A COMPARATIVE STUDY OF AUDIO SONGS CLASSIFICATION METHODS BASED ON LOW AND HIGH LEVEL FEATURES

[1]Nandanwar Sneha Ashok, [2]Bhavesh Tanawala, [3]Hemant Vasava
[1]Department of Computer Engineering, [2]Assistant Professor, [3]Assistant Professor
[1]Department of Computer Engineering,
Birla Vishvakarma Mahavidhyalaya College, Anand, 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. RESEARCH METHODOLOGY

Researchers have shown that humans are not only able to recognize the emotional intentions used by musicians but also feel them. When we listen to various music we normally tend to experience a change in blood pressure, heart beat rate etc. In this paper we try to present a method to classify music based on the mood of the user. Our work is based on extracting or estimating the inherent mood in the musical piece by extracting various parameters like *beat, tempo, rhythm, timbre, pitch, tone, sound level, genre and vibrato* and tagging it to the song. We then use this information to classify the songs in a media player.[10]

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	**Humorous**	**Lyrical**	**Dreamy**
Joyous	**Playful**	**Leisurely**	**Yielding**
Gay	**Whimsical**	**Satisfying**	**Tender**
Happy	**Fanciful**	**Serene**	**Sentimental**

Cheerful Bright Pathetic	Quaint Sprightly Delicate Light Graceful Vigorous	Tranquil Quite Soothing Exhilarated	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]

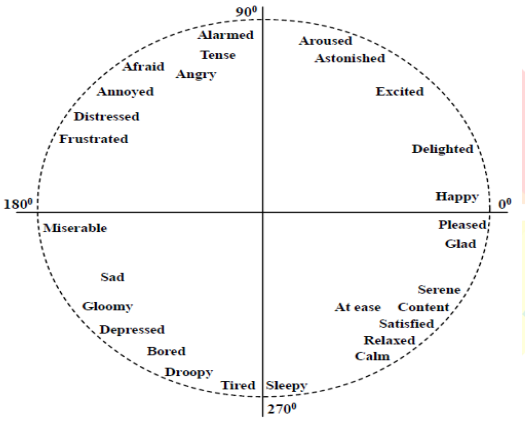

Fig2: Russell's Circumplex emotion model [3]

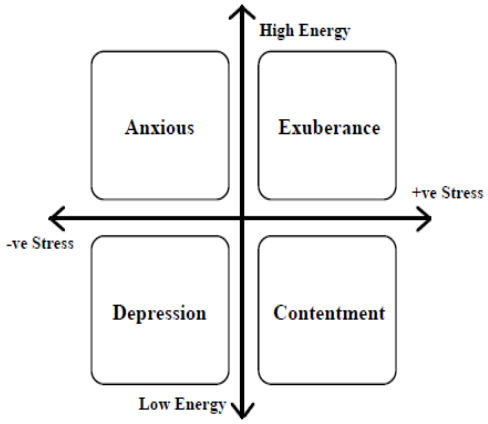

Fig3: Thayer's dimensional model[3]

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]
Here, feature extraction refers to high and low level features and these features are divided into two parts, segmental features and suprasegmental features. Segmental features are the individual sounds or tones that make the music; which includes acoustic structures such as amplitude, and pitch and Suprasegmental features are the foundational structures of a music, such as melody, tempo and rhythm.[1]

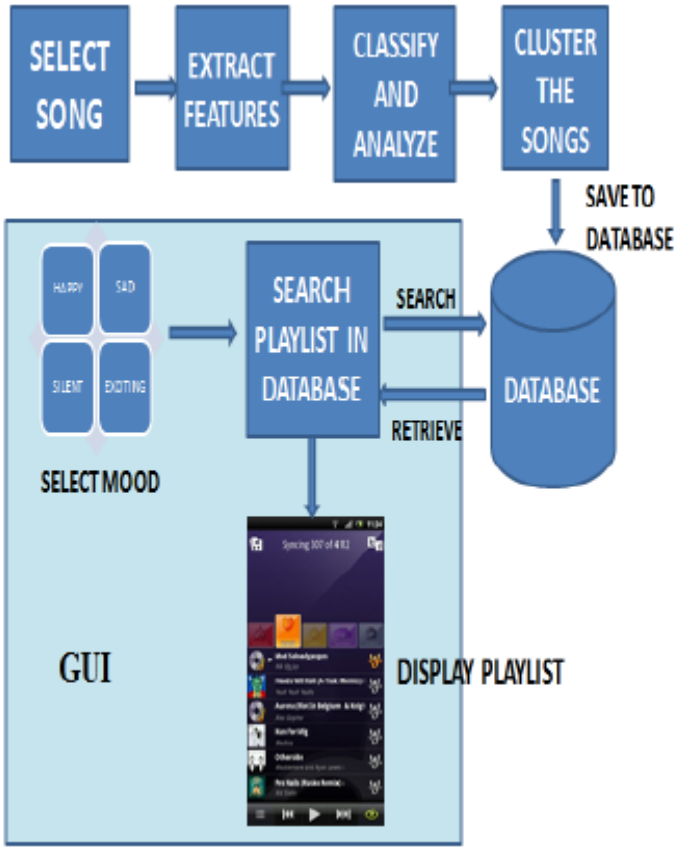

Fig4: System Architecture [2]

Feature extraction

Timbre features such as intensity, pitch and amplitude are extracted from the music piece and measures derived from them. The timbre features captures the tonal quality of sound that are related to different instrumentation. Timbre" is actually the quality of a musical note or tone. [15] Each individual feature is described as given below:

Intensity: The Intensity feature of a song, also known as dynamics clearly gives an indication of the rate of loudness or calmness of music and also is a indicator to measure stress in the composition of music.

Amplitude: Amplitude of the sound waveform in time domain can be quantized by the root mean square (RMS) of the signal.

Pitch: The pitch of a sound is dependent on the frequency of vibration and the size of the vibrating object. This feature corresponds to the relative lowness or highness that can be heard in a song.

Tempo features are the speed or pace of the musical piece. The pace is determined by the frequency of the beats; hence tempo is measured in BPM (beats –per-minute).[1]

Rhythm: By rhythm features, we can extract some details about whether the music emotion is positive or negative. Fast songs tend to be more happier than slow ones and we can extract rhythm features including Beat Sum, Strongest Beat and Strength of Strongest Beat.[2]

Fluctuations: Fluctuations in music indicate the rhythmic periodicities. The fluctuations are estimated by first computing a spectrogram and then FFT is performed on each band. The sum of the resulting spectrum leads to a rhythmic periodicities i.e. fluctuations.[1]

MFCC: Mel Frequency Cepstral Coefficients (MFCC) is one of the most normally used feature extraction method in music recognition and the need of Mel Frequency Cepstral Coefficients can be taken as one of the standard method for feature extraction. The use of about 20 MFCC coefficients is common in ASR, although 10-12 coefficients are often considered to be sufficient for coding speech.[6]

IV. Results and Discussion

Classification: The classification of mood is based on the Thayer's model of mood. First of all, the features are extracted for determining the mood. Now considering only one feature and deducing conclusions on the mood of the song is not reliable, and the values of one feature may wrongly represent the mood of the music piece, all the described features are considered. Next, the mood of the music piece is selected by any classification method that will be selecting a particular mood depending on the threshold level determined. The threshold values of the features were determined to represent a particular mood.[2]

Classification Algorithm

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: When once all the features extracted and then normalized, we can use Support Vector Machine model for classification. the libsvm [4]library is used. When preliminary tested, the best results were achieved by the C-SVC method with the RBF kernel (Radial Basis Function).Then this configuration is taken as our algorithm. We implemented a grid search algorithm as suggested [5]. We take the parameters that obtains the best accuracy using a 10-fold Cross Validation on the training set and when the optimal parameters are found, we train a SVM model and use it to predict the mood categories.[7]

Bagging of Random forest Algorithm: Here in this it is shown that ensemble by generating an ensemble of Random Forests using bootstrap aggregation also known as bagging, the randomly sampled data attributes are split on the basis of "Gini Index" which has shown better results when working with CART trees for growing Random Trees. Gini index considers a binary split for each attribute, a weighted sum of the impurity of each resulting partition is calculated for each split. This approach shows rise not only in accuracy of classification of music datasets as compared to traditional Random forest approach, but also has shown a consistent better performance as compared to other classification techniques.[15]

Table1: Classification accuracy using audio features for each category [6]

Moods	SVM	RandForest	k-NN(k=3)
Angry	98.1%	95.4%	69.5%
Happy	81.5%	55.9%	55.9%
Sad	87.7%	86.2%	55.0%
Relaxed	91.4%	91.2%	61.8%
Mean	**89.8%**	**87.6%**	**60.5%**

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.