Study of Emotion Detection in Tunes Using Machine Learning

Rishika Shetty, Shweta Kasbe, Kimaya Jorwekar, Dharti Kamble, Prof. Makarand Velankar

Information Technology, MKSSS's Cummins College of Engineering Pune

Abstract- The main objective of this paper is to study possible emotions generation in listener's mind due to listening of tunes. Such emotions can be detected automatically using the audio features such as zero crossing, compactness, spectral centroid, spectral Flux, spectrum Roll off and Beat Histogram etc. We have explored machine Learning algorithms such as SVM (support vector machine) and ANN (Artificial Neural Network) for classification. The proposed technique of emotion detection is done in two parts as feature extraction and classification of tunes using machine learning techniques. We have studied different tools for extracting the features of tunes. These extracted features can be further given to the classifiers to categorize the emotions.

Index Terms- SVM (support vector machine), ANN (Artificial Neural Network), Emotion detection, Feature extraction, Audio tools

I. INTRODUCTION

Music is a form of art which plays a very important role in an individual's life. It has the power to change anyone's state of mind, instantly. Every tune has a different emotion which will be captured using different audio features. We have studied different features used by researchers and emotions extracted. We have selected useful features to detect specific emotions in music.

We propose to categorize the tunes into four categories namely

1 .Sad
2. Happy
3. Peppy
4. Calm

The goal of this paper is to study and develop automatic music emotion detection system. We have analyzed different musical features in order to map them into four categories of emotion: Sad, Happy, Peppy and Calm. Further we have studied use of SVM (support vector machine) and Neural Networks for classification.

II. MODELS PROPOSED

Much work has been carried out in the past by researchers regarding emotion recognition using music and different models used by them are described in this section. We have described ANN (Artificial Neural Network) and SVM (Support Vector machine) and Valence-Arousal model which are prominently used by researchers and are widely used for solving problems in various domains.

2.1 ANN Model for Music Emotion Recognition

ANN models are basically feed forward neural network (FFNN) models which try to map an input vector onto itself. It consists of an input layer, an output layer and one or more hidden layers. The number of units in the input and output layers are equal to the size of the input vectors. The number of units in the hidden layer is less than the number of units in the input or output layers. The middle layer is also the dimension compression layer (Layers are shown in Fig 4.3). The activation function of the units in the input and output layers are linear, whereas the activation function of the units in hidden layer can be either linear or nonlinear. The performance of ANN models can be interpreted in different ways, depending on the problem and the input data. If the data is a set of feature vectors in the feature space, then the performance of ANN models can be interpreted either as linear and nonlinear principal component analysis (PCA) or distribution capturing of the input data.[1]

2.2. SVM Model for Music Emotion Recognition

Support vector machine (SVM) is based on the statistical learning and quadratic programming[1]. The aim of SVM classifier is to devise a computationally efficient way of learning good separating hyper planes between different classes in a high dimensional feature space. SVM is used to identify a set of linearly separable hyper planes which are linear functions of the high dimensional feature space. [1][9][10]

2.3 Russell/Thayer's Valence-Arousal model

Russell/Thayer's Valence-Arousal model is the most noted dimensional model [2], in this emotion exists on a plane along independent axis. In this model the High to low indicates arousal (intensity) and positive to negative indicates valence (appraisal of polarity). Figure2.3 shows the space where some regions are associated with distinct mood categories. In this paper, we assume that the task of music emotion detection is to automatically find the point in the VA plane which corresponds to the emotion induced by a given music piece [2][3][9][10].

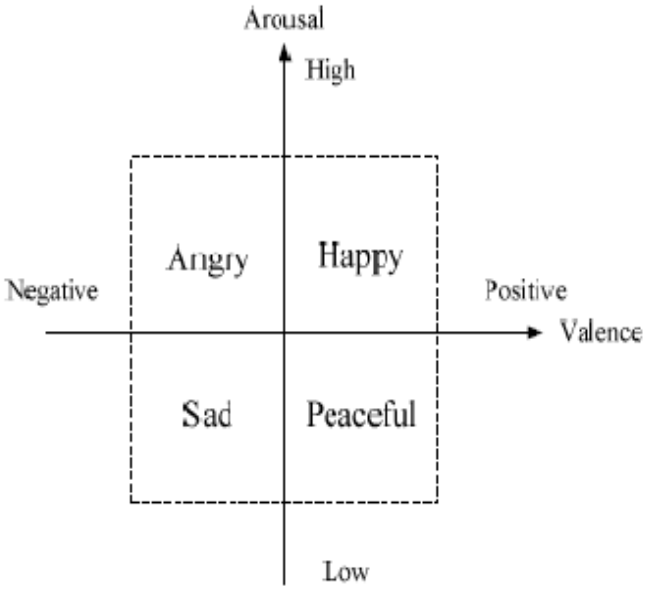


Fig.2.3 Different regions correspond to different categorical emotions. [2]

III. Features for Emotion Detection

We can use various low level features of music for automatic emotion detection. These features can be extracted by various digital signal processing methods. We have described few of the important features described by different researchers in their work. We propose to use some of them for emotion detection in our work.

3.1 Zero Crossing

One of the most important features used in music emotion detection and information retrieval. Basically, it's the count of number of times the signal (time based) crosses zero frequency.[5] This is a time domain based feature and it provides us information about frequency of the musical signal which is also considered as perceived pitch information of music.

3.2 Compactness

Compactness is intuitively understood as the degree to which elements of a set are close together in a set. It sums over frequency bins of an FFT. This provides an indication of the noisiness of the signal spectrum.[5][7] This is frequency domain feature and is very useful to understand or predict possible noise in signal.

3.3 Spectral Centroid

Spectral Centroid, usually associated as the measure of the brightness of a sound, of a spectral frame is defined as the average frequency weighted by amplitudes, divided by the sum of the amplitudes.[7] It is calculated as the weighted mean of the frequencies present in the signal. It is a frequency domain feature and also referred as median of spectrum.

3.4 Spectral Flux

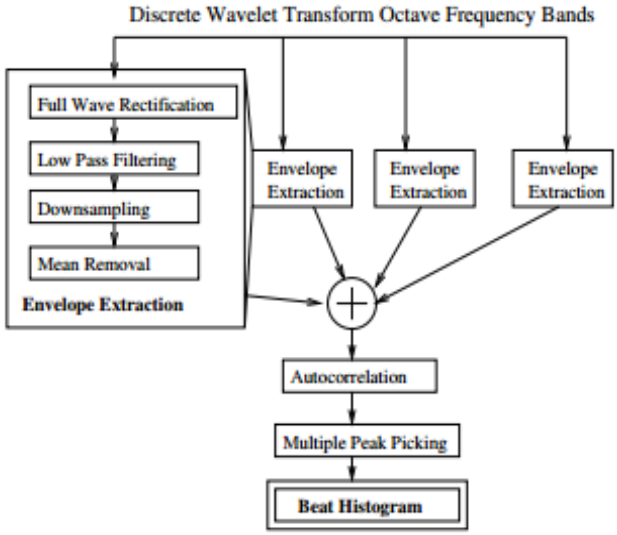
Spectral Flux is defined as the spectral correlation between two adjacent windows. It's the degree of change of spectrum between windows.[7] It does not depend on overall power. It is usually calculated using Euclidian distance method or 2-norm method.

3.5 Spectrum Rolloff

Spectral rolloff is defined as the frequency where 85% of the energy in the spectrum is below this point. It is often used as an indicator of the skew of the frequencies present in a window.[7]

3.6 Beat Histogram

This feature auto correlates the RMS for each bin in order to construct a histogram representing rhythmic regularities. This feature is used as a base feature for determining best tempo match. [5][7]



3.6 Beat Histogram Calculation Diagram

3.7 Beat Sum

It is sum of all bins in the beat histogram. This is a good measure of the importance of regular beats in a signal [7].

3.8 Strength of Strongest Beat

How strong the strongest beat in the beat histogram is compared to other potential beats. [7]

3.9 Strongest Beat

It is the strongest beat in a signal, in beats per minute, found by finding the highest bin in the beat histogram. [7]

3.10 RMS (Root Mean Square)

RMS is calculated on a per window basis. It is defined by the equation:[5]

$$RMS = \sqrt{\frac{\sum_n^N x_n^2}{N}}$$

Where N is the total number of samples provided in the time domain. RMS is used to calculate the amplitude of a window.

3.11 Fraction of Low Amplitude Frames

Spectral Flux is defined as the spectral correlation between adjacent windows. It is often used as an indication of the degree of change of the spectrum between windows. [5]

3.12 Power Spectrum

Power spectrum is a measure of the power of different frequency components. [7]

3.13 MFCC

MFCC is a short-term spectral-based feature contains much information. This section describes the process of extracting MFCC from the given input music signal [1]. The mel-frequency cepstrum is good for recognizing structure of music signals as reviewed from previous papers and in modelling the subjective pitch and frequency content of audio signals [4].

MFCC feature extraction process :

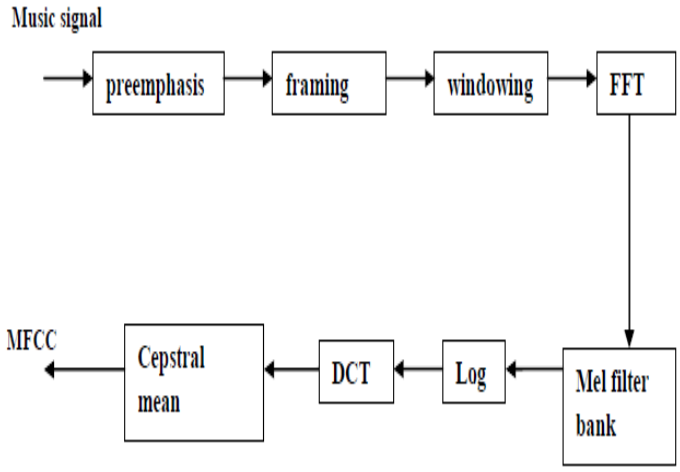


Fig 3.13 MFCC Feature Extraction.[8]

- Pre-processing

The continuous time signal is sampled at sampling frequency. At the first stage in MFCC feature extraction is to boost the amount of energy in the high frequencies. This pre emphasis is done by using a filter. [8]

- Framing

It is a process of segmenting the music samples obtained from the analog to digital conversion (ADC), into the small frames with the time length within the range of 20-40 ms. Framing enables the non stationary music signal to be segmented into quasi-stationary frames, and enables Fourier Transformation of the audio signal. It is because, audio signal is known to exhibit quasi-stationary behaviour within the short time period of 20-40 ms. [8]

- Windowing

It is the process to window each individual frame, in order to minimize the signal discontinuities at the beginning and the end of each frame. [8]

- FFT

Fast Fourier Transform (FFT) algorithm is ideally used for evaluating the frequency spectrum. FFT converts each frame of N samples from the time domain into the frequency domain. [8]

- Mel Filter bank and Frequency wrapping

The Mel filter bank consists of overlapping triangular filters with the cutoff frequencies determined by the center frequencies of the two adjacent filters. The filters have fixed bandwidth and linearly spaced centre frequencies on the Mel scale. [8]

- Log

The logarithm has the effect of changing multiplication into addition. Therefore, this step simply converts the multiplication of the magnitude in the Fourier transform into addition. [8]

- Discrete Cosine Transform

It is used to orthogonalise the filter energy vectors. Because of this orthogonalization step, the information of the filter energy vector is compacted into the first number of components and shortens the vector to number of components. [8]

IV. Proposed Work

4.1 General Discussion

There are various algorithms which can be used for classification such as K-NN, SVM, and Neural Network. We will be focussing on two of them mainly SVM and Neural Network which has better accuracy and should work great while classifying the categories.

4.2 SVM (Support Vector Machine)

A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyper plane. In other words, given labelled training data (supervised learning), the algorithm outputs an optimal hyper plane which categorizes new examples. [9][10]

An SVM can even handle Non-linearly separable data by adding a slack Variable to allow misclassification

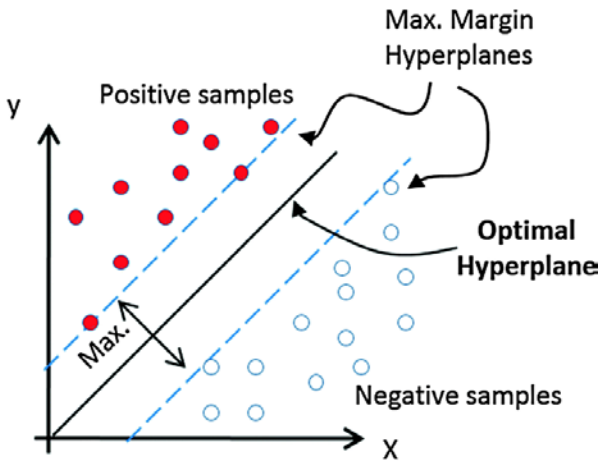


Fig4.2 SVM schema

4.3 ANN (Artificial Neural Network)

Dr. Robert Hecht-Nielsen defines a neural network as “a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs “.

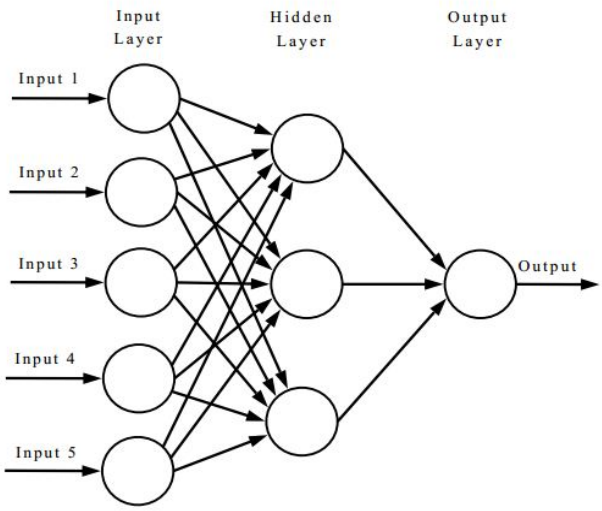


Fig. 4.3 ANN Schema

An artificial neural network is composed of many artificial neurons that are linked together according to specific network architecture. The objective of the neural network is to transform the inputs into meaningful outputs.

V. TOOLS

We have referred to various tools used by researchers in the literature. We have decided to focus on open source free available tools to be focused for time being. Tools which we studied and would be useful in order to extract the features from an audio file are Jaudio and Praat.

5.1 Jaudio

Jaudio is an open source software package for extracting features from audio files as well as for iteratively developing and sharing new features. The features can then be processed by other packages such as Weka, M2K or ACE (Autonomous Classifier Engine). These extracted features can then be used in many areas of MIR research, processing with machine learning framework such as ACE. Jaudio provides a comprehensive solution to the problem of the duplication of work in programming feature extraction. This system permits general use of a large number of features in a fashion that is both easy to use and extensible [5].

5.2 PRAAT

Praat is a public domain, widely used speech analysis toolkit that is supported on a variety of platforms (e.g., Windows, Macintosh, Linux, and Solaris). It provides a variety of value data structures, such as Text Grid, Pitch Tier, and Table, to represent various types of information used for extracting prosodic features [6]. It provides a built-in programmable scripting language for calling Praat's commands and extending its capability. Additions to Praat functionality can be immediately adopted into the tool. This is especially useful for incorporating new prosodic features [6]

VI. CONCLUSION

In this paper, we have considered four emotions happy, peppy, calm, sad. The music dataset can be collected from various websites. Jaudio and praat tools are used to extract the features from music tunes. ANN and SVM can be used as classifiers for recognizing the emotions. We propose to use Jaudio for feature extraction and ANN for classification and test results on the bench mark as perceived emotions by humans. This proposed work is an outcome of detailed study of literature and hands on experience with tools. We may incorporate new methods and algorithms to improve the performance of emotion detection system.

ACKNOWLEDGEMENTS

Acknowledgement is the only way through which we express our deep and profound gratitude to the people who directly or indirectly contribute inspiring guidance, keen interest and continual encouragement in completing this paper. We express our sincere thanks to researchers, our college, MKSSS's Cummins college of Engineering, our principal Dr. Madhuri Khambete and our HOD Mrs. Madhura Tokekar for the support and all those who have provided us valuable guidance towards the completion of this paper.

REFERENCES

[1] N.J. Nalini, S. Palanivel, Emotion Recognition in Music Signal using AANN and SVM, International Journal of Computer Applications (0975 – 8887) Volume 77– No.2, September 2013.

[2] Yu-Hao Chin, Chang-Hong Lin, Ernestasia Siahaan, and Jia-Ching Wang, Happiness Detection in Music Using Hierarchical SVMs with Dual Types of Kernels

[3] Sebastian Napiorkowski, Music Mood Classification- an SVM based approach Topics on Computer Music (Seminar Report) HPAC - RWTH - SS2015.

[4] Gursimran Kour, Neha Mehan, Music Genre Classification using MFCC, SVM and BPNN, International Journal of Computer Applications (0975 – 8887) Volume 112 – No. 6, February 2015

[5] Daniel McEnnis, Cory McKay, Ichiro Fujinaga, Philippe Depalle, JAUDIO: A FEATURE EXTRACTION LIBRARY, c 2005 Queen Mary, University of London.

[6] Zhongqiang Huang, Lei Chen, Mary Harper, An Open Source Prosodic Feature Extraction Tool, (Z. Huang, 2006).

[7] Cory McKay, jAudio: Towards a standardized extensible audio music feature extraction system, 2006

[8] E. Vijayavani, S. Lavanya, P. Suganya, E. Elakiya Emotion Recognition Based on MFCC Features using SVM, International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 4, April 2014.

[9] Janet Marques and Pedro J. Moreno, A Study of Musical Instrument Classification Using Gaussian Mixture Models and Support Vector Machines, CRL 99/4 June 1999

[10] T.N.CHARANYA, R.VIJAYALAKSHMI, Music emotion recognition using support vector machines and regression approach, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 1, January 2015

AUTHORS

First Author- Rishika Shetty, Cummins College of Engineering, rishika.shetty@cumminscollege.in

Second Author – Shweta Kasbe, Cummins College of Engineering, shweta.kasbe@cumminscollege.in
Third Author - Kimaya Jorwekar, Cummins College of Engineering, kimaya.jorwekar@cumminscollege.in
Fourth Author - Dharti Kamble, Cummins College of Engineering, dharti.kamble@cumminscollege.in
Fifth Author– Prof. Makarand Velankar, Faculty Information technology, Cummins College of Engineering makarand.velankar@cumminscollege.in