E-MUSIC

JANANI PRABHA S[1], ANTONY JENITH S[2], DHAKSHANAMOORTHY V[3], CHINMAYA S[4]

[1,2,3]UG Scholar, Department of Computer Science and Engineering, SNS College of Technology, Coimbatore, India.
[4]Assistant Professor, Department of Computer Science and Engineering, SNS College of Technology, Coimbatore, India.

Abstract: *The objective of this project is to detect emotion and select music to be played based on the detected emotion. Music can be a powerful tool to describe human emotion here in this project that can help the user to play music based on stated emotion or detected one. The project builds an application, that is extremely easy to use and also easy in a technical way of usage. The second objective is to use a Convolutional Neural Network with high accuracy throughout train as much as possible of data, and also test the results to check for any errors because if so, the data should be trained once again the methodology of solving this problem is to build a fully functional app that solves this problem, starting from the front end there an easy and understandable interface anyone can use. On the back end the main algorithm in this project is to build a Convolutional Neural Network to help in the goal of achieving high accuracy rate, because the Convolutional Neural Network is the best in the science of building any network that works with images. A fully functional app that built to solve this problem (Desktop Only) and also trained almost 28000 images with different states of emotions with a very high accuracy rate which is "85%" for training and "83%" for testing rate, the application is successfully suggesting music by suggesting single songs that fits any user's emotion*

Keywords: Machine learning, emotion recognition, Convolution Neural Network.

1. Introduction

Music is an art form and activity whose medium is Sound organized in time. Music plays a very important role in improving an individual's life as it is important medium. As now in today's world, multimedia system and their technology is increasing and advancement of this system is also increasing. Today in the field of music, many new technologies are introduced to world. Various music players are developed with option like quick forwarding, reversing the song back, repeating mode shuffling variable playback speed and many other options are increasing in this field. Though all these options satisfy the user's basic need. But yet user has got the task to select his particular song on the browser and choose the songs depending on his current behaviour or his emotional state. So, behavioural and emotional states are classified by the user's facial expressions or with his audio analysis and feature management which is used to detect user's current situation and that particular state are classified and accordingly the playlist is generated. It automatically captures the photographs of the users based on the recognizing the facial expression, mood is classified. In the same manner audio analysis and feature management is done to provide a group of state based on the playlist. And these inputs are provided to classify the user behavioural and emotional state and automatically desired playlist is generated accordingly.

2. Literature Review

2.1. Machine Learning

Machine learning is a field of computer science that uses statistical techniques to give computer systems the ability to "learn" (i.e., progressively improve performance on a specific task) with data, without being explicitly programmed. The name machine learning was coined in 1959 by Arthur Samuel. Evolved from the study of pattern recognition and computational learning theory in artificial intelligence, machine learning explores the study and construction of algorithms that can learn from and make predictions on data– such algorithms overcome following strictly static program instructions by making data-driven predictions or decisions, through building a model from sample inputs. Machine learning is employed in a range of computing tasks where designing and programming explicit algorithms with good performance is difficult or infeasible; example applications include email filtering, detection of network intruders or malicious insiders working towards a data breach, optical character recognition (OCR), learning to rank, and computer vision. Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, Where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning. Machine learning can also be unsupervised and be used to learn and establish

baseline behavioural profiles for various entities and then used to find meaningful anomalies. Within the field of data analytics, machine learning is a method used to devise complex models and algorithms that lend themselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, 3 engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover "hidden insights" through learning from historical relationships and trends in the data. Effective machine learning is difficult because finding patterns is hard and often not enough training data are available; as a result, machine-learning programs often fail to deliver. Our work is based and worked and it is run on python runtime library so we refer[1] this reference gives python contribution for scientific area, and our player plays the music according to the type of emotion and it is based on support vector machine learning technique so we refer [2] which introduce some concepts of machine learning[3] We refer this to understand the language for data mining and machine learning.[4]This video link gives brief idea for emotion recognition in python.[5]This video link gives out brief idea of facial recognition in python. We need a dataset for our player to detect type of emotion, we need to generate it so we refer [6] for data set generation,[7] for emotion recognition with python and open CV.

2.2. Steps in Machine Learning

Machine learning is a field of computer science that gives computers the ability to learn without being programmed explicitly. The power of machine learning is that you can determine how to differentiate using models, rather than using human judgment. The basic steps that lead to machine learning and will teach you how it works are described below in a big picture: 1. Gathering data 2. Preparing that data 3. Choosing a model 4. Training 5. Evaluation 6. Hyper parameter tuning4 7. Prediction.

2.3. Gathering Data

Once you know exactly what you want and the equipment's are in hand, it takes you to the first real step of machine learning- Gathering Data. This step is very crucial as the quality and quantity of data gathered will directly determine how good the predictive model will turn out to be. The data collected is then tabulated and called as Training Data.

2.4. Data Preparation

After the training data is gathered, you move on to the next step of machine learning: Data preparation, where the data is loaded into a suitable place and then prepared for use in machine learning training. Here, the data is first put all together and then the order is randomized as the order of data should not affect what is learned. This is also a good enough time to do any visualizations of the data, as that will help you see if there are any relevant relationships between the different variables, how you can take their advantage and as well as show you if there are any data imbalances present. Also, the data now has to be split into two parts. The first part that is used in training our model, will be the majority of the dataset and the second will be used for the evaluation of the trained model's performance. The other forms of adjusting and manipulation like normalization, error correction, and more take place at this step.

2.5. Choosing a model

The next step that follows in the workflow is choosing a model among the many that researchers and data scientists have created over the years. Make the choice of the right one that should get the job done.

2.6. Training

After the before steps are completed, you then move onto what is often considered the bulk of machine learning called training where the data is used to incrementally improve the model's ability to predict. The training process involves initializing some random values for say A and B of our model, predict the output with those values, then compare it with the model's prediction and then adjust the values so that they match the predictions that were made previously. This process then repeats and each cycle of updating is called one training step.

2.7. Evaluation

Once training is complete, you now check if it is good enough using this step. This is where that dataset you set aside earlier comes into play. Evaluation allows the testing of the model against data that has never been seen and used for training and is meant to be representative of how the model might perform when in the real world.

2.8. Parameter Tuning

Once the evaluation is over, any further improvement in your training can be possible by tuning the parameters. There were a few parameters that were implicitly assumed when the training was done. Another parameter included is the learning rate that defines how far the line is shifted during each step, based on the information from the previous training step. These values all play a role in the accuracy of the training model, and how long the training will take. For models that are more complex, initial conditions play a significant role in the determination of the outcome of training. Differences can be seen depending on whether a model starts off training with values initialized to zeroes versus some distribution of values, which then

leads to the question of which distribution is to be used. Since there are many considerations at this phase of training, it's important that you define what makes a model good. These parameters are referred to as Hyper parameters. The 6 adjustment or tuning of these parameters depends on the dataset, model, and the training process. Once you are done with these parameters and are satisfied you can move on to the last step.

2.9. Prediction

Machine learning is basically using data to answer questions. So, this is the final step where you get to answer few questions. This is the point where the value of machine learning is realized. Here you can Finally use your model to predict the outcome of what you want. The above-mentioned steps take you from where you create a model to where you Predict its output and thus acts as a learning path.

3. Project Analysis

3.1. E-Music:

This project develops a Convolutional Neural Network. The CNN was selected because this project is images-based solution, there are many ways to solve this problem through them, like body language, voices techniques. Facial Expressions is a way to detect emotions and that's the methodology implemented through the Convolutional Neural Network and the science itself. This project is a trial to implement and solve one of the most important problems that no one around notice, music is taking a very huge part of our everyday routine and we choose what to listen to base on many things on them is Emotions. Emotions and music are nearly close in their structure and the way they built with are almost near. In this project is a simple implementation that applies playing music based on any user's Emotion, maybe he/she is happy, sad, nervous, neutral, etc. This is one of the hardest problems that not many researchers around solved, so this project is a real hard work to solve this problem in a very effective way. Also, in this project we are studying Emotions in most of the parts, what is Emotions? How it can be constructed? And also, how to detect them? Because knowing these things will me it easy develop the app and also help the algorithms know what to learn and why from the beginning. Among the history the constructions of the emotions are very different from history to another and from group of people to another, that's what makes it not easy to solve, and at the beginning it will start be studying very small group of people and expanding to the others by the time.

3.2. Problem Statement

Music listeners have tough time creating and segregating the play-list manually when they have hundreds of songs. It is also difficult to keep track of all the songs: sometimes songs that are added and never used, wasting a lot of device memory and forcing the user to find and delete songs manually. Users have to manually select songs every time based on interest and mood.

User's also have difficulty to re-organize and playing music when play style varies. Currently in existing application, music is organized using play-list, and playlist songs cannot be modified or altered in one click. Users have to manually change or update each song in their play-list every time. The sequence of songs in a play-list might not be the same every time, and songs that a user wants to listen frequently might not be given priority or might be left out from the list. Currently, there are no applications that allows users to play songs on-the-go without selecting songs manually or from a play-list. Motivation As a music lover, I've always felt that music players should do far more things than just playing songs and allowing users to create play-lists. A music player should be intelligent and act according to user's preferences. A music player should help users organize and play the songs automatically without putting much effort into selection and re-organization of songs. The Emotion-Based Music Player provides a better platform to all the music listeners, and ensures automation of song selection and periodic updating of play-lists. This helps users organize and play songs based on their moods. The player should also give recommendation for users to change songs on-the-go. It calculates song weight based on EMO-algorithm (discussed in chapter 3 4) to help users have more customized and organized play-lists.

3.3. Proposed System

Here we propose a Emotion based music player (Emo Player). Emo player is an music player which play songs according to the emotion of the user.. It aims to provide user preferred music with emotion awareness. Emo player is based on the idea of automating much of the interaction between the music player and its user. The emotions are recognized using a machine learning method Support Vector Machine (SVM) algorithm. In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. It finds an optimal boundary between the possible outputs. The training dataset which we used is Olivetti faces which contain 400 faces and its desired values or parameters. The webcam captures the image of the user. It then extract the facial features of the user from the captured image. The training process involves initializing some random values for say smiling and not smiling of our model,

predict the output with those values, then compare it with the model's prediction and then adjust the values so that they match the predictions that were made previously. Evaluation allows the testing of the model against data that has never been seen and used for training and is meant to be representative of how the model might perform when in the real world. According to the emotion, the music will be played from the predefined directories. Advantages of Proposed System • Users don't want to select song manually. • No need of playlist. • Users don't want to classify the songs based on the emotions.

4. SYSTEM SPECIFICATIONS

4.1. Hardware Requirements

The most common set of requirements defined by any operating system or software application is the physical computer resources, also known as hardware. The hardware requirements required for this project are: • Intel i3 • 4GB RAM • Webcam • Speaker

4.2. Software Requirements

Software Requirements deal with defining software resource requirements and prerequisites that need to be installed on a computer to provide optimal functioning of an application. These requirements or pre-requisites are generally not included in the software installation package and need to be installed separately before the software is installed. The software requirements that are required for this project are:

• Python 3.6

• Open CV 3.1

• Tensorflow

• Keras

• Pycopy-webbrowser

• Numpy

• Os-Sys

• HTML

•JS & CSS

5. PROJECT DESCRIPTION

5.1. Existing System:

The current most popular emotion based music player is stereo mood. It lacks capabilities in the sense that the user needs to type in what he is feeling, rather than using computer vision to determine his emotion. To solve the problem of emotion recognition a lot of work has been done in the past. To extract and determine the emotion of a user, we need to extract features from an image and use them against a trained data set to classify the input and determine the emotion. In the existing system, we need to set mood manually like happy or sad and then get your mood automatically by analysing a periodical camera capture. Limitations of existing system • It requires the user to manually select the songs. • Randomly played songs may not match to the mood of the user. • User has to classify the songs into various emotions and then for playing the songs user has to manually select a particular emotion.

5.2. Need:

The proposed model is able to extract user's expression by capturing of facial expressions of the user using camera and thus will detect user's emotion. The proposed system will only depends on the image captured and the song will play according to the mood specified by analysis and recognition of the image.

5.3. Proposed Methodologies

As outlined the work is divided in to two major parts: Image Processing, which is used for recognizing human emotions; and mp3 music file analysis used to extract song information and classify songs.

5.4. Past Work

The current most popular emotion based music player is Stereo Mood. It lacks capabilities in the sense that the user needs to type in what he is feeling, rather than using computer vision to determine his emotion. To solve the problem of emotion recognition a lot of work has been done in the past. To extract and determine the emotion of a user, we need to extract features from an image and use them against a trained data set to classify the input and determine the emotion.

5.5. Feature Extractors

A feature extractor is an application which extracts important points in an image. Different works have been done in the field of Computer vision for feature extractors, the most prominent ones being Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF). Each of these has different impacts on classifying the emotion of the user. I initially work with a new technique called Binary Robust Independent Elementary Features (BRIEF) before moving onto techniques such as SURF and SIFT.

5.6. Classifiers and Prediction

After extracting features from an image set of training and testing data, a feature classifier is needed to sort out and classify the testing data with relevance to the training data. A Support Vector Machine (SVM) is the

most predominantly used classifier to tackle the emotion recognition problem. For experimental purposes I use an SVM and a Naive Bayes Classifier.

5.7. Facial Emotion Recognition

Several approaches have been proposed to classify human affective states. The features used are typically based on displacements of specific points or spatial locations of particular points; this technique is known as Facial Action Coding System(FACS). In an approach taken by Liu et al in, he presents an algorithm for classification of brain electrical signals in human emotions. This algorithm was based on the model of fractal dimension. He proposed a bi dimensional Valence - Arousal approach, where by the six emotions are divided into different categories Black et al explored the use of local parameterized models of image motion for recovering and recognizing the non-rigid and articulated motion of human faces. They used these parametric models to extract the shape of the mouth, eyes and eyebrows. They achieved a high success rate of 95% to classify Happy, 90% to classify Anger and a 100% success rate to classify the Sad emotions. On the other hand the approach used by Yacoob and Davis in which facial expressions are recognized in image sequences using statistical properties of the optical flow with only very weak models of facial shape. In this project several approaches are considered, including a Principal Component Analysis (PCA) approach, using multiple Facial Action Units and different feature extractors with clustering approaches. Each of these approaches is used with different classifiers to determine the emotion of the user with the accuracy tested on a data set of 40 subjects each. Project is about an algorithm which generates the playlist by selecting the music. The playlist is generated on the basis of user's behavioural and emotional state. The system tries to provide an interactive way for the user to carry a job of generating the playlist. The proposed system here simply uses the algorithm by its proposed modules such automatically capturing the picture, recognizing the facial expression, analysis of audio feature management, analysing users state and interfacing the user and these modules work on the basis of different mechanism which are carried out and works in such a manner that the desired output that is the playlist is generated.

The working can be started as follows:

• The proposed system works by providing user interface which further prompts the user and detects its state by using algorithm.

• After analysing the interface files are detected, and images are capture automatically. Facial expression is recognised using the concept of image processing.

• The recognition is extracted and is then classified accordingly, which is further use to analyse behavioural and emotional state.

• After this the limited set of types based on the audio analysis and the feature management values are to be processed.

• In the next step playlist is generated which divides the songs as per the further analysis and selecting the matched songs are generated in the playlist and from this generated playlist songs are played. The architectural view for this proposed system is given below:

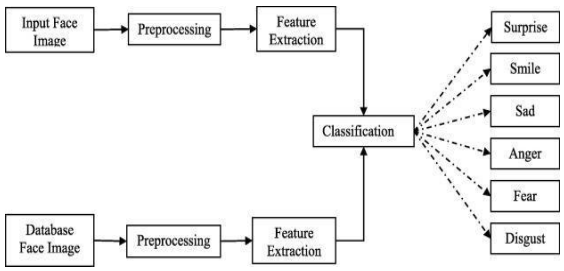

Fig.1. System Architecture

6. WORKING MODULES

6.1. Automatic capture and Input Photographs

In this module camera is opened and with help of image processing system camera automatically capture the photographs of the user and takes the real time images which is provided for further use.

6.2. Facial expression Reorganization

After capturing the images, the recognizing of that particular image is processed i.e., recognition of facial expression of user.

6.3. Analysing behavioural and emotional state

While detecting the images state recognition is generated with help of image processing system and required state is analysed for further generating a playlist.

6.4. Generating the playlist

Different types of songs are generated but accordingly the playlist is generated.

6.5. Audio analysis and feature management

In this module a list of songs are generated in the form of input. The input is then matched with the output.

6.6. Objectives

The key objectives of this project can be split into two parts, the recognition of the emotion of the user and music analysis. The project is centred and focused more towards different approaches to emotion recognition and the impact of each technique used. The emotion recognition stage is heavily based on image processing and machine learning. The music analysis is done by reading the MP3 metadata of a music file.

6.7. Emotion Recognition

The key aim of this section is to implement and analyse various techniques to extract features and classify the emotion of a person. The image processing step requires turning the image to grayscale and resizing it. This is followed by extracting multiple features using different techniques and adapting different classifiers to determine the mood of the user. Using these different methods and techniques, an analysis is made to determine the best solution for the emotion recognition problem based on my project.

6.8. Music Analysis

Using the bit stream from mp3 files, we extract metadata to determine the required information for each particular song. Using the determined emotion, create a playlist of songs for the user.

6.9. Evaluation:

To evaluate the quality of this app it will be very hard to say that the app is always working properly, many time the app crashes at the run and many times the app crashes while suggesting an image, so the app as a normal functionalities is build right and also working right, but not perfect. In the next part, there will be a discussion regarding the results and the accuracy rate.

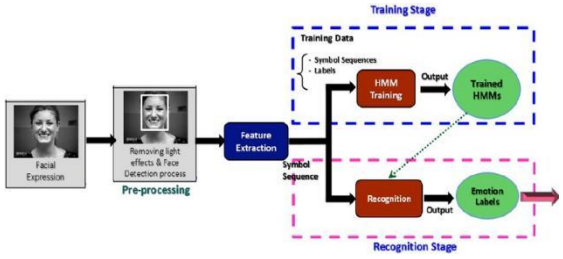

Fig.2. Working Process

6.10. Results and Discussions

Results and Discussions In the experiment, the images dataset was very huge with almost 27000 images or even more and there are 7 main songs with different varieties genres and tempos. The accuracy rate for the app was 85% which somehow good but it needs more training to be more accurate. The app is working properly with almost 90% of the running trials, which is also a good results, but when it comes to detecting the face in real time working also properly and very well, but when it comes to detecting the app at the beginning did the recognition so many times wrong but by more training and more editing through the Convolutional Neural Network scripts the detection is improving every time the app do moretraining. To say that the app is running is a perfect thing but it has to make sure that the CNN is implemented right, with the layer needed to build a complete fully functional Convolutional Neural Network starting from the input layer which always work properly, to convolutional layer which also important to the basic 37 feature selection process and then the pool layer which turns every negative value within the matrix to zero which means detecting the black nodes within the images or the matrix, all of the above layers are implemented right, the last layer which is the most critical layer, it is also implemented right and it is used to extract the very advanced features from the images.

6.11. Future Work

There are two limitations facing this project the first one is developing mobile app and the second one is to encrypt and make user's data safe from being stolen. Developing mobile app will be the smartest way in solving this problem because it is easy use and can be in any user's pocket all time. Then developing a secure app is also important as a future work because there is a risk when it comes to user's data, this can be done by encrypting the image itself after receiving it as input from the user.

References

[1] "An Intuitive Explanation of Convolutional Neural Networks," 2016. [Online]. Available:https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets/. [Accessed 2021].

[2] K. Cherry, "Overview of the 6 Major Theories of Emotion," 06 May 2019. [Online].Available: https://www.verywellmind.com/theories-of-emotion2795717.

[3] Shlok Gilda, Husain Zafar, Chintan Soni and KshitijaWaghurdekar, "Smart Music Player Integrating Facial Emotion Recognition and Music," Department of Computer Engineering, Pune Institute of Computer Technology, Pune, India, Pune, India, 2017.

[4] Sushmita G. Kamble, and Asso. Prof. A. H. Kulkarni , "Facial Expression Based Music Player," in Facial Expression Based Music Player , Jaipur, India, 2016.

[5] EDM Prod, "TL; DR Music Theory: Pentatonic Scales," EDM Prod, 2019. [Online].Available:

https://www.edmprod.com/tldr-music-theory-pentatonicscales/.

[6] M. Chingshun Lin, "Music Emotion Recognition Based on two level support vector machine", in 2016 International Conference on Machine Learning and Cybernetics, Jeju, Korea, 2016.

[7] MATLAB , "What Is Deep Learning?," MATLAB, 2019. [Online]. Available: https://www.mathworks.com/discovery/deep-learning.html. [Accessed 2019].

[8] G. S. Dan Duncan, "Facial Emotion Recognition in Real Time," Dan Duncan, Gutam Shin. [9] OpenCV, "About OpenCV," OpenCV, 2019.[Online].Available:https://opencv.org/about/.

[10]https://www.pyimagesearch.com/2018/09/10/keras-tutorial-how-to-get-started-with- keras-deep-learning-and-python

[11]https://towardsdatascience.com/face-detection-recognition-and-emotion-detection-in- 8-lines-of-code-b2ce32d4d5de