

Chatbot Song Recommender System

Prof Rahul Naitam¹, Akash Dahat², Akshit Patil³, Ksitij shanware⁴, Ashish Anasane⁵

¹Assistant Professor, ^{2,3,4,5} Students,
Department of Information Technology & Engineering
Wainganga College of engineering and management, Nagpur, India, 441108

rahulnaitam.wcem@gmail.com

Received on: 03 April, 2023

Revised on: 25 April,2023

Published on: 27 April,2023

Abstract- *In this era of technological advancements, music recommendation based on mood is much needed as it will help humans relieve stress and listen to soothing music according to their mood. In this project, we have implemented a chatbot that recommends music as well as movies based on the user's mood. The objective of our application is to identify the mood expressed by the user and once the mood is identified, songs are played by the application or a list of movies are displayed in the form of a website according to the choice made by the user and also his current mood. Our proposed system is implemented as an application which can be run on the user's desktop and its main focus is to reliably determine the user's mood. Human computer interaction (HCI) has a lot of importance in today's world and the most popular concept in HCI is recognition of emotion from facial images. In this process, the frontal view of the facial images is utilized so as to detect the mood from the images. Another important factor is the extraction of facial elements from the user's face. We have used the Haar Cascade Algorithm for accurately detecting the user's face in the live webcam feed and the CNN Algorithm is used to detect the emotion being expressed by the user from the facial features. Facial attributes like the arrangement of the mouth and eyes are used in order to detect the mood of the user.*

Keywords—*Haar cascade, mood detection, mood based recommendation, CNN*

I -INTRODUCTION

Emotion detection is an important process in our project which requires accuracy and this can be done

effectively with the help of facial expressions which is how humans understand and interpret an emotion. Research shows that when a person's facial expressions are read, it can actually vary your interpretation of what is being spoken and it can also control how the conversation turns out. Humans are capable of perceiving emotions which is exceedingly important for a communication to be a success and hence in a typical conversation almost 93% of communication depends on the emotion being expressed.

In our project, the process of emotion detection of the user is done with the help of facial images that are captured through the live webcam feed. Happy, sad, angry, fear, surprised, disgust, and neutral are the seven basic emotions -common to humans, and they are identified by the various expressions of the face as depicted in Fig 1. In this project we aim to find and implement an effective way to identify all these emotions from frontal facial emotion. The positioning and shape of for example the eyebrows and lips are used by the application so it can understand and interpret the facial attributes that make up the expression and thus the emotion being expressed by the user. Fig 2 demonstrates how various facial features are taken into consideration to identify the emotion.

The Chatbot module of the application makes use of AI techniques for its implementation. Our chatbot is rule based which is the AI methodology used to design a

simple Chatbot.

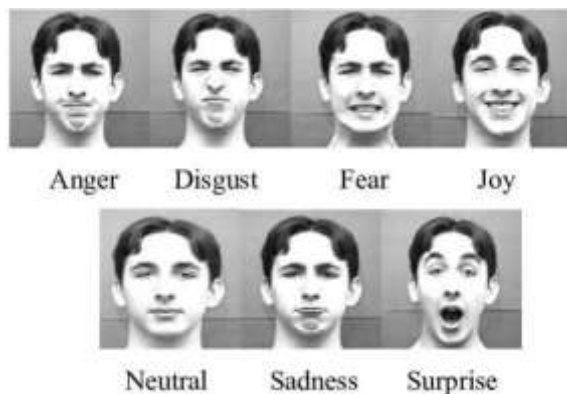


Fig. 1. The seven basic emotions

We have made use of rule based chatbot as our application required a simple chatbot. The emotion detection module utilizes Deep Learning algorithms for identifying the face of the user in the input image and then accurately determine the emotion displayed on the user’s face. It implements two algorithms, the Haar Cascade Algorithm is used for identifying Active Appearance Model (AAM) technique wherein various points on the face are located forming a feature such as eye, eyebrows and mouth. These collected points are then used to generate a data file that provides the necessary details about the model points identified and hence detects the facial expression that is given as input to the AAM model.

II- METHODOLOGY

The application developed in our project is called “MoodBot”, the application primarily is a Chatbot application which incorporates the emotion detection module. The emotion detection module is used for identifying the emotion expressed by the user and hence making it essential to the application as it provides the entertainment in the form of Music and Movies according to the user’s mood. The application consists of three main modules: Chatbot, Mood detection and Music/Movie recommendation. Fig 3 illustrates the block diagram for the working of the application presented here.

As shown in Fig 3, once the application is opened the user’s screen displays the chatbot window, which acts as the base of the application. The chatbot application named MoodBot provides the user with three options. The first

one being chatting, that is the user can chat with the chatbot using the textbox to type in the message and then click on the send button to send the message. Second option is to click on the “My Mood” button, upon which the chatbot application will start the emotion detection process. The last option is to simply quit the application.

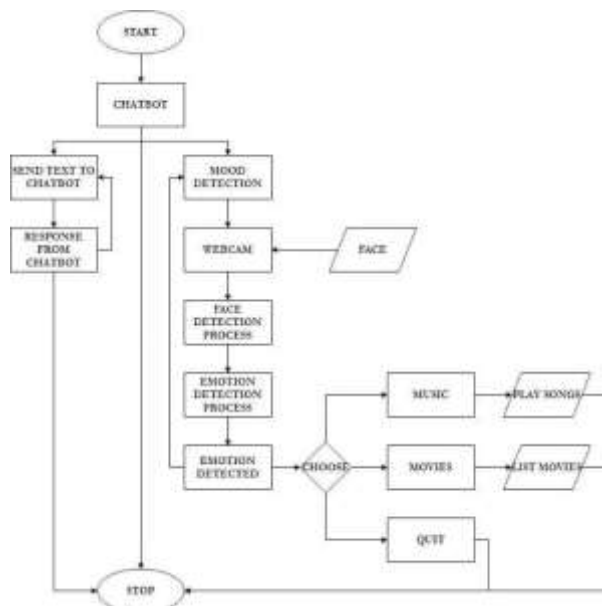


Fig. 3. Block Diagram of the Proposed system

When the user selects the My Mood option the application starts the emotion detection process. The emotion detection process involves using the webcam/camera to capture the face and passing it to the face

detection process where the face is analyzed. It is then passed on to the emotion detection process which analyzes the face features and classifies the emotion into one among the seven classes.

Once the current mood of the user is detected, the application uses a pop-up window to display the user’s mood identified by the application and provides the user with three choices. The first one is music, when this is selected the application will start playing songs based on the user’s mood. The second option is movies, when this option is selected the application opens the Movie for Your Mood Website, a specially designed website which displays a list of movies appropriate for the user’s current mood. The last option is to quit the application. Every time the user feels a change of mood, all he needs to do is click on

the My Mood button and the application will do the rest. Also, the user can continue to do other tasks on the computer, as the music will continue to play in the background.

A. Artificial Intelligence

Artificial intelligence (AI) is described as simulating human intellect in machines so that they are capable of thinking like humans and imitate their actions as they're programmed to do so. The term AI can basically be associated with any machine that indicates the presence of attributes related to a human mind like learning and problem solving. AI is an interdisciplinary science comprising various outlooks, yet the breakthroughs happening in machine learning and deep learning are creating a paradigm shift in virtually every sector of the tech industry. To this date, Neural networks and fuzzy logic (FL) are the more commonly and frequently used AI technologies.

B. Chatbot

Chatbot also known as chatterbot, is a popular AI application which allows it to be incorporated and used via any significant messaging applications and it prompts human conversation using voice commands or text chats or both. Machine Learning and NLP (Natural Language Processing) are the more often used AI technologies for developing chatbot applications. The chatbot module in our proposed system implements Rule based Chatbot which follows a list of predefined rules for answering the queries the user has listed. Rule-based chatbots are mainly used by basic applications that are trained to answer the questions according to the rules.

Rule-based chatbots might be incapable of interpreting complicated conversations, which is its only setback. It is able to only execute the tasks it has been programmed to do unless the developer decides to add in more upgrades. The future of chatbots includes it being equipped with Emotion AI and advanced sentiments analytics in order to understand and interpret the conversations more like human beings.

C. Deep Learning

Deep learning basically belongs to a larger class of machine learning techniques based on artificial neural networks with representation learning. It comes under

machine learning in artificial intelligence (AI) that has networks capable of learning unsupervised from data that is unstructured or unlabeled. It is also referred to as deep neural learning or deep neural network. The main strength of deep learning algorithms is in learning processes and it produces a high degree of intelligence to systems based on them. In deep neural networks, the deep refers to the aspect that multiple layers of processing lead to the transformation of the input data be it images, speech, or text into certain output useful for making decisions. In deep learning, a computer model is able to learn how to classify tasks directly from images, text, or sound.

Deep learning models can achieve high levels of accuracy, sometimes surpassing human-level performance. A large set of labelled data and neural network architectures with many layers is used to train the model. Deep learning algorithms achieve recognition accuracy at higher levels than ever before. Recent advances in deep learning have improved a great deal where deep learning outperforms humans in some tasks like classifying objects in images.

A. Face Detection – Haar Cascade Algorithm

Face identification is among the most basic applications utilized in face recognition technology and it plays a vital role

in emotion detection. Our application first needs to identify the face of the user in order to recognize the emotion being expressed. Face detection is a type of application that comes within "computer vision" technology. It is a technique wherein they design algorithms and train them so that it is capable of locating the faces or objects accurately within the images. The images could be captured in real time or from pictures, our application gathers images of the user from each frame of the live webcam feed. A popular use of this technology is in airport security systems and in smartphones for locking and unlocking it using face ID.

The application we have developed makes use of the Haar Cascade Algorithm for the face detection process. Haar Cascade is a machine learning object and face detection algorithm whose objective is to detect objects or faces in a video or image. This algorithm is developed on the notion of features presented by Paul Viola and Michael Jones in their paper "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001.

Here a cascade function is given training with the help of many positive and negative images and this approach is based on machine learning. After this it is used to recognize objects in other images and this same concept can be extended for identifying faces in images.

Additionally, the Face detection process uses classifiers, that are algorithms whose objective is to detect whether there is a face present (1) or no face present (0) in an image. Immense training has been given to these classifiers in order to identify the faces with the help of several images so as to achieve greater reliability. Our project makes use of OpenCV library of python that consists of two sorts of classifiers, Haar Cascades and LBP (Local Binary Pattern). Out of these two our application uses the Haar Cascade Classifier for the face identification process.

Initially, many positive images that contain faces and negative images that do not have any faces are required by the algorithm for training the classifier. Next, we choose facial attributes from the image so they can be extracted. We first fetch the Haar Features. A Haar feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums. Each feature is a single value obtained by subtracting the sum of pixels under white rectangle from the sum of pixels under black rectangle. Fig 4 shows how the Haar features are selected and evaluated.

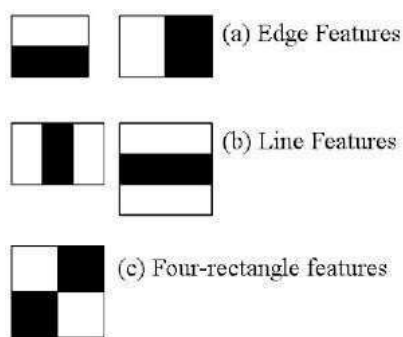


Fig. 4. Haar Features

The Adaboost method is used for the selection of the best features out of 150000+ possible features and it also trains the classifiers that make use of them. To do this, all the features are applied to all the training images and then it

finds the best threshold for each of these features that helps in classifying the faces as positive and negative. The features with the least error rate are selected as they best classify the face and non-face images. The final classifier is a weighted sum of the weak classifiers, it says weak as these on their own are unable to classify the image but when combined with others form a strong classifier. Cascade of Classifiers is used to check a possible face region as it is known that a major part of the image is non-face region. The features are grouped into different levels of classifiers and then applied one by one on a window. When a window fails the first level, it is discarded and the remaining features aren't applied to it. In case it passes, the second level of features is applied to it and the same process is continued. When the window has passed all levels then it is concluded that it is a face region and in this way the classifier is taught to distinguish between face and no face.

A. Emotion Detection – CNN Algorithm

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which takes an image as input, assigns importance (learnable weights and biases) to the various aspects in the image based on which it is capable of differentiating one from the other. This deep learning architecture is commonly used and sought-after. In comparison to other classification methods ConvNet requires lesser pre-processing. In earlier methods filters are hand-engineered but ConvNets are capable of learning these filters/characteristics with enough training.

CNN is computationally efficient. Special convolution and pooling operations are used by it and it also conducts parameter sharing which allows the CNN models to run on any device, making them universally appealing. When using CNNs, there is no need to identify the features required for the classification of images, thereby eliminating the need for extracting the features manually. CNN works by extracting features directly from the images. The relevant features are not pre trained because they are learned while the network trains on a collection of images.

The Convolutional Neural Network is known for its work with image data and it is highly used and recommended for identifying what an image could be or what the image contains. The basic CNN structure is as follows: Convolution -> Pooling -> Convolution ->

Pooling

-> Fully Connected Layer -> Output. Fig 6 illustrates how a typical convolutional neural network works. Our application makes use of such a structure for the emotion detection process. Once the Haar Cascade method has successfully detected the presence of a face in the image taken as input, it is then handed over to the CNN structure for identifying the mood expressed by the user.

In CNN, convolution is used to create feature maps using the original data and pooling is down-sampling which is commonly in the form of max-pooling, and here a region is selected for which the maximum value is calculated and this value becomes the new value for that entire region. Fully Connected Layers are typical neural networks, where all nodes are fully connected.

The mood or the emotional state of the user can be interpreted using his facial expressions which are just the positioning of the facial muscles. The application developed here will be trained to interpret and distinguish among the seven fundamental emotional states which includes happy, sad, fear, neutral, angry, surprised, and disgust. The convolutional neural network is trained with the help of FER2013 dataset and different hyper-parameters are used to refine the model.

Fer2013 is basically used to identify and classify the emotion. It is an open source dataset which is used to group all the faces according to the emotion depicted in the facial expression into one among the seven classes (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral). It has two main columns – emotion and pixel. The emotion column consists of a key (0-6) that represents the emotion being expressed in that image.

It is known that all images are made up of pixels which are represented with the help of numbers. All colored images have three color channels — red, green, and blue — and each channel is represented by a grid or 2-dimensional array. A number between 0 and 255 is stored in each cell of the grid, thus denoting the intensity of that cell. When the alignment of the three channels happens, the image appears as perceived normally by us.

The dataset we have considered consists of images that are 48x48 in dimension. There exists only one channel as all the images in the dataset are gray-scale. The image data is extracted and arranged into a 48x48

array. To normalize the data, it is converted into unsigned integers and then divided by 255 as it is the maximum possible value a single cell can have and also it ensures that all the values lie between 0 and 1. This application is implemented as a desktop

1. Next the Usage column will be checked and the data will be stored in different lists, where one will be used to train the network and the other to test it. To create a Sequential Convolutional Network, we have made use of Keras which is implemented on top of TensorFlow in our project. The emotion detection process uses a four-level CNN framework.

First, the primary expressional vector (EV) is extracted and generated finding the various relevant facial points of importance, using CNN. The changes happening in the expression are directly connected to EV. On receiving the image which serves as the input data to each of the convolutional layers, it changes it and the result of which is passed on to the next level. In this way CNN is used to recognize the mood of the user.

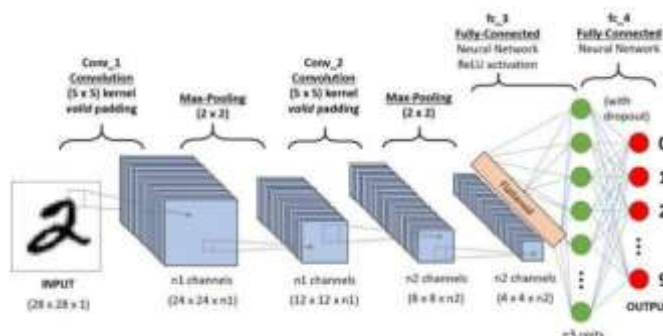


Fig. 5. CNN Structure

A. Music/Movie Recommendation

Once the emotion has been detected and classified into one of the seven categories, a popup box is displayed on the user's screen with the options music and movie. He can make a selection among the two and accordingly music will be played or a list of movies will be displayed. If the user chooses music, a song is played based on the emotion detected. The songs will play one after the other, until the user asks the chatbot to detect the mood again.

When the user selects the movie option, he is redirected to a movie website that is specially designed for this application and classifies the movies based on mood. It is a simple movie website that displays the

user's mood for which the list of movies is suggested on the top along with the website name. Only the movies appropriate for the mood are listed on site and they're further categorized into three categories: Latest Movies, Top Rated and Recommended. Each movie listed on the website provides the user with the movie poster and some basic details of the movie such as the genre, rating and also an IMDB link of the movie in case the user is interested and wants to know more about the movie.

III -CONCLUSION

Chatbots are one of the most important advancements of AI Technology. Our project successfully combines this technology with the human's need for entertainment in the form of Music and Movies. In this age and time of technology, such an application would serve the purpose of helping humans relax and relieve their stress. The MoodBot application developed in our project is a simple chatbot that allows users to choose music and movies according to their mood. The application is implemented as a desktop application, thereby being available to the user whenever required. When the user chooses the music option, songs appropriate to his mood are played. And when the choice is movies, the application opens a specially designed website that recommends movies to him as per his current mood.

The future enhancements for the application could be automatic detection of the user's mood which could be initiated when the user opens the application thereby eliminating the need for the "My Mood" button. The emotion detection module could be executed at regular intervals to check for change in the user's emotion and in case there's a change ask the user to make a choice among music or movies.

Other enhancements could be a fully functioning website like IMDB dedicated to suggest movies depending on the mood and also previous choices of the user. Also, could add like and dislike options to the songs being played so that the user's liked songs are played often and the disliked songs are not played.

REFERENCES

[1] G. M. Mate, Nikhil Wadekar, Rohit Chavan, Tejas Rajput, Sameer Pawar, Mood Detection with Chatbot using AI-Desktop Partner, April 2017.

- [2] Ai Thanh Ho, Ilusca L. L. Menezes, Youstra Tagmouti, E-MRS: Emotion-based Movie Recommender System.
- [3] Manish Dixit, Sanjay Silakari, A Hybrid Facial Feature Optimization Approach using Bezier Curve, IEEE, 2015.
- [4] Fatima Zahra Salmam, Abdellah Madani, Mohamed Kissi, Facial Expression Recognition using Decision Trees, 13th International Conference Computer Graphics, Imaging and Visualization, 2016.
- [5] Jae Sik Lee, Jin Chun Lee, Music for My Mood: A Music Recommendation System Based on Context Reasoning, 2006.
- [6] Renuka R. Londhe, Dr. Vrushshen P. Pawar, Analysis of Facial Expression and Recognition Based On Statistical Approach, May 2012.
- [7] Dolly Reney, Dr. Neeta Tripaathi, An Efficient Method to Face and Emotion Detection, Fifth International Conference on Communication Systems and Network Technologies, IEEE, 2015.
- [8] Shan C, Gong S, McOwan PW. Facial expression recognition based on local binary patterns: a comprehensive study. *Image Vis Comput.* 27(6):803–816 (2009).
- [9] S. Jagannatha, M. Niranjanamurthy, and P. Dayananda, "Algorithm Approach: Modelling and Performance Analysis of Software System", *Journal of Computational and Theoretical Nanoscience (American Scientific publishers)*, December 2018, Volume 15, Issue 15, PP. 3389–3397
- [10] Enrique Correa, Arnoud Jonker, Micha'el Ozo, Rob Stolk, Emotion Recognition using Deep Convolutional Neural Networks, June 30, 2016.
- [11] Y. Lv, Z. Feng, and C. Xu, Facial expression recognition via deep learning. In *Smart Computing (SMARTCOMP)*, 2014 International Conferenc