Musical User Preference Modelling

Alex Condotti, Prateek Gaddigoudar, Vamsi Ghorakavi

Carnegie Mellon University

Electrical & Computer ENGINEERING

Introduction

Problem

- Current music recommendations take into account music labels such as genre and artists (feature-independent).
- Streaming services like spotify gather audio features from the music they provide.
- We want to only use audio features along with previously listened to songs in order to recommend the next song.

Motivation

We would like to understand why a person likes a song. People may like songs based on the artist, genre, popularity, or just from the music itself. We would like to quantify how much a person likes the music based on how it sounds, and use this knowledge to recommend similar songs that a user will enjoy.

Related Work

- The bulk of our project is based on [1] Content-Filtering, which discusses several models and APIs to provide data about music preferences and features in the track They use a Hybrid Gaussian Mixture Model to recommend songs based on user preferences and audio features.
- We also examined current techniques for music recommendations, as described in [2] Collaborative filtering, which details methods used by Spotify for content suggestion.

Dataset

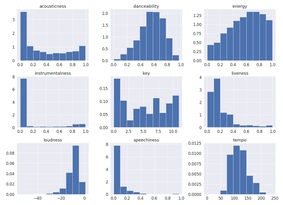
We used last.fm and Spotify API:

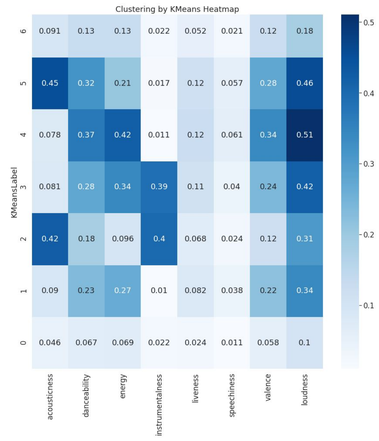
- last.fm LFM-1b dataset[3] contains user preference lists for tracks, genres and artists. It was sampled between 2013 and 2014. So the tracks available were not current. The bulk of our project is based on [1] .
- Spotify API [4] was used to generate audio features as danceability, energy, acousticness, liveness, valence etc for selected input tracks.

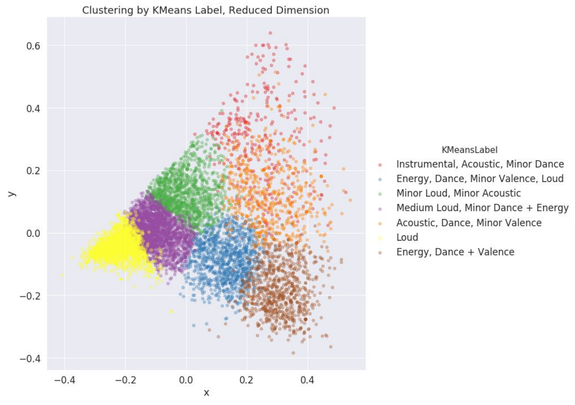
Feature Selection and Modeling

Feature Engineering

In order to choose the features for our models, we first scaled them to normalize the distances. Then we viewed the distributions of all the features on the tracks. We selected features that gave optimal clusters.







Modeling

Clusters

- Optimal number of clusters varies greatly between users based on the songs they prefer.
- We found the most reliable way to determine the number of clusters was to use the gap statistic.

Results

- In order to determine how our model performed, we looked at the distance from the cluster to the center of the cluster. The average of the distances was 0.087.
- In addition, the average number of clusters that we found per user was 12.34
- A better way to evaluate the model would be to create an application, and determine if a user likes the songs.

Future Work

- Combine feature based model with genre and artist to make a better classifier.
- Test with users to determine if feature based recommendations work satisfactorily.

References

1 Zangerle E, Pichl M. (2018) “Content - Based User Models: Modeling the many Faces of Musical Preference.” Proceedings of 19th ISMIR Conference. Paris, France.

2 Savar B, Karypis G, Konstan J, Riedl J. (2001) “Item - Based Collaborative Filtering Recommendation Algorithms.” www10. Hong Kong.

3 http://www.cp.jku.at/datasets/LFM-1b/

4 https://api.spotify.com