Analysis of Popular Song Lyrics across Genres

COLX523 Group 8 Project Proposal

By Meiyu, Aswin, Rakesh, and Ziggy

Goal

Build a corpus of lyrics from different musical genres and analyze the themes, the patterns of language usage and linguistics features from the most popular songs (top charting song) within those genres over the past decades.

Source of data

- Top 100 charts by genre and year, collecting song titles and artists and lyrics
  https://en.wikipedia.org/wiki/Billboard_Year-End_Hot_100_singles_of_1980, etc.
- Spotify API to collect music features
  https://developer.spotify.com/documentation/web-api
- Genius.com to search and collect lyrics
  https://genius.com

Features

- Genre
  e.g: Pop, Rock, Hip Hop, R&B, Country, Latin Pop, etc.
- Years
  e.g: decades from 1970s to 2020s
- Language - English
- Music features
  e.g: duration, danceability, acousticness, etc.
- Lyric features
  e.g: readability, coherence, number of named entities, keywords (most common words minus stopwords)

Data Format

The corpus will be stored as a JSON file, with one document per song containing all features. There are approximately 200 words per document, and nearly limitless amount of lyric data available, which is about or easily over the brown-sized corpus.

Primary Annotation

- Genres
- Named Entity Tagging

Product

We will develop a web app which allows for searching through the corpus by keyword, with a given set of keywords which return results.

Stretch Goals

- Adding a feature to generate a wordcloud from corpus keywords, with the ability to filter data by decade or genre.
- Building an index of named entities in songs.
- Adding a generative model using Markov Chains or GPT3 to generate new songs in the style of a particular genre and decade.
- The ability to search for artists using the Spotify API to get features from their most popular lyrics.
- The ability to generate songs in the style of a particular artist using a mix of the generative model and Spotify API search.
- Adding a feature to search a song on Spotify and get its genre alignment/predictions for both sound and lyrics

Hypotheses

- Different genres will contain different numbers of named entities. *For example, Hip Hop lyrics are more likely to reference prominent figures than Pop lyrics.*
- Different genres of lyrics may have distinct writing styles. *For example, rap songs tend to follow certain rhyme patterns whereas rock music may avoid rhyme intentionally, certain genres of music like to tell stories etc.*

Further Evaluation

- Creating a fun application for generating new songs
- Analysing trends in lyrics over time
- Identifiying key themes in music in a particular decade
- Identifying rhyme schemes and/or predominant stylistics in different genres
- Finding stylistic trends in music using Spotify's music features
- Analysing correlation between musical and lyrical style (i.e: valence of the sound vs valence of the lyrics)