

# Multimedia Recommendation System using Collaborative Filtering Techniques

Shruthi k

4<sup>th</sup>sem Mtech

Dr. Ananda Kumar K R

Professor

Dept of CSE, SJB Institute of Technology, Bangalore-560060, India.

**Abstract-** Now-a-day's recommendation systems are very popular. One of the most popular recommender systems called collaborative filtering techniques (CF). Music brings people together; it allows us to experience the same emotions. Currently musical genre classification is done manually and requires even the trained human ear considerable effort. Therefore clustering songs automatically and then drawing valuable insights from those clusters is an interesting problem and can add great value to music information retrieval systems. Most of the work in this field has involved extracting the audio content from audio files. This paper explores technique to cluster songs based on Audio Attributes such as artist name, title, year, album types by using K-Means algorithm. I experiment with different sets of attributes and genres to cluster the music files and using collaborative filtering techniques to recommend music to users.

## I. INTRODUCTION

Recommender system typically provides a user with a list of recommended items may be interested in, or predict how much might prefer each item. Recommender systems have recommended to user interest items and thus similar of that interest item will recommend, and are applied in a variety of applications. The most popular recommendation system ones movies, music, news, books, research articles, search queries, social tags, and products. However, there are also recommender systems for experts, jokes, restaurants, financial services, life insurance, persons (online dating), and Twitter followers.

Recommender system consists of three types of filtering mechanism namely: Collaborative filtering, Content Based and Hybrid filtering(see figure1).

Collaborative filtering based on collecting and analyzing a large amount of information on user's behaviors, activities or preferences and predicting what users like based on their similarity to other users. Content based filtering methods are based on a description of the item and profile of the user's preferences. Hybrid is a combining of collaborative and content based filtering techniques. In these filtering techniques collaborative filtering method are considered in our work.

The collaborative filtering algorithms are categorized as: Memory based Recommendation generalizes from memory based data at the time

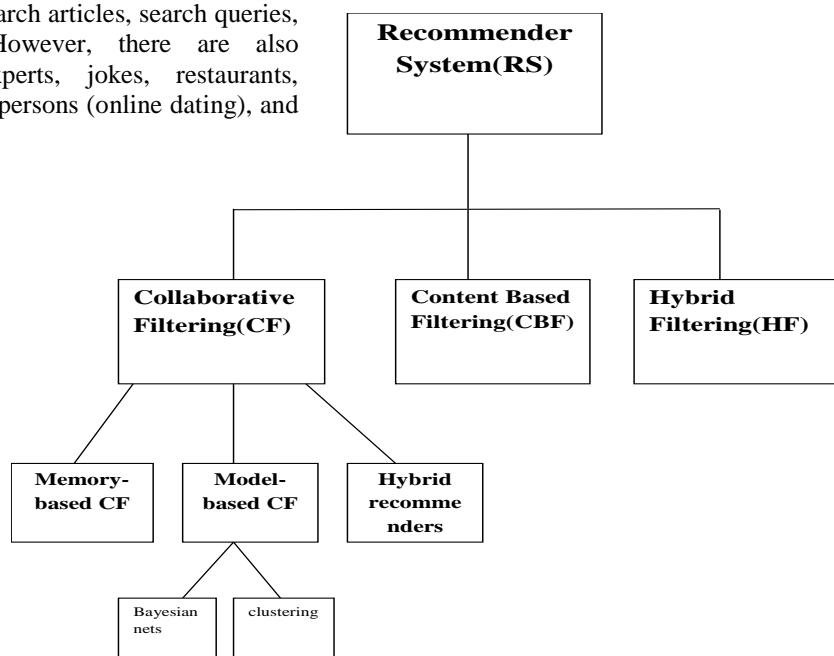


Figure1: Classification of Recommender System

of making memory based learning it is also called as lazy learning. In memory based learning users are categorized into groups based on their interest. Model based collaborative filtering is a two stage process for recommendations in the first stage model is learned offline in the second stage a recommendation is generated for a new user based on the learned model. Our project work only considering the Model Based Recommendation system. Model-based CF algorithms, such as Bayesian models, clustering models, and model dependency networks, have been investigated to solve the shortcomings of memory-based CF algorithms. In which model based CF algorithms choosing a Clustering models in our project work.

1) Bayesian Belief network collaborative filtering: Bayesian Belief network is a directed acyclic graph (DAG).

2) Clustering collaborative filtering algorithms:

Clustering is a grouping of data or dividing a large data set into smaller data sets of some similarity. Clustering methods can be categorized into three types namely:

Partitioning methods, Density-based methods and Hierarchical methods. A commonly-used partitioning method is k-means algorithm which has two main advantages such as relative efficiency and easy implementation. Density-based clustering methods typically

Search for dense clusters of objects separated by sparse regions that represent noise.

Partitioning methods: Given a database of  $n$  objects, a Partitioning method constructs  $K$  partitions of data, where each partitions represents a cluster  $K$  less than equals to  $n$ .

K-Means Algorithm: Where each cluster is represented by the mean value of objects in the cluster.

K-Medoids Algorithm: Where each cluster is represented by the objects located near the center of the cluster. In our project am going to consider the k-means algorithm for large data sets to cluster the data based on attributes after by using CF techniques to recommend the items for user interest. K-means is the simplest and most popular classical clustering method that is easy to implement. The classical method can only be used if the data about all the objects is located in the main memory. The method is called K-means since each of the  $K$  in K-means refers to the number of groups , that we want to put our objects into .It is also called the centroid method, initially the randomly selected the  $k$  clusters centers and compute the Euclidean distance of each object in the dataset from each of the centroid. Our recommender system is avoids overload multimedia contents in the mobile devices and the user's therefore don't waste appropriate time searching for multimedia contents that they are/might be interested in. However introducing Multimedia Recommender Systems by using K-means clustering algorithm. To recommend what the user search particular information that only relevant most top of the data will be recommended to users. In a variety of applications clustering algorithms have been used. Now a day's the volumes of collected data is increasing in all domains. Because of so much data available, the development of algorithm is necessary to extract useful information from the vast stores. Data mining uses many clustering techniques for finding interesting information from the large data set. Other applications of clustering are image processing, pattern

recognition, and classify documents on the web for information discovery.

## II. RELATED WORK

Music classification is an interesting problem with many applications, from Pandora to dynamically generating images that complement the music. However, the music information retrieval field the challenging task is a music genre classification. They investigate various machine learning algorithms, including k-nearest neighbor (k-NN), k-means, multi-class SVM, and neural networks to classify the following four genres: classical, jazz, metal, and pop [1]. K-Means clustering is a clustering algorithm that classifies or groups objects into a specified number of clusters. Initially,  $k$  cluster centers are randomly selected from the given data set and each data point is assigned to the cluster of the nearest cluster center. Each cluster center is then recalculated to be the mean value of its members and all data points are reassigned to the cluster with the closest centroid. This process is repeated until the distance between consecutive cluster centers [2].

When a user cannot remember the title of a song, or its related details, the most direct and convenient method to search for the song is by humming. This search method is particularly important when a user does not have access to operate the audio device. The background of the user often influences the genres of the songs being searched and use the information from a user's search history, as well as the properties of genres common to users with similar backgrounds, to estimate the genre the current user may be interested in based on a probability calculation [3].

## III. SYSTEM ARCHITECTURE

### A. Frame work

#### Module 1:

The audio file is given as input which can be an mp3, wav, .csv file format and it's going to be processed by the system and given as an input to the feature extractor (see figure2).

#### Module 2:

Using jAudio feature extraction tool the selective audio features are extracted. A feature vector comprising all of these features is extracted for each of the music clip.

#### Module 3:

K-means clustering algorithm is used for clustering songs into two clusters-artist name, year. The K-means algorithm is elaborated below:

#### 3.1 K-means Algorithm:

K-Means clustering is a clustering algorithm that classifies or groups objects into a specified number of clusters. Initially,  $k$  cluster centers are randomly selected from the given data set and each data point is calculated minimum distance by Euclidean distance and assigned to the cluster of the nearest cluster center. Each cluster center is then recalculated to be the mean value of its members and all data points are reassigned to the cluster with the closest centroid. This process is repeated until the distance between consecutive cluster centers.

The Algorithm is implemented as follows:

Step1: Initialize the center of the clusters. Here decide number of cluster k. Then: 1. Initialize the center of the clusters  $\mu_i$  = some value,  $i=1\dots k$ .

Step2: Attribute the closest cluster to each data point.

Step3: Set the position of each cluster to the mean of all data points belonging to that cluster.

Step4: Repeat steps 2-3 until convergence

Module 4:

The database stores the cluster the songs based on attributes such as artist name, title, year. Again using collaborative filtering techniques search the database what user requires that list of top items will be displayed in GUI screen.

These all modules will be process in see the below figure2.

The main advantages of k-means algorithm are large data set can be divided into group of cluster so save memory space according to simple Euclidean distance metric. In large data set it's not so easy to search exact information in it so if we cluster the dataset into number of cluster can easily search required data using collaborative filtering techniques.

#### IV. RESULT

To recommend multimedia for the user by using collaborative filtering techniques and all music folder can be group of cluster based on k-means algorithm. Decide the number of cluster k based on attributes of music data set such as artist name, title, album types, year, duration etc, cluster centers are stored in database then user will searching a songs based on cluster centers it will display the list of songs in a GUI screen.

#### V. CONCLUSION

This paper has presented a music information retrieval system (MIR). Mainly our project is using k-means algorithm and collaborative filtering techniques. K-means for using cluster the data for music files and to recommend the

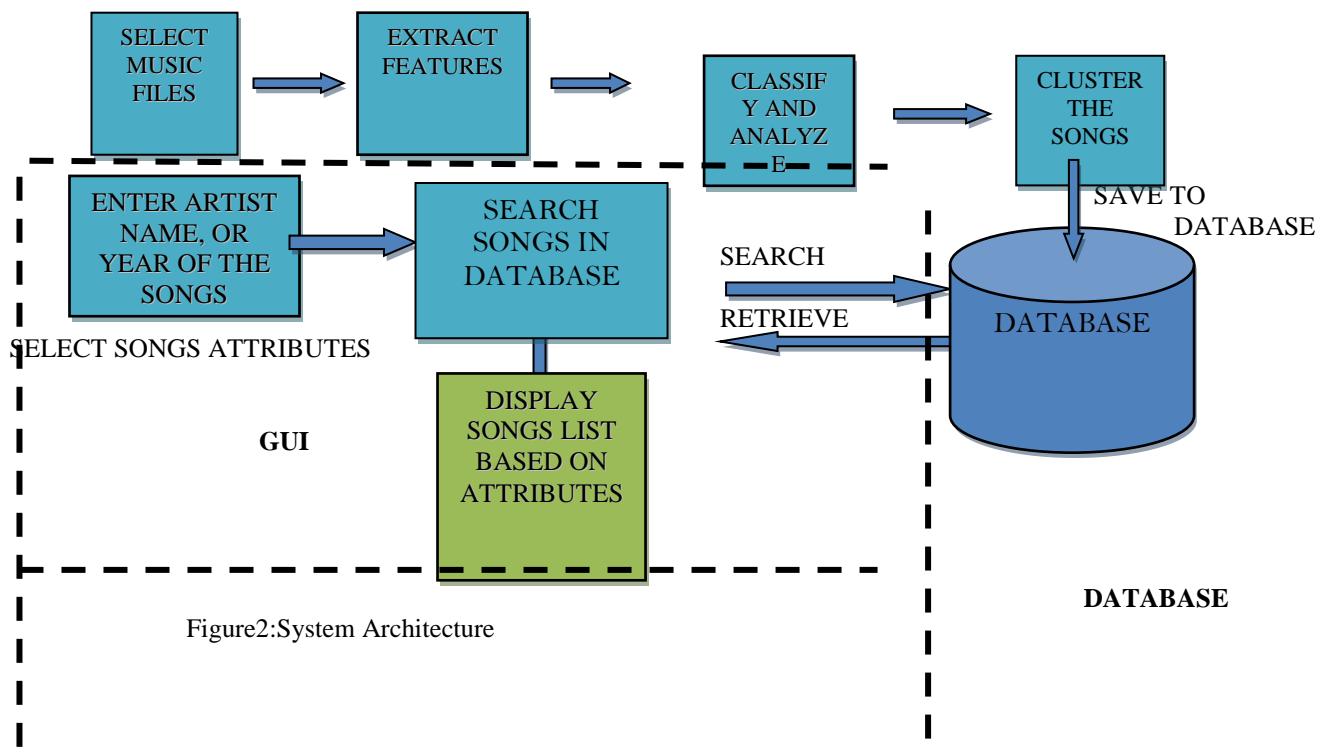


Figure2:System Architecture

songs for the user by approaching CF techniques.k-means algorithms to increase the accuracy and reduce the computation time. We

Study the similarity measure applied to cluster the artists based on genre. We had to determine how to represent cluster centroids and how to update to better centroids each iteration. To solve this, we chose to represent a centroid as if it were also a multi-variate Gaussian distribution of an arbitrary song

We conclude that the similarity measure is possible to be performed not only in music category but also in other categories, as long as the model based and the data structures are available to be constructed, thus the similar hierarchy as in music category can be used as the basis of similarity measure.

## REFERENCES

- [1] Musical Genre Classification of Audio Signals, George Tzanetakis and Perry Cook, *IEEE Transactions on Speech and Audio Processing*,
- [2] Ning-Han Liu, "Effective Results Ranking for Mobile Query by Singing/Humming Using a Hybrid Recommendation Mechanism" *IEEE transactions on multimedia*, vol. 16, no. 5, august 2014.
- [3] J. B. MacQueen (1967): "Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297.
- [4] F.Xia, N. Y. Asabere, A. M. Ahmed, J. Li, and X. Kong, "Mobile multimedia recommendation in smart communities: A survey," *IEEE Access*, vol. 1, pp. 606\_624, 2013.
- [5] Sanjog Ray, Anuj Sharma "A Collaborative Filtering Based Approach for Recommending Elective Courses",Information Systems Area,{sanjogr, f09 anujs}@iimdr.ac.in
- [6] Abhishek Sen" Automatic Music Clustering using Audio Attributes", International Journal, ISSN : 2319-7323/ Vol. 3 No.06 Nov 2014.
- [7] Xinx Wang, David Rosenblum, Ye Wang "A Daily, Activity-Aware, Mobile Music Recommender System", MM'12, Nara, Japan,October 29–November 2, 2012.
- [8] <https://sourceforge.net/projects/jaudio/>
- [9] Jiale Sheny Meng Wangz Shuicheng Yan? Peng Cui "Multimedia Recommendation: Technology and Techniques", SIGIR'13, July 28–August 1, 2013, Dublin, Ireland.
- [10] Tapas Kanungo," An Efficient k-Means Clustering Algorithm: Analysis and Implementation", *IEEE transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, july 2002.
- [11] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Adv. Artif. Intell.*, vol. 2009, Article ID 421425, Jan. 2009.
- [12] Li, T. et al. 2009. Music Clustering with Features From Different Information Sources. *In IEEE Transactions on Multimedia*, Vol. 11, No. 3, pp. 477-485.
- [13] Hong, J. et al, 2008. Tag-based Artist Similarity and Genre Classification. *Proceedings of IEEE International Symposium on Knowledge Acquisition and Modelling Workshop*. Wuhan, China, pp. 628-631.
- [14] Fu, A., Lu, G., Ting, K.M., Zhang, D.. "A Survey of Audio-Based Music Classification and Annotation" *IEEE Transactions on Multimedia*. <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5664796&tag=1>.
- [15] Mandel, M., Ellis, D."Song-Level Features and SVMs for Music Classification" <http://www.ee.columbia.edu/~dpwe/pubs/ismir05-svm.pdf>.
- [16] Berenzweig, D. P. W. Ellis, and S. Lawrence. Using voice segments to improve artist classification of music.In *AES 22nd International Conference on Virtual, Synthetic and Entertainment Audio*, 2002.
- [17] B. Whitman, G. Flake, and S. Lawrence. Artist detection in music with Minnowmatch. In *IEEE Workshop on Neural Networks for Signal Processing*, 2001.
- [18] N. C. Maddage, C. Xu, and Y. Wang. Singer identification based on vocal and instrumental models. In *17th International Conference on Pattern Recognition (ICPR)*, 2004.
- [19] Beth Logan and Ariel Salomon. A music similarity function based on signal analysis. In *ICME 2001*, Tokyo,Japan, 2001.