A SOFT COMPUTING APPROACH FOR INTELLIGENT MUSIC RECOMMENDER SYSTEM ON THE BASIS OF MOOD AND SPEECH ANALYSIS

Robotics and Intelligent System Design Lab

Indian Institute of Information Technology and Management Gwalior, Madhya

Pradesh-474010, India

Ritu Tiwari[1], Pradip Swarnkar[2], Khushbu Suryavanshi[3], Monika Mahor[4] and Harsh Kumar Singh[5]

[1]tiwariritu2@gmail.com
[2]swarnakar@gmail.com
[3]khusic4@gmail.com
[4]monikamahore@gmail.com
[5]hs4056@gmail.com

***Abstract*—Along with the rapid expansion of digital music, managing and searching for the willing songs have become very difficult and tedious task which has come as a challenge to the software developers to provide the user with the recommendations user is really interested in.Since the available music repository is so vast and consists of different kind of music, selecting some particular songs of ones own choice from this big ocean is somewhat hectic.So there comes the role of a Recommendation system, guiding users through this big ocean of songs so that user can get to choose what he/she exactly wants to listen.The objective of this research is to study and to develop a music recommender system which recommend songs to the user on the basis of some solid algorithms and techniques.Analysis of the emotional mood and user's speech are the main highlights of the work which are based on neural network clustering including some content-filtering algorithms, user's listening history, popular songs are the main techniques that has been implemented in this project along with some content filtering algorithms and neural network clustering.**

***Keywords*—Music Recommendation, neural-network clustering, mood based classification, speech analysis, content filtering algorithms, speech emotion analysis**

.

I. Introduction

As the area of digital music is rapidly expanding and along with this, song anonymity is also increasing.So there is a need of an intelligent system that can manage user choices and preferences and provide a wise suggestion of songs as per a user's own taste.The research work aims to develop a recommender system providing the user with accurate and satisfactory recommendation list of songs on the basis of the current emotional state of the user by analysing his mood and speech and optimizing the list further on the basis of content-filtering algorithms such as popularity,listening history and genre to deliver a user friendly, accurate and efficient music recommendation system.Speech emotion analysis covers the science of recognition and synthesis of audio signals plus covers the vocal behaviour of the user to predict the mark of effects. The measurable voice parameters reflects the current state the user is currently experiencing.For example, arousal associated with an anger state produces change in respiration and increase in muscle tension which causes the vibration of the vocal folds and tract shape, affecting the characteristics of the speech which in turn can be used by the listener to predict the current state of the user, in the same way recommendation system can extract some features like Spectral Centroid, Strength Of Strongest Beat, MFCCs, Fraction Of Low Energy etc. from the audio clip and on the basis of some algorithms and clustering can predict the respective state of mind of the user.

II. Related Work

The main literature used here is the classification, prediction and consideration in recommendation with the help of some algorithms to get some correct results.Mood and speech analysis are the new advancements in the field of music information retrieval system that artificially scans the human

mind on the basis of the fluctuation in his voice as well as in his moods.Recommendation is a vast field of research which includes rapid development per year with highly accurate results so it is necessary to take into account as much factors as possible related to the user to make the final recommendations.

k mean clustering has been used to predict the moods on the basis of numerical features associated with each song and speech analysis is also based on extraction of feature of input speech.Audio features are the numerical values associated with every audio signal that helps differentiating one audio from the other , there exists a large number of features but it is recommended to use least possible features to present effective recommendations.[1] gives a deep insight of the existing features and why they are used and how they are calculated from a raw audio file in that field of audio analysis for example, K mean clustering is used to predict the moods on the basis of numerical features associated with each song.

After many researches it has been proved that emotions can be classified on the basis of dimensions there exists two-Dimensional-Arousal-valence model proposes that different emotions generate from separate neural networks.[4] used 3-D resonance arousal valence model, the two-Dimensional-Arousal-valence model becomes three-Dimensional when an additional dimension i.e resonance gets added.In order to differentiate even nearly related moods like fear and anger, in this model two dimensions are same as the previous model only third dimension depends upon musicology like continuity-discontinuity,melody-harmony contrast.Cold start, sparsity, big data, these are some demerits of using collaborative filtering therefore emotion based content filtering has been used.

Musical genre is basically a statistical property of an audio signal such as rhythmic structure, instrumentation etc automatic genre classification as given in [6].Intensive analysis on relevant audio features is performed in [12].2-path structure designed processes the prosodic and spectral features in parallel.[16] gives an idea about the melody extraction algorithms and aims to produce frequency patterns corresponding to the pitch of the dominant melody of an audio recording.Recommendations based on some more aspects other than evaluating the relevant feature has been consider in [15] such as some illegal songs or other illegal contents.This comes with two types of recommenders out-of-the box. The first uses traditional algorithms, like Content Based and Collaborative Filtering that can be used as baseline for comparison. The next recommenders is geared towards overcoming the cold start problem by first exploiting information provided during registration (Annual, Country, and City Charts recommender) and secondly leveraging knowledge from social networks (Social Neighborhood and Social Tags recommender), Annual Charts Recommender.Other referred papers are, [2], [13], [8], [18], [16], [5].

III. Methodology

Here, we briefly describe the approaches used to recommend the songs of user's choice. Mentioned below are the techniques applied.

A. Recommendation on the basis of user's previous listening history

We used Million Song Dataset provided by Kaggle consists of the songs user already had listened to, the dataset has triplets containing user ID, song ID, and count(the number of times user had listened to that song).Songs corresponding to a each user's ID have been extracted and stored to maintain each user's listening history.

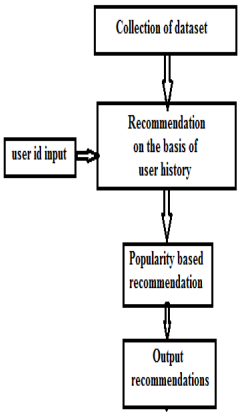

Fig. 1. Flow diagram of recommendation approach for dataset 3

B. Popularity based recommendation

This approach is the most basic and conventional algorithm where the system recommends those songs which are the most popular ones and the popularity of songs is calculated on the basis of number of times that song has been played by various users in the whole dataset.So the most played song is the most popular one and least played is least popular so the tracks been presented to the user are on the basis of decreasing order of popularity.

C. Recommendation using k-nearest neighbour algorithm

K nearest neighbour is an algorithm that takes a case and compares that case with the all pre-existing cases using a similarity measure using distance functions.Calculating distance

of a case from other all available cases to find the nearest neighbour is the key component of KNN algorithm and which can have strong effect on performance.Some of the distance measures used in KNN are Euclidean distance, Manhatton distance, Chebyshev distance.

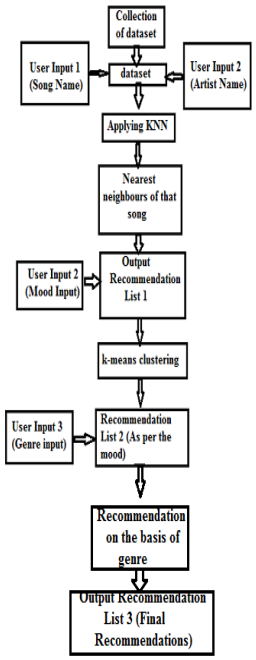

Fig. 2. Flow diagram of recommendation approach for dataset 2

D. K-Means Clustering

The k-means algorithm takes the input parameter k,and partitions a set of n objects into k clusters so that the resulting intracluster similarity is high but the intercluster similarity is low.Cluster similarity is measured in regard to the mean value of the objects in a cluster, which can be viewed as the cluster's centroid or center of gravity. In our project we have divided our dataset into 12 clusters on the basis of calculated distance(chebyshev) from centroid on the basis of various features extracted from the songs and assigning the appropriate mood to each cluster afterwards.

E. Speech emotion analysis

Speech emotion analysis covers the science of recognition and synthesis of audio signals plus covers the vocal behaviour of the user to predict the mark of effects. The measurable voice parameters reflects the current state the user is currently experiencing.For example, arousal associated with an anger state produces change in respiration and increase in muscle tension which causes the vibration of the vocal folds and tract shape, affecting the characteristics of the speech which in turn can be used by the listener to predict the current state of the user, in the same way recommendation system can extract some features like Spectral Centroid, Strength Of Strongest Beat, MFCCs, Fraction Of Low Energy etc. from the audio clip and on the basis of some algorithms and clustering can predict the respective state of mind of the user.

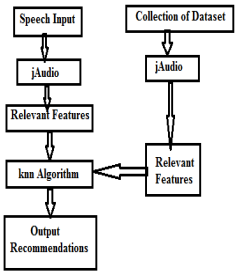

Fig. 3. Flow diagram of recommendation approach for dataset 3

F. Content-based-filtering(genre based music recommendation)

Further the songs have been classified on the basis of genres after the application of KNN algorithm.The system filters the songs again on the basis of different genres like classical pop and rock, pop, folk, hip hop, metal, soul and reggae, punk and classical as per the user's choice.

IV. Dataset Details

A. Dataset 1(Kaggle's million song dataset challenge)

The dataset consists of logs of user's listening pattern of songs.The contents are as follows: (i)

1) User Id
2) Song Id
3) Song count i.e. the number of times user has listened the song.

Dataset provided by Kaggle's million song dataset challenge was released by Columbia University Laboratory for the best possible evaluation of offline music recommendation systems. The dataset is open, everything is possibly known and freely available.This consists of around million songs in the form of triplets(userID, songID, play count) collected from the history of over one million users and metadata of millions of songs, metadata is all about the features extracted from the other features. But the users are anonymous here and thus information about their personal liking and timestamps of listening events are not available.
Reference: https://www.kaggle.com/c/msdchallenge/data

B. Dataset 2(Million Song Dataset)

Million song dataset is freely available group of extracted features and metadata of over 1 million contemporary popular songs.The Million Song Dataset is a collaborative project between The Echo Nest and LabROSA.The dataset is huge(280 GB) so subdataset of 10,000 songs(1.8 GB) can be used for a quick taste.Some of the features given in the dataset are as follows: (i)

1) genre
2) track id
3) artist name
4) title
5) loudness
6) tempo
7) time signature
8) key
9) mode
10) duration
11) avg timbre
12) var timbre

Reference:
http://labrosa.ee.columbia.edu/millionsong/pages/getting-dataset

C. Dataset 3(Dataset for speech emotion analysis)

In order to meet the requirements of some additional features to analyse the voice signals, the dataset is designed with some additional popular songs and their required features to make predictions.The features are extracted by using jAudio software with the Million song Dataset.Some of the features given in the dataset are as follows: (i)

1) Spectral centroid overall standard deviation
2) Root mean square
3) Fraction of low energy windows
4) Strongest beat
5) Zero crossing
6) Method of moments etc

Here k nearest neighbour algorithms is used to find nearest neighbours of input speech of user.Features of input speed is calculated through jAudio feature extraction software then knn is applied to the dataset of songs.So the nearest neighbours on the basis of speech analysis i.e., songs on the basis of current mood of user are then recommended.

V. Results

A. Output recommendations for dataset 1

The first recommendation list is the outcome of the first content-based-algorithm that is on the basis of popularity and user's listening history.All popular songs that are not present in user's listening history are extracted along with the history of the user.

B. Output recommendations for dataset 2

Final recommendation list for the following user successive inputs:
song name: Love Song
artist name: Sara Bareilles
mood choosen: Happy
genre: Classic Pop and Rock

TABLE I. Final Output recommendations

title	artist name
Talking Bout Love	Martha and The Vandellas
New York City Here I Come	Albert Hammond
Mercenary Territory [Live Album Version]	Little Feat
Just Comin' Down With Love	Jeannie C. Riley
The Love I Never Had	Tavares
Your Imagination (Album Version)	Brian Wilson
Rich Get Richer	The O'Jays
Tu Dis Que Tu M'aimes [How Deep Is Your Love?]	Jonatha Brooke
The Answer Is Yes	RODNEY CROWELL
You're Such A Beautiful Child	Archie Bell and The Drells
Introducing The O'Jays	The O'Jays

C. Output Recommendations for dataset 3

For speech emotion analysis, input is taken in the form of user speech as for sample input the output recommendations are shown in the table below.

input: Speech5(anger)

The output recommendations are as follows:

VI. Conclusion and Future Scope

The Collaborative filtering methods on user histories are not enough to produce an effective recommendation systems.Thus this system combines all advancements as well as traditional algorithms to put the best result in front.The system successfully able to classify the big dataset on the basis of moods and speech.It involves classifying the song on the basis of some value(text string) entering into the system and also on the basis of, by just giving a token of speech.The effect of features extracted through jAUDIO were successfully reflected in the results produced.The accuracy of the project is up to the mark and can be applied in the real life world and further accuracy can be enhanced by increasing the dataset and evaluating the audio features more precisely. Overall, system is fulfilling all the requirements enlisted earlier:

1) Recommendations related to the input song.
2) Recommendations on the basis of mood.
3) Filtered the songs on the basis of genre, popularity and listening history.

TABLE II. OUTPUT RECOMMENDATIONS FOR SPEECH EMOTION ANALYSIS

title
Natural Born Killer.wav
Livin on a Prayer.wav
With You.wav
Michael Jackson - Rockin Robin.wav
Michael Jackson - Ben.wav
ACDC - 01 - T.N.T..wav
ACDC - 01 - T.N.T..wav
LimpBizkit- Nookie.wav
Michael Jackson - P Y T (Pretty Young Thing).wav
Michael Jackson - You Rock My World.wav
Limp Bizkit - Show Me What You Got.wav
14 adams - summer of 69!s.wav
14 adams - summer of 69!s.wav
ACDC - 18 - Heatseeker.wav
High Voltage.wav
Limbizkit - Mission Impossible 2.wav
Michael Jackson - Dirty Diana.wav
Michael Jackson - Give In To Me.wav
Michael Jackson - Blame It On The Boogie.wav

4) Recommendations on the basis of analysing audio speech or audio signal.

FUTURE SCOPE: Future work can be extended by making the system fully dynamic and implementing the search algorithms by using voice clip as input by search engines and various music service providers, providing an appropriate and accurate results.

REFERENCES

[1] A study of audio classification on using different feature schemes with three classifiers,2010 International Conference on Information, Networking and Automation (ICINA)

[2] Content-based music audio recommendation, Cano, Pedro and Koppenberger, Markus and Wack, Nicolas, Proceedings of the 13th annual ACM international conference on Multimedia, 2005 ACM.

[3] Music recommendation, Celma, Oscar, Springer, 2010.

[4] Emotional States Associated with Music: Classification, Prediction of Changes, and Consideration in Recommendation, Deng, James J and Leung, Clement HC and Milani, Alfredo and Chen, Li, ACM Transactions on Interactive Intelligent Systems (TiiS), 2015.

[5] Are the emotions expressed in music genre-specific? An audio-based evaluation of datasets spanning classical, film, pop and mixed genres, Eerola, Tuomas, Journal of New Music Research, 2011.

[6] Automatic Musical Genre Classification of Audio Signals, George, Tzanetakis and Georg, Essl and Perry, Cook, Proceedings of the 2nd International Symposium on Music Information Retrieval, Indiana, 2001.

[7] Data mining: concepts and techniques: concepts and techniques, Han, Jiawei and Kamber, Micheline and Pei, Jian, Elsevier, 2011.

[8] Experimental studies of the elements of expression in music, *Hevner, Kate*, The American Journal of Psychology, 1936.

[9] jAudio: Towards a standardized extensible audio music feature extraction system, McKay, Cory, Course Paper. McGill University, Canada, 2009.

[10] A mood-based music classification and exploration system, Meyers, Owen Craigie,Massachusetts institute of technology, 2007.

[11] EmuPlayer: Music recommendation system based on user emotion using vital-sensor, Nguyen, TL, Kanagawa, Japan: Keio University, 2010.

[12] A new approach of audio emotion recognition, Ooi, Chien Shing and Seng, Kah Phooi and Ang, Li-Minn and Chew, Li Wern, Expert Systems with Applications, 2014.

[13] Music recommendation, Celma, Oscar, Springer, 2010.

[14] A literature review and classification of recommender systems research, Park, Deuk Hee and Kim, Hyea Kyeong and Choi, Il Young and Kim, Jae Kyeong,Expert Systems with Applications, 2012.

[15] MUSE: A Music Recommendation Management System, Przyjaciel-Zablocki, Martin and Hornung, Thomas and Schätzle, Alexander and Gauß, Sven and Taxidou, Io and Lausen, Georg.

[16] Melody extraction from polyphonic music signals: Approaches, applications, and challenges, Salamon, Justin and Gomez, Eva and Ellis, Daniel PW and Richard, Guilhem, Signal Processing Magazine, IEEE, 2014

[17] Which emotions can be induced by music? What are the underlying mechanisms? And how can we measure them?,Scherer, Klaus R, Journal of new music research, 2004

[18] Neuroscience projections to current debates in emotion psychology, Scherer, Klaus R, Cognition & Emotion, 1993.

[19] Emotional states generated by music: An exploratory study of music experts,Scherer, Klaus R and Zentner, Marcel R and Schacht, Annekathrin, Musicae Scientiae, 2002.

[20] Recognizing human emotional state from audiovisual signals, Wang, Yongjin and Guan, Ling, Multimedia, IEEE Transactions on, 2008.