Proc. of Int. Conf. on Intelligence Computing & Information Technology , ICIT

Music Emotion Detection: Transfer Learning Approach

Pallavi Jaiswal[1] and Prof. Bhavana Tiple[2]
[1] Maharashtra Institute of Technology,Department of Computer Engineering, Pune, India
Email: pallavijaiswal2330@gmail.com
[2] Maharashtra Institute of Technology,Department of Computer Engineering, Pune, India
Email: bhavana.tiple@mitpune.edu.in

***Abstract*—Music plays an important role in day to day life. It has power to stimulate various emotions within us. In the field of psychological studies emotion detection has become important research topic. Emotion detection in music will support bonding between human and music. This will not only help to understand psychological state of mind but also provide improved music information retrieval system. We have selected term Raaga as the basic model of our work were different emotions are associated different Raaga. In current work we are using Indian Classical music database. In the proposed system, music features are selected and extract to map the relationship between music and emotion. Transfer learning is the upgrading version of machine learning. Were our current task is formed by using knowledge learned while implanting related domain task. Thus, in the proposed system we will be using transfer learning method to map various emotions from music.**

***Index Terms*— Music, Transfer Learning, psychological state.**

I. INTRODUCTION

Music information retrieval is hot research topic which helps to classify the psychological state of human mind. The difficulty in music classification is to tag each music file on the basis of genre, mood, etc. it is tough task to determine the emotions in music as users are interested in different types of music.in our work we are focusing in Indian Classical music. Emotions have always been the important entity of the human experience and basic material for thousands of books and stories, but they have been interesting topic in science and technology too. Emotions are essential concept in human communication. To recognize emotions is also interesting topic in psychological studies and behavioural science. Thus, automated systems will improve quality and speed where most of work is processed manually. A major assumption in machine learning approach is that training and testing data must be in same feature space or domain and same distribution. Emotions vary from person to person. Hence, it becomes complex task to detect emotions. Documenting and tagging of data of all possible emotions would take more time and money. Due to large variety of data emotion detection is big problem. In this paper, emotion detection and transfer learning are studied and represented.

A. Indian Classical Music

Indian classical Music's basic component is Raag, Swara and Rasa. These raga is associated with different

Grenze ID: 02.ICIT.2018.2.505

pitch and melody. Raaga is the combination of swara i.e Sa, Re, Ga, Ma, Pa, Dha, Ni, Sa. Indian classical music specifies nine types of moods associated with Raaga. These are Shringar, Hasya, Karuna, Anger, Veer, Bhayanaka, Veebhasta, Adbhuta and Shanta, These moods are called as Raasa in Indian Classical Music. Indian Classical music has classified into two types i.e. Carnatic Music and Hindustani Classical music. Carnatic music is basically related to religious songs or Bhakti songs. Hindustani classical music is North Indian folk songs.

TABLE I. NINE EMOTIONS IN INDIAN CLASSICAL MUSIC

Raasa	Short Description
Shringar	Love
Hasya	Humor
Karuna	Pathos
Rudra	Anger
Veer	Heroism
Bhayanaka	Terror
Veebhatsa	Disgust
Adbhuta	Wonder
Shanta	Calm

The above table depicts various emotions in Indian classical music. Shringar rasa gives the feeling of love. Hasya rasa gives the feeling of joy or humor .Karuna rasa gives the feeling sadness. Rudra rasa gives the feeling of angerness. Veer rasa gives the feeling of bravery or heroism. Bhayanaka gives the feeling of anxiety, fear. Veebhasta gives the feeling of disgustness. Adbhuta gives the feeling of surprise or wondering. Shanta rasa gives feeling of peace and calm.

II. EMOTION WHEEL

Understanding human emotions is a complex task. Various tools and techniques have been designed across various domains to understand and conclude human emotions. Different emotion-mapping models were proposed to detect and measure human emotions. There are variety of emotions and basically it is classified in three categories i.e. Positive, Negative and Neutral. [4]In the book, “The expression of Emotions in the Man Animals” explains the connection between human emotions and movements[11].

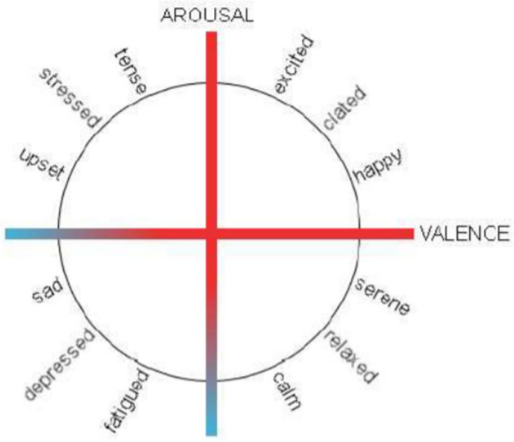

Figure 1. Circumplex Model

The Fig 1. Shows the circumplex model,2013. The circumplex model specifies that how human feelings are isolated in a two-dimensional circular space model. This includes valence and arousal dimensions. Arousal is

represented in vertical axis and valence in horizontal axis, while the center of the circle represents a neutral valence or neutral emotions and a medium level of arousal. These models have been most commonly used to test the words related to emotions, facial expressions with emotions, and affective states. Circumplex model was represented in different ways according to the use of application and emotions required. In dimensional emotion theory, emotion are the combination of arousal and valence which depicts the degree of pleasantness and intensity of mood. Example excited emotion could be thought as high arousal and positive valence. This wheel combines various emotions according to behavioural and evolution mechanisms [4].Example love is mixture of joy and trust which falls under Happy emotion. By referring this model we can manually plot basic nine emotions i.e. rasa of Indian classical music in four quadrants.

III. Emotion Detection

Question arises in mind that, How to recognise emotions? Initially emotions classification comes naturally within us. While discussing regarding emotions one thing crosses in mind i.e. facial expression. In social messaging applications like Whatsapp or Facebook we share emotions by emoticons (emoji or smileys). Body gestures can classify deep emotions into positive and negative emotions. Speech is very important part for human communication. The challenge is to link emotional state and conversations (Human Conversation). Hence it is difficult to classify emotion in speech and it is more difficult to detect emotion in Music. The problem with all emotional content is to measure psychological state linked to emotions. Music evolved various emotions which depends on psychological state of human mind. If a person is frustrated he will feel a song as irritating one but same person is in good mood he will enjoy that songs. Hence it is very difficult to map emotions with music. To map audio features with emotions will initially refer basic features for different emotions like tempo and pitch. For example audio features of shringar rasa there will be fast tempo and normal pitch. This framework especially uses Indian Popular Hindi songs. We have studied various data classification algorithms in order to understand, train and test the model representing the moods of these audio songs. We have studied speech emotion recognition system that how emotions are associated with audio.

IV. Review of Literature

Emotion, is deep conscious experience characterized by intense mental activity. It is response to scenario which indirectly shows psychological state of human mind.Sinno Jialin Pan, Qiang Yang[1] describes transfer learning concept and its method. It depicts that knowledge gained while implementing framework is used in target domain. Various methods like Feature based, instance based, and relational parameter based method is explain in depth. In transfer learning we don't need to start from scratch. We can refer related domain model for labelling the data which saves time for tagging of data.Felix Weninger, Bjrn Schuller[2],proposed model which provides indexing and retrieving the most interesting and important events of the action movie to attract the viewer by using the classified audio categories. audio features is extracted and make the model by using audio feature vector and classify the audio class to detect and recognize video scenes by using SVM.As an outcome Various action are classified by using SVM-HMM classification.[3] Pan, S. J., and Yang, Q,various features are classified. Classification is done by Standard Gausian Framework. Different features are used like MFCC, PA , AFTE etc. Then there frequency values are evaluated. With help of standard Gaussian process these frequency values are compared. After comparison it is confirmed that these features are used for classification of audio file, whether it is speech or music.[4] Priyanka Lokhande, Bhavana Subhash Tiple depicts comparisons of outcomes using various machine learning approaches like Deep belief network, Support Vector Regression, Fine neural networks etc.[5] Adnan Mehmood Bhatti, Muhammad Majid, Bilal Khan, user listens music by using EEG headset. To select features for classification brain signals are recorded via EEG headset. SVM is used for detecting the emotions. Age and genre wise emotions are detected. Music plays a very important role in our life as it reveals various emotions. Plenty of work has been done on emotion detection from western music and less in Indian music using traditional machine learning approach. So we find it a challenging task to see how we can utilise the relevant done on emotion detection and take it further to process using transfer learning. After literature survey we had refer audio features like MFCC, Spectrum and temporal features to map emotions with their feature values.

V. Transfer Learning Approach For Music Emotion Detection

Transfer learning is one of the type of machine learning where it recurrently uses constraints that are used to share knowledge between the source and target domains. The objective of transfer learning is to save resources, time and money on tagging the data. Example is pre-training of neural network in NLP (Natural Language Processing) for embedding of word. Another example is multimedia web mining.Transfer Learning saves time on labeling the data.We can labeled limited set of data. It uses trained CNN. Eventually we don't need to trained network for feature learning It provides decent accuracy rates even if training sets are of limited length.Several studies has been done on Music Information Retrieval system.[5] Hamel proposed to learn music features using embedding of mel-spectogram, genre, and tags of music file. Transfer learning refers knowledge of refers knowledge of domain related task to current task in order to implement model faster than traditional machine learning. In this approach initially we had studied speech emotion detection model were features of audio are almost same in speech and audio. Only difference is in speech we had vocal distribution but in audio there are melody in background and vocal distribution. Based on this data is being labeled according to nine emotions. The study by Deng, J depicts automatic emotion in whispered speech were autoencoder audio features are transferred between test and training data to evolve common features from it.

VI. Proposed System

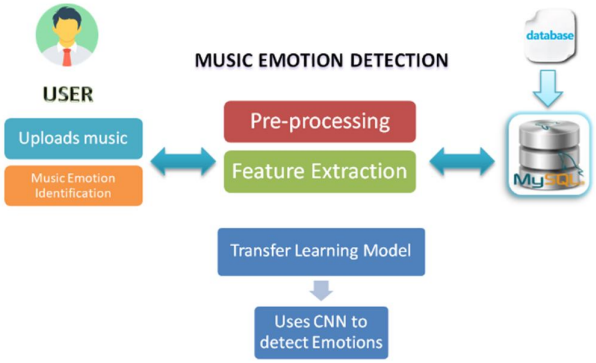

Figure 2. Proposed System

In proposed system we are using transfer learning approach. Initial task is to label data. This will help to extract and select features for predicting the music emotion. As discussed earlier that we will be using Convolutional neural network (CNN). Two layers of CNN are used. CNN are almost similar to ordinary neural network. A CNN is made up of layers and each layer of CNN converts fixed functionalities into Differentiable function that may or may not have any parameters. To avoid overfitting i.e failure of fitting in the network dropout function is used and data for future observation large amount data sets needs to be used in order to classify various emotions with audio features. The above fig shows the mechanism of detecting emotions by extracting audio features and transfer learning model. Below points represents the step of system architecture.

A. Preprocessing

Initially user will upload audio file. Audio clip can be in mp3 format or wav format. System will convert this file into .wav format file by using medio.io converter. System converts music file into waves which is the lossless format for audio formatting and it helps to extract best audio features from music file. Audio file is then segmented into 30 secs file in order to get analysis on data for further processing.

B. Feature Extraction and Selection

After getting pre-processed data, audio features are extracted using Audio feature extractor. In this phase we upload set of audio files then extractor will extract relevant features of audio. We have got features like Melody contour, Temporal feature, pitch feature and energy features. For Features selection we had used on of the data mining tool to select features for training and testing purpose.

C. Transfer Learning Model

This is the main processing unit. We had refer Speech emotion detection model for labelling the data. Depending on the pitch value and temporal value our data is been labelled. CNN is used for classification of music emotion. It has two layers. First layer has charge of doing heavy lifting i.e to select high level of feature without which rest functionalities is impossible. Initially will input selected audio features and weights to input layer of CNN with one bias. Network will start learning and calculated the net input function. To avoid linearity in classification we had use RELU activation function. While convolving to take iterations we had used zero padding and max pooling function and by using features emotions are labelled . For training purpose we had used 65 songs and for testing purpose we had use 29 songs. We are considering 24 frames each frame consist of three parameters i.e. energy value, pitch, time and music file. By using 2x2, 3x3 kernel size iteration is done. Final output gives the classified labelled of emotions on audio file. Thus CNN itself trains and select the features to label emotion classifier.

VII. CONCLUSIONS

The Music emotion detection in music is very important topic in science and technologies. Our objective is to identify the emotion associated with particular music. In this work we will be mainly focus on audio features to get the efficient result. Existing work on Music emotion detection is studied. Almost the work on music emotion detection is done by using Western music. In this framework our data sets is Indian music. As emotions are vary from person to person. A classified framework is developed to identify music emotion by using transfer learning approach. We have used CNN to classify the features into emotions. Transfer Learning approach helps to implement fast as compared to machine learning. This approach provides good accuracy results using limited set of database.

REFERENCES

[1] Sinno Jialin Pan, Qiang Yang ”A Survey on Transfer Learning”, 2010 IEEE Transactions on Knowledge and Data Engineering Volume: 22 Issue: 10.

[2] Felix Weninger, Bjrn Schuller,”On-Line Continuous-Time Music Mood Regression With Deep Recurrent Neural Networks”Felix Weninger On 02 April 2015

[3] Pan, S. J., and Yang, Q”Sentiment Analysis Where AI Meets Psychology”, 2012 IEEE Transactions on knowledge and data engineering.

[4] Priyanka Lokhande, Bhavana Subhash Tiple,” A Framework for Emotion Identification In Music: Deep Learning Approach”, International Conference on Intelligent Computing and Control Systems ICICCS 2017

[5] Adnan Mehmood Bhatti, Muhammad Majid, Bilal Khan, Human Recognition and analysis in response to audio music using Brain signals”,ELSEVIER 2016.

[6] Laurier, Herrera,”Audio Music Mood Classification Using Support Vector Machine”,2010 IEEE Transactions on knowledge and data engineering .

[7] Rajshri Pendekar, S. P. Mahajan, Rasika Mujumdar, Pranjali Ganoo,”Harmonium Raga Recognition”,IJMLC 2013 Vol.3(4): 352356 ISSN: 2010-3700I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[8] Eduardo Countinho, Jun Deng, Byjorn Schuller,”Transfer Learning Emotion Manifestation across Music and Speech.”,IEEE 2014 conference paper. DOI:10.1109/IJCNN.2014.68898K. Elissa, “Title of paper if known,” unpublished.

[9] Ringeval, F., Schuller, B., Valstar, M., Jaiswal, S., Marchi, E., Lalanne, D., Cowie, R., and Pantic, M. Av+ ec 2015: The first a□ect recognition challenge bridging across audio, video, and physiological data. In Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge (2015), ACM

[10] Peng, S., Yun, J., Li, Z., and Minghai, X. Speech emotion recognition using transfer learning. IEICE TRANSACTIONS on Information and Systems 97, 9 (2014), 2530–2532.

[11] Dhall, A., Ramana Murthy, O., Goecke, R., Joshi, J., and Gedeon, T. Video and image based emotion recognition challenges in the wild: Emotiw 2015. In Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (2015), ACM, pp. 423– 426.