matsubara (EXTENDED ABSTRACT)

Alexander Stradnic

Final Year Project on the Use of Sequential Ordering for Creating Playlists

University College Cork

Abstract In this Final Year Project I present the use of Case-based Recommendation in creating music playlists, along with providing an implementation of this CBR system. Playlists are sequences of songs arranged in a particular order. Using this information, context can be gained from previous playlists in order to build a new playlist from a given seed song or starting list. This project is based on the below papers [1, 2].

1   Introduction

My project is on creating playlists using Recommender Systems, focusing on using Case-based Sequential Ordering.

Case-based Recommendation(CBR):   The approach in which a playlist is generated from a seed song/shorter playlist using sequential patterns learned from a dataset of existing playlists.

CBR is a method of recommendation which can be used in contexts in which a meaningful order or sequence to objects is desired or useful. There can also be a large variety of possible values (such as songs in this case). A set of playlists are then selected from the total dataset based on relevance and other criteria forming the Case Base. From this Case Base of playlists $\mathcal{C}$, two main values are computed for each playlist:

1. The Attribute Variety

2. The Coherence of the Playlist

   Variety   The variety of each playlist in $\mathcal{C}$ is calculated based on the repetition of song features within a set distance.

   Coherence   The coherence of each playlist in $\mathcal{C}$ is calculated based on how related it is to the input song.

2   Datasets

The data was acquired from Spotify's Million Playlist Dataset. Processing was done on the dataset running in SQLite. The data was downloaded from Spotify's servers, structured as a set of JSON files. This was then converted into an SQLite database to improve performance and maintain consistency the with earlier similarity-based algorithms which used a shortened collection from MusicBrainz called the Million Song Dataset.

The processing itself is done using *Python*. This is due to the amount of useful libraries such as *Pandas* and *Numpy*.

3 Playlist Generation using CBR

4 Testing

A variety of CBR systems will be tested with varying weightings of Coherence and Variance, along with simpler pseudorandom and similarity-based algorithms. These will then be ranked by survey.

References

[1]   Enric Plaza Claudio Baccigalupo. "Case-based Sequential Ordering of Songs for Playlist Recommendation".

[2]   Barry Smyth. "Case-Based Recommendation".