

|  |
| --- |
| Music Recommendation Lastfm |
|  |
| February 28  Authored by: Ajay Parihar, Caroline Kadic |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Introduction**With the prevalence of the internet, mobile devices and commercial streaming music services, the amount of digital music available is greater than ever. Sorting through all this music is an extremely time-consuming task. Music recommendation systems search through this music automatically and suggest new songs to users. Music recommendation systems have been developed in commercial and academic settings. Industrial systems such as Pandora and Spotify have achieved commercial success, but they have not reached perfection. Song observes that from an academic perspective, “the development of music recommender systems is still at a very early stage.” The challenge of music recommendation is to create a system that can continually find appealing new music while minimizing the amount of user effort involved.The amount of music available online on music store is huge amount. Itunes: 18 million songs **Amazon**: 17 million songs  LastFM is a well-known music streaming service that has open-sourced some of their data sets. In this project we tried multiple recommendation algorithms to provide recommendation to users based on different characteristics. Problem  Music Characteristics  Some of the difference in recommendation system for music compared to other industry are:   * user preferences might be implicit * repetitive consumption * instant feedback * context(morning, work, afternoon)   Recommendation Problem  1. Prediction problem – estimation of the items’ likeliness for a given user  2. Recommend a list of N items – assuming that the system can predict likeliness for yet unrated items  Findings  We tested multiple recommendation systems and found User based with the least RMSE, also we tested content based and it gave us pretty good result. We have to reduce the size for item based and it can be a factor for bad prediction.  Below are some prediction and Classification accuracy for  Collaborative Recommendation System   |  |  | | --- | --- | | Recommendation System | Prediction | | User Based | 0.65 | | Item Based | 0.98 | | Hybrid | 0.987 |   Content-Based Recommendation System  The main benefit of using tags were that the features were already defined by users,  which was not likely to be commonly used enough to use as a prominent feature. To deal with this problem, we grouped all the used tags from the user\_tagged\_artist data, and used the 100 most commonly used tags, across all users.  We ran similarity matrix on artists and results are pretty close to on lastfm. As you can see in the below example Madonna & Kylie Minogue are quite similar but Madonna & Radiohead are different.  > cosine(madonna, kylie\_minogue)  [,1]  [1,] 0.9573262  > cosine(madonna,radiohead)  [,1]  [1,] 0.1502968  Hybrid Recommendation System  We also tried hybrid recommendation system combining the result of both User & Item based but it didn’t give us best result.  Cluster Based Recommendation System  We also tried cluster based analysis, the below diagrams shows the higherical relationship and recommendation for the users.  https://raw.githubusercontent.com/aparihar11/recommendation_systems/master/Data/img/Dendogram.png  We ran cluster analysis and below were the results for the users.    Methodology  There are different methodology available for recommendation and being used in different sectors like retail, news, music etc.  General Approaches:  There are multiple recommendation techniques. The two main approaches are collaborative and content based filtering.   * Collaborative filtering   Collaborative systems generate recommendations by comparing ratings of items between different users. Collaborative systems do not attempt to model the actual item being recommended; they only analyze users’ responses to those items.   * Content-based filtering   Instead of comparing user ratings, content based systems construct models of the items being recommended and compare them to the preference model of the current user.   * Hybrid systems   Content-based filtering is like a car with high acceleration but low top speed. It can give effective recommendations quickly because it doesn’t need a lot of users, but the effectiveness of those recommendations won’t improve after the system does gain many users. Collaborative filtering is the opposite; it has high top speed but low acceleration. As such, the two approaches can be combined into a hybrid system that draws on the strengths of each. |
| Data Exploration  The data consists of user tags, artists list etc. Looking at the data we found the below things.  Most popular artists  We ran search to find the top artists and founds Lady Gaga, BS, Rihanna as the most popular artists.  > head(popular\_artists)  # A tibble: 6 x 2  # Groups: name [6]  name n  *<chr>* *<int>*  1 Lady Gaga 611  2 Britney Spears 522  3 Rihanna 484  4 The Beatles 480  5 Katy Perry 473  6 Madonna 429  Top music genres  We did a wordcloud on the genres and found the top genres were:   1. Rock 2. Alternative 3. Pop 4. indie |
|  |