“Good Composers Borrow, Great Ones Steal!” Executive Summary

Reggie Bain Emelie Curl Larsen Linov Tong Shan Glenn Young

Introduction and Objective

Throughout history, composers and musicians have borrowed musical elements like chord progressions, rhythms, lyrics, and melodies from each other. Our motivation for this project is born of a fascination with this phenomenon, which comes in various forms like covers and sampling and, of course, extends to less legal examples like unconsciously or intentionally copying the work of another. Even famed and highly regarded composers like Johann Sebastian Bach, Antonio Lucio Vivaldi, Wolfgang Amadeus Mozart, and Franz Joseph Haydn are not innocent of borrowing from their contemporaries, predecessors or even recycling their own works. More modern examples include John Williams cutting and pasting passages from Igor Fyodorovich Stravinsky’s ballet *The Rite of Spring* to complete his score for *Star Wars: A New Hope*. Similarly, in 2015, in a high-profile court case, defendants and artists Robin Thicke and Pharrell Williams were ordered to pay millions of dollars in damages for copyright infringement to Marvin Gaye's estate as well as give songwriting credit to Gaye, considering they borrowed from Gaye’s "Got to Give it Up" when writing their hit "Blurred Lines." Our project aimed to use deep learning to assess the similarity between musical clips to potentially establish a more robust and empirical way to detect music plagiarism.

Stakeholders

- Artists looking to ensure others do not plagiarize their work;
- Record companies and courts looking to have an objective measure of song similarity; and
- Companies looking to better classify and recommend wider arrays of music to listeners.

KPIs

- The model minimizes the “triplet-loss” function between a triplet of (anchor, similar, different) audio files to the degree that it will differentiate between similar and different audio files.
- The model beats a baseline of calculating the triplet-loss between audio files without feeding those audio files into the model. We'll call this the "no-model" baseline.
- Study the nature of famous plagiarism cases and how these songs may be related in a way detectable by various metrics typically used in audio analysis.

Methodology

Our data primarily came from the Million Song Dataset (MSD) and several Kaggle datasets where authors paired known songs with clips or full recordings accessed through various APIs. In some cases, we wrote scripts to obtain song previews that we could not otherwise acquire. Due to restrictions on file size, storage, and time, we ultimately focused on 10s-30s clips of songs from a 50K song subset of the MSD available on Kaggle. Using audio analysis techniques such as calculating log-mel spectrograms, audio augmentation, and transfer learning, we established a way to compare the similarity between audio files to detect potential plagiarism. During training, we focused on minimizing the “triplet-loss,” which aims to maximize the distance between different embeddings while simultaneously minimizing the distance between similar embeddings. These embeddings are referred to as the anchor, positive, and negative. We explored a wide array of models, including building a CNN from scratch and fine-tuning various pre-trained models such as ResNet-18 and transformer models such as Audio Spectrogram Transformer.

Results

We developed a model to analyze the similarity between songs by directly analyzing the audio. The best-performing model was the ResNet-18 fine-tuned on a dataset of 10K triplets (anchor, positive, negative) of songs. It beat the baseline by about 84.83%, doing a much better job of separating the song embeddings in the 128-dimensional embeddings space. Positive songs were generated using augmentations of the anchor songs, and lower layers of the ResNet were frozen, leaving only the last residual block and fully connected layer. We also used our best-performing model to make a Streamlit demo showcasing similarity scores between predetermined songs. It allows you to upload your own audio to test the model as well. Ultimately, we feel that we achieved our KPIs and created a model that shows promise going forward.

Limitations and Future Directions

We faced several key roadblocks, particularly a lack of computing power and storage. Audio data sets are complex and extremely large, which limits us to training smaller datasets for fewer epochs. We deployed our model on a different set of songs and covers of those songs with limited success, but the model showed promise for having such limited data and training. In addition to increased computing power/storage, improving data quality is a top priority. Other priorities include aggregating different recordings of the same song to create more difficult triplets for the model to learn, studying embeddings using different audio processing techniques, and experimenting with additional model architectures. We also want to train a plagiarism classifier head on top of our existing embeddings model.