HiReME: Video Interview Bot

Rani Bane[1] Yashoda Eknarayan[2] Prasad Ambekar[3]

[1-3] Information Technology Engineering, Padmabhushan Vasantdada Patil Pratishthan's College of Engineering, Maharashtra, India

---***---

Abstract- *With the advancement of technology, human intelligence is overpowered by the machine learning and artificial Intelligence. Each and every field using this immense growth of technology for the betterment of their businesses, as this helps in cutting cost, time and manpower, providing the more accurate and precise results, within a blink of an eye. So, why not to use it in the most important aspect of a business? That is, employee hiring or job recruitment process. As we know, there are lots of job portals, virtual interview environment providing software's are available to ease up the hiring process. But even with this, there are lots of challenges that the company and employees both faces during the entire procedure, such as monitoring numerous interviews, biased results, keeping records etc. So, there is need of completely automated system which will help companies to schedule virtual interviews and help employees with unbiased results, with minimal manual work. Therefore, in this paper we introduced a Deep Learning based system for analyzing interviews for personality prediction of candidates, saving both time and money and making a company professional's life much easier. For this, recorded videos will be analyzed based on Facial Expression, Prosodic Information, Sentiment Analysis of Response, Speech (e.g., word counts, topic modeling).*

Keywords: - **Convolution Neural Network, Facial Expression Detection, Prosodic Features, Sentiment Analysis, Voice Analysis.**

1. INTRODUCTION

Recruitment is the process of hiring employees for a specific job post, as per the requirement of a company. The Recruitment process has multiple candidate filtering rounds such as tests and GD's which can easily be conducted without having continuous monitoring. But after this, there will be an interview round which is most important round in the recruitment cycle and hence require continuous monitoring, thereby increasing cost. Personality prediction is the main goal of this interview rounds. With the increment in population, to perform this recruitment process, companies need to have large number of manpower, lots of time and much increased cost. As we know India is the second highest populated country in the world and major part of this population is of age group 20 to 30, which seeks out for a job. So, to serve these many numbers of available candidates with the handful of vacancies, there is a need to have an automated system that will cut the hiring expenses, and enhance the recruitment procedure providing accurate results with the minimal amount of time. Also with this cost-effective system, the companies can generate the unbiased and precise results which will be beneficiary for both company and candidates. Therefore this paper introduce the system which helps with the interview round by personality prediction using Facial Expression detection (Neutral, Happy, Sad, Angry, Surprise, Fear), Prosodic Information (e.g., syllables, pauses, etc.) and Sentiment Analysis (Positive, Negative, Neutral).

2. PROPOSED METHOD

Once the candidate applies for a job from any source his record is stored in the company database and the data is stored in the website database as well. From the website, the company sends a link to the candidate, which consists of a unique code (Access Key) for the candidate to login into the website. The candidate is provided with two links on the website, one for a practice test and another to start the interview. The practice test link gives the candidate an overall idea of how the interview will be conducted on the online platform. While in the Start Interview link the actual interview is conducted. In the interview, system asks questions to the candidate and then a video of the candidate is recorded while he is giving his answer to the question. The System then analyses the video and a displays a detailed report of the candidate's sentiment analysis, Emotion analysis, Voice Analysis. This report is visible to the company on their page. The company can then compare the report of all the candidates and select the best candidate for their organization.

3. USER INTERFACE

3.1 Web Application

Website is created for a company to post the test with which questions to be asked to a candidate for Video Interview are added to the database. Unique ID (Access key) and password is generated for the candidates using which they can login to the system and appear for an interview. For candidates to record videos, video recording facility is available on website. These videos are then analyzed and its result is viewed on website. Website is created using Flask framework.

3.2 Animated Speaking Character

To enhance the interview process and give candidate a real feel of interview, an animated speaking character is used as Interviewer.

It consists of following parts-

- Text To Speech Conversion: Libraries PYTTSX3 is used to convert text into speech. Windows API provides voices through Sapi5, and by using this engine Voice 0(Male) and voice 1(Female) are the two voices that we can access. Amongst which Voice 0(Male) is used, Then by using speak function speech is obtained.

- Animated Character: It is created using HTML, CSS and JavaScript.

3.3 Video Recording

For this, VideoCapture from cv2 is used. Once the candidate's interview is started, bot will ask question to the candidate one by one. Candidate needs to record the answer by using recording button available on the web application. After asking each question the recording will automatically start after "5 sec" if 'Start Recording' button is not clicked. These videos will be further processed by the system.

4. ANALYSIS

4.1 Facial Expression Analysis

CNN (convolution neural network) is the artificial neural network which helps in image classification. In CNN, neuron of single layer is connected to the small region or part of the layer before it. Hence using CNN, weights that needs to be handle or process will be less and it will require less numbers of neurons.

CNN classifies the image by considering pieces, called as Features. It compares this feature with an input image, by roughly matching them at roughly matching position to get more précised results.

For facial expression detection CNN is used. For this, recorded videos are converted into number of KeyFrames to get the multiple images. Then these images are fed to CNN to detect the facial expression. Modules used are Subprocess module, OS module, CV2 module for video capture, Keras module to analyze facial expression, Matplotlib for plotting of graphs, Haar-Cascade-Files for defining facial key points, etc.

Fer2013 (Facial Expression Recognition 2013), a dataset from Kaggle, is used for facial expression recognition. Using this dataset, the emotions that the system detected are 'Angry', 'Fear', 'Happy', 'Sad', 'Surprise' and 'Neutral'. The dataset is divided into training dataset and testing dataset. Whenever the system gets any image, it will detect the facial expression of that image by comparing it with the images of the training dataset. System reads an image using the values at each pixel, and using CNN these pixels are processed further for generating results. Images are fed to the system. Then they are converted into greyscale images.. From this greyscale image, the face is then detected. Then we resized the images into the scale of 49*49*1. After resizing the detected face, scaling of an image is done. This image then undergoes multiple layers to detect the emotion of that image. As shown in Fig-1.

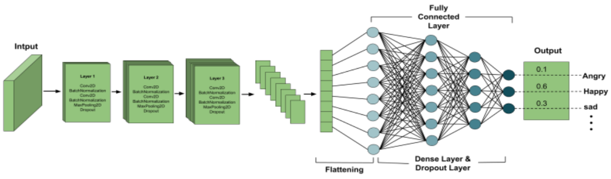

Fig.-1: Facial Expression Detection Model

- Conv2D Layer: In this layer, the input image is classified using feature map. CNN classifies the image considering small pieces of matrix (M*M) called as features, these features are then mapped onto the input image having pixel matrix of N*N. After mapping, pixel value at the feature is multiplied to the corresponding pixel value of an input image. Once the multiplication of all the pixel values from the feature and the input image is done, all the values from the generated matrix are added together and then after dividing this sum by the number of pixels, filter value is obtained. This feature is moved throughout an image. After placing all the values, of all the filters we get the simplified matrix (N*N).

- Batch Normalization Layer: Before sending the output matrix which is generated from convolution layer to the Pooling Layer, Batch Normalization is done. This helped converting the large pixel values to pixel values of decimal points, which made the further process easy.

- MaxPooling2D layer:

  In MaxPooling layer, we have considered the window size of W; this window is moved throughout an input image. This layer considers the maximum pixel value from the window and neglects all other values. Hence this layer shrinks the input image, so that it can be fed to the Fully Connected Layer. So, the N*N matrix obtained from previous layer of each filters, is reduced to PL1*PL1 matrix. This generated PL1*PL1 matrix is then again subjected to all above layers twice, to shrink the image further in PL2*PL2 matrix, and then this matrix is fed to Fully Connected layer.

- Flatten Layer: In this layer, all the values of all the features are stacked up into a single list or vector. This list values are then considered for detecting facial expression.

- Dense Layer (Fully Connected): In this layer, each neuron of a layer is connected to all other neurons of the layer before it; hence it is called Dense Layer or Fully Connected Layer. Actual classification is done at this layer. The vector that was generated from Flatten layer is used here for classification. In training dataset different values of such vector or list will be high for different expression .Hence, considering index of high values from training dataset, and then comparing those values with the vector values at that index of an input image, its expression is classified.

- Dropout Layer: This layer is used to prevent overfitting. As we know the images can have large pixel values, and when we feed these values to the Fully Connected network, it will require large number of weights in the first hidden layer itself, and this leads to the overfitting. Hence, to avoid this Dropout Layer is used.

For generating the actual results, we have considered epochs= 60. Epochs are the number of images that system considers to predict the result. To get more precised result the epoch value can be increased and the final result is generated.

Result:

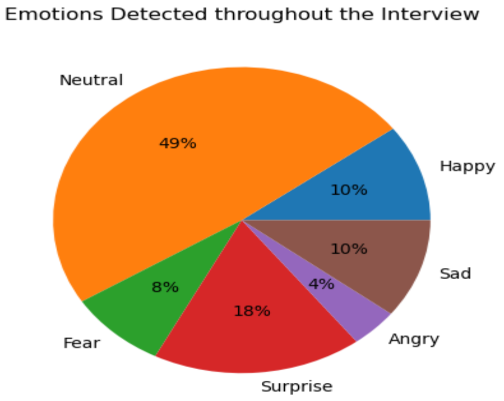

Fig.-2: Result of Facial Expression Detection

4.2 Voice Analysis

For voice analysis, Librosa library is being used. In this, the analysis is done considering multiple parameters. Using voice analysis we recognized, Gender and mood of speech, Pronunciation posteriori probability score percentage. We detected and counted number of syllables, number of fillers and pauses. We measured the rate of speech (speed),speaking time (excl. fillers and pause),total speaking duration (inc. fillers and pauses), ratio between speaking duration and total speaking duration, fundamental frequency distribution mean, fundamental frequency distribution SD, fundamental frequency distribution median, fundamental frequency distribution minimum, fundamental frequency distribution maximum, 25th quantile fundamental frequency distribution and 75th quantile fundamental frequency distribution. After considering all these parameters voice is analyzed.

Result:
Total Syllables= 70
Total Pauses= 3
Speech Rate= 3 # syllables/sec original duration
Articulation Rate= 4 # syllables/sec speaking duration
Speaking Duration= 17.3 # sec only speaking duration without pauses
Original Duration= 24.3 # sec total speaking duration with pauses
Balance= 0.7 # ratio (speaking duration)/ (original duration)
Pronunciation_posteriori_probability_score_percentage = 80.15
Mood: speaking passionately

4.3 Sentiments Analysis

Sentiment analysis is a technique used in Natural Language Processing (NLP) and is used to determine whether given data is positive, neutral or negative. NLP is a field of artificial intelligence that is used for the interaction between machines and human language. For Sentiment Analysis, we have used the TextBlob library. TextBlob is an open-source library for performing Natural Language Processing (NLP) tasks which include part-of-speech tagging, translation, noun phrase extraction, sentiment analysis, classification, etc. It has a sentiment lexicon in the form of an XML file which it leverages to give both polarity and subjectivity scores.

Modules used are TextBlob for sentiment analysis, MatPlotLib for plotting of graph, pandas for reading CSV file and regex to eliminate unwanted text. Once the candidate answer which is stored in a CSV file is added to a DataFrame and then the unwanted text is removed from the sentences. Then the DataFrame is passed through a TextBlob module in which the polarity score and subjectivity score of the sentence is generated.

- Polarity: Polarity score is a float value ranging [-1.0, 1.0] where -1.0 being very negative and 1.0 being very positive. This score is used to check the sentiment of a sentence.
- Subjectivity: Subjectivity score is a float value ranging [0.0, 1.0] where 0.0 being very objective and 1.0 being very subjective.

This Polarity score is used for analysing whether the sentence is negative, neutral or positive. Once the analysis is completed the result is used to generate a graph of the total number of negative, neutral and positive sentences in an answer and an overall calculation is done to check whether the answer is negative, neutral or positive.

Result:

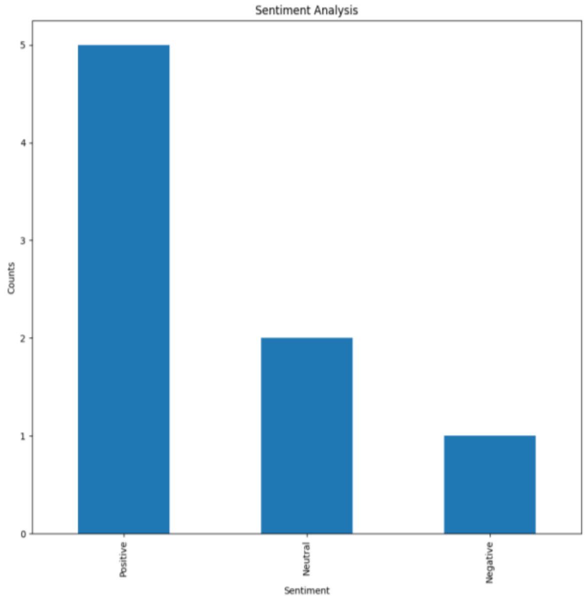

Fig.-3: Result of Sentiment Analysis

5. CONCLUSION

In this paper, we have proposed a method for interview process, which can be used efficiently for personality prediction of candidates. By using well trained CNN and proper dataset, result can be generated with more precision.

Proposed system conduct the Interviews virtually using video recordings, it will then process videos for Facial expression Recognition, Voice analysis and sentiment analysis. The system itself performed personality prediction of candidates by analyzing Facial Expression with the 64% of precision, sentiments with 56% of precision and by analyzing voice. This automated video interview system will help analyse parameters such as candidate's facial expression, voice and sentiments which are enough to describe one's personality. This will then help the recruiters to evaluate the candidate at their convenience, by comparing how each candidate answered the same set of questions, by just viewing at the results provided by the system.

ACKNOWLEDGEMENT

We would like to thank our project guide, Prof. Darshana Tambe and our project coordinator Prof. Vijaya Sagvekar- Bijoor (Professors at Faculty of Information Technology, Vasantdada Patil Pratishthan's College of Engineering and Visual Arts) for guiding us in every stage of the project development starting from Requirement Analysis to the actual Implementation of the system. We would also like to thank Dr. Seema Ladhe (Head of Information Technology Department, Vasantdada Patil Pratishthan's College of Engineering and Visual Arts) for sharing their perspectives and domain experience which helped in many aspects of this study. Moreover, we would like to express our gratitude and appreciation to our friends who have helped and supported us in this study.

REFERENCES

[1] A.T. Rupasinghe1, N.L. Gunawardena2, S. Shujan3, D.A.S. Atukorale4." Scaling Personality Traits of Interviewees in an Online Job Interview by Vocal Spectrum and Facial Cue Analysis." 2016 International Conference on Advances in ICT for Emerging Regions (ICTer): 288 – 295

[2] Himanish Shekhar Das1 · Pinki Roy1" Optimal prosodic feature extraction and classification in parametric excitation source information for Indian language identification using neural network based Q-learning algorithm" International Journal of Speech Technology,2018.

[3] Sarah S. Alduayj, Phillip Smith." Sentiment Classification and Prediction of Job Interview Performance." 978-1-7281-0108-8/19/$31.00 ©2019 IEEE

[4] Hung-Yue Suen1, Kuo-En Hung1, Chien- Liang Lin2." Intelligent video interview agent used to predict communication skill and perceived personality traits" Suen et al. Hum. Cent. Comput. Inf. Sci. (2020) 10:3

[5] Muhammad Abdullah, Mobeen Ahmad, Dongil Han." Facial Expression Recognition in Videos" May 16, 2020 at 00:40:26 UTC from IEEE Xplore.

[6] Lei Zhao, ZengcaiWang, Guoxin Zhang." Facial Expression Recognition from Video Sequences Based on Spatial-Temporal Motion Local Binary Pattern and Gabor Multiorientation Fusion Histogram" Mathematical Problems in Engineering Volume 2017, Article ID 7206041, 12 pages.

[7] Mohsen Fallahnezhad, Mansour Vali, Mehdi Khalili." Automatic Personality Recognition from Reading Text Speech" 2017 25th Iranian Conference on Electrical Engineering (lCEE). 978-1-5090-5963-8/17/$3l.00 ©2017 IEEE.

[8] Iftekhar Naim1, M. Iftekhar Tanveer2, Daniel Gildea1, and Mohammed (Ehsan) Hoque." Automated Prediction and Analysis of Job Interview Performance: The Role of What You Say and How You Say It" ROC HCI, Department of Computer Science, University of Rochester, ROC HCI, Department of Electrical and Computer Engineering, University of Rochester,2015