

10/10/2022

MUSI 8803

Computational Musicology

Kelian (Mike) Li, Xuedan (Annie) Gao, Shimiao Liu

Midterm Proposal

Introduction

problem/question

The project will investigate the relationship between Chinese/Han traditional music and Chinese pop music. In particular, the project will compare the melody characteristics between a han folk music dataset and a Chinese pop music dataset. The goal is to answer whether the melodies of Chinese pop music are influenced by the melodies of Han traditional music.

motivation

China has a long music history and unique music styles. However, many argue that commercial popular music in China today is no longer rooted in Chinese/Han traditional music. Functional harmony was not established in Chinese/Han traditional music, so it has to be borrowed from Western music. However, the question of whether the melodies of Chinese popular music are rooted in Chinese/Han traditional music remains unanswered.

hypothesis (or hypotheses)

The distribution of melodic features in Chinese popular music has some similarities to that in Chinese traditional music.

We define the melodic features to be the interval distribution of a song, the note density of a song, and the rhythmic patterns of a song.

Corpus description

Traditional Chinese music is vaguely defined as the music of Chinese people except for popular music or in the style of other regions. In practice, people consider Chinese

music composed by 1840 (or 1912) to be traditional Chinese music, plus the music composed later than 1840 (or 1912) but in the same style.

The availability of traditional Chinese/Han music datasets is more limited. Many existing dataset focuses on the instrumental music of a single instrument. The project will use the Han songs in the Essen Folksong Collection. This subcollection contains 1225 songs in Humdrum format. This dataset is verified by the research community, and it is considered as traditional Chinese/Han music. We will use convenience sampling and utilize the entire dataset.

The beginning of Chinese popular music was established by both Western and Japanese colonial power and nationalist composers. The later development diverged because of political divisions, but the cultural communication channels had always existed between Mainland China (PRC), Taiwan (ROC), Western colonies (Hong Kong and Macau), and overseas Chinese. Since the reform and open-up of Mainland China, cultural communication has generally been uninterrupted. Thus, this project considers Chinese pop music as a whole despite the regions of its creation. For this project, Chinese popular music is defined as songs created by Chinese in style considered as Chinese popular music by the majority of Chinese music listeners.

The dataset for Chinese popular music is called the [POP909](#) dataset. It contains multiple versions of the piano arrangements of 909 Chinese popular songs created by professional musicians. The main body of the dataset contains the vocal melody, the lead instrument melody, and the piano accompaniment for each song in MIDI format. The project only focuses on the vocal melody. This dataset is verified by the research community, and it is considered as Chinese popular music. We will use convenience sampling and utilize the entire dataset.

Although the two datasets are professionally collected and recognized by the research community, there could be issues for this particular project. First of all, the data format needs to be converted. There are tools to convert between humdrum format and midi format, but lossless conversion is not guaranteed. Secondly, the POP909 dataset was originally collected for the purpose of music generation. There may be unknown problems to be used for music analysis. Thirdly, because convenience sampling is applied to both datasets due to the total data size, there may be disproportionality within the datasets unadjusted. Fourthly, because the definitions of Chinese popular music and traditional Chinese/Han music are essentially based on the overall opinions of Chinese people, some songs can be included in both categories.

Analysis plan

What is/are the musical parameter(s) that need(s) to be analyzed/searched/modeled? (E.g., rhythm patterns, chord changes, melody). Given that, how might you need to pre-format your data? Will you analyze your feature in counts? Proportions? Or as a function of time/duration? (E.g., counting pitch classes largely ignores duration unless you “timebase” or “slice” them first).

These Parameters we will use are interval, note density, and rhythm patterns. We're going to delete the instrumental part of the dataset, keep the melody track and convert it from MIDI to Humdrum format. We then use the command line and humdrum functions to calculate the proportion of each parameter (regardless of duration) and compare the proportion of these parameters in the two datasets. We plan to analyze the data in 4-bar units.

We plan to analyze the proportions of each parameter. By doing this will help us to compare the similarity of the melodic features between Chinese popular and traditional music.

If your hypothesis requires operationalization that impacts how you do the analysis(e.g., “...more repetitive than...”) you must include your operationalization in this analysis plan section (e.g., “we will examine how frequently 3-, 4-, and 5-note patterns recur in each song using a window that slides every 1 note.”)

We can use the humdrum command to calculate the distribution of intervals, note density, and rhythm patterns in the two datasets to judge whether traditional Chinese music and modern pop music have similarities in melodic characteristics. For example: in these two datasets, we will examine how frequently different intervals(e.g. pure fifth, major third, etc.) occur in each category using a window that slides every four bars; how the note density (sparse or dense) distribute in each songs; what are the most common rhythmic patterns in Chinese popular/traditional music, such as quarter notes, triplet, etc..

computational musicology tools which (if any) is likely best suited to your analytical problem? Thinking about the “shape” and complexity of your expected data may help here.

The computational musicology tools will be Music21 or the Humdrum toolkit depending on the information loss of humdrum-to-midi conversion and midi-to-humdrum conversion.

statistical test

Do you have a linear/continuous question (e.g., is something changing over time? Changing over frequency/pitch height?), or do you have things that fall into categories (e.g., composer X uses melodic riff 'a' more than composer Y) and you wish to examine differences between the categories? If the latter, how many categories?

We have things that fall into categories and we wish to examine differences between the categories. There are two categories: Chinese popular music and Chinese traditional Han music. We have three dimensions, interval, note density, and rhythmic pattern. We plan to analyze the similarity of these three features between Chinese popular music and traditional han music.

Is the data you are collecting continuous or categorical? (E.g., pitch height in semitones would be considered continuous, whereas chord labels would be categorical).

Interval and note density are continuous. Rhythmic pattern is categorical. We will analyze the data by selecting 4 bars as a unit.

Timeline

Write up a timeline (with deadline dates!) for your full analysis plan from midterm through final presentation date. Preference is for a table-style with overlapping boxes to help illustrate parallel tasks. It should include (these items may overlap):

- Corpus pre-processing and cleaning (max 2 weeks) – what data do you need? What format do you need it in? What data needs to be extracted or added?

What data:

Deadline date: 10/10/2022

Melodies of Chinese popular music and melodies of traditional Chinese/Han music.

- Han Chinese folksong from The Essen folksong collection dataset.
- POP909 dataset.

What format:

Convert midi to

humdrum. <https://www.humdrum.org/Humdrum/commands/midi.html>

What to extract:

Extract the melody by manually deleting instrumental midi tracks.

- Reading in the data and unifying the corpus if necessary
Deadline date: 10/17/2022
- Coding the analysis (include as many events/steps as possible)(min 2 weeks)
Deadline date: 10/31/2022
- Some initial/basic graphing of analytical outcomes
Deadline date: 11/07/2022
- Setting up data for testing
Deadline date: 11/14/2022
- Performing statistical tests
Deadline date: 11/14/2022
- Final graphs / tweaking / cleaning-up
11/21/2022
- Writing final paper
11/28/22
- Final presentation (final exam slot)
12/05/22