

# Music Recommender System

Shefali Garg(11678), Fangyan Sun(EXY1329)

February 28, 2014

## 1 Motivation

With the rise of digital content distribution, people now have access to music collections on an unprecedented scale. Commercial music libraries easily exceed 15 million songs, which vastly exceeds the listening capability of any single person. With millions of songs to choose from, people sometimes feel overwhelmed. Thus, an efficient music recommender system is necessary in the interest of both music service providers and customers. Users will have no more pain to make decisions on what to listen while music companies can maintain their user group and attract new users by improving users' satisfaction.

In the academic field, the domain of usercentric music recommendation has always been ignored due to the lack of publicly available, open and transparent data. Million Song Dataset Challenge provides data which is open and largescale which facilitates academic research in usercentric music recommender system which hasn't been studied a lot.

## 2 Description

Our study is based on Million Song Dataset Challenge in Kaggle. We would like to do a large, personalized music recommendation system with the goal of predicting the songs that a user is going to listen. We learn from users listening history and full information of all songs (metadata, audio content analysis and standardized identifiers). Our goal is make our system largescale and more personal to users.

## 3 Project approach

We propose following algorithms for our task-

1. **Popularity based**

The most trivial recommendation algorithm is to simply present each song in descending order of its popularity skipping those songs already consumed by the user, regardless of the user's taste profile.

2. **Same artist greatest hits**

This simply produces the most popular songs by artists that the user has already listened to. This gives some level of personalization in the recommendation system.

3. **Collaborative Filtering**

It can be either user-based or item-based. In user-based recommendation, users who listen to the same songs in the past tend to have similar interests and will probably listen to the same songs in future. In the item-based recommendation strategy, songs that are often listened by the same user tend to be similar and are more likely to be listened together in future by some other user.

4. **Latent factor model**

The ratings are deeply influenced by a set of factors that are very specific to the domain (e.g. genre, artist). These factors are in general not obvious and we need to infer those so called latent factors from the rating data. Users and songs are characterized by latent factors and a latent factor model such as Singular Value Decomposition (SVD) can decompose rating matrix into the product of a user feature and an item(song) feature matrix.

## 4 Dataset

We would be using the database provided by Kaggle competition, refer <https://www.kaggle.com/c/msdchallenge/data>

It includes metadata (e.g., artist identifiers, tags,etc) , audio content analysis and standardized identifiers.

## 5 References

McFee, B., BertinMahieux,T., Ellis, D. P., Lanckriet, G. R. (2012, April). The million song dataset challenge. In Proceedings of the 21st international conference companion on World Wide Web (pp. 909916).ACM.

Aioli, F. (2012). A preliminary study on a recommender system for the million songs dataset challenge. PREFERENCE LEARNING: PROBLEMS AND APPLICATIONS IN AI, 1.

Koren, Yehuda. "Recommender system utilizing collaborative filtering combining explicit and implicit feedback with both neighborhood and latent factor models." U.S. Patent No. 8,037,080. 11 Oct. 2011.