MUSIC BY MOOD: CUSTOMIZED MUSIC RECOMMENDATION SYSTEM USING DCNN

REEVA ANDIPARA[1] , JAHNVEE JOSHI [1]

[1] Students of Vyavasayi Vidya Pratishthan Engineering College, Rajkot, Gujarat, India

ABSTRACT— *Facial expression is the means of non-verbal communication. It depicts the emotional status of a person. Using this emotional status, personalized songs can be suggested to improve user listening experience. Facial Expression based music recommendations aims at scanning and interpreting the data and accordingly recommending a list of songs based on the parameters provided. Associating the user data (facial expression) with the music is a difficult task. In this paper, we propose a customized music recommendation system (CMRS) to correlate the user data and the music. To accomplish this correlation, we have used deep convolution neural network. Extracting the mandatory input from human face is done directly from the camera. This input is then used to deduce the information like emotional status of the user. For this, the detection of face is done and the output is then given to emotional recognition module that uses Convolution Neural Networks (CNN) for recognizing the emotion of the detected face. CNN uses the facial expression and recognizes one to the 7 possible moods: happy, sad, angry, disgust, neutral, surprise and fear. This data is then used to get a list of songs that are in accordance with "mood" obtained from the input provided earlier. To achieve this, Deep Convolution Neural Networks (DCNN). This approach is used to extract the latent features from music data (e.g., metadata) for classification. So, the DCNN uses the user-data and the music data for classifying the songs based on user's mood [1]. The songs are classified and as per the mood of the user and then the recommendation is done in recommender module (CMRS). CMRS recommends songs to the user by mapping the emotion of the user to the song, considering the preference of the user.*

Keywords—Convolution Neural Network, Music Recommender System, Emotion Recognition, Deep Convolution Neural Networks, Face Detection

I. INTRODUCTION

In the current years, the music industry has shifted towards digital distribution through various online music stores and streaming services such as Spotify, Gaana and Google Play. Due to this, automatic music recommendation has become a progressively pertinent problem: it allows listeners to learn new music that matches their tastes, and enables online music stores to aim their wares to the right audience. Although recommender systems have been studied broadly, the problem of music recommendation in particular is complicated by the variety of different styles and genres, as well as social and geographic factors that influence listener preferences and their current mood [10]. The number of dissimilar songs that can be recommended is large, especially when recommending specific songs. Many recommender systems rely on usage patterns: the combinations of songs that users have listened to or rated provide information about the users' preferences, and how the songs relate to each other [2].

Convolutional Neural Networks (CNNs) is successfully used in the field of image classification. The input image to the system can be captured using a camera. This image undergoes image enhancement, where images are mapped based on the tone with low contrast to restore the original contrast of the image. That image will be input of the emotion recognition and using CNN it will give one of the moods as output. With help of some basic features of face, like eyebrow, lip etc. mood will be classified in seven categories. Those seven categories are: Neutral, Happiness, Surprise, Fear, Disgust, Anger and Sadness [3][5].

Mood	**Music**		
	Song	***Album***	***Artist***
Angry	Damaged I	Damaged	Black Flag
Happy	Uptown Funk	Uptown Funk	Mark Ronson
Surprise	I gotta Feeling	Black Eyed Peas Invasion Of I Gotta Feeling - Megamix E.P.	Black Eyed Peas
Sad	Breath Me	Colour the Small One	Sia
Fear	In Twenty Years or So	Pure Comedy	Father John Misty
Disguist	Reject	Nimrod	Green Day
Neutral	What a Beautiful day	Tales Of Ten Men	Brett Every

CNNs classify images, cluster them by similarity and perform object recognition. They are primarily used for image processing but can also be used for other types of input such as audio. A characteristic use case for CNNs is where you feed the network with images and the network classifies the data. CNNs start with an input "scanner" which is not intended to parse all the training data at once [8].

CNN compares the images piece by piece. The features are the pieces that CNN looks for. The filters are moved through an image in the first layer known as Convolution layer and convolution is performed. Rectified Linear Unit (ReLu) activates a node only if the input is above a certain quantity, while if the input is below zero then the output is also zero, but when the input rises above a certain threshold, it shows a linear relationship with the dependent variable. Next comes the pooling layer. This layer is used to shrink the image, i.e. reduce the size of the image. Stacking up of all layers is done after pooling layers. Fully connected layers are obtained when all shrinked images are put in a single list. This is the final layer and the actual classification is done in this layer.

The deep convolution neural network (DCNN) is used to classify the songs on the basis of latent features of the songs that is obtained from metadata of the songs [9]. Recommender systems are used to generate and provide suggestions for the user according to their preferences. In the Customized Music Recommendation System (CMRS) we are using a hybrid recommender system which is a combination of two different recommender strategies in different ways to benefit from their complementary advantages. In this case combination of Content- Based filtering (CBF) and Collaborative Filtering (CF) is used.

Collaborative Filtering

This approach works on the assumption that people who agreed in the past will agree in the future, and that they will like similar kinds of items as they liked in the past [7]. The system generates recommendations using only information about liking of different users or items. Collaborative filtering methods are classified into two models: memory-based and model-based [4]. A vital benefit of this approach is that it does not rely on machine analyzed content and therefore it is capable of precisely recommending complex items such as music without requiring any "understanding" of the item itself.

Content Based Filtering

It is based on a description of the item and a profile of the user's preferences. These methods are best suited to situations where there is known data on an item, but not on the user. Content-based recommenders treat recommendation as a user-specific classification problem and learn a classifier for the user's likes and dislikes based on an item's features. These algorithms recommend items which are similar to those that a user did liked in the past, or is examining in the present. It does not rely on a users' temporary profile. Features mined from the user-generated reviews are improved meta-data of items because as they reflect features of the item like meta-data, mined features are broadly concerned by the users [11].

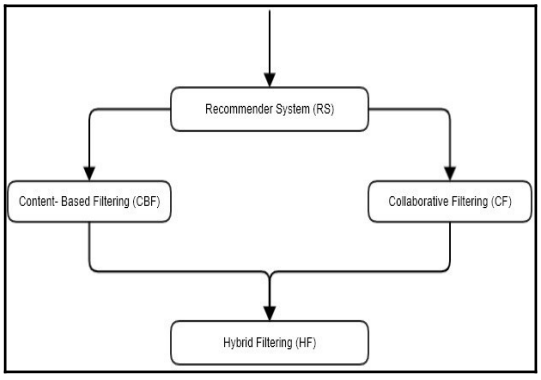

Fig. 1. Hybrid Filtering

With exponential rise in music content, people now have access to music collections on unprecedented scale, as 1 million tracks are released every six weeks [12], there is requirement of good recommender system (RS), which will suggest song not only according to mood but also according to penchant of an individual.

II. PROPOSED WORK

For recommending music on the basis of mood of the user, emotion recognition is done using CNN approach. CNNs are used to stimulate the human brain when scrutinizing visuals. Thus, a CNN is used to construct a model which successfully classifies emotions in seven moods: happy, sad, angry, neutral, disgust, fear and surprise.

- Graphical Input: This is the front-end of the proposed architecture. User's face acts as an input which is seized through the webcam. This input serves as a medium that provides us with the user information in the form of facial expression. User's emotions are identified through their expressions.
- Face Detection: Face detection is done using classification. Classification is a practice of categorizing the given set of values in different classes on the basis of their common features. Supervised learning can be used for classification. In these techniques the training dataset is associated with their consequent labels.
- Emotion Recognition: Now, the labelled dataset of faces is available that corresponds to one of the 7 different moods as mentioned. These datasets are then converted into vectors which are convolution neural networks (16- layer convolutional neural networks). CNN is a neural network that contains input layer, output layer and several other hidden layers**.** The hidden layer consists of an activation function that is a ReLu. Fully connected layers are obtained when all shrinked images are put in a single list. This is the final layer and the actual classification is done in this layer. This process will result in the classification of the image in what category they fall which in our case is the expression of the user and we get the intended output.
- DCNN: So, from the emotion recognition module, we get the mood and emotion of the user. The DCNN approach uses the user-data like mood and metadata like genre, artist name, producer etc. presented in the songs to classify the songs. The DCNN approach stores the songs according to their classification as a part of music data available in form of metadata. This set of classified songs list is given as an input to the CMRS.

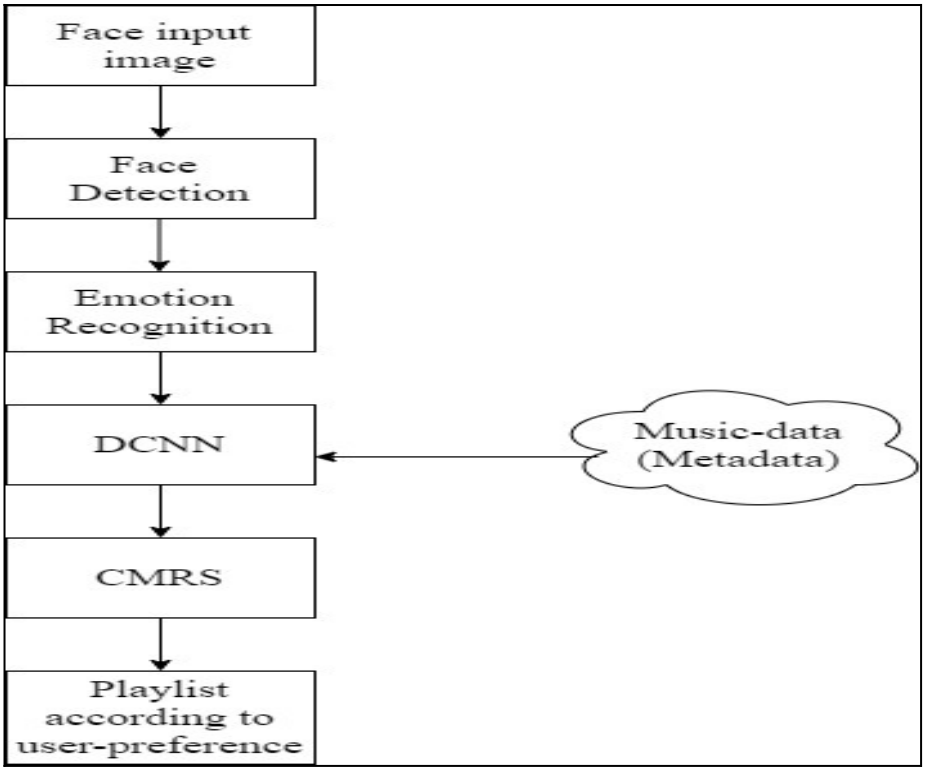

Fig. 2. System Architecture

- CMRS: CMRS is the core layer where the recommendation is deployed. This is aprocessing layer. The actual recommendation is done in this layer. The set of output obtained after the DCNN is collected and the customized songs are recommended to the user by hybrid approach of recommendation. The recommendation is done with the combination of CF and CBF.

III. Conclusion And Future Scope

This customized music recommendation system recommends songs on the basis of the user emotions at a particular time and according to their preferences. We request users to provide their facial expression as an input for recognizing the emotional status of the user which is given as an input to the DCNN for the classification of songs. DCNN provides us with the list of the songs that are mapped by the mood of the user to the music data. CMRS results the user with the songs as per their preference and their mood. Thus, CMRS reduces the user efforts of generating the playlists. It will provide users with the better delight by providing them with the most appropriate songs according to their current emotions.

Future Scope:

- Development in android application for better comfort.
- Can be used even for determining the emotions of mentally challenged people.
- Determine the drowsy mood while driving.

IV. Abbreviations

- CNN: Convolutional Neural Networks
- DCNN: Deep Convolutional Neural Networks
- RS: Recommendation System
- CF: Collaborative Filtering
- CBF: Content Based Filtering
- CMRS: Customised Music Recommendation System

Acknowledgment

A special thanks to our Guide Prof. Sagar Virani who supported us in completion of the paper. We would also like to thank our Head of Department of Computer Engineering Department, Dr. Tejas Patalia for his motivation and support. Lastly we would also like to thank all the faculty members and non-teaching staff member of the Department of Computer Engineering, Vyavasayi Vidya Pratishthan Engineering College.

References

[1] Ashu Abdul, Jenhui Chen, Huan-Yuan Lion, Shun Hao Chang “An Emotion- Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach” MDPI Journals Applied Science Volume 8, Issue 7

[2] Aaron van den Oord, Sander Dieleman, Benjamin Schrauwen, “Deep Content Based Music System” NIPS Paper

[3] T. Kanade, J. F. Cohn, and T. Yingli, "Comprehensive database for facial expression analysis, " in Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, 2000, pp. 46- 53.

[4] Kim, H. T., Kim, E., Lee, J. H., & Ahn, C. W. (2010,
April). A recommender system based on genetic algorithm for music data. In 2010 2nd International Conference on Computer Engineering and Technology (Vol. 6, pp. V6-414). IEEE

[5] F. Abdat, C. Maaoui and A. Pruski, “Human- computer interaction using emotion recognition from facial expression”, 2011 UKSim 5th European Symposium on Computer Modelling and Simulation

[6] Shravanthi K, Sneha Rao, Swati Joshi, Vandana, "A Personalized Music Recommendation System"

[7] Anand Neil Arnold, Vairamuthu S., "Music Recommendation using Collaborative Filtering and Deep Learning"

[8] Abhijeet B. Benke, Shubham S. Jadhav, Swapnil A. Joshi "Emotion Based Music Player Using Facial Recognition"

[9] Shlok Gilda, Husain Zafar, Chintan Soni, Kshitija Waghurdeka "Smart Music Player Integrating Facial Emotion Recognition"

[10] Mirim Lee, Jun-Dong Cho. "Logmusic: Context- Based Social Music Recommendation Service on Mobile Device", Ubicomp '14 Adjunct, September 13- 17, 2014, Seattle, WA, USA.

[11] Elena Kirzhner "Machine Learning. Explanation of Collaborative Filtering Vs Content Based Filtering", May 8 2018, [Online] Available:

https://codeburst.io/explanation-of-recommender- systems-in-information-retrieval-13077e1d916c

[12] Journal of Musical Things "How many songs are added to streaming music everyday" [online] Available: https://www.ajournalofmusicalthings.com/youll-be- stunned-at-how-many-songs-are-added-to-streaming- music-services-every-day-i-was/