FMA Project Proposal: Genre Classification and Popularity Regression

Daniel Amin, Kiwook Kwon

Data Science Program
The Graduate Center: CUNY
365 Fifth Avenue. New York, NY 10016
damin@gradcenter.cuny.edu,
kkwon@gradcenter.cuny.edu

Kuan Chen, Stephanie Luk

Data Science and Engineering Program
The City College of New York: CUNY
160 Convent Avenue. New York, NY 10031
kchen009@citymail.cuny.edu,
sluk000@citymail.cuny.edu

Motivation

The music industry is one of the biggest industries existing today, ranging from music streaming platforms such as Spotify, Youtube Music, etc. to record labels, music distributors, advertisers, and so on. Most of the time these corporations receive thousands of songs per day to be reviewed for business. Spotify and other streaming platforms need genre classification in order to build their playlists and recommendations; record labels, distributors and advertisers receive demos and need genre classification as well to know how to properly market the music. In some cases, record labels and advertisers also scout for music with potential to be signed and marketed, in this case, usually done manually by humans. The Music Information Retrieval (MIR) is also a growing field of research with many real-world applications. With the massive amounts of music in the world there is a huge push to be able to browse, search, and organize large music collections efficiently.

A music genre is a conventional category that identifies some pieces of music as belonging to a shared tradition or set of conventions [1]. Music can be divided into different genres in many different ways. The artistic nature of music means that these classifications can and are often subjective and controversial, and some genres may overlap.

This project will be an application-oriented project with two potential tasks. One will be to see how well we can classify the genre of a song given the audio file. The second task will be a prediction of the popularity of a song using a regression based on the audio files. The motivation of these tasks is to improve the automation of these processes in the music industry.

Method and Intended Experiments

The Free Music Archive (FMA) is an interactive library of high-quality, legal audio downloads which was introduced in 2017 and launched in 2019[1] with the purpose of offering free access to new music to the public [2]. We will use the dataset of the UCI Machine Learning Repository[2]. The dataset is a large collection of public copyright licensed audio files. It has four sizes available:

1. fma_small.zip: 8,000 tracks of 30s, 8 balanced genres (GTZAN-like) (7.2 GiB)
2. fma_medium.zip: 25,000 tracks of 30s, 16 unbalanced genres (22 GiB)
3. fma_large.zip: 106,574 tracks of 30s, 161 unbalanced genres (93 GiB)
4. fma_full.zip: 106,574 untrimmed tracks, 161 unbalanced genres (879 GiB)

[1] https://www.freemusicarchive.org/
[2] http://archive.ics.uci.edu/ml/datasets/FMA%3A+A+Dataset+For+Music+Analysis

We plan to use the fma_small data. The dataset includes metadata and features for all audio tracks. We will first conduct an exploratory data analysis (EDA) to explore the relationships of the included audio features. We will look into and get a general understanding of the audio features: chroma, tonnetz, mel frequency cepstral coefficient (MFCC), spectral centroid, spectral bandwidth, spectral contrast, spectral rolloff, root mean square energy, and zero-crossing rate. The audio files will be converted into spectrograms, a visual representation of spectrum of frequencies over time, as features for modeling purposes. For the genre classification task, the spectrograms will be fed into a Convolutional Neural Network (CNN). We may utilize a Short-term Fourier Transform (STFT) spectrogram of musical signals [3] as the input of our network to get a baseline before implementing a CNN. Some previous works suggest using a convolutional recurrent neural network (CRNN) for music tagging [4]. For the popularity regression task, we intend to try a CNN and Recurrent Convolutional Neural Network (RNN). We will use different fine-tuning techniques such as truncating the last layer (softmax layer), adjusting learning rates, batch size, etc., to determine whether it could improve the performance of the models.

References:

[1] J. Samson. "Genre". URL: https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000040599

[2] M. Defferrard, K. Benzi, P. Vandergheynst, and X. Bresson. "FMA: a Dataset for Music Analysis". URL: https://arxiv.org/pdf/1612.01840.pdf

[3] K. Choi, G. Fazekas, M. Sandler, and K. Cho. "Convolutional Recurrent Neural Networks for Music Classification". URL: https://arxiv.org/pdf/1609.04243.pdf

[4] L. Feng, S. Liu, J. Yao. "Music Genre Classification with Paralleling Recurrent Convolutional Neural Network". URL: https://arxiv.org/pdf/1712.08370.pdf