Multi-Feature Song Cover Detection for Dynamic Playlist Generation

Sahil Ahuja SAHILAHU@SEAS.UPENN.EDU
Daniel Moreno MORENOD@SEAS.UPENN.EDU
Helen Qu HELENQU@SEAS.UPENN.EDU

Abstract

We outline a pipeline to detect if a given music track is a "cover version" of a given reference track. We accomplish this by extracting features from each track's chromagram and learning both a SVM and neural network classifier to detect cover songs by cross-correlating these features. We then run a genre classifier on the test track in order to generate a playlist of different covers songs of the same genre. Our classifiers have an accuracy of **X%** on the covers80 database and **X%** on the SecondHandSongs database. (Accuracy will be added in final report)

1. Introduction

A cover song is a reinterpretation of a musical piece with a different instrumentation, musical character, genre or other different features. Covers are unique in the sense that while key elements of the original composition prevail, the overall experience of listening to a cover is completely different than listening to the original version.

The impact of automatic song cover detection is wide-reaching. At the consumer level, finding covers for songs is useful because different covers may cater to individual tastes better than the original tracks. Cover song detection is also a building block in tackling larger problems in the field such as the automatic organization of vast music libraries and the generation of relevant, personalized music recommendations. More broadly, an accurate song cover detection algorithm can provide insight into how humans encode information about a musical track.

While humans can easily detect which elements of the original composition remain and which have been altered, it is a far more interesting problem to train a classifier to accomplish the same task. To do so reliably, we need a method for representing music tracks that is robust to changes in tempo, instrumentation and general musical style.

1.1. Extracting Features from Tracks

In order to analyze and extract features from the music tracks, we opted for a beat-synchronous chroma representation. A chroma representation relates the twelve pitch classes {C, C#, D, D#, E, F, F#, G, G#, A, A#, B#} across all octaves in order to capture the harmonic and melodic characteristics of a piece of music. This representation is robust to changes in timbre and instrumentation. We use existing libraries that employ short-time Fourier transforms in order to determine a tracks' chromagram, or mapping of its chroma features.

By specifically determining the beat-synchronous chroma features, we are able to also account for variations in tempo between the reference and the test tracks. This robustness makes beat-synchronous chroma representation a great candidate for encoding tracks with the purpose of song cover detection.

1.2. Dynamic Playlist Generation

After determining that a given reference/test pair is an original/cover pair, we want to take the system a step further and categorize the cover based on its genre. Genre classification of covers can help us realize one of the many potential applications of this technology – personalized music recommendations and self-organizing music libraries along various dimensions. Given the genre classification of the cover, we can add it to a running playlist of covers of that genre. Different kinds of covers serve different kinds of tastes, and we want to be able to deliver relevant covers to interested users in a clean, dynamic manner. For example, country music lovers can access an ever-growing playlist of country-style covers and can expand their musical repertoire from just original country songs.

University of Pennsylvania, CIS 419/519 Course Project.

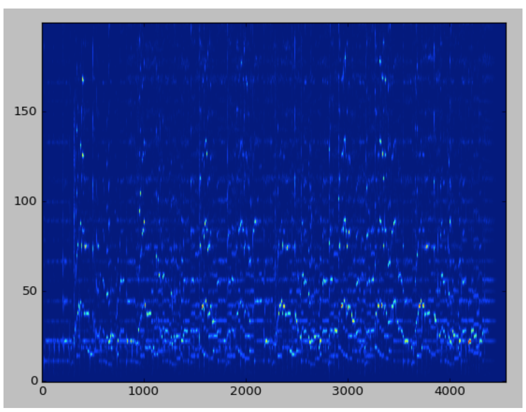
2. Overview of Our Method

2.1. Libraries and Datasets

We are training and testing our SVM on the train and test sets of the SecondHandSongs, a cover songs dataset compiled and maintained by LabROSA at Columbia University. We are evaluating the comparative performance of our cover song detection algorithm against existing algorithms by running it on the covers80 dataset, a collection of 80 original/cover song pairs and a standard benchmark for performance on this task. Currently, the system with best performance, Hydra, has achieved approximately an 84% accuracy on covers80. Both of these datasets also will be useful for the genre detection mechanism. We also plan to use madmom, a Python audio signal processing library specially tailored for music information retrieval tasks, to help us generate our beat-synchronous chromagrams.

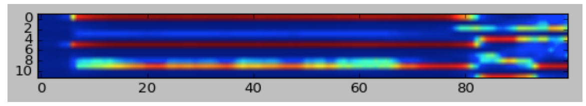
2.2. Feature Detection / Feature Normalization

In order to classify each song, we have created a generic $load\_script.py$ file which takes as input a song, in any format such as $.mp3$ or $.wav$ and creates and outputs its spectrograph and chroma. The spectrogram is created by the madmom library. The following example of a spectrogram was created for the song Yesterday, by the Beatles:



The spectrogram is a visual representation of frequencies over time, and these intensities of the sound frequencies are given as a numpy matrix, so they will be used as input for the SVM.

Secondly, and probably more useful going forward according to literature on cover detection, is chroma analysis. Chroma analysis dumps the song's frequency into 12 buckets according to the 12 semitones, therefore ignoring other qualities of the song that vary between covers. We also use the madmom library to calculate chroma and it is returned to us as a numpy matrix with time on one axis and the 12 buckets on the other. The following is a pictoral representation of the first 100 time seconds of the same Yesterday track:



Going forward, we will use these methods in conjunction with SVM and neural network analysis to detect the covers.

2.3. Genre Classification

We plan to employ a similar system to genre classification – extracting relevant musical features and training a classifier to categorize an audio signals genre based on differentiation in those features. We choose the features to analyze based on previous work by Tzanetakis et al.s paper on genre classification. They advise focusing on two sets of features – musical surface features, including measures of spectral brightness, shape, and change; and rhythm features, calculated by processing the audio signal with a Discrete Wavelet Transform. Differences in this combination of features are most indicative of genre.

References

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.375.1310&rep=rep1&type=pdf

https://www.ee.columbia.edu/~dpwe/pubs/EllisP07-coversongs.pdf

http://digitalmusics.dartmouth.edu/~mcasey/bregman/

http://www.ee.columbia.edu/~dpwe/talks/coversongs-2007-04+audio.pdf

http://madmom.readthedocs.io/

http://labrosa.ee.columbia.edu/millionsong/secondhand

http://labrosa.ee.columbia.edu/projects/coversongs/covers80/

http://ismir2001.ismir.net/pdf/tzanetakis.pdf