Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018)
IEEE Xplore Compliant - Part Number:CFP18J06-ART, ISBN:978-1-5386-0807-4; DVD Part Number:CFP18J06DVD, ISBN:978-1-5386-0806-7

Music Recommendation Based on Content and Collaborative Approach & Reducing Cold Start Problem

Parmar Darshna
Computer Engineering
Birla Vishwakarma Mahavidyalaya
Baroda, India
chandani_simle@yahoo.com

Abstract— **Nowadays, Recommendation becomes the most popular area for many researchers. The main aim of recommendation is to provide meaningful suggestions to users for particular item based on users past interest and behaviors towards items. There are two most popular recommendation algorithm is 1) Content-Based Filtering 2) Collaborative Filtering. Content-Based method recommends music based on user data. Collaborative method uses rating and content sharing between different users to recommend music. Here, to provide music recommendation by content-based method music subjective features Speechiness, loudness, Acoustiness etc. are analyzed. The extracted features are stores into database by using Kmean clustering algorithm. For Content-based method, whenever user fires query to database music feature attribute value compares with clusters centroid. Once attribute value match, music can be recommended to user as Content-based method. For collaborative method, rating given by user to particular music is considered and adjusted cosine similarity is used to find similarity between user-user. Once similarity found, prediction rating algorithm is used to provide recommendation to user. Cold-start is most common problem for new user. Here, most popular tracks are recommending to user to solve it..**

Keywords—Recommendation, subjective features, content-based approach, collborative approach, cold-start

I. INTRODUCTION

Data Mining is a process of acquiring meaningful data from a huge number of data [1]. Nowadays, Internet has become the lead source for everyone for getting information about anything. The web is the place where billions to millions information sources are available as per user need. As there are huge amount of datas are available over web, the problem of information overload appears. Because of Information overload the mining results can take more time to search the required information or even can provide wrong prediction as sparsity of data [2].

To overcome information overloading problem, information filtering systems are required. Recommendation is subsystem of information filtering. This system filters required information from large number of available information over web and provides suggestion as per user preference. Many of search engines such as Google etc. have partially solved the information overload problem but there is a lack of prioterization and personalization. So recommendation is the personalized way of predicting items as per user need from users past history. The recommendation system helps users for getting required information within a very less time and saves time [3]. Recommendation can be provided for a number of items such as book, movies, music, article, product etc.

II. BACKGROUND

A. *Content – Based Recommendation System*

Content based recommendation system provides prediction based on user or item information and past interests of user. Content-based filtering method examines users past interests for particular item. Upon examines the user interests, the system provides recommendation for the items that have highly similar kind of features related to user interest or items accessed in past [4].

B. *Collaborative Recommendation System*

This technique analyzes a large amount of data collected from user responses to an item as rating in past and recommends items to user. Here, analyzing item content is not necessary and information is shared between two users so that can provide surprising recommendation which user may pretend to be interested. The base of this method depends on relationship between user and items and also on rating feedback matrix where each element representing a specific rating on a specific items [6].

III. RELATED WORK

A content-based personalized music filtering system learns the user's preferences by mining the melody patterns from the music objects in the user's access history. Using these melody patterns, a melody preference classifier is then constructed for each user. An incoming music object will be recommended to the user if it is classified into the preferred class. In this system, only the pitch information is considered for feature extraction. Ignoring other information, e.g., duration and loudness, provided in the music objects limits the system to deal with other kinds of user preferences [6].

Ringo is a pioneer collaborative music recommendation system. In Ringo, each user is requested to make ratings for

some music objects. These ratings constitute the personal profile. For collaborative recommendation, only the ratings of the users whose profiles are similar to the target user are considered. A music object will be recommended based on the weighted average of the ratings considered [7].

Gaana uses collaborative and content based filtering systems. Gaana app analyzes the content of its entire application and recommends features based on the user's preferences or interests. If the app suggests the items and preferences which the user has already used in the past it leads to overspecialization. Collaborative filtering (CF) systems compute correlations between users; they predict song / album ratings for the current user, based on the song/ album ratings provided by other users, who have preferences for the songs which are highly correlated to the current user [8].

Spotify is a digital music service that gives you access to millions of songs. Spotify provides access to over 30 million songs, with more music being added every day. As of June 2016, Spotify has 100 million monthly active users, and as of September 2016, it has 40 million paying subscribers [9]. It uses Content-based filtering algorithm for Spotify Radio services and Collaborative filtering Algorithm to discover weekly playlist using Implicit Matrix Factorization method [10].

Pandora utilizes a classification system that is the heart of their service. Pandora recommends by matching up the user's artist and song likes with other songs that are similar. The greatest challenge for Pandora is classifying songs in their database and building their musical taxonomy. To accomplish this, Pandora employs a team of trained musicians who perform a manual classification on each song before adding it to their database. The musicians spend their workdays listening to a collection of songs and tagging each according to approximately 400 musical attributes. Once songs are properly classified in the database Pandora compares the description of musical tastes of a station selected by an individual user with the classification of the songs in the music database. This comparison returns a collection of songs that drive the playlist. The key to Pandora's method of recommending music from its database is to utilize an efficient and effective proximity measure algorithm to determine the neighborhood of music to play on a station [11].

IV. Proposed Methodology

In this paper, there are two methods are used 1) Content-Based Approach 2) Collaborative Approach. The System Block Diagram contains A. Feature Extraction Module B. K-Mean Clustering Module C. User Playlist D. Recommendation Module and Database.

A. Feature Extraction Module

This module extracts feature attributes from music object using Spotify API. It extracts 9 feature attributes from music object such as Dancibility, Energy, Loudness, Speechiness, Acoustiness, Instrumentalness, Liveness, Valence and Tempo. These features are extracted by giving TrackID of music object to Spotify website.

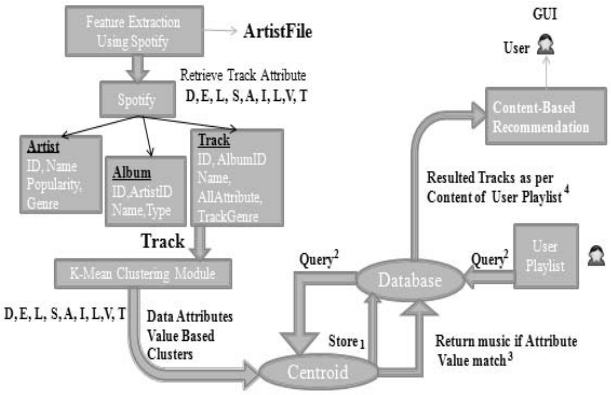

Fig. 1. Content-Based Recommendation Diagram.

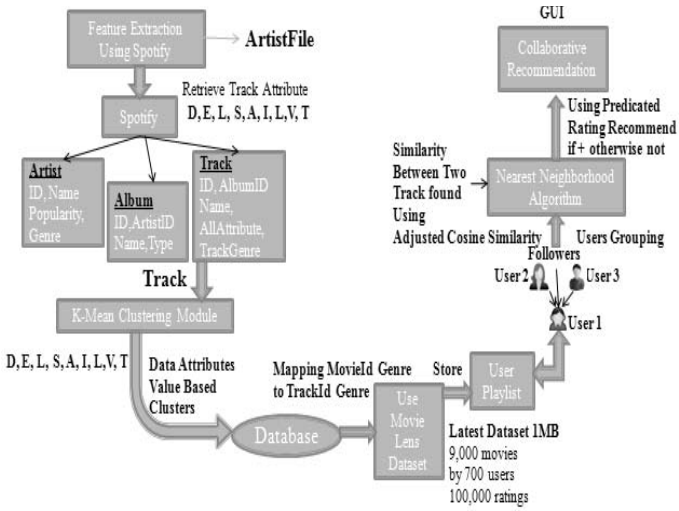

Fig. 2. Collaborative Recommendation Diagram.

It provides web developers SpotifyID by registering to website. Though usage of SpotifyID user can access the website and as website contains a million of songs, each song contains TrackID through which feature attributes of music object is retrieved in JSON (JavaScript Object Notation) format using a link https://api.spotify.com/v1/audio-features.

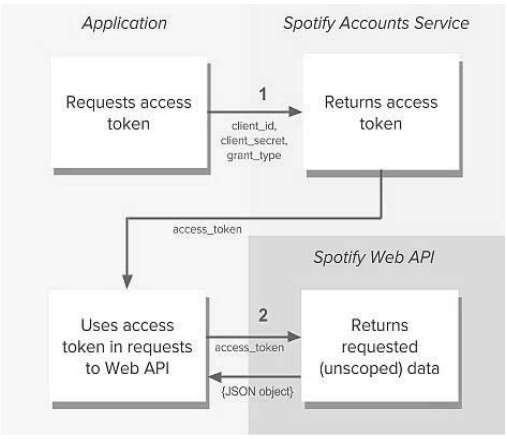

Fig. 3. Process through Spotify API to get attributes features of music object [9].

Music Feature Attribute Information

1) Dancibility - 0.000000 to 1.000000 - Combination of Rhythm, Beat & Overall Regularity. 0.0 Least danceable and near to 1.0 most danceable.

2) Energy - 0.000000 to 1.000000 - Fast, loud, noisy track represents measure of intensity and activity.

3) Loudness – -60 to +60 – Represents quality of sound and correlates to physical amplitude.

4) Speechiness - 0.000000 to 1.000000 - Detects presence of Spoken Words.

5) Acousticness - 0.000000 to 1.000000 – Detects whether the track is acoustic or not.

6) Instrumentalness - 0.000000 to 1.000000 – Detects whether track contains no vocal.

7) Liveness – 0.000000 to 1.00000 – Detects presence of audience.

8) Valence – 0.000000 to 1.000000 – High Valence relates to happy and Low valence relates to sad tracks.

9) Tempo – 0.000 to 0.250 – Speed of average beat duration.

B. K – Mean Clustering

Kmean is partioning based algorithm used to cluster music objects according attribute value wise.

K – Mean Algorithm

Clustering Concept [12]: The objects within a group need to be similar to one another and different from the objects in other groups.

Algorithm [12]:

1. Select k points as initial centroids.
2. Repeat
3. Form k clusters by assigning all points to the closest centroid.
4. Re-compute the point of each cluster.
5. until the centroids don't change

$$d(i,j)=\sqrt{|x_{i1}-x_{j1}|^2+|x_{i2}-x_{j2}|^2+...+|x_{ip}-x_{jp}|^2}$$

Fig. 4. Euclidean Distance Formula for finding similarity between two tracks [9].

C. User Playlist

This module contains a list of music object in user profile. The Tracks downloaded from Spotify website is stored into database along with 9 feature attribute value.

D. Recommendation Module

1) Content-Based Filtering Approach: For Content-based filtering method, to recommend music object the similarity between music object attribute value and cluster centroid attribute values is matched As per outcome of K – Mean algorithm different cluster centroids are made.

Upon finding similarity between music object and cluster centroid attribute value, music object which is not in the playlist of user would be recommended to user. When user firing query to database, music objects with same cluster attribute value are returned.

2) Collaborative Filtering Approach: For Collaborative filtering method, music object with ratings are considered for recommendation. Here, adjusted cosine similarity is used to find out similarity between two tracks. Upon finding similarity between two tracks prediction rating algorithm is used to provide recommendation to user.

Grouping of two users are required. Here, Each User has their Followers, Tracks and Rate provided to track. Rating and Followers is base of Collaborative Recommendation. At least two Followers are grouped together, so also known as User Grouping.

Nearest Neighborhood Algorithm

Nearest Neighborhood Algorithm is used to find out similarity between two tracks of Users and Followers using Adjusted Cosine Similarity [13]. The algorithm calculates the adjusted cosine similarity between all pairs of Tracks:

$$w_{ij}=\frac{\sum_{u\in U_{ij}}(r_{ui}-\bar{r}_u)(r_{uj}-\bar{r}_u)}{\sqrt{\sum_{u\in U_{ij}}(r_{ui}-\bar{r}_u)^2\sum_{u\in U_{ij}}(r_{uj}-\bar{r}_u)^2}}.$$

Fig. 5. Adjusted Cosine Similarity between songs i and j [13]

- U is the set of users (Uij is the set of users that have rated both i and j).
- rui is the rating for the item i, given by user u.
- bar(ru) is the average rating for a user u.

Predicted Rating

This similarity is combined with the intersection of the songs that a given user has rated and each song's neighborhood in order to predict the ratings the user would give to all other songs [13]. The songs with the highest predicted ratings should be recommended to the user [13].

$$\hat{r}_{ui}=\frac{\sum_{j\in R_u\cap N_i} w_{ij}r_{uj}}{\sum_{j\in R_u\cap N_i}|w_{ij}|}.$$

Fig. 6. The predicted rating of item i for user u [13]

- Ru is the set of all rated songs by user u. Ni is the neighborhood of songs for song i.
- wij is the similarity from Fig 5.

Database/Dataset

Dataset has built by downloaded from Spotify website. There is a file name ARTISTNAME.DAT in which you need to put any of artist name if information is available on Spotify then will give all data related to it else says artist information is not available. There are total 70 Artist in which 45 Artists are random with mixed top, hindi genre, pop genre, dance pop,

canedian pop, chrismas pop, post-teen pop, desi, latin, latin pop, Indian pop etc. Most music is belongs to pop genre. Another 10 Artist from Country Blue Genre 10 Artists from Jazz Instrumentalists 5 and Vocalists 5 and 10 Artists from Folk Bollywood. There might be possibility that some artist will not have any genre and here limit of Album is 10.

There are 70 Artist with 656 Album with 639 distinct with 9529 Tracks. After performing k mean 155 clusters are made.

For Content-Based, Output of Kmean Clusters are used. Each of Tracks in User Playlist has music feature attributes. Ex. Dancibility, Energy etc**.** When a User query into Database, Track (Attribute Value) match with Cluster Centroids (Attribute Value). If match, then Tracks which are not in User Playlist recommend to user.

User of MovieLens Dataset

The dataset has downloaded from MovieLens Website. It contains 100004 ratings and 1296 tag applications across 9125 movies. These data were created by 671 users between January 09, 1995 and

October 16, 2016. This dataset was generated on October 17, 2016. Users were selected at random for inclusion. All selected users had rated at least 20 movies [15].

Mapping Movie Genre to Music Genre

To map movie genres I used below measures which is similar to music genre according instruments and have the most common attributes related to movie genre.

1. Pop [loudness, acousticness][16]: Action, Adventure, Animation, Children's, Crime, Mystery, Horror, Sci-Fi, Thriller, War, Western

2. Jazz [liveness, Speechiness, energy][17]: Comedy, Fantasy

3. Blue [liveness, Speechiness][18]: Drama**,** Documentary, Romance

4. Filmi [acousticness, liveness][19]: Film-Noir, Musical

User Interface

For front-end Java Jdk-7 with Eclipse Neon and for Back-end Microsoft SQL Server 2012 is used. To provide Graphical User Interface Tomcat Server is used. After Login User has two toggle buttons: 1) User Playlist 2) Follower 3) Content – based recommendation 4) Collaborative recommendation

Fig. 6. UI for Content Based Recommendation

Fig. 7. UI for Collaborative Recommendation

If there is no music in playlist for new user, then based on popularity of tracks music would recommend to user to solve cold-start problem.

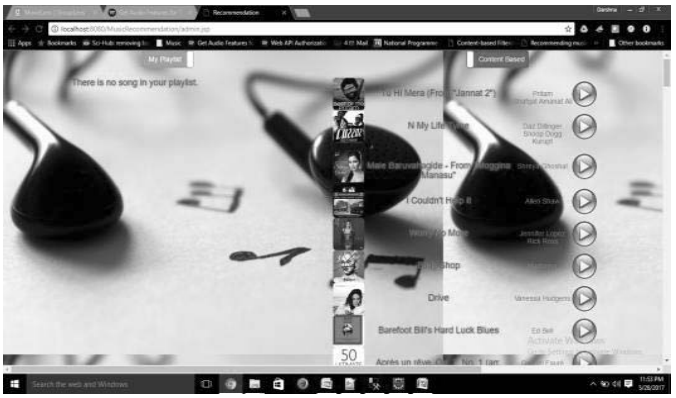

Fig. 8. UI for Reducing Cold Start Problem

IV. Result Analysis

TABLE I. COMPARISION OF KMEAN OUTPUT

Kmean Output	**Kmean Output**	**Kmean Output**	**Kmean Output**	**Kmean Output**
20 Artist	41 Artist	70 Artist	70 Artist	70 Artist
3154 Tracks	3346 Tracks	9925 Tracks	9925 Tracks	9925 Tracks
84 Cluster	89 Cluster	155 Cluster	155 Cluster	155 Cluster
332068 Miliseconds	51410 Miliseconds	355134 Miliseconds	357610 Miliseconds	365563 Miliseconds

V. Conclusion

For Music Clustering, All music Track has been clustered with total 9 attributes by 155 clusters. As per observation, if k mean run for more than one time accuracy decreases for the same dataset. A cluster creation time is only depends on similarity of tracks (attribute value) and not on number of tracks. *Ex.* Dataset with 3154 tracks creates 84 cluster centroids while Dataset with 5000 tracks takes 27 cluster centroids.

For Content-based Recommendation, When a query is performed on Cluster Centroid into database, attribute value of

cluster centroid is matched with track attribute value. If match, then recommend to user. If no music is available in user playlist then, according to popularity track music would recommend to user (Cold-start problem solution).

For User Grouping Based Recommendation, Nearest Neighborhood Algorithm is used using Cosine Similarity and Predicted Rating to recommend track. Here, MovieLens dataset is used to provide huge number of rating for User Playlist. To provide proper classification Movie Genre is mapped to Music Genre. So, Quality of recommendation is achieved.

V. Future Work

For large dataset recommendation process take more time to execute as sequential approach, so will make parallel approach using Hadoop Environment for future perspective.

References

[1] Jiawei Han and Micheline Kamber, "Data Mining – Concepts and Techniques," Second Edition

[2] Information Overload- https://en.wikipedia.org/wiki/Information_overload.

[3] Information filtering system - https://en.wikipedia.org/wiki/Information_filtering_system

[4] HUNG-CHEN CHEN, ARBEE L.P. CHEN, "A Music Recommendation System Based on Music and User Grouping," Springer Science + Business Media, Inc. Manufactured in The Netherlands, Journal of Intelligent Information Systems, 24:2/3, 2005, 113–132.

[5] Megha Jain, "ALGORITHM FOR RESEARCH PAPER RECOMMENDATION SYSTEM," International Journal of Information Technology and Knowledge Management, July-December 2012, Volume 5, No. 2, pp. 443-445.

[6] Kuo, F.F. and Shan, M.K., "A Personalized Music Filtering System Based on Melody Style Classification," In Proceedings of IEEE international Conference on Data Mining, 2002, pp. 649–652.

[7] Sharadanand, U. and Maes, P., "Social Information Filtering: Algorithms for Automating 'Word of Mouth'," In Proceedings of CHI'95 Conference on Human Factors in Computing Systems, 1995, pp. 210–217.

[8] Study of Recommender System - www.ganna.com ! My First Blog - https://tejaswinidevappa.wordpress.com/2015/01/20/study-of-recommender-system-www-gaana-com/

[9] Music for everyone – Spotify - www.spotify.com

[10] Frederik Pruifer, 2016 Music Recommendation at Spotify – How Spotify Recommedns music

[11] Michael Howe, Pandora's Music Recommender

[12] Zijun Zhang, "K – Means Algorithm, Cluster analysis in Data Mining" 2012.

[13] Stephanie Smith, Sarah Jones, Karl Bo Lopker, "What's Next? Music Recommendation System".

[14] MovieLens- Wikipedia - https://en.wikipedia.org/wiki/MovieLens

[15] MovieLens | GroupLens - http://grouplens.org/datasets/movielens/latest/

[16] Pop music - https://en.wikipedia.org/wiki/Pop_music

[17] Jazz - https://en.wikipedia.org/wiki/Jazz