

Ivette M Tapia

Capstone 1 Proposal

Songs Recommender System using the Million Song Dataset

Project Description

A music streaming service wants to determine: *What song(s) should be recommended to specific service users given a large catalog of popular songs to choose from?* This is an important question, because users have very disparate tastes in music and if a music service does not play or make visible songs a user likes often enough, they are more likely to disengage from the music service. Disengagement from the music service has impacts on the service's advertising and fee revenue.

Methodology

For this problem, I will use either a content-based approach or a collaborative filtering approach. In the content-based approach, the recommendation model examines the properties of the content and recommends content to the user that is similar to what the content they have consumed previously.

In the collaborative filtering recommendation approach, the items recommended to an user are those preferred by similar users. As such, the recommendations are based on similarity measures between users and/or items. Moreover, a hybrid of the content-based and collaborative filtering approaches may be possible. While a recommender system can also be knowledge or demographically based, these are likely not appropriate given the dataset.

Dataset

I will use the Million Song and Taste Profile Subset datasets to build the recommender system for songs. The Million Song dataset contains a collection of 48 song audio features such as tags, loudness, danceability, year released, similar tracks, tempo etc. for 1M popular songs. The Taste Profile Subset is a separate, supplementary dataset to the Million Song dataset that contains user, song and play counts for 1.019M users and 48.4M user-song-count triplets. This dataset is provided and maintained by LabRosa at Columbia University. The data is collected from the EchoNest device.

Dataset Website: <https://labrosa.ee.columbia.edu/millionsong/>

Dataset Citation: Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The Million Song Dataset.

Project Deliverables

The deliverables for this project are: (1) code notebook with a complete implementation of the recommendation system for songs in the python language, and (2) a slide deck explaining the problem at hand, recommendation system implementation procedure, and results.