

IEOR - 4571 Personalization Theory and Application: Project Outline
Gaurav Singh, Carolyn Silverman, Cindy Wu, Maura Fitzgerald

- For Part II of the project, we will build a music recommendation system using the following kaggle dataset:
<https://www.kaggle.com/c/kkbox-music-recommendation-challenge/data>
 - The task of the challenge is to “predict the chances of a user listening to a song repetitively after the first observable listening event within a time window was triggered.”
 - Key properties of dataset include:
 - A training set with first observable listening event for each unique user-song pair
 - Song information in unicode, including genre categories, artist name, etc.
 - User information, including city, gender, age, etc.
 - Output result of 1 or 0 for each song
- We will explore three different state-of-the-art algorithms to build recommender systems in R [mostly from scratch]:
 1. Collective factorization or Factorization machines [with regularization] use one hot encoders for users, items and other input data to generate features and use regularization to curb overfitting.
 2. Content-based recommender system that utilizes the song and artist metadata like song_length, genre, artist_name, composer, lyricist, language. Since the use of public data is encouraged for this competition, we intend to incorporate readily available external data from the web see if it increases the performance of the model.
 3. SVD++ : build components of explicit and implicit feedbacks to from multiple features of the input [item and users [if available]]

We intend to measure the effect of hybridization on the accuracy measure. Specifically, we will build hybrid models using all combinations of the three models described above. We will tune hyperparameters and evaluate our models using many of the same approaches as Part I (e.g. cross-validation, accuracy, coverage, etc), and we will compare the performance of each model to that of traditional approaches like collaborative filtering, as well as a baseline random recommender system.

The purpose for the recommendation system is to drive user engagement with our listening service and maximize ad revenue. Since nearly every listening service utilizes some sort of recommendation system, a more accurate predictive model will keep our service competitive in the field.