Mood Based Music Player

Anuja Arora, Aastha Kaul, Vatsala Mittal
Computer Science Department
Jaypee Institute of Information Technology
Noida, India
anuja.arora29@gmail.com, aasthakaul19@gmail.com, vatsala.mitt@gmail.com

Abstract— **The ability of music to produce an emotional response in its listeners is one of its most exciting and yet the least understood property. Music not only conveys emotion and meaning but can also stir a listener's mood. This paper will study various algorithms based on classification to provide a clear methodology to i) classify songs into 4 mood categories and ii) detect users mood through his facial expressions and then combine the two to generate user customized music playlist. Songs have been classified by two approaches; by directly training the models namely KNN, Support Vector Machines (SVM), Random Forest and MLP using selected audio features and by predicting a songs arousal and valence values using these audio features. The first approach attains maximum accuracy of 70% using MLP while the latter achieves accuracy of 81.6% using SVM regression. The face mood classifier using HAAR classifier and fisher face algorithm attains precision of 92%.**

Keywords—mood; classification; Multi Layer Perceptron; SVMregression; Valence; Arousal

I. Introduction

Listening to music nowadays has become a day to day activity. There are a huge number of categories of music a person to listen to. Human emotions are related to music so much since we choose to listen to a song that relates to our mood at a particular time. Several studies on Music Information Retrieval [14, 15, 18, 19] have also been carried out in recent decades.

Facial expressions are a great indicator of the state of mind of an individual. It is indeed the most natural and basic way to express emotions [1, 2, 11]. In spite of this strong correlation, most of today's music software is still devoid of providing the facility of mood-aware playlist generation.

Emotional meaning of the music is subjective and thus, it depends upon many factors like place, tradition and culture whereas the mood category of a song varies depending upon several psychological conditions. Music listeners, collectors or psychologists may use mood wise music widely to categorize their music collection, or help soothe their clients. Despite such extensive use, this field of research is unexplored by many, thus classification task becomes much more difficult, yet important.

The music database keeps on increasing as the audio data also increases in the digital world. There has been some development in creating archives for these kind of database. Because of the progression in innovation, and the regular increase of the web , the network related with music increment continuously. This has prompted an extensive database of music, which surely is difficult to categorize manually on the basis of moods of the music. Subsequently, there is a need to create a less time consuming technique such a big assignment. The adjustments in meanings of a mood after some time have additionally lead to an expanded trouble for its categorization. For instance, the music we listen today is such a great amount of not quite the same as what it was 20 years prior.

In this research, we study methods for classifying moods with the help of features of an audio file and propose approaches for this research by using machine learning classifiers. DEAM dataset has been used for mood classification. It has more than 2800 songs annotated with 4 moods: Happy, Sad, Angry and Relax and with their valence and arousal values. The idea behind this paper is that give attention towards on how good audio features are to predict moods of a audio file. Also, we validate the performance of result in order to predict valence and arousal values using these audio features

The following segment gives a review of a portion of the present work done for music and mood arrangement. In section 3, we explain the methods and approaches undertaken for this research. Section 4 portrays the results of the investigations directed and the experiments done. Section 5 is an overview of the research and states future work which can be done.

II. State of Art

The problem of Music Emotion Recognition is a very interesting field of study and poses a lot of application. This area gives an outline of the related and well known work done by specialists in the field of audio and mood recognition of audio files.

In 2005 Wieczorkowska et al [3] published a paper in which their goal was to aid the user to find piece of music for specific moods. They classified the 327683 .mp3 songs dataset into 6 emotions using KNN and general accuracy in test in parallel yielded 37% correctness. In 2008 [21] used a regression approach to the problem of MER and achieved 64% accuracy for arousal and 59% for valence. In 2012 Yading

978-1-5386-9436-7/19/$31.00 ©2019 IEEE

Song et al [22] did evaluation of music features for the task of MER. Data set of 2904 songs that had been tagged with one either "happy", "sad", "angry" or "relaxed", and SVM was used as the training algorithm. It showed spectral features outperformed the rest of the acoustic features.

In the same year [16] did Music Emotion Classification with 903 songs of 5 different categories. SVM was chosen as the classification algorithm with 10 cross validation resulting in achieving F-measure of 47.2% with precision of 46.8% and recall of 47.6%. In 2014 Aathreya et al in [16] used multilayer neural network for classifying songs based on moods. In 2015 JacekGrekow [12] used 3 different audio tools during emotion detection. Dataset with 4 emotions was trained and tested by various algorithms like KNN, Random Forest in WEKA. The best results were obtained after applying attribute selection for data from jAudio. In the same year Braja Gopal et al [5] prepared two systems; first was to detect the valence and arousal for each song and second was mood classification system of hindi songs and achieved maximum F-score of 72.32.

In 2016 [17] RenuTaneja et al extracted audio features like tempo, beats, RMSE using jAudio to form clusters of 4 different emotions. Kee Moe Han et al [13] took the average of emotions given by 15 people as the emotion of the song and a classifier was trained with the same. Audio features such as pitch, timbre etc. were then extracted and the emotion of the music signal could then be classified through the dataset with a probabilistic classifier. Recently in 2017 a music system using facial recognition [20] was made by V. R. Ghule et al.

III. Methodology

There are different techniques which can be used for classification so as to fit properties of the database and which then leads to different outcomes with respect to actual ground truth data. In this section we explain details of the all the techniques used and all the approaches selected for experimenting for classifying moods of audio files

A. Model

In the following approach we convert audio files of different moods to numerical values using features of these files in the python language. Feature extraction method is applied to retrieve different kind of audio features such as harmonic features, spectral, rhythm, energy and chroma vectors of the audio files, and this paper explores through many algorithms based on classification so as to suggest a new approach to classify and detect moods. We investigate basic classification models such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Multi-Layer Perceptron (MLP) and Random Forest. Training of the dataset is done using the above mentioned models and then using these trained models predictions are made on test set so as to provide results. Arousal and Valence values are a major factor for improving the training models. Valence is positive or negative affectivity, whereas arousal measures how calming or exciting the information is. Now rather than using these audio features directly, we predict the valence and arousal values using SVM Regression with different kernels. These can be then mapped into the 2-D Valence-Arousal Space to identify the mood.

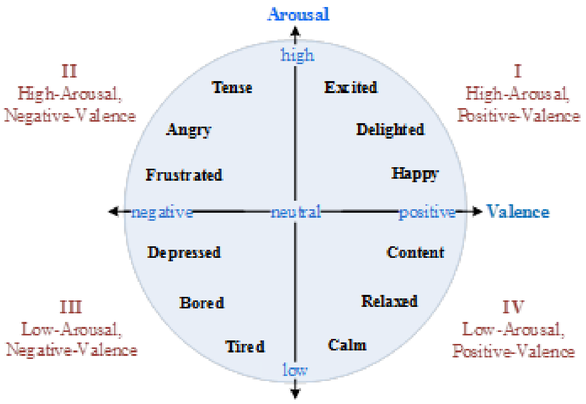

Fig 1. Two Dimensional Valence Arousal Space[23]

PyAudioAnalysis and librosa library [6, 7, 8, 9] in the python language was used for extracting features from the audio files. The total numbers of features extracted are 36. The size of each audio file is approximately between 5mb-10mb and they are 30-60 seconds long. Details of all the features is given in Table 1

TABLE I. EXTRACTED FEATURES

Feature ID	Feature Name	Description
1-2	Root Mean Square Energy (RMSE)	Compute root-mean-square (RMS) energy for each frame, either from the audio samples
3	Total Beats	Defines the total beats
4	Tempo	Estimate the tempo (beats per minute)
5-6	Harmonic	Extract only the harmonic component.
7-8	Chroma STFT	Compute a chromagram from a waveform or power spectrogram.
9-10	Chroma CQ	Constant-Q chromagram.
11-12	Chroma CENS	Computes the chroma variant "Chroma Energy Normalized", CENS
13-14	Melspectrogram	Compute a mel-scaled spectrogram.
15-18	Mfcc	Mel Frequency Cepstral Coefficients form a cepstral representation where the frequency bands are not linear but distributed according to the mel-scale.
19-20	Spectral Centroid	The center of gravity of the spectrum.
21-22	Spectral	Compute pth-order spectral bandwidth

	Bandwidth	
23-24	Spectral Contrast	Compute spectral contrast
25-26	Spectral rolloff	The frequency below which 90% of the magnitude distribution of the spectrum is concentrated.
27-28	Poly_features	Get coefficients of fitting an nth-order polynomial to the columns of a spectrogram.
29-30	Tonnetz	Computes the tonal centroid features (tonnetz).
31-32	Zero Crossing Rate	The rate of sign-changes of the signal during the duration of a particular frame.
33-34	Percusive	Extracts percussive elementsfrom anaudio time-series.
35-36	Frames_to_time	Extracts percussive elements from an audio time-series.

B. *Feature Selection*

Not all features affect musical mood in the same way and hence we need to carefully select the features we intend to use for the purpose of mood detection from audio tracks. Audio features are put into 4 suitable dimensions namely: Dynamic, Harmony, Rhythm and Spectral. Using Feed Forward selection it is clear that spectral, dynamic and harmony features used together help achieve the best accuracy.

TABLE II.DIMENSIONS OF AUDIO FEATURES

Dimension	Audio Features
Dynamic	rmse _mean and rmse_std
Rhythm	Total Beats and Tempo
Harmony	Harm_mean, harm_std, chroma_stft_mean, chroma_stft_std, chroma_cq_mean, chroma_cq_std, chroma_cens_mean, chroma_cens_std
Spectral	Melspectrogram_mean, melspectrogram_std, mfcc_mean, mfcc_std, mfcc_delta_mean, mfcc_delta_std, cent_mean, cent_std, spec_bw_mean, spec_bw_std, contrast_mean, contrast_std, rolloff_mean, rolloff_std, poly_mean, poly_std, tonnetz_mean, tonnetz_std, zcr_mean, zcr_std

C. *Classification*

For classification the paper presents two approaches. First, to test and predict directly by using audio feature and four base classification algorithms: Random Forest (RF), Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Multi-Layer Perceptron (MLP) are employed. The dataset was split into training and test where 80% of the data was dedicated to the training set and the remaining 20% was for test set. Fig. 1 shows the overall flow for first approach. Initially, python library has been used to extract features from take audio emotion data set. Further, feature evaluation is performed in order to provide error free feature content to classification models. Four above mentioned classification models have been used.

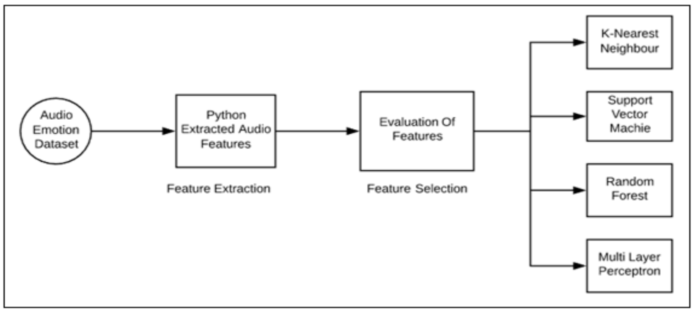

Fig 1. Mood Classification Approach (KNN: K-Nearest Neighbor, SVM: Support Vector Machine Classifier, RF: Random Forest Classifier, MLP: Multi-Layer Perceptron)

Second, use the selected audio features to predict the valence and arousal values using SVM regression with 3 different kernels- Linear, Poly and rbf. These are then mapped into 2D Space to identify the mood. Fig. 2 shows the flow for the Valence-Arousal approach.

After feature selection is done the important and highest ranking features are used to train models and then the confusion matrix is created with the help of the predicted moods and the known moods of the audio samples in the test set. Calculation of accuracy is done using the V-A model and confusion matrix is formed as the final outcome.

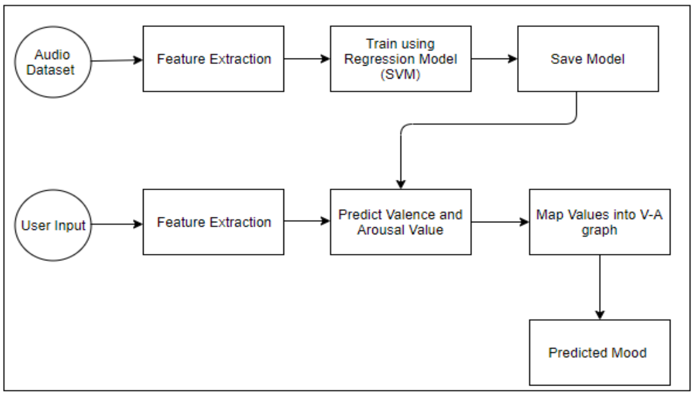
Audio Dataset
Feature Extraction
Train using Regression Model (SVM)
Save Model
User Input
Feature Extraction
Predict Valence and Arousal Value
Map Values into V-A graph
Predicted Mood

Fig 2.Valence-Arousal Classifier Model

D. *Facial Mood Detection*

For detecting a user's mood this paper makes use of facial expressions. The pre-trained HAAR frontal-face classifier is used for detecting a user's face on screen. Before training the model it is necessary we preprocess and standardize the images by keeping only the face portion of the image and by turning it into black and white.

To create our model we use fisherface algorithm and collect 16 images per mood category in a 5 second time span. Once the model is trained successfully it can be used to detect a user's mood. After detecting the users mood confusion matrix is plotted and precision was then calculated using it.

IV. EXPERIMENTS AND DISCUSSION

Features which were extracted from the audio files are divided into two sets train and test set respectively. Numerous performances of the classification models have been calculated with the help of training the data and predicting results of test set. Performance was improved for predicting the correct mood by doing numerous attempts on classification models.

A. Data

Classification of the mood of an audio file is very much dependent on what kind of dataset is being used and what features have to be extracted. DEAM dataset is used for the research being done. The MediaEval Database for Emotional Analysis of Music (DEAM) is a diverse dataset annotated with mean valence and mean arousal values along with the mood of the audio file. Metadata including, song title, genre and artist is also provided.

DEAM [10] dataset contains more than 2800musical audio excerpts belonging to four moods. The four moods available in the datasets are: Happy, Sad, Angry and Relax. Categories such as Tense, Excited, and fear as moods of the songs are not available in this dataset. The training dataset consists of audio files which are of type .wav(waveform) and average duration of audio files were around 45 seconds, for 4 moods. Table III shows the number of audio files dedicated for each mood across the training set of the DEAM dataset.

For mood detection of user a dataset of 448.jpeg images is manually created where in all the 4 mood classes has equal images. Table IV shows the number of files of mood across the face mood detection dataset.

TABLE III. AUDIO FILES IN DATASET

MoodCategory	Number of Audio Files
Happy	753
Sad	759
Angry	637
Relax	749

TABLE IV. IMAGES IN DATASET

MoodCategory	Number of Audio Files
Happy	112
Sad	112
Angry	112
Relax	112

B. Classification Results

The results of models with valence and arousal are better than results of models which directly uses the audio features. Also it is observed that not all features contribute in a similar fashion to the task of mood classification and hence feature selection is crucial to this problem.

Table V shows the evaluation of audio features using MLP as the classifier. Table VI shows the accuracy scores of different models that trained directly with audio features. To evaluate the kernels in SVM Regression we calculate the accuracy scores, shown in table VII. Table VIII shows the confusion matrix for face mood detection. The precision can be obtained using the confusion matrix.

TABLE V.EVALUATION OF AUDIO FEATURES

Features	Accuracy
Dynamic	38.83
Rhythm	27.38
Harmony	35.35
Spectral	61.52
Harmony and Rhythm	46.79
Harmony and Dynamic	43.32
Spectral and Dynamic	60.13
Spectral and Harmony	61.52
Spectral and Rhythm	65.68
Rhythm and Dynamic	33.10
Spectral, Dynamic and Rhythm	61.98
Spectral, Dynamic and Harmony	70.88
Spectral, Dynamic and Rhythm	49.04
All features	61.35

TABLE VI.MOOD CLASSIFICATION USING AUDIO FEATURES

Learning Model	Accuracy
KNN	60.08
SVM	56.50
RF	55
MLP	70.88

TABLE VII.EVALUATION OF KERNELS IN SVM REGRESSION

Kernel	Accuracy
Rbf	81.6
Linear	59
Poly	76.4

TABLE VIII.CONFUSIN MATRIX

Actual/Predicted	Happy	Sad	Angry	Relax
Happy	101	11	0	0
Sad	3	97	0	12
Angry	9	0	102	1
Relac	0	5	0	107

It is deduced that Spectral, Dynamic and Harmony features used together gave better results than using all the features. For mood classification by model trained with audio features it is seen MLP outperforms the rest and achieves an accuracy score of 70.88%. For predicting valence and arousal values using SVM regression it is seen Rbf kernel provides the maximum accuracy of 81.6% and hence should be used. Using Table VIII the precision for the face mood classifier is calculated, which comes out to be 92%.

V. CONCLUSION AND FUTURE SCOPE

Through the course of our exploration to recognize mood of audio files utilizing audio features , we have concluded some critical outcomes. Firstly, we conclude that not all features contribute equally to the task of mood classification and hence it is imperative to perform feature selection. For the classifiers trained on audio features it is seen MLP (70.88%) has the best performance for the classification task than the rest for music mood classification by decreasing over-fitting and applying different parameters. The task of predicting valence and arousal values to be used in music mood classification is done using SVM regression. 3 different kernels are used and rbf (81.6%) performs the best out of the three. The audio features extracted could be sufficient to predict moods but the models could still be improved by applying feature extraction on every single audio file and not seeing them as whole since every audio file would have different features which are most important part of that audio file, which may lead to an increase in the accuracies obtained. For the task of user's mood detection, the model was trained with HAAR frontal-face classifier and fisherface algorithm. This model has a precision of 92%.

To make our work progressively solid and usable we may change classifier mixes to enhance the less positive cases in results. Be that as it may, so as to improve the execution of our classifiers all the more effectively, refining the list of features is the most important factor. Region and culture from which the audio file belong is also an important factor on classifying the mood of the audio file.

REFERENCES

[1] A. Lehtiniemi and J. Holm, "Using Animated Mood Pictures in Music Recommendation", In *16th International Conference on Information Visualisation*.(2012)

[2] A. S.Dhavalikar and Dr. R. K. Kulkarni, "Face Detection and Facial Expression Recognition System" ,*International Conference on Electronics and Communication System* (ICECS -2014).

[3] A. Wieczorkowska, P. Synak, R. Lewis, & Z. W.Raś, "Extracting emotions from music data." In *International Symposium on Methodologies for Intelligent Systems* (2005, May) (pp. 456-465). Springer, Berlin, Heidelberg.

[4] Aljanaki, Anna, Y. H. Yang, and M. Soleymani." Emotion in music task at MediaEval 2015.", In *MediaEval 2015 Workshop, Wurzen, Germany*,(2015).

[5] B. G. Patra., D. Das and S. Bandyopadhyay. "Music emotion recognition system.", In *Proceedings of the International Symposium Frontiers of Research Speech and Music (FRSM-2015)*. 2015.

[6] B. G. Patra, P. Maitra, D. Das, and S. Bandyopadhyay."Feed-Forward Neural Network based Music Emotion Recognition.", In *MediaEval 2015 Workshop, Wurzen, Germany.*2015.

[7] B. G. Patra, P. Maitra, D. Das, and S. Bandyopadhyay."Feed-Forward Neural Network based Music Emotion Recognition.", In *MediaEval 2015 Workshop, Wurzen, Germany.*2015.

[8] B. G. Patra, D. Das, and S. Bandyopadhyay. "Unsupervised approach to Hindi music mood classification.",In*Mining Intelligence and Knowledge Exploration*. Springer International Publishing, 2013.62-69.

[9] B. G. Patra, , D. Das, and S. Bandyopadhyay. "Automatic Music Mood Classification of Hindi Songs", In *3rd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP-2013)*. 2013.

[10] DEAM dataset:TheMediaEval Database for Emotional Analysis of Music[Online] Available : http://cvml.unige.ch/databases/DEAM/

[11] F. Abdat, C. Maaoui and A. Pruski, "Human-computer interaction using emotion recognition from facial expression",In*UKSim 5th European Symposium on Computer Modelling and Simulation* (2011).

[12] J. Grekow, "Emotion Detection Using Feature Extraction Tools." In *International Symposium on Methodologies for Intelligent Systems,*(2015, October), (pp. 267-272). Springer, Cham.

[13] K. Han, T. Zin& H. M. Tun, "Extraction Of Audio Features For Emotion Recognition System Based On Music", In *International Journal Of Scientific & Technology Research* , (JUNE 2016).

[14] L. Lu, D. Liu, and H.-J. Zhang, "Automatic mood detection and tracking of music audio signals", In *IEEE Trans. Audio, Speech, Lang. Process.*, (Jan. 2006), vol. 14, no. 1, pp. 5–18.

[15] M.-Y. Wang, N.-Y.Zhang, and H.-C. Zhu, "User-adaptive music emotion recognition," In *Proc. Int. Conf. Sig. Process.*, (2004), pp. 1352–1355.

[16] R. Panda & R. P. Paiva., "Music emotion classification: Dataset acquisition and comparative analysis." In *15th International Conference on Digital Audio Effects* ,(2012). (DAFx-12).

[17] R. Taneja, A. Bhatia, J. Monga& P. Marwaha, "Emotion detection of audio files." In IEEE *Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on* (2016, March). (pp. 2397-2400).

[18] T.-L. Wu and S.-K.Jeng, "Extraction of segments of significant emotional expressions in music," In *Proc. Int. Workshop Comput. Music Audio Technol.*, (2006), pp. 76–80.

[19] V. Carvalho and C. Chao, "Sentiment retrieval in popular music basedonsequential learning," *Proc. ACM SIGIR*, 2005

[20] V. R. Ghule, A. B. Benke, S. S. Jadhav, S. A. Joshi," Emotion Based Music Player Using Facial Recognition", In *International Journal of Innovative Research in Computer and Communication Engineering*, (February 2017), Vol. 5, Issue 2.

[21] Y. H. Yang, , Y. C. Lin, , Y. F. Su, & H. H. Chen, "A regression approach to music emotion recognition.", In *IEEE Transactions on audio, speech, and language processing*, (2008) *16*(2), 448-457.

[22] Y. Song, S. Dixon, & M. Pearce, "Evaluation of Musical Features for Emotion Classification." In *ISMIR* ,(2012, October). (pp. 523-528).

[23] 2D Valence Arousal Space[Online] Available:https://www.researchgate.net/figure/The-2D-valence-arousal-emotion-space-Russell-1980-the-position-of-the-affective_fig1_254004106