International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

MULTIMODAL HIERARCHICAL ATTENTION NEURAL NETWORK LOOKING FOR CANDIDATES BEHAVIOUR WHICH IMPACT RECRUITER'S DECISION

[1]Dr.K.SUDHAKAR, [2]CH.SIRI CHANDANA REDDY, [3]CH.SHIVANI, [4]CH.NEHA

[1]Assistant Professor, Department of Electronics and Communication Engineering, Malla Reddy Engineering College For Women, Maisammaguda, Dhulapally Kompally, Medchal Rd, M, Secunderabad, Telangana.

[2,3,4]Student, Department of Electronics and Communication Engineering, Malla Reddy Engineering College For Women, Maisammaguda, Dhulapally Kompally, Medchal Rd, M, Secunderabad, Telangana.

ABSTRACT

Automatic analysis of job interviews has gained in interest amongst academic and industrial research. The particular case of asynchronous video interviews allows to collect vast corpora of videos where candidates answer standardized questions in monologue videos, enabling the use of deep learning algorithms. On the other hand, state-of-the-art approaches still face some obstacles, among which the fusion of information from multiple modalities and the interpretability of the predictions. We study the task of predicting candidates performance in asynchronous video interviews using three modalities (verbal content, prosody and facial expressions) independently or simultaneously, using data from real interviews which take place in real conditions. We propose a sequential and multimodal deep neural network model, called Multimodal HireNet. We compare this model to state-of-the-art approaches and show a clear improvement of the performance. Moreover, the architecture we propose is based on attention mechanism, which provides interpretability about which questions, moments and modalities contribute the most to the output of the network. While other deep learning systems use attention mechanisms to offer a visualization of moments with attention values, the proposed methodology enables an in-depth interpretation of the predictions by an overall analysis of the features of social signals contained in these moments.

I. INTRODUCTION

In today's competitive job market, organizations face the challenge of identifying candidates whose skills and behaviors align with their specific needs and culture. Traditional recruitment methods often rely heavily on resumes and interviews, which can overlook valuable insights from a candidate's online presence and behavioral traits. To address this challenge, our project proposes a Multimodal Hierarchical Attention Neural Network (MHANN) designed to analyze and evaluate candidate behavior across various data modalities. By integrating diverse information sources—such as resumes, social media profiles, and interview feedback—this innovative approach aims to provide a holistic view of candidates. The hierarchical attention mechanism within the network enables the model to prioritize and weigh different

behavioral indicators, ensuring that the most relevant aspects are highlighted in the recruitment process.

The use of advanced neural network architectures not only enhances the accuracy of candidate assessments but also streamlines the decision-making process for recruiters. As organizations increasingly leverage data-driven approaches, our project seeks to contribute valuable insights into how candidate behavior influences recruiters' decisions, ultimately leading to more informed hiring outcomes. This integration of technology and behavioral analysis marks a significant advancement in the field of recruitment, promising to enhance the efficiency and effectiveness of talent acquisition strategies.

II.EXISTING SYSTEM

Previous work on automatic analysis of job interviews also differ on the corpora used to train and evaluate the systems. These corpora differ on several points, namely: the type of interview (face-to-face or asynchronous), the settings of the interview (real position or simulation), the origin of the labels (practitioners, specialists, or non-specialists), and the size of the corpus. In fact, interview settings and type matter, as it could influence a candidate's behavior during the interviews [35], [56]. Secondly, as hireability is a complicated label to provide, it is important to know who labeled the data. Finally, the number of candidates affects reliability and method. Table 1 contains a summary of the corpus of job interviews used in previous work. Notice that only two other databases were collected in a real setting [30], [47]. Apart from the database of the present study, the only database comparable in terms of candidates is the one we previously collected in [30]. We had to proceed with this second collection for two reasons: first, for legal and privacy reasons, many videos from the first database were not available anymore, as the expiration date has been reached.

Second, the previous dataset did not include a timestamp of the spoken words, making it impossible to merge low level modalities. In summary, the database used in this study consists of 5,148 real asynchronous video interviews of candidates for sales positions, assessed by practitioners. Recent advances drastically reduce the time spent manually coding behavioral cues. Tools are now available to automatically code vocal [18] or visual [5] cues. Moreover, advances in automatic speech recognition have enabled researchers to obtain the automatic transcription of the candidate's verbal content. As asynchronous job interviews are videos, features from each modality (verbal content, audio, and video) have to be extracted frame by frame in order to build a classification model. Audio cues consist mainly of prosody features (e.g., fundamental frequency, intensity, mel-frequency cepstral coefficients) and speaking activity (e.g., pauses,
silences, short utterances) [48], [57]. Features derived from facial expressions (e.g., facial action units, head rotation and position, gaze direction) constitute the most extracted visual cues [13]. In order to describe the verbal content, researchers have used lexical statistics (e.g., number of words, number of unique words), dictionaries (Linguistic Inquiry Word Count) [57], topic modeling

[45], bags of words or more recently word or document embedding [11], [43].

Disadvantages

• The complexity of data: Most of the existing machine learning models must be able to accurately interpret large and complex datasets to detect candidate Behavior.

• Data availability: Most machine learning models require large amounts of data to create accurate predictions. If data is unavailable in sufficient quantities, then model accuracy may suffer.

• Incorrect labeling: The existing machine learning models are only as accurate as the data trained using the input dataset. If the data has been incorrectly labeled, the model cannot make accurate predictions.

III.PROPOSED SYSTEM

In this paper, we base our work on a specific dataset of on-demand video interviews which have been all conducted in real recruiting campaigns. This gives us a realistic setting and an ad-hoc notion of hireability, based on the candidates that were indeed labeled as such. In that sense, we propose a multimodal hierarchical attention neural network that aims to predict hireability based on facial expressions, prosody, and verbal content extracted from video answers of candidates. This neural network fuses the modalities at regular time interval, overcoming the problem of noisy ASR while fusing at a fine scale. Moreover, three components based on attention are used to understand 1) which questions and answers are considered more important by the automatic system, 2) in an answer, which moments are considered more remarkable by the automatic system, 3) in these moments, which modalities contribute the most. To evaluate this system, we, in collaboration with an HR company, have collected a new database constituted of more than 5,000 real asynchronous job video interviews for more than 400 real open positions. Our results highlight that 1) multimodal models perform better than standalone monomodal models, 2) performance of multimodal architecture drop when fusing at word level with automatic transcripts, 3) a regular time interval fusion can overcome this limitation. Finally, we led a study on the proposed attention mechanisms to understand if the moments highlighted by the system are indeed relevant. For that purpose, we compare these moments with random moments drawn from the same answer, characterizing their differences using the original features. Lastly, we investigate if these moments are carrying more information than random moments with regards to the task of predicting hireability.

Advantages

i) **Baseline using integrated features.**We apply standard statistical functions to collapse time dimension and to obtain one fixed vector for each modality. The mean, standard deviation, minimum, maximum, sum of positive gradients, and sum of negative gradients are applied to the audio and video sequences. The average pooling is used for the language modality. This approach is one of the most used amongst previous works [45], [48], [58].

ii) **Baseline using bag of audio/video/language words.** We chose to compare our model to [13]'s bag of audio and video words: we run a K-means algorithm on all the lowlevel frames in our dataset. Then

we take our samples as documents, and our frames' predicted classes as words, and use a "Term Frequency-Inverse Document Frequency" (TFIDF) representation to model each sample.

iii) **HireNet Monomodal Model.**We assess our model in the monomodal setting by removing the multimodal answer encoder. Monomodal input sequences are used and, for audio and video modalities, values are smoothed with a time window of 0.5 s and an overlap of 0.25 s as it has been done originally in [30]. Finally, we compare the performance between the use of the additive attention mechanism and the proposed contextual attention mechanism.

IV. MODULES

Service Provider

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Status, View Ratio, Download Trained Data Sets, View All Remote Users.

View and Authorize Users

In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

Remote User

In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like REGISTER AND LOGIN, PREDICT TYPE, VIEW YOUR PROFILE.

V. CONCLUSION

The Multimodal Hierarchical Attention Neural Network (MHANN) has demonstrated significant potential in analyzing candidates' behaviors that influence recruiters' decisions. By integrating various data modalities—such as resumes, social media profiles, and interview feedback—this model provides a comprehensive understanding of a candidate's suitability for a position. The hierarchical attention mechanism allows for nuanced weightage across different features, ensuring that the most relevant behavioral indicators are highlighted during the evaluation process. The insights gained from this project can enhance the recruitment process, making it more data-driven and efficient. By leveraging advanced neural network architectures, recruiters can make informed decisions that align with organizational goals and cultural fit. Future research could explore the application of this framework in diverse industries and its adaptability to evolving recruitment challenges.

VI. REFERENCES

1. Vaswani, A., Shankar, S., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (NeurIPS).

2. Yang, Z., Yang, D., Dyer, C., He, X., & Socher, R. (2016). Hierarchical Attention Networks for Document Classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association

for Computational Linguistics: Human Language Technologies (NAACL-HLT).

3. Zhang, Y., & Wang, M. (2019). Multimodal Deep Learning for Emotion Recognition in Conversations. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP).

4. Liu, Z., Wu, Y., & Wang, Y. (2020). A Survey on Deep Learning for Recruitment. Journal of Computer Science and Technology, 35(3), 471-493.

5. Zhao, X., & Huang, Y. (2021). The Impact of Social Media on Recruiters' Decision Making. Journal of Human Resources Management, 39(2), 100-115.

6. Chen, Z., Liu, X., & Zhang, Y. (2022). Behavioral Analytics in Recruitment: A Neural Network Approach. International Journal of Information Systems, 16(1), 25-39.

7. Kowsari, K., & Brown, D. (2020). Deep Learning for Recruitment: An Overview. Human Resource Management Review, 30(4), 112-121.