ECE 225A: Probability and Statistics in Data Science using Python

Final Project Plan

Pop Music Chord Progression Analysis

Hao-Wen Dong
A53309488
hwdong@ucsd.edu

Chun-Jhen Lai
A53310726
c8lai@eng.ucsd.edu

We plan to analyze the statistical properties of chords and chord progressions in over 5000 pop songs available on the HookTheory platform.[1] In particular, we are interested in analyzing the prior probabilities for different chords (i.e., the probability that a certain chord presents), the transition probabilities between chords and chords (i.e., the probability that a certain chord A is followed by a certain chord B) and the most frequent chord progressions.

Data—HookTheory is a community-based platform where users can upload the lead sheets (i.e., melody and chord progressions) for different songs. In particular, we will use the version compiled and provided on a GitHub repository.[2] Specifically, this dataset contains 18986 music segments from 11380 songs of 4956 artists. Genre tags are also provided.

Method—First, we will extract the chord progression from the lead sheet of each music segment. Specifically, we will consider only the sequence of chords and discard the duration information for each chord. Second, we will then analyze the above-mentioned statistical properties, including prior probabilities, transition probabilities and most frequent chord progressions. Moreover, as genre tags are available in this dataset, we are also interested in how these statistics differ from genre to genre with an eye to reveal some interesting trends on the usage of chords and chord progressions in different genres. Finally, if time permits, we would like to apply the n-gram model on this dataset and build a predictive model for chords. This can be useful for intelligent chord recommendation in music composition software.

[1] https://www.hooktheory.com/

[2] https://github.com/wayne391/lead-sheet-dataset