.g., pitch, rhythm, and timbre) that vary significantly across genres is complex and requires sophisticated feature extraction techniques.
- **Dataset Size and Diversity:** The *GTZAN dataset* is commonly used but is relatively small and limited in its diversity. We plan to supplement it with data from the *Free Music Archive (FMA)* and *MuseScore*, but this will introduce additional challenges around ensuring consistent genre labels, feature extraction, and avoiding overlap between tracks from different datasets.

- **Model Performance:** Choosing machine learning models that can effectively classify genres with varying audio characteristics while also being computationally efficient will be a challenge, particularly when combining multiple datasets.

Dataset

We will use the *GTZAN Music Genre Dataset*, which contains 1,000 audio tracks evenly distributed across 10 genres. To supplement this dataset and increase its diversity, we will incorporate:

- The *FMA Small Dataset*, which consists of 8,000 tracks across 8 genres, several of which overlap with GTZAN. This will provide us with additional training data and help address the limitation of GTZAN's small size.
- *MuseScore*, where we will manually extract and convert *MIDI files* or rendered audio tracks into usable data, focusing on expanding genres like *Classical* and *Jazz*.

This combined dataset will enhance the model's ability to classify a wider variety of musical styles.

Methods and Algorithms

We will extract key audio features such as *Mel-Frequency Cepstral Coefficients (MFCCs)*, *chroma features*, and *tempo* using the *Librosa* library. These features will serve as the input for training the following machine learning models:

- *k-Nearest Neighbors (k-NN)*, which will classify tracks by comparing them to their closest neighbors in the feature space.
- *Support Vector Machines (SVM)*, which will use hyperplane separation to classify genres with more complex boundaries.
- *Random Forests*, which will aggregate decision trees to improve classification performance and avoid overfitting.

We will explore existing implementations of these algorithms but also experiment with tuning hyperparameters and optimizing the models to handle the combined datasets effectively. We aim to improve model performance by refining feature selection and handling any dataset imbalances caused by the integration of FMA and MuseScore data.

Background Reading

To guide our methodology and approach, we will examine:

- Tutorials and articles on audio feature extraction using *Librosa*.
- Research papers on machine learning models for music genre classification, particularly those from sources like *arXiv* and *IEEE*, to provide context on existing work in this area.

Evaluation Plan

We will evaluate the performance of our models using the following metrics:

- *Accuracy*, *Precision*, *Recall*, and *F1-score* will be the primary quantitative metrics used to evaluate classification performance across genres.
- *Confusion matrices* will provide qualitative insight into which genres are being confused by the models.
- We will generate plots of training and testing accuracy to monitor model performance over time and use *cross-validation* to ensure generalization and avoid overfitting.

We expect that supplementing the *GTZAN dataset* with tracks from *FMA* and *MuseScore* will result in improved robustness and generalizability of our models. We will take measures, such as employing audio fingerprinting, to avoid duplicate tracks across datasets and ensure the uniqueness of each data point in the training set.