Proposal of Music Recommendation System

Overview

This project explores the "hetrec2011-lastfm-2k" dataset from "http://www.last.fm" which contains social networking, tagging, and music artist listening information from a set of 2K users from Last.fm online music system and builds a music recommendation system for users. The data set can be found via the link:

https://grouplens.org/datasets/hetrec-2011/

Goals

The recommendation system will **give a ranked list of recommended artists** to users based on

1. Collaborative Filtering: Recommend the artists that are in a user's friends' favorite artists list to the user.
2. Content-based Recommendation: Based on users' listening history, find the pattern of users' listening behavior, recommend the artists list that best matches a user's listening behavior.

Methods

For the Collaborative-filtering approach:

The favorite artists list of each user can be obtained by the tagging records of each user which are already provided by the dataset. The user_artists.dat dataframe contains users' hitting times for artists. The information provided by the favorite artists list and the user_artists.dat dataframe tells us how favorable an artist is to a user.

The user-friend data frame provides the connection among users.

SQL-typed manipulation among these dataframes will provide a user's friends' favorite artists, hence give a recommendation list of artists. We will use pandas package in Python to achieve this though.

For the Content-based approach:

We will use the technique of matrix decomposition along with stochastic gradient descent to make regressions on users' ratings (hitting times) on artists. We will then sort a user's ratings from high to low, which gives a favorite artists list of the user.

Reports

The primary function of this project is to give a recommendation list of artists of a user. The client will input his/her own UserID. Our recommendation system will give a recommendation list consisting of the wished number of artists. We will also compare our recommendation list to a user's actual favorite artists list provided in the data source to test on the accuracy of our recommendation system.