IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AUDIO SONGS CLASSIFICATION METHODS BASED ON FACIAL EMOTION USING LOW AND HIGH LEVEL FEATURES

[1]Nandanwar Sneha Ashok
[1]Assistant Professor
[1]Department of Computer Science and Engineering,
ITM Vocational University,Vadodara, India

Abstract: Music is an amusing sound that leads us to experience peace and happiness. One of the fine arts today is music. Music digitalization provides us to access easily the different kinds of music. We often choose to listen to a song or music which best fits our mood at that moment, by means of features of music the mood of a song can be expressed. Most of the music software present today are lacking for providing the facility of mood-aware play-list generation. Manually choosing songs list suiting a specific mood or occasion by music listeners may be time consuming, that can be avoided by annotating songs with the relative emotion category. Here, work involves analysis of various features in order to learn, train and test the model representing the moods of the audio songs. Some of the basic components are to be considered for music emotion classification audio feature consists of feature set belong to groups dynamic, rhythmic, spectral, and harmonic. Support Vector Machine (SVM) is used as classifiers to classify the music mood recognition.

IndexTerms – Mood Models, Timbre, Tempo, SVM, KNN

I. INTRODUCTION

It is quoted once by a well-known German philosopher Friedrich Nietzsche "Without music, life would be a mistake". Songs has always been a fundamental factor of recreation of human life. Songs are not just useful for entertainment, but also studies have shown that listening the correct music does play an important role in healing; rejuvenating and even inspiring human mind in challenging situations such as is widely studied and demonstrated by the field of Music Therapy [13].Music Therapy: Music therapy is used to cure a patient using music as a medicine. Music therapist uses music and all of its facets like physical, emotional, mental, social, aesthetic and spiritual to help patients**.** Music Industry: In music industry, the mood of the song plays a vital role in its perception. Hence, Profits can be maximized by determining the mood of the top selling song. As the amount of available music-related information increases, the challenges of organizing and analyzing such information become paramount. Within few clicks music is now available at everyone's finger tips which in the olden days were limited to live concerts, performances or radio broadcasts. Music has now become very easily accessible and available to everyone. As Music database is increasing day by day it would not be wrong to say that we might hear repeated music pieces. With the amount and variety of music forms available easily, humans do not always listen to the similar type of music all the time. Everyone has their interests, favorite artists, albums, music type. To simply put, everyone have their personal choices and more importantly, even these choices might differ from time to time. What is needed is an additional parameter or search filter, in this case “Mood”, which signifies the emotion of that particular music piece. Here, we study the relation between mood and music by music emotion detection and classification method. [15]

II. MOOD MODEL

Two approaches studied generally for mood(emotions) models are –

Categorical approach: It comprises of various separate fundamental moods, for example, glad, pitiful, outrage thus on however these moods are shifted rapidly, since the essential moods are irregular.
Hevner’s categorical model: The famous categorical approach is **Hevner’s affective** checklist where eight clusters were laid out in a circle. The categorical approach focuses mainly on distinguishing different emotions from music.

Merry Joyous Gay Happy Cheerful Bright Pathetic	Humorous Playful Whimsical Fanciful Quaint Sprightly Delicate Light Graceful Vigorous	Lyrical Leisurely Satisfying Serene Tranquil Quite Soothing Exhilarated	Dreamy Yielding Tender Sentimental Longing Yearning Pleading Plaintive
Sad Mournful Tragic Melancholy Frustrated Depressing Gloomy Heavy Dark	Robust Empathic Martial Ponderous Majestic Exalting	Triumphant Dramatic Passionate Sensational Agitated Excited Impetuous Restless	Spiritual Lofty Inspiring Dignified Sacred Solemn Sober Serious

Fig 1: Hevner's affective model [3]

Dimensional approach: This classifies emotions along several axes such as valence (pleasure), arousal (activity), potency (dominance) and this is generally the most commonly used approach.[15] Songs can display varying emotions. During the collection of the corpus, it was observed that many of the songs display positive emotion such as excited, pleased, relaxed etc and also negative emotions such as afraid, frustrated, depressed etc. For clarity, taxonomy can divided into two classes. One class represents the happy mood class consisting of all positive emotions and other class represents sad mood class consisting of all negative emotions.[17]

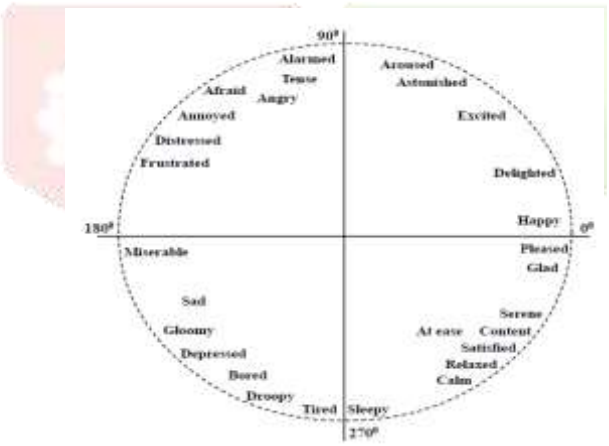

Fig2: Russell's Circumplex emotion model [3]

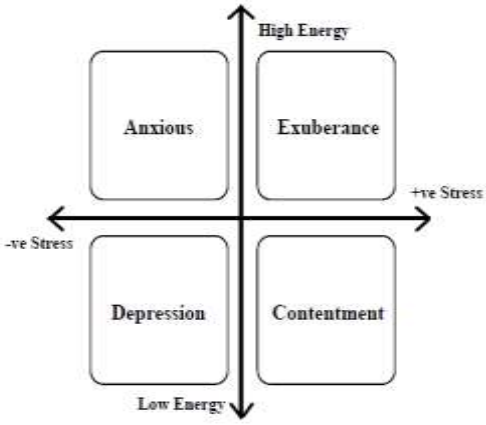

Fig3: Thayer's dimensional model[3]

Russell's Circumplex model of affect: It is a well known two-dimensional model called the Valence-Arousal Model proposed by Russell. One axis to represents the arousal level which shows the intensity in the form of high (active) and low values (inactive) and the other axis to represent valence, which shows polarity ranging from positive (happy) to negative (sad).[17]

Thayer's model: In this model, mood can be derived from two factors: Energy (High/Low) and Stress (Positive/Negative) this terminology can be divided into four clusters: Anxious, Contentment, Depression and Exuberance.[17]

III. EMOTION DETECTION

The planned method can extract user's facial expression [1] and options to discover the mood of the user.

Once the emotion is detected, listings of songs appropriate to the mood of the user are going to be conferred. It aims to provide higher enjoyment to the music lovers in music listening. Within the model, following mood areas unit included: Happy, Sad, Relax and Angry. The input to the model continues to be pictures of users that area unit additional processed to see the mood of user. The system can capture the image of the user at the start of appliance. The picture area unit captured using digital camera. The image captured antecedently can be saved and passed to the rendering part. The mood of user might not be same once some time: it might or might not be amendment. Therefore, the image is captured once each determined interval of your time and that image are going to be forwarded to next phase [1].

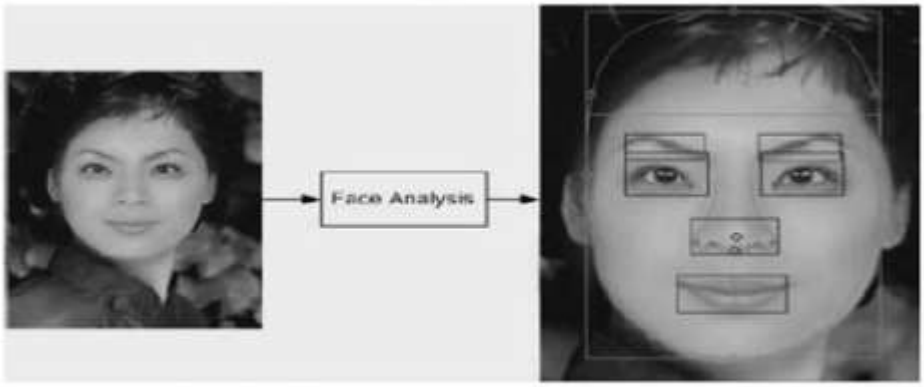

Fig3: Feature point detection

Emotion mapping and Playlist generation

Classified songs are mapped to the user's mood. The system was developed when bearing on the Russell's 2-D Valence-Arousal Model and Emotion wheel, when the mapping procedure is complete, a play list of relevant songs are generated. Similar songs are sorted along whereas generating the play list. Similarity between the songs was calculated by examination songs over some time intervals, features value corresponding to audio file were compared to the values corresponding to audio files matching to same category label. Initially, a playlist of all songs matching to the appropriate class is generated.

User Emotion	Music Mood
Neutral	Calm, Refreshing
Happy	Happy, Elated
Sad	Serene, Soothing
Surprise	Excited , Energetic

Table 2: Emtion Mood Mapping

IV. PROPOSED SYSTEM

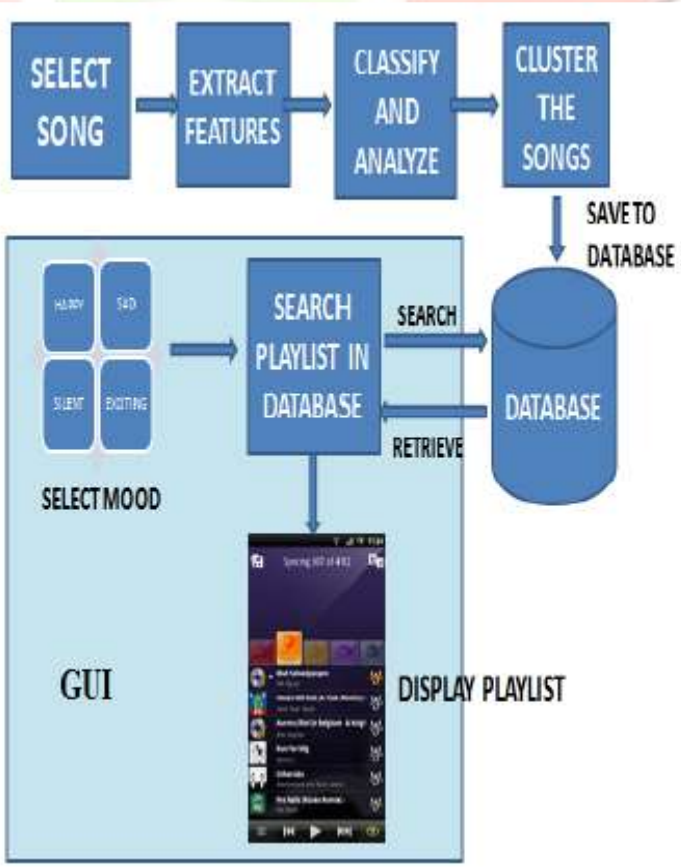

Fig4: System Architecture [2]

Feature Extraction
For the feature extraction technique every clip is split into frames. The extracted choices conjure two categories: Timber and Tempo. The set will represent mood information to point and play terribly purposeful role for mood detection.

For Pitch Detection:
MFCC Technique –
Mel Frequency Cepstral Coefficients (MFCC) is one altogether the foremost normally used feature extraction technique in music recognition and conjointly the necessity of Mel Frequency Cepstral Coefficients is taken as one of the standard technique for feature extraction. The good thing about regarding twenty MFCC Coefficients is common in ASR, though 10-12 Coefficients are sometimes thought of to be enough for writing speech.

Zero Cross Rate –
Zero Cross Rate indicates the quantity of times that a proof crosses the horizontal axis i.e, the speed at the signal changes from positive to zero to negative to zero to positive.

Classification:

The proposed method uses two different classifiers that are K-nearest neighbors and SVM classifier and performs a comparative Analysis.
K-means Algorithm: The goal of this algorithm is to divide M points in N dimensions into K clusters so that the within-cluster sum of squares is minimized. The algorithm needs as input a matrix of M points in N dimensions and a matrix of K initial cluster centers in N dimensions. The commonly used procedure is to search for a K-partition with locally optimal within-cluster sum of squares by moving points from one cluster to another. [2]

SVM: Support Vector machines are built and supported the ideas of call planes that outline call boundaries and a choice plane is one that separates between a group of objects having totally different category memberships. Once the features and extracted and normalized, we train support vector machine model. We tend to use libsvm library, in keeping with preliminary test the effective results are achieved by the CSVC methodology with the RBF Kernel (radial basis function). Consequently we tend to use this configuration in our formula. The to make decision that values to chose for the value C and therefore of the kernel operated, we implement a grid search formula like one urge[9]. We tend to keep parameters that acquire the most effective accuracy using 10-fold Cross Validation on the coaching set. Finally once the best parameter square measure found, we train a SVM Model and use it to predict mood model.

K-nearest neighbors:

The K-nearest neighbor or (k-NN) may be non-parametric technique that is employed for classification and regression. The input consists of k nearest neighbors examples within the feature house. The output depends on whether or not k-NN is employed for regression or classification. In k-NN classification, the output may be a category member. Associate in nursing object is classed by a majority vote of its neighbors (k maybe positive whole number, k<1) . If k=1, then the thing is just assigned to the category of that single nearest neighbor. K-NN comes below instance-base learning, or lazy learning, wherever the operate is barely approximated domestically and every one analysis is delayed till classification. The k-NN formula is among the only of all machine learning algorithms it may be helpful to weight the contributions of the neighbors in order that the nearest neighbors contribute a lot of the common than the a lot of distant ones[9].

V. EXPERIMENTAL RESULTS

Training and Testing

The datasets in every stage were subjected to a variety of assorted existing classification algorithms beneath various runs and folds. These algorithms showing a bias towards solely specific category labels or acting terribly low were discarded thereby subjecting the dataset to a 80%-20% training-testing split learning and analysis for the algorithms.
Confusion Matrix: The columns of the confusion matrix represent the predictions and the rows and represent the actual class. Correct predictions always lie on the diagonal of the matrix.

$$\begin{bmatrix} TP & FN \\ FP & TN \end{bmatrix}$$

Wherein, the True Positives (TP) indicates the amount of instances of a category that were properly foreseen, True Negatives (TN) indicates the amount of instances NOT of a selected category. False Positives (FP) indicate the amount of instances that are incorrectly foreseen happiness there to class and False Negatives (FN) indicate the amount of instances that were incorrect.

V. CONCLUSION

It can be concluded that audio features of Bollywood Music were successfully mapped with their respective moods and it is observed that the success rate of detecting the mood accurately for Indian Bollywood music is 70% and the success rate falls (30%) when detecting mood in western music. The Bagging of Random Forest approach performed much better as compared to other decision tree based algorithm and especially in case on analysis of Indian popular music unlike western music where SVM and neural network algorithms overshadowed the classifier accuracy. The classification performance achieved seems to be satisfactory so far thus making it useful for use in real applications.[15]

REFERENCES

[1] Aathreya S. Bhat, A. V. (2014). An Efficient Classification Algorithm for Music Mood Detection in Western and Hindi music using Audio Feature Extraction. Fifth International Conference on Signals and Image Processing, (p. 6). Bangalore, India.
[2] Amey Ujlambkar, O. U. (2014). Mood Based Music Categorization System for Bollywood Music. International Journal of Advanced Computer Research, (p. 8).
[3] Aniruddha M. Ujlambkar, V. Z. (2012). Automatic Mood Classification Model for Indian Popular Music. Sixth Asia Modeling Symposium. Bali, Indonesia: IEEE.
[4] Chih-Chung Chang, C.-J. L. (2013). A Library for Support Vector Machines. Taiwan.
[5]Chih-Wei Hsu, C.-C. C.-J. (2016). A Practical Guide to Support Vector Classification. Taiwan.
[6] Cyril Laurier, J. G. (2008). Multimodal Music Mood Classification. Seventh International Conference on Machine Learning and Applications.
[7] Cyril Laurier, P. H. (2007). Audio music mood classification using support vector machine. Retrieved from Reseachgate.
[8] D, P. D. (2017). Music Classification Based On Mood Recognition. International Journal of Advance Engineering and Research Development, 3.
[9] Grekow, J. (2018). Audio features dedicated to the detection and tracking of arousal and valence in musical compositions. Journal of Information and Telecommunication, 13.
[10] Hampiholi, V. (2012). A method for Music Classification based on perceived Mood Detection for Indian bollywood music . International Journal of Computer and Information Engineering, (p. 8).
[11] Jia-Min Ren, M.-J. W.-S. (2015). Automatic Music Mood Classification Based on Timbre and Modulation Features. IEEE Transactions on Affective Computing (p. 12). IEEE.
[12] Mrs. Tiple, A. H. Music Mood Detection.
[13] Nikolay Mirenkov, K. K. (2008). Quality of Life Supporters Employing Music Therapy. AINAW '08 Proceedings of the 22nd International Conference on Advanced Information Networking and Applications - Workshops, (p. 6). Washington, USA.
[14] Pooja V. Janse, S. B. (2014). A Comparative Study between MFCC and DWT Feature Extraction Technique. International Journal of Engineering Research & Technology, 4.
[15] Satyapal Yadav, A. S. (2017). Music Based Mood Classification. International Journal of Computer Trends and Technology (IJCTT) , 9.
[16] Shlok Gilda, H. Z. (2017). Smart Music Player Integrating Facial Emotion Recognition and Music Mood Recommendation. IEEE WiSPNET 2017 conference. (p. 5). Pune, India: IEEE.
[17] Swati Chauhan, P. C. (2017). Music mood classification based on lyrical analysis of Hindi songs using Latent Dirichlet Allocation. International Conference on Information Technology (InCITe). Noida, India: IEEE.