International Journal of Engineering Research in Computer Science and Engineering (IJERCSE) Vol 3, Issue 3, March 2016

Big Data analytics for Designing Music Recommendation System

[1]K.V.S.N.Rama Rao, [2] N.Vamsi Krishna Reddy, [3] M. Monika, [4] M. Akhilesh, [5] M. Vinay, [6]Ch. Chandrasekhar
[1][2][3][4][5][6] Department of CSE, MLR Institute of Technology, Dundigal, Hyderabad.
[1] kvsnramarao@yahoo.co.in, [2]vamsi.mana1@gmail.com, [3]monika.meenagari@gmail.com, [4]m.akhilesh21@gmail.com [5] vinaykumarmlr41@gmail.com, [6] chanducool38@gmail.com

***Abstract*:-- With the growth of the World Wide Web, a large amount of music data is available on the Internet. When a user searches for a track, the existing system displays a vast amount of related data. In that displayed data, there may be irrelevant data which do not match user's interest. Hence, user has to search rigorously for the track which is a time taking process. So, here we design music recommendation system using big data analytics by performing content-based, collaborative and statistic-based recommendation, which are based on the favorite degrees of the users to the music groups.**

***Index Terms*--- Data analytics, music recommendation system, user interest, social networks**

I. INTRODUCTION

Big data analytics is the process of analyzing large data sets containing different data types -- i.e., big data -- to reveal concealed patterns, unknown correlations, customer preferences and other useful business facts. The analytical results can lead to more effective marketing, new profit occasions, better customer services, improved operational effectiveness, competitive advantages over peer organizations and other business assets.

The primary goal of data analytics is to help companies make more effective business decisions by allowing data operators, forecast modelers and other analytics professionals to anatomize abundant transaction data, as well as other forms of data that may be untapped by current business intellect programs.

That could include Web server logistics and Internet data, social media and network activity reports, text from customer e-mails and survey replies.

The two main routes for a recommendation system have been discussed in the literature: the content-based filtering approach and the collaborative filtering approach. In the content-based filtering approach, recommends items that are Similar to the items the user preferred or queried in the past.

It relies on product features and textual item descriptions. Based On the user profiles, the system recommends only the data items that are highly relevant to the user profiles by computing the similarities between the data items and the user profiles.

In this approach, the representation of data items and the description of user preferences in profiles are main outcomes which dominate the efficiency of recommendation. Instead of comparing the data items and the user profiles, the collaborative approach computes the same qualities between the user profiles. Users of similar profiles will be grouped together to share data in their profiles. The main goal of the collaborative approach is to make recommendations amidst the users in the same group.

In the collaborative filtering approach, the system may have large possibility to recommend abrupt data items by the nature of data sharing. Some systems use both content-based and collaborative filtering. The system analyzes the used social networks to derive the individual profiles and compares the user profiles to group users for collaborative recommendation service. In past years, the music becomes more popular due to the elaboration of the technology. Various kinds of music around us turn into complex and huge.

In addition to searching required music objects for users, it becomes necessary to develop a music recommendation service. The Music Recommendation System (MRS) is a website which provides the service of music recommendation based on music data collection and user interests. There have been many researches in the field of MRS, such as content-based music filtering system with revise user profile, the perspectives of user profile information in music recommender systems and the music recommendation based on music data collection and user interests.

A content-based music filtering system with a revisable user profile is using a decision tree in a content-based music filtering system. The perspectives of user

profile information is describing existential research into the agents influencing the trade-off between the sensed benefits of customization and the privacy 'costs' experienced by users.

Instead of textual descriptions, the music recommendation based on music data grouping and user interests considers the perceptual properties of music objects, such as pitch, duration, and loudness, which can be extracted from the music objects. For users, the preferences are derived from the access histories and recorded in profiles.

They have proposed the Content Based method to recommend the music objects that belong to the music groups the user is recently interested in. To know the recent interests of the user, they classify the latest transactions in the access history. The Content Based method recommends recently hot music to users according to the access history of users.

II. PROPOSED SYSTEM

In addition to searching expected music objects for users, it becomes necessary to develop a recommendation service, which provides reliable results to an individual. Any music application or systems now-a-days gives unrelated or a vast related data which becomes overhead for the user during music search process, which is very hard task to find the music required by the user at that point of time.

While an individual is searching for a song or music which he like to listen then the existing system will result the content of melody music of many albums which the user does not like, hence to overcome such kind of problems a music recommendation system is needed to give a personalized user search results.

We design the Music Recommendation System (MRS) to provide a personalized service of music recommendation to any user. The music objects are first analyzed. For each polyphonic music object, the representative track is first determined, and then certain features are extracted from this track.

Here, the user personalized data of user likes/dislikes, shares, comments, searching process via facebook, Google. Also time based estimation of song listening is considered for a better music recommendation system. We design music recommendation system by analyzing the music objects. The representative track is first determined and then certain features are extracted from the track**.** The content-based, collaborative and statistic based recommendation methods are proposed.

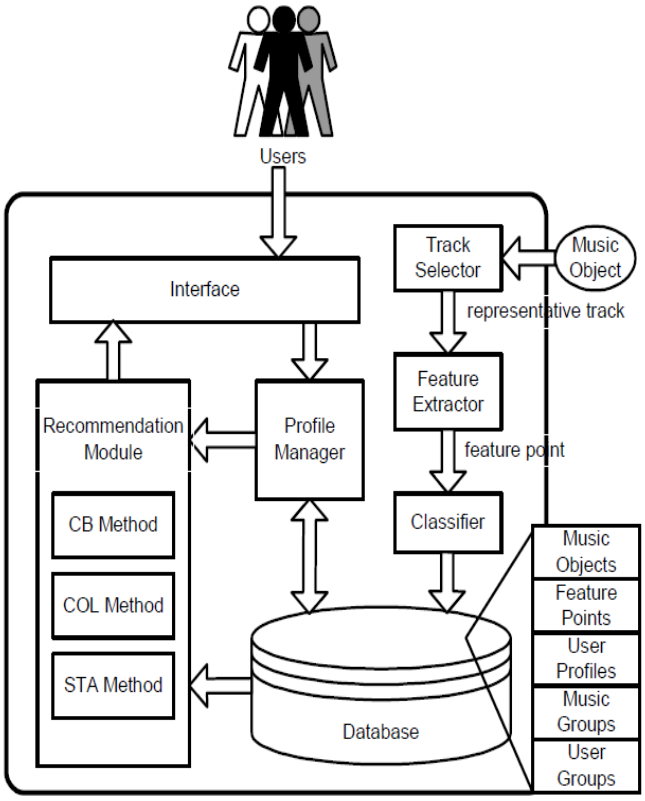

Fig1: Music Recommendation System

The User first enters through interface and accesses the profile and search for a track of any type of music. The application will result the track which he searched for along with some recommendations. Internally, a music object identifies the type of the track from the track selector. Now, the genre of the song is identified from the feature extractor. The extracted features are analyzed and classified accordingly in the data base.

Data base consists of music objects, feature points, user profile, music groups and user groups. Thus the profile manager will access the database and displays the requested results. The recommendations will be presented based on content based, collaborative, hybrid-based techniques.

2.1 Collaborative filtering

A collaborative recommender system tries to predict the utility of items a user, based on items previously rated by the other users who are similar. For example, when recommending books, a collaborative recommending system tries to find other users who have a history of agreeing with u (e.g. they tend to buy similar books or give similar ratings for books). Collaborative based recommender system can be memory-based or model-based.

Memory-based methods uses heuristics to make rating predictions based on the entire collection of items previously rated by users. That is, the unknown rating of an item user combination can be estimated as an aggregate of ratings of the most similar users for the same item. A k-nearest neighbor approach is used, i.e.., we find the k other users that are most similar to our target users. Various approaches can be used to compute the similarity between users. This approach defines both classification and regression of the query point. It is the easiest approach and computation is performed until classification. For both regression and classification weights are provide to the nearest and the distant objects, in this way we can define the nearest or the neighbors of the query point. The nearest neighbors contribute more to the average than the distant ones. K-nearest neighbor approach doesn't rely on the any fixed parameters, its completely a non-parametric approach and relies on means, variance etc.,

The most popular approaches use either pearsons correlation coefficient or cosine similarity. A weighted aggregate can be used, which adjusts for the fact that different users may use the rating scale differently. Model-based collaborative recommender systems use a collection of ratings to learn a model, which is then used to make rating predictions. For example, probabilistic models, clustering, Bayesian networks, and other machine learning techniques have been used.

2.2 Content Based filtering

Content Based filtering is a technique in information retrieval system and a query is used in IRS to retrieve the data . It is also referred as cognitive filtering, which recommends items based on compilation between the content of the objects and user profile. The content of each item is represented as a set of descriptors or terms, the words that occur in a document. The user profile is represented with the same terms and built up by analyzing the content of items which have been seen by the user.

Problems have to be viewed while implementing a content-based filtering method. When terms are allocated automatically a method has to be chosen that can take out these terms from music objects. Second, the terms have to be shown such that both the user profile and the objects can be compared in a significant way. Third, a learning algorithm has to be selected which is able to know the user profile based on seen items and can make recommendations based on this user profile.

A standard approach for term classification and prefer single words from documents. Some of the learning processes also represent the user profile as one or more vectors in the same multi dimensional space which makes it simple to compare the documents and profiles.

Through this content based system two things can be found such as recognizing the files, which the user has already expressed preference and the files which matches with the user profile. For example, if a user likes to purchase shirts with collar and according to his previous purchases it is found that the user prefers dark colors over bright colors. Keeping these previous purchase history the system can suggest a black color shirt with collar as this product is more likely to buy. Recommendation can also be given by taking feedback from the users on their likes/dislikes, preferences, search patterns etc.,

Francesco, the writer of Recommender system handbook, stated that the content based filtering requests the user to view and rate the data items. Based on these ratings a separate profile of user's interest can be prepared. This profile is helpful in displaying the user's interest. According to him, there are three steps in recommendation system.

1. Content analyzer: Main role is to represent the content and analyze the content and extracts the features of the data using feature extraction technique.

2. Profile learner: It collects the user's preferences, simplifies it and creates the user profile.

3. Filtering components: It tries to relate the user profile with items profile and recommends the data that suits best.

III. MERITS OF CONTENT BASED FILTERING

A. User Independence: Collaborative filtering needs user ratings to make the user list but content based filtering analyzes the user data and data items for recommendation.

B. This does not have the Cold Start Problem.

Demerits

A. Limited content analysis: User profile should be strong enough to get accurate recommendations.

IV. RELATED WORK

The music items in the database of the Music Recommendation System (MRS), as well as the incoming music items, are factors for music recommendation. When a new music item is fit in the database, it goes over the track selector and the feature extractor. According to the extracted features, the incoming music item is properly allocated to certain music collection by the classifier function block. When the user access a music object from the music list or the recommendation results, the account or the profile manager will record the item data into the access history.

The information of each accessed music item, i.e., the access time, the item ID, the corresponding music collection which the item belongs to, and the corresponding transaction is noted in the access history. Note that the transaction ID is single level increasing.

Many researchers made their attempt to prevent cold start problem in collaborative filtering by including content-based filtering.

Arbee L.P. Chen et al. have proposed the Content Based method to recommend the music objects that belong to the music groups the user is recently interested in. Instead of textual descriptions, they consider the perceptual properties of music items, such as pitch, duration, and loudness, which can be directly extracted from the music objects. For users, the preferences are derived from the access histories and recorded in profiles.

An interface or a website should be created for accessing the songs from any music application by considering and applying the required filtering algorithms like collaborative filtering, content based filtering and hybrid filtering.

These algorithms act as filters in providing recommendations and search results required by the user according to his interests which are gathered and recorded in the database through classification of music items and objects such as, genre, pop, rock music types are analyzed and classified into different types accordingly.

A music tracks can be ranked based on similarities and this way it can be stored in a homogeneous class, making music extraction process a bit simpler. That is when a music track is requested, depending upon its genre it's searched in that particular class and then the requested track is displayed.

Some of the researchers who incorporated both content-based and collaborative filtering are viewed in area of research.

McFee explained to find similarity on a sample of collaborative filter approach using an artist-level similarity approach. Here a bag-of-words are used for the music track feature extraction.

Weston et al. has investigated "Collaborative retrieval" which means the problem rose when a user's item is used as query item for another user. They had optimized their item scoring function for finding features in music. They also used the bag-of-words technique to identify the music features on large data

Both McFee and Weston et al. optimized their models using a ranking loss. They have opted to use quadratic loss functions instead, because they found their optimization to be more easily scalable.

In an approach proposed by Cano, the similarities of a song track to other are computed are based on meter, timber, rhythm patterns and tempo. Furthermore some of the tracks are classified depending upon the emotion of the song.

HU has proposed "Weighted Matrix factorization" for finding the no. of times a song is listened by a user. This WMF has been used in many content based filtering methods. Lieberman has developed Letizia, a user interface, which tracks and records the user information on the web which includes user search requests and browsing history. From those records, it tries to find the web pages of user interests. It provides rating facility for each web pages and collects the ratings from the user. Based on this it makes a list of user interests.

Pazzani created Syskill & Webert, which is a software agent that records and tries to determine the user's interested things on web.

A neural network model is proposed by Jennings and Higuchi, which collects the information on the news articles. This neural network is formed as a result of news articles which the user had read or ignored.

V. CONCLUSION AND FUTURE WORK

Recommending and personalization are important approaches to combating information over-load. Machine Learning is an important part of systems for these tasks. Creating a good relational database of music, making relations between artists, albums, musical genres and

époques, could greatly expand the capabilities of a recommender algorithm to help users discover new music.

REFERENCES

[1] Jiawei Han, Micheline Kamber and Jian Pei for data mining concepts and techniques, third edition

[2] Y.Hijikata, K. Iwahama and S. Nishida, "Content-based Music Filtering System with Editable User Profile," in Proc. of ACM Symp. On Applied Computing, Kennesaw, Georgia, 2006, pp. 1050-1057.

[3] E. Perik and B. de Ruyter, P. Markopoulos and B. Eggen, "The Sensitivities of User Profile Information in Music Recommender Systems," in Proc. of the 2nd Annual Conf. on Privacy, Security and Trust, New Brunswick, Canada, 2004, pp. 137-141.

[4] P. Cano, M. Koppenberger and N.Wack: "An industrial strength content-based music recommendation system," In Proc. of 28th ACM SIGIR, pp. 673, 2005.

[5] Yajie Hu, Mitsunori Ogihara "A music recommendation system based on human behaviour" in 12th International Society for Music Information Retrieval Conference (ISMIR 2011)

[6] Brian McFee, Luke Barrington, and Gert R. G. Lanckriet. Learning content similarity for music recommendation. IEEE Transactions on Audio, Speech & Language Processing, 20(8), 2012

[7] Aaron van den Oord, Sander Dieleman, Benjamin Schrauwen for "Deep content-based music recommendation".

[8]] L. Barrington, R. Oda, and G. Lanckriet, "Smarter than genius? Human evaluation of music recommender systems," in Proc. 10th Int. Conf. Music Inf. Retrieval, 2009

[9]] O. Celma. "Music Recommendation and Discovery in the Long Tail". PhD thesis, Universitat Pompeu Fabra, Barcelona, 2008

[10]] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. The million song dataset. In Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR), 2011

[11] Philippe Hamel and Douglas Eck. Learning features from "music audio with deep belief networks". In