Authors: J. Hamski, V. Nguy
Document: Project 4 Proposal
Date: July 7, 2016

Project 4 Proposal

Business Goal:

The business goal of our proposed Project 4 will use social context to provide recommendations for a random user based on their contacts.

Data Source:

We will continue to use the Last.fm dataset that contains:

- Listeners and a count of listens for particular artists
- LIsteners social graph

Audience:

The audience for this recommender system will be people who want to music recommendations as if it were still 2006 - since that's when the dataset records were created! If we approached this from a product-development point of view, we could view this as an attempt to provide recommendations to a user who has social connections to Last.fm users but a sparse listening record.

Contextual Element:

We will simulate a new user by creating a new user who is a 'neighbor' of a group of interconnected listeners. This will simulate a 'friend' joining Last.fm. We will then use a combination of collaborative filtering and social context to provide this simulated user a list of song recommendations heavily influenced by their friends.

Execution Plan:

Dataset Preparation
Since this is our fourth project with the Last.fm dataset, we're going to take a more enhanced data-cleaning approach due to previous results.

- Filter out artists with non-english characters in their names.

- Filter out artists which were listened to by less than some small percent of the users in the dataset or some other method of trimming the long distribution of extremely obscure artists.

Collaborative Filtering

Our first and second projects used user-based collaborative filtering, but recommendations tended to be dominated by popular artists regardless of simulated inputs. Therefore, we will implement a collaborative filtering technique that penalizes extremely popular artists.

Social Graph Calculation

We will use the social graph data to determine the closeness of users and ‘boost’ artist recommendation scores from collaborative filtering when two users are connected.

Recommender Validation

We will evaluate the recommender’s ability to predict existing user preferences by calculating root mean square error and creating a confusion matrix.

We will simulate new users without any data but with social connections to see how it affects their recommendations.