EMOTION based PERSONALIZED RECOMMENDATION SYSTEM

Prof. Vanita Babanne[1], Mrunal Borgaonkar[2], Mrunali Katta[3], Prajakta Kudale[4], Vaishnavi Deshpande[5]

[1]*Professor, Dept. of Computer Engineering, RMDSSOE Pune, Maharashtra, India*
[2,3,4,5]*Student, Dept. of Computer Engineering, RMDSSOE Pune, Maharashtra, India*

---***---

Abstract - *Emotions are an integral part of human psychology and are of various kinds. To differentiate these is one of the objectives of our proposed system. The proposed system is based on the recommendation framework of video contents taking into consideration human emotions. In existing work, the recommendation system is based on the user's previous history or log. The recommendation technique used is not fully personalized or does not consider the user's current point of interest. Hence the solution to this is a personalized content recommendation based on emotional characteristics. In correspondence to our system which also considers aspects such as age, gender, and emotional characteristics in real-time. Our approach is to overcome the accuracy problem in existing applications [1].*

Key Words: LBPH, Emotion detection, Content-based method, Recommendation.

1. INTRODUCTION

Emotions arise as a function of physiological arousal state which is both subjective and private to an individual. Facial expressions play an important role in detecting human emotions and current state of mind. Machine Learning can detect emotions by capturing facial expressions and learning what each expression means. It also applies that knowledge to new information presented.

In recent years, machine learning has achieved success in computer vision and recommendation systems. It has benefits on machine learning-based recommendation algorithms and has made remarkable progress in the following three aspects:

- Powerful representation learning capability
- Collaborative / Content-based filtering
- Deep interaction between features

Recommender systems have traditionally relied on data-centric descriptors for content and user modeling. In recent years we have witnessed an increasing number of attempts to use emotions in different ways to improve the quality of recommender systems. The Automatic recommendation is one of those considerable techniques. Various media sharing platforms are the experimental areas that are automatic machines to collect human-like characteristics. Humans express non-verbal and involuntary channels in many ways, mainly facial expressions. By calculating human facial expressions, we are going to develop an emotion-based video recommendation system.

2. EXISTING SYSTEM APPROACH

Mahesh Babu Mariappan, et.al [2] presented FaceFetch a novel context-based multimedia content recommendation system that understands a user's current emotional state through facial expression recognition and recommends multimedia content to the user. The system can understand a user's emotional state. They provided desktop as well as mobile user interface and multimedia content such as music, movies, and other videos of user interest from the cloud. Shangfei Wang et.al [3] has introduced a general framework for video affective content analysis, which includes video content, emotional descriptors. In this paper, the author's current research was in both direct and implicit video affective content analysis. Their main focus was on direct video affective content analysis. As a result of these developments, video affective content analysis is becoming increasingly important. The goal of video affective content analysis is to automatically tag each video clip by its affective content. Abhishek Tripathi et.al [4] in this paper, they harness the potential of the said two techniques and propose EmoWare (emotion-aware), a personalized, emotionally intelligent video recommendation engine. It uses a context-aware collaborative filtering approach. The users' non-verbal emotional response towards the recommended video is captured through facial expressions. And analysis is done using decision-making and video corpus evolution with real-time feedback streams.

3. PROPOSED SYSTEM

We are proposing a real-time emotion-based recommendation system. This system works as a personalized assistant to the user. Since it's a real-time feature extraction and recommendation system, we extract facial features through video inputs from users. The facial features are calculated through LBP and Haar Cascade algorithms, which helps in detecting the emotion of the user. Facial expressions are produced by the often movements of muscles of the face in a manner and fashion that could change the position and shape both absolute and relative of facial features such as eyes, eyelids, nose, lips, cheek muscles [2].

This system can be used as a personal assistant for entertainment purposes. There are many such systems developed for the recommendation, what makes our system stand out of this is, it is categorized based on personal choices and age groups. Personal choices can change based on gender and age. Hence, based on emotion as well as age and gender, users get recommended videos.

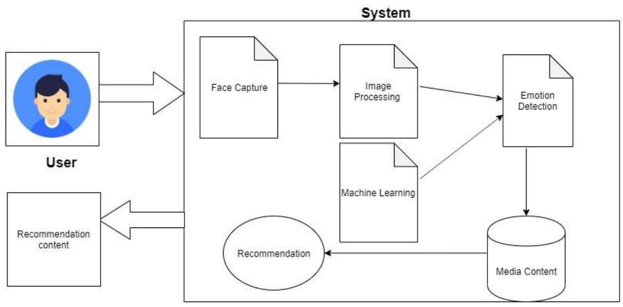

Figure 1: Architecture of the System

4. METHODOLOGY

A. Face Acquisition

Face detection is carried out using Haar cascade classifiers, LBPH [7], and OpenCV. Haar cascade is a classifier that detects faces and body parts in an image. It can be trained to identify almost any object in an image. A set of classifiers called Haar features, processes every region of the input image. Firstly, collect the Haar features. Consider a specific region of image or video frame, then algorithms sum up pixel intensities and positive and negative images are created. After every region is processed, the difference between sums is found out, to extract features from it. The Haar features are responsible for the accuracy of the classifier. LBPH algorithm considers 9 pixels at a time and constructs a histogram. It has parameters such as Radius, Neighbours, Grid X, Grid Y [5]. Using these coordinates, the face is detected.

B. Feature Extraction

It is a process where feature points are located on the face once the face is detected. The facial points mainly are eyes, eyebrows, lips, cheeks, etc. The movement of all these feature points is recorded for further detection. The feature extraction is normalized by scaling, translation, and rotations. This ensures a high level of extraction. Translation fixes the points in the image and checks whether they are aligned. Scaling is done so that the distance from the camera is fixed, size is fixed. Not every video frame or image in the dataset has these same as the input. Rotation is done on the image only when it is necessary. In other situations, such as the head is tilted on one side or difficult to extract features, rotations are done and facial expressions are recorded.

C. Age and Gender Detection

The process of automatically detecting the age of a person from an image using a webcam is implemented in three stages:

- Detect face from an input stream
- Extracting face region
- Apply the age detection algorithm to detect the age of the person

For stage one, to detect the face of a person, Haar cascades are used. As the system is for a single user, we consider only one face at a time. The system is capable of producing bounding boxes for faces in which the face of the user will contain. For best accuracy, deep learning-based face detection is most preferred.

We are using the following age ranges:

(0-2, 4-6, 8-12, 15-20, 25-32, 38-43, 48-53, 60-100)

These are non-contiguous since the detection of age is mainly based on the appearance of the person. Hence these categories give the best results. To detect age from a video stream, first, we filter out a false positive face in the frame. Also, age classification results will be accurate if the face is near to the camera. A convenience function is used which accepts a frame, localizes a face, and predicts age. We send blobs through CNN, which results in detection and filter out confidence. For faces meeting minimum confidence, region coordinates are extracted, and age is predicted using above age ranges.

For Gender detection, we used the following algorithm and the Adience dataset.

Algorithm Gender Detection

Input: labeled supervised data for gender detection G and unlabeled data for recognition R.

Output: Predict the data, test labeled data T using Machine Learning Model M.

Steps: -

1. Label the data and divide the mixed labeled data for the detection of gender G.

2. Normalize the data for R.

3. Divide G and R separately.

4. Create model M_G and M_R for Data G and Data R resp

5. For each M_i do

6. Pre-process the data

7. For ("male", "female") do

8. Predict the real detected face and categorize either male or female

9. End for

10. End For

11. Return

D. Emotion Detection

The most important function of the system is the detection of emotions. All the above-generated results such as Face acquisition and feature extraction are used to detect the emotion of the user.

Various emotions can be detected:

- Happy
- Angry
- Sad
- Fearful
- Disgusted
- Neutral
- Surprised

Based on the above results, the trained model tries to predict the emotion of the user.

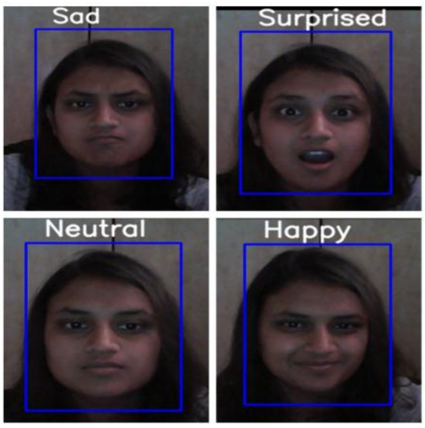

Figure 2: Emotion Detection

E. Recommendation

The content-based approach is all about users or items to be recommended. Considering our system, emotion-based recommendation, additional information is age and gender. Content-based methods are used to build a model, observe interactions of the user with an item. In a model-based collaborative method, the model is trained so that it can reconstruct interactions of the user with the item. Then the user gets suggestions based on this model. In content-based methods, the model is provided with content to represent the user. This model may have high bias and low variance. If our classification is based on user features, that means modeling, optimization, and computing is done by the item. In this case, the probability of each user to like a particular item is considered. The model is trained with the data a lot of users have interacted with. That means even if we apply more robust methods, it is not more personalized. And if item features then, modeling and computation are done by the user this is also known as user-centered. Here the probability of a user to like each item is considered. That means this much personalized than the previous method.

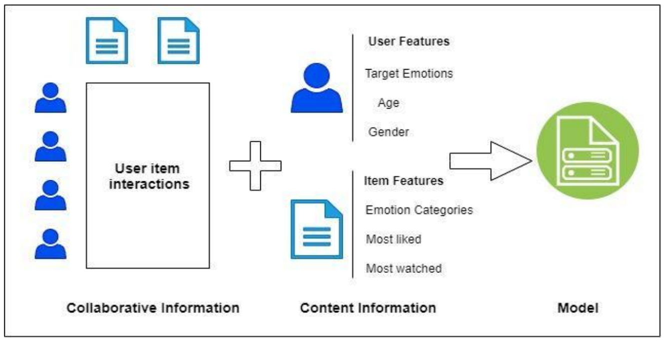

Figure 3: Content-based recommendation

F. Database Connectivity

To provide security to the application and user data, we provide database connectivity to the application. Python Sqlite3 used to store user login data. Users can log in to the system using a username/ email address and password. Only this information is stored in the database, not others.

G. Django

Django is an open-source framework for web applications. Using Python and Django, application deployment is faster. It is based on the MVT (Model-View-Template) architecture [6].

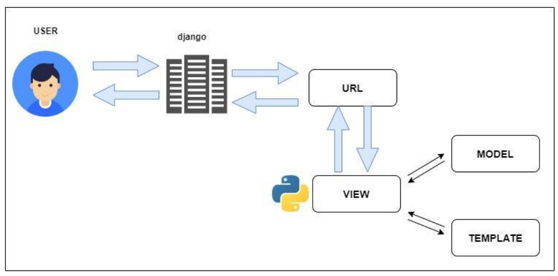

Figure 4: Django MVT Architecture

5. EQUATIONS

I. Face detection using LBPH

Divide the window into cells of 3x3 pixels. For each pixel in a cell, compare the pixel to each of its neighbors from top left, middle-left to top-right. Follow the pixels clockwise. If the center pixel's value is greater than the neighbor's value, then write "1". otherwise, write "0" as shown in the figure. In this way, an 8-digit binary number is generated which needs to be converted to decimal. [5]

Example

6	5	2
7	6	1
9	8	7

Threshold

1	0	0
1	X	0
1	1	1

Weights

1	2	4
128	X	8
64	32	16

Pattern = 11110001

LBP=1+16+32+64+128=241

C = (6+7+8+9+7)/5-(5+2+1)/3=4.7

$$LBP = \sum_{n=0}^{7} s\left(\mathrm{i}_n - \mathrm{i}_c\right) 2^n$$

Where

ic = value of the center pixel (xc, yc),

in = value of the eight surrounding pixels,

function s(x) is defined as:

$$s(x) = \begin{cases} 1 & if\ x \geq 0 \\ 0 & if\ x < 0 \end{cases}$$

II. Content-based method for recommendation

In our system for each user, we are training a simple linear regression.

$$X_i = \mathrm{a}_{X_i} \frac{1}{2} \sum_{(i,j) \in E} \left[(X_i)(Y_j) - M_{i,j}\right]^2 + \frac{\lambda}{2}\left(\sum_k (X_{ik})^2\right)$$

Where M denotes the user-item interaction matrix. X is a matrix of row vectors that represents user coefficients to be learned. And Y is a matrix of row vectors representing features of items. 'i' is the user which is fixed, and this computation is over (user, item) pair, here 'a' stands for the argument of the minimum.

Now consider interaction matrix M(n x m), factorize matrix such that,

$$M \approx X.Y^T$$

$$user_i \equiv X_i \quad \forall_i \in \{1, \ldots, n\}$$

$$item_j \equiv Y_j \quad \forall_j \in \{1, \ldots, m\}$$

Now we try to minimize the error,

$$(X,Y) = \mathrm{a}_{X,Y} \sum_{(i,j) \in E} \left[(X_i)(Y_j)^T - M_{i,j}\right]^2$$

Adding regulation factor gives,

$$(X,Y) = \mathrm{a}_{X,Y} \frac{1}{2} \sum_{(i,j) \in E} \left[(X_i)(Y_j)^T - M_{i,j}\right]^2 + \frac{\lambda}{2}\left(\sum_{i,k} (X_{ik})^2 + \sum_{j,k} (Y_{jk})^2\right)$$

We can see that the basic factorization can be expended to a complex model.

On top of this, if our system is based on numeric outputs then we can increase the quality of the output in much easiest manner such as error measurement or mean square method.

i. Mean Square Method:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} \left(y - \underline{y}\right)^2$$

ii. Root Mean Square Method:

$$RMSE = \sqrt{MSE}$$

We can binarize the model so that it can be used for the test dataset. This model then used for new and personalized video recommendations to the system.

6. DATASET INFORMATION

a) Adience Dataset

For the detection of age and gender, the dataset used is the Adience dataset, which is a collection of unfiltered faces for age and gender prediction [8]. The data included in this collection is intended to be as true as possible to the challenges of real-world imaging conditions. There are a total of 26,580 number of photos and 2,284 number of subjects. Total number of labels are 8 i.e. (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60-100) also gender labels are present.

b) FER 2013 Dataset

The FER2013 database was introduced during the ICML 2013 Challenges in Representation Learning. FER2013 is a large-scale and unconstrained database collected automatically by the Google image search API. It is an open-source dataset with 35.887 grayscale, 48x48 sized face images, and different labels such as:

- 4593 images- Angry
- 547 images- Disgust

- 5121 images- Fear
- 8989 images- Happy
- 6077 images- Sad
- 4002 images- Surprise
- 6198 images- Neutral

It contains 28,709 training images, 3,589 validation images and 3,589 test images [9].

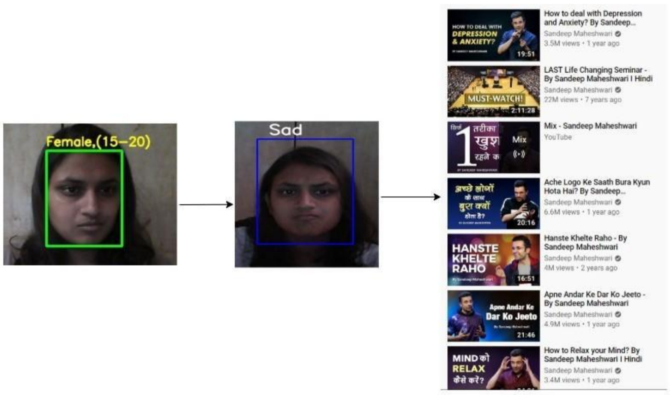

Figure 5: Recommendation using Age, Gender, Emotion Detection

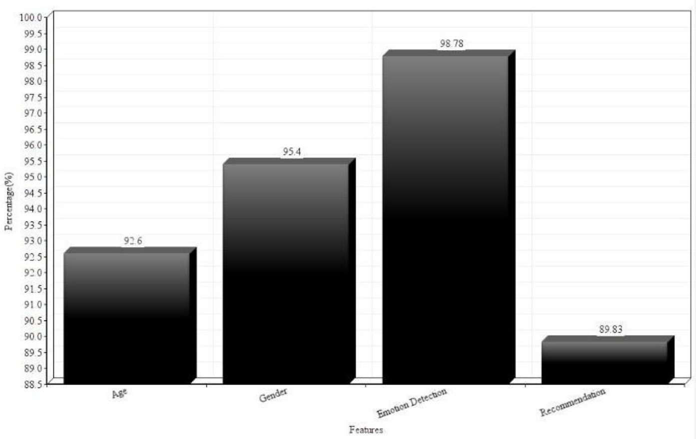

Figure 6: Accuracy of the system

CONCLUSION

We have developed this Personalized Video Recommendation (PVR) System [1] as an assistant to a user for entertainment purposes. For the system, we used Haar cascade and Local Binary Patterns Histogram (LBPH) algorithms for face detection, feature extraction, and emotion detection. FER 2013 dataset is used for emotion detection and age, gender detection Adience dataset is used. The system is a web application hence Django framework is used for development. A content-based method is used for the recommendation of videos. Videos will be recommended depending on the emotion detected by the system. Also, the age and gender of the user are considered to make it more personalized. Videos that are more relevant to target emotion, most liked or interacted by other users are recommended through the internet. Video categories are music videos, motivational speeches, quotes, movies, cartoons, humor, action, lifestyle.

REFERENCES

[1] M. Borgaonkar, M. Katta, P. Kudale, V. Deshpande, Prof. V. Babanne, "PVR System: Personalized Video Recommendation", IJRECE, Vol. 7, Issue 4, pp. 302-305, Oct-Dec2019.

[2] M.B. Mariappan, M. Suk, B. Prabhakaran, "FaceFetch: A User Driven Multimedia Content Recommendation System Based on Facial Expression Recognition", 2012 IEEE International Symposium on Multimedia, vol.1, pp. 84-87.

[3] S. Wang, Q. Ji, "Video affective content analysis: a survey of state-of-the-art methods", IEEE TRANSACTIONS ON AFFECTIVE COMPUTING, Vol. 6, pp. 410-430, oct-dec 2015.

[4] A. Tripathi, T.S. Ashwin, R.M. R. Guddeti, "WmoWare: A Context-Aware Framework for Personalized Video Recommendation Using Affective Video Sequences", IEEE Access, vol. 7, pp. 51185-51200, 2019.

[5] A.P. Singh, S.S. Malvi, P. Nimbal, G.K. Shyam, "Face Recognition System Based on LBPH Algorithm", IJEAT, Vol. 8, Issue 5S, pp.26-30, May 2019.

[6] S. Roy, Prof. M. Sharma, Dr. S.K. Singh, "Movie Recommendation System Using Semi-Supervised Learning", IEEE 2019 Global Conference for Advancement in Technology, Bangalore India, Oct 18-20, 2019.

[7] S. Singh, M.G. Mohiuddin, S. Pandit, "Face Detection and Recognition by Haar Cascade Classifier, Eigen Face and LBP Histogram".

[8] https://talhassner.github.io/home/projects/Adience/Adience-data.html#agegender

[9] https://medium.com/@birdortyedi_23820/deep-learning-lab-episode-3-fer2013c38f2e052280#:~:text=fer2013%20is%20an%20open%2Dsource,7%20emotions%2C%20all%20labeled%2D.