International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 6 June 2023, pp: 41-46 www.ijaem.net ISSN: 2395-5252

Interview Based Human Behavior Analysis withMachine Learning

Anju Raj E., Farzana T.
M.tech Cse Department, MES Collage of Engineering ,Kuttipuram
Assistant Professor M.tech Cse Department, MES Collage of Engineering ,Kuttipuram
Kerala, India

Date of Submission: 01-06-2023 Date of Acceptance: 10-06-2023

ABSTRACT—A job interview is an interview consisting of a conversation between a job applicant and a representative of an employer which is conducted to assess whether the applicant should be hired. Interviews are one of the most popularly used devices for employee selection. A job interview typically precedesthe hiring decision. It is used to evaluate applicant's knowledge, skills, abilities, and behavior in order to se- lect the most suited person for the job. Recruiters make their opinion, on the basisof both verbal and nonverbal communication of an interviewee. Our behavior and com- munication in daily life are cross-modal in nature. Facial expression, hand gestures and body postures are closely linked to speech and hence enrich the vocal content. Nonverbal communication plays an important role in what I saying and what am ac- tually mean to say. It carries relevant information that can reveal social construct of a person as diverse as his personality, state of mind, or job interview outcome; they convey information in parallel to our speech. In this paper, I am presented an auto- mated, predictive expert system framework for the computational analysis of HR Job interviews. The system includes analysis of facial expression, language and prosodic details of the interviewees and thereby quantifies their verbaland nonverbal behavior. The system predicts the rating on the overall performance of the interviewee and on each behavior traits and hence predict their personality and hireability. ii

Index Terms—HR Interviews, Behavior Analysis, Cross-modalanalysis.

I. INTRODUCTION

Nonverbal communication is the transmission of messages or signals through a nonverbal platform such as eye contact, facial expressions, gestures, posture, use of objects and body language. In face-to-face communication, our nonverbal be- havior conveys information about our personality and traitsin addition to our speech. Nonverbal communication is an unconscious process and hard to manipulate, therefore playsan important part in the outcome of the employment interview. The nonverbal communication includes visual cues, audio cues, and language or lexical cues. Visual cues or facial cues can have a powerful effect on interview performance. Our physical appearance and how we dress are only part of visual cues which are important for an interview. Along with this, visual cues also include our facial expressions (smiling, neutral, surprise) during the conversation. Hand gestures, eye contact, body orientation also contributes to the visual cues during communication.

Another nonverbal cue is the audio cues or paralinguistic cues which refers to the vocal (prosodic) characteristics and features. Not only what the interviewee says but how he/she says can make a lot of difference in the outcome of the interview. Voice modulation, variation in pitch, the rate of speech, intensity etc. are some of the important vocal cuesthat may break or make your chances to be hired.Monotone voice, narrow pitch range are some characteristics that are not preferred in an interview

This paper addresses the challenge of automated compu- tational analysis of cross-modal communication, including visual, language and prosody of job interview videos, to create a predictive expert system that helps predict hireability and rate different trait of the interviewees. The results are then displayed on the dashboard for the interviewer.

This predictive system can be used for screening, hiring can- didates for a job interview and act as an expert support system for hiring managers to take appropriate decisions. This system can also be used for a fully automated screening of candidates in online interviews as

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 6 June 2023, pp: 41-46 www.ijaem.net ISSN: 2395-5252

shown in the figure 1.1 . This can help with initial shortlisting of suitable candidates for further domain specific evaluation. From figure 1 the steps includes

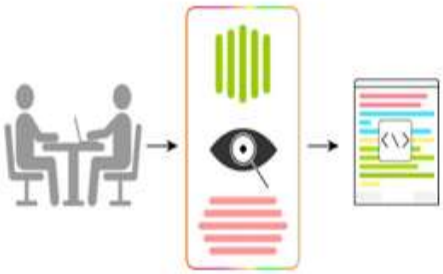

Fig. 1. Job Interview analysis model. From left to right

Interview Recording, Feature Extraction,Prediction Model for classification and rating of different traits of individuals. From

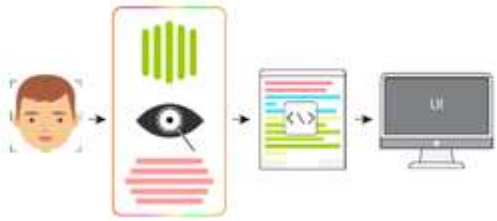

Fig. 2. Automated online interview. From left to right

figure 2 the steps are Interview taken by candidate online, Audio, Visual and Language Feature Extraction,To predicthireability, Results displayed.

II. LITERATURE SURVEY

A. State of the art

1) Can nonverbal cues be used to make meaningful per- sonality attributions in employment interviews?: : This pa-per studies the relationships between nonverbal cues and interview per- formance ratings and also examines the role of personality attributions. Conscien- tiousness attributions explain the relationship between visual cues and interview rat- ings, extraversion attributions mediate the relationship between vocal cues and inter- view ratings. Neuroticism attributions had a suppressing effect for both visual and vocal cues. Three types of nonverbal cues are studied - dynamic, static, and paralinguistic cues. Dynamic cues are easily changed such as eye contact, body orientation, smiling, gesturing, and head movement in [2]. Static cues are demographic variables and physical attractiveness. Paralinguistic cues refer to vocal characteris- tics such as speech rate, volume, tone, and pausingThe Brunswik lens model shows the potential impact of personality attributions when both channels of information are used. Coefficients for the model are computed by regressing the combined personality attributions on both dependent variables, job performance on one side of the lens model and expert interview rat- ings on the other side of the lens model. [**?**, 2]his suggests that interview raters and job performance raters almost equally use personality attributions when making their rat- ings.Understanding the impact of effects of personality in interviews has great orga- nizational relevance. Extraversion has been shown to be a much-desired managerial trait. One conclusion which can be made is that no matter how much an interview is

2) Social signaling: Predicting the outcome of job inter- views from vocal tone and prosody : What does it take to succeed in a job interview? Recruiters, career coaches and social psychologists alike have highlighted the role of speaking style, confidence and demeanor in the face-to-face interview process. Past work suggests that non-linguistic verbal cues play an important role in the outcomes of interviews, and that social signaling measures are quite effective in predicting behavioral outcomes in different social interactions (e.g. nego- tiations, dating). In this paper, we quantify the non-linguistic speaking style of engineering school students in practice job in- terviews, using features extracted from their vocal tone and prosody. Face-to-face communication in humans is a highly complex process involving verbal content, non-verbal signal-ing, gestures and body posture. While verbal communication is explicit and is easily interpreted, non-verbal communication is subtle and implicit. Nonetheless, it has been well established that both channels of communication affect conversational dynamics and influence the relationships between individuals. In non verbal communication subtle aspects of speech like tone, intensity, pitch etc. are categorized as non-verbal paralinguistic communication, and observed in cases where a person is described as 'driving the conversation' or 'setting the tone' of the conversation [3].Face-to-face interviews play pivotal role in the hiring process for companies and help the recruiter evaluate the candidate's skills, motivation and personality traits. There are different types of job interviews, including broad or screening interviews (e.g. 'why do you wantto work for this company?'), behavioural interviews (e.g. 'give me an example when you managed multiple projects'), tech-nical

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 6 June 2023, pp: 41-46 www.ijaem.net ISSN: 2395-5252

interviews (related to particular job skills) amongst a few Screening interviews are usually conducted by Human Resource (HR) managers, and are meant to gauge the overall motivation, attitude and aptitude of the candidate.

*3)*Multimodal analysis of body communication cues in employ- ment interviews : Gestures are an essential compo- nent of body communication as they are used to enrich the vocal content and aid listener comprehension by augmenting the atten- tion, activating images or representations in the listener's mind, and increasing the recall of what is beingsaid. Job hireability impressions and self-rated personality can be predicted using body communication cues[4]. Speaking status can be used to improve the prediction performance of personality and hireability In order to analyse the predictive validity of body posture with respect to self-rated personality traits and hireability impressions, a regression problem was defined which aims at predicting the exact hireability and personality scores, where each social variable is consideredas an independent regression task.

Gestures are an essential component of body communica- tion as they are used to enrich the vocal content and aid listener comprehension by augmenting the attention, activating images or representations in the listener's mind, and increasing the recall of what is being said . Moreover, restraining people from gesturing strongly affects the speakers' fluency. Body posture is another important component of body commu- nication; various emotions such as fear, sadness, or happiness have been shown to be correctly inferred from a person's pose. In conversations, body posture can be used as markers during a conversation: for instance, changes of body posture can precede a long utterance and may be kept for the duration of the speaking turn [18]. In this sense, both gestures and postures are inherently multimodal, in that they do not only occur in the visual modality, but are conditioned on the speaking status (i.e., audio modality) of the person. For this reason, we believe that it is necessary to consider the speaking status when analyzing posture and gestures. Except communication and conscience, all hire-ability measures are significantly correlated with each other

*4)*Real Time Sign Language Recognition using PCA : The Sign Language is a method of communication for deaf-dumb people. This paper[5] presents the Sign Language Recognition system capable of recognizing 26 gestures from the Indian Sign Language by using MA TLAB. The proposed system having four modules such as: pre-processing and hand seg- mentation, feature extraction, sign recognition and sign to text and voice conversion. Segmentation is done by using image processing. Different features are extracted such as Eigen values and Eigen vectors which are used in recognition .The Principle Component Analysis (PCA) algorithm was used for gesture recognition and recognized gesture is converted into text and voice format. The proposed system helps to minimize communication barrier between deaf-dumb people and normalpeople.

III. PROPOSED METHOD

This section details down the process of analysis on the recorded interviews and accordingly extracted features. Subse- quently, we present the methods required to extract the audio, visual and language cues from the recorded data.

1) Audio Cues : The major concern while extracting audio cue features is on the prosodic features. The preliminary step is to detect the individual who is speaking during the interview,the interviewer or the interviewee. Whilst during this process, we also need to segment the speech and non-speech chunks. Secondly, we need to make sure, that, the existence of any noise should be eliminated from the audio, if present. For audio cues, we extract the prosodic features of Pitch, Band energy, and the Speech Rate respectively.

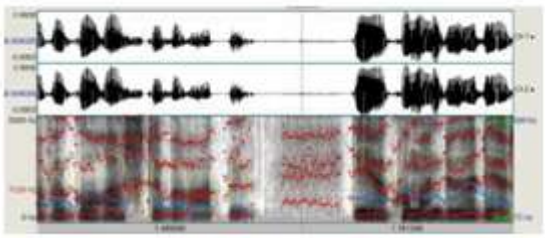

Fig. 3. Audio cue wave format

The tool utilized for this purpose is PRAAT, an open-source audio analysis tool for extracting but not limited to the above mentioned prosodic features. I first split the audio files into different chunks, wherein these chunks consist the answers from the interviewee and construct the collection of prosodic features of these splitted chunks.pyaudio python libraries are used.

*2)*Visual cues : Visual cues are the facial features of the interviewee during the interview. For every video frame, facial features of the

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 6 June 2023, pp: 41-46 www.ijaem.net ISSN: 2395-5252

interviewee are extracted. The first stepin this is to detect the face itself in the frames. The Haar Cascade classifier for face detection can be used to do perform this task of face detection. Different facial cues extracted are facial expression, gaze, and head nods. There are different methods to detect facial expressions. Machine Learning is one of the approaches to detect facial expressions. Since it's a classification problem we can classify facial images using ML into different expressions like neutral, joy, sadness, contempt, anger, disgust, fear and surprise.

Second visual cue extracted is head nods. This is used to detect Yes and No during when signaled using head nods. The third cue in facial feature extraction is the eye gaze. I know theimportance of eye contact is an important nonverbal behavior, Hidden Markov model is used for head shaking. it is used to monitor feedback, reflect cognitive activity. Then extract the amount of time the interviewee looks at the interviewer while answering a question or during the course of the interview. Apart from these three cues we also extracted some more cuesusing facial action units. I used the dlib library to detect facial landmarks. Facial landmarks help us localize different regions of the face like eyes, eyebrows, nose, mouth, and jawline.

Fig. 4. The blue dot, representing the center of the head, for keeping thetrack of head nods

Fig. 5. Eye movement tracking indicators

from figure 6 represents these 68 points that maps the facial structure on the face. These points are then used to calculate more facial features like eyebrow raise distance from the eye, eye-opening distance, lip corner distance etc.Eye haarcascade algorithm is used..

Fig. 6. Dots plotted on a facial Figures to help understand the production ofvarious expression according to the movements of these points.

*3)*Language/Lexical Cues : Here use "VADER" standing for Valence Aware Dictionary and sentiment Reasoner for verbal cues/ lexical cues. The VADER categories include negative emotion terms (e.g., sad, angry), positive emotion terms (e.g., happy, kind) and various content word categories (e.g., anxiety, insight). It is one of the most used lexicons for analyzing lexical cues, and especially when it comes to sentimental words.

*4)*Predictive Model : After extracting features from the facial cues, audio cues, and verbal cues, we create a single feature vector joining the features from the facial, audio and verbal cues for each interview. This finally generates the dataset. This dataset has ratings for the judges as well. This dataset is used against model prediction to find the similarity between the prediction and actual judges score. I am used a regression model to predict the hireability and overall score of the interviewed candidate.

A. System Architecture

The system has two modules, Input, and Analysis respec- tively. The input module is used to automatically collect video and audio from the input devices and store the data into a suitable data storage. The analysis module takes thedata (recorded interviews) singularly to perform the analysis on them and thereby extract features from audio, video and language cues. The proposed system diagram for the interview analysis is represented by Figure.

After the features have been extracted from each source, they are then used in the prediction model for predicting the ratings for each specific trait for the interviewee, ultimately producing the prediction of hireability of the interviewee. The results from the analysis and predictions are then displayed onthe dashboard for the hiring manager to see

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 6 June 2023, pp: 41-46 www.ijaem.net ISSN: 2395-5252

B. DataSet Description

1) Interview Collection- Data Gathering : The initial stage of this research is collecting interview dataset. Hence, accord- ingly, I conducted mock interviews for an open position. A data set consisting of some videos have been created.Interview

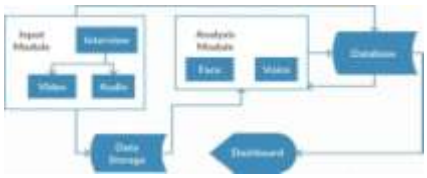

Fig. 7. System Architecture

is taken in a realistic setting. The setup is fairly simple, the hardware utilized is a camera, which would face the interviewee and a single unidirectional microphone, for the interviewee. Two chairs for the seating and a table for the positioning and placement of the camera. The camera is used to collect facial expressions and body cues while the microphone is used to collect audio conversations during the course of the interview. The primary medium of interaction between the interviewer and the interviewee is considered to be with the English language.

The interviews conducted were undertaken by professional and trained experts having industrial experience in a profes- sional environment. Before these interviews are conducted, a questionnaire is prepared consisting of 7 questions which are classified into different personality characteristics. Seven basic traits, each of them is being supported by a question from the questionnaire. This mapping is essential; thus, each question helps to determine that specific trait of the candidate. Refer to the following table, which portrays the mapping of each trait with the question constructed:

2) Annotation- Data Labelling: Next step in dataset build- ing is labeling of the interviews, for which a chair of independent judges rate all the interviews manually. This process is essential so as to get the ground truth about the interview statistics. A rating form was circulated to the chair who watched the entire interview of every candidate and accordingly assigned scores for different assessment questions in the form. Since human ratings are subjective in nature, the chair consisted of 6 judges, for them to rate the interview and finally taking the average of the final ratings.

IV. IMPLEMENTATION AND EVALUATION

The proposed system produced a variety of results, froma single personality trait to the overall hireability recommen- dation of the candidate. We produced graph plots for every aspect, from internal structures, models to the final output,so as to represent the accuracy and working of the system constructed

A. User Interface Results

User Interface outputs are purely choice based, depending upon the requirements of the system user. The UI that we designed for this project allowed an interviewer to observe in real time; a smile score variation graph. it represents the real- time smile score variation shown to the HR manager in the

Fig. 8. Smile Score Variation Graph as visible to the interviewer on theirUser Interface UI. As the interview proceeds the graph changes depending on the user's smile intensity.

Fig. 9. An interviewee's final scores displayed to the interviewer on their UIdashboard at the end of the interview.

From figure 9 when an interview ends, the system accord- ingly makes its predictions, and thusly produces a series of ratings to the interviewer. These ratings belong to various parameters; as visible, Calmness, eye contact, excitement, focused, friendliness, et .Not only the individual parameters but the system also predicts an overall and recommended ratings of the interviewee. These parameters are predicted with

different inputs provided by the interviewee in a combination.

V. CONCLUSION

The system developed here uses a multimodal approach, we have used Audio, Video as well as Lexical cues for preparing the prediction model. This modal simplifies the task of an organization for conducting HR based interviews, by allowing the interview process to be automated, carrying out the tasks of evaluating the verbal and nonverbal signatures of the interview candidate. Both are equally important when it comes to HR interviews. One of the most important case to consider isthat the existing systems do not propose a multimodal nature existence. They are either dependent on audio or on video cuesfor analysis.

The dataset was generated with the help of video recording sessions of interviews of a number of candidates, they were then manually rated by a chair of judges for the sake of ground truth, so a comparison could be made when the system predictsthe scores and ratings for the candidates. While generatingdata sets it became a necessity to ensure that the video quality was maintained, the candidate's facial points were visible in the recordings, not only that, but, we had to make sure that any kind of noise recorded, whilst during the interview was supposed to be minimized if not eliminated.

REFERENCES

[1]. P. Tripathi, P. P. Watwani, S. Thakur, A. Shaw and S. Sengupta, "Discover Cross-Modal Human Behavior Analysis", ,IEEE accss 2018.

[2]. DeGroot, T., Gooty, J "Can Nonverbal Cues be Used to Make Meaningful Personality Attributions in Employment Interviews?. ", J Bus Psychol 24, 179–192 ,2009.

[3]. V. Soman and A. Madan, "Social signaling: Predicting the outcome of job interviews from vocal tone and prosody", in ICASSP. Dallas, Texas, USA: IEEE, 2010.

[4]. Nguyen, A. Marcos, M. Marron, and D. Gatica, "Multimodal analysis of body communication cues in employment interviews ", , In ACM International Conference on Multimodal Interaction (ICMI), 2013.

[5]. A. Madan, R. Caneel, S. Pentland ""VibeFones: Socially aware Mobile Phones", (ISWC), Switzerland, 2006.

[6]. A. Pentland, J. Curhan, J. Khilnani, M. Martin, N. Eagle, R.Caneel, A.Madan, ""A Negotiation Analyzer", Santa Fe, NM, Oct. 25-27 2004

[7]. S. Feese, B. Arnrich, G. Troster, B. Meyer, and K. Jonas."Detecting posture mirroring in social interactions with wearable sensors. ", IEEE International Symposium, 2011.