Automatic Playlist Generation
What makes a good playlist?

Problem Statement

Music recommender systems (MRS) have recently exploded in popularity thanks to music streaming services like Spotify, Pandora and Apple Music. By some accounts, almost half of all current music consumption is by the way of these services. While recommender systems have been around for quite some time and are very well researched, music recommender systems differ from their more common siblings in some characteristically important ways: the duration of the items is less (3-5 min for a song vs 90 minutes for a movie or months/years for a book or shopping item), the size of the catalog of items is larger (10s of millions of songs), the items are consumed in sequence with multiple items consumed in a session, repeated recommendations have a different significance since listening to the same song as part of different playlists may be ok, and consumption occurs passively i.e. in the background. Music Recommender Systems then require different approaches from traditional recommender systems.

One of the major problems in Music Recommender Systems is the station/playlist generation problem. At its heart, the playlist generation is about finding the set of songs to recommend to best extend the experience of a listener in the midst of a playlist. By suggesting appropriate songs to add to a playlist, a Recommender System can increase user engagement by making playlist creation easier, as well as extending listening beyond the end of existing playlists.

Data Resources Students will focus on the million playlist dataset, but in their exploration might use other publicly available playlist data, individual song data and/or generate data for contextual information.

  - Large (5.4 GB) Playlist dataset from spotify
  - Created in 2018
- Million Song Dataset – https://labrosa.ee.columbia.edu/millionsong/lastfm
  - Song Database (plus similar artists/songs) from Last.fm
  - Generated in 2010/2011 (but you can use last.fm to regenerate)
  - Could be useful for song based contextual metadata
  - Database of Lyrics
  - May come in handy for NLP/Textual based metadata
- Generated Contextual information via Student Team UI
Questions to ascertain playlist listener emotional state
Questions to determine the “purpose” of the playlist e.g. romantic, study, etc.

Project Goals

Automatic Playlist Generation — create a model for song discovery on the basis of the base playlist and user/context information that might be important to quality of a playlist. Some of user/context information might include intent, emotion, location, playlist “purpose” (driving/road trip, studying, etc). Use the developed model(s) for automatic playlist generation.

Cold Start Problem — A variant of the problem of automatic playlist generation, the cold start problem involves creating a model to find good choices of songs for new playlists with relatively few prior playlist entries.

Psychological Factors in RecSys — Create tools (e.g. student team built UI for playlist data collection) to gather data and construct models to evaluate how psychological factors and listener state can influence (and improve) Music Recommender Systems.

Opportunity

Students have the opportunity to submit their work as part of the ACM 2018 RecSys Challenge (https://recsys.acm.org/recsys18/challenge). If their contribution is accepted, students will get to present at the RecSys Challenge 2018 Workshop at the 12th ACM Conference on Recommender Systems in Vancouver in October 2018.