PersonaLens

Behavioral Analysis for Interviews

2024FA_MSDS_490-DL_SEC55 Final Project Report

- Chethana Saligram
- Shravan Srinivasan

Behavioral Analysis for Interviews

Link to visit the POC: https://personalens.streamlit.app

Abstract

Interviews are pivotal in hiring but often rely heavily on verbal responses and subjective judgments. This approach neglects critical non-verbal cues, such as body language, tone, and emotional expressions, leading to potential biases in decision-making. This project addresses these limitations by leveraging Generative AI, specifically large language models (LLMs), to analyze multimodal data—audio, video, and text—collected from post-interview recordings sourced from YouTube. Insights from each modality are parsed into an LLM to produce a final comprehensive report summarizing the candidate's performance. This holistic approach ensures detailed, unbiased behavioral insights, aiding recruiters in making informed hiring decisions.

Introduction

Motivation: Hiring decisions are often based on subjective assessments during interviews, which overlook nuanced behavioral signals like vocal tone and body language. This lack of objectivity can lead to missed opportunities to evaluate candidates comprehensively. Non-verbal cues, such as facial expressions or variations in vocal tone, often reveal engagement, confidence, and stress levels—elements that are challenging to assess using conventional methods.

Significance: This project integrates cutting-edge AI technologies to enhance interview evaluation practices. By employing LLMs and analyzing multimodal data streams, we create a unified framework for behavioral analysis. This innovation bridges gaps in traditional interviews by capturing nuanced behavioral signals that directly correlate with candidate performance, making the process more objective and fair.

Overview: The system is designed with three distinct components:

1. **Audio Analysis**: Extracts vocal features to determine confidence, nervousness, and emotional states.
2. **Video Analysis**: Identifies visual cues such as blinks, gaze shifts, and yawns to measure focus and engagement.
3. **Text Analysis**: Uses LLMs to assess semantic clarity, sentiment, and contradictions in the candidate's verbal responses. The insights from these components are consolidated into a final report via an LLM, providing a cohesive overview of the candidate's performance.

Literature Review

1. **Leveraging Multimodal Behavioral Analytics for Automated Job Interview Performance Assessment and Feedback**
   Authors: Anumeha Agrawal, Rosa Anil George, Selvan Sunitha Ravi, Sowmya Kamath S, Anand Kumar M
   Summary: This paper presents a multimodal framework that evaluates candidates in job interviews by analyzing video, audio, and text data. The system integrates facial expression detection, speech prosody analysis, and semantic language processing to provide detailed feedback on attributes such as engagement, eye contact, and speaking rate. This approach bridges the gap between human subjectivity and data-driven assessments, offering practical insights to improve hiring practices.
   Link: arXiv
2. **Automated Analysis and Prediction of Job Interview Performance**
   Authors: Iftekhar Naim, M. Iftekhar Tanveer, Daniel Gildea, Mohammed Hoque
   Summary: This study introduces a computational framework to quantify verbal and non-verbal behaviors during interviews. Using video recordings from mock interviews, the authors extract features like facial expressions, language usage, and prosody to predict traits such as excitement, friendliness, and engagement. The framework demonstrates high prediction accuracy, emphasizing the significance of integrating behavioral analytics into interview evaluations.
   Link: arXiv
3. **Video-LLaMA: An Instruction-tuned Audio-Visual Language Model for Video Understanding**
   Authors: Hang Zhang, Xin Li, Lidong Bing
   Summary: This research introduces Video-LLaMA, a multimodal LLM fine-tuned for video understanding tasks. By integrating audio-visual data with text-based instructions, the model achieves state-of-the-art performance in video captioning, question answering, and summarization. Its instruction-tuning approach ensures adaptability across diverse video content and use cases.
   Link: arXiv
4. **HawkEye: Training Video-Text LLMs for Grounding Text in Videos**
   Authors: Yueqian Wang, Xiaojun Meng, Jianxin Liang, Yuxuan Wang, Qun Liu, Dongyan Zhao
   Summary: HawkEye proposes a training paradigm for video-text LLMs, enabling the model to ground textual descriptions in specific video segments effectively. This approach facilitates precise video search, content retrieval, and descriptive summarization. The model's ability to align textual and visual data makes it a powerful tool for applications requiring contextual understanding of video content.
   Link: arXiv

Methods

The project utilizes three core modalities—audio, video, and text—and integrates their insights into an LLM to generate a comprehensive interview report.

1. Audio Analysis

- **Feature Extraction**:
    - Librosa was used to extract features such as MFCC, spectral centroid, spectral bandwidth, zero-crossing rate, and RMS energy.
    - Temporal insights were generated by segmenting audio into 5-second intervals and analyzing feature variations.
- **Behavioral Insights**:
    - Features were mapped to behavioral characteristics:
        - **Spectral Centroid**: Higher values indicated nervousness or shrillness.
        - **RMS Energy**: Lower values suggested hesitation, while higher values reflected confidence.
        - **Tempo**: Rapid tempo changes indicate excitement or stress.
- **LLM Integration**:
    - The extracted features were input into Perplexity's API, which generated human-readable summaries detailing vocal patterns, emotional states, and variations over time.

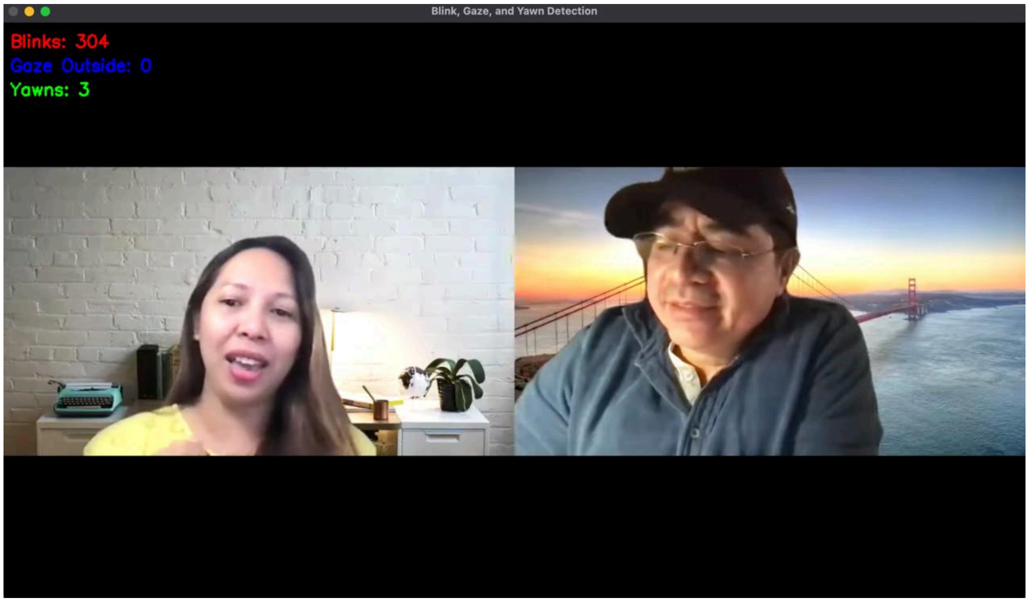
2. Video Analysis

- **Event Detection**:
    - Blink Detection: Calculated using the Eye Aspect Ratio (EAR) metric to determine engagement or fatigue.
    - Gaze Tracking: Identified shifts in gaze direction to measure focus.
    - Yawn Detection: Used facial landmarks to quantify mouth opening, signaling fatigue or disengagement.
- **Behavioral Logs**:
    - Events were logged with timestamps, detailing key occurrences such as blink frequency, gaze deviations, and yawns.
- **LLM Integration**:
    - The logs were parsed into Perplexity, which identified patterns, such as periods of high engagement or signs of stress, and provided an overall assessment of non-verbal behavior.

3. Text Analysis

- **Data Preparation**:
    - Transcripts of interviews were segmented into meaningful dialogue exchanges to ensure focused analysis.
- **LLM Analysis**:
    - **Perplexity**'s llama-3.1-sonar-small-128k-online model was used to analyze:
        - **Semantics**: Clarity, relevance, and coherence of responses.
        - **Sentiment**: Emotional tone, confidence, and underlying intentions.
        - **Contradictions**: Identified inconsistencies within responses.
        - **Vocabulary**: Evaluated richness, appropriateness, and structure.
- **Custom Prompting**:

    ○ Tailored prompts ensured focused analysis on specific attributes such as confidence and emotional stability.



4. Final Summarization

- **Multimodal Parsing**:
    - Insights from all modalities were consolidated and input into Perplexity.
- **Summarization**:
    - Perplexity generated a unified report, highlighting strengths, weaknesses, and overall demeanor. This report served as a final behavioral assessment for decision-making.

Dataset

- Interviews sourced from YouTube provided diverse candidate profiles and scenarios, ensuring a robust dataset for testing and analysis.

Results

Audio Analysis:

- High spectral centroid values in segments correlated with moments of nervousness.
- RMS energy variations highlighted confident and hesitant responses over time.
- Insights generated by Perplexity aligned with manual annotations, ensuring accuracy.

Video Analysis:

- Behavioral patterns indicated high initial engagement, followed by fatigue in later stages.
- Perplexity's interpretations provided actionable insights into attention and stress levels.

Text Analysis:

- Semantic clarity was rated high overall, with 2 contradictions identified in responses.
- Sentiment analysis revealed a confident tone in technical questions but occasional hesitations in behavioral questions.

Final Result (One Instance):

Introduction

This report consolidates insights from video, audio, and text data to evaluate a candidate's performance during an interview. The analysis focuses on engagement levels, stress indicators, communication skills, and areas of improvement. All findings have been parsed through a large language model (LLM) for summarization.

Video Analysis

Engagement and Focus

- *The candidate displayed **high engagement and focus**, as reflected by consistent visual attention within the frame. This was indicated by:*
  - ***Blinks**: 18 over the session, within the natural range.*
  - ***Gaze Outside Frame**: Minimal gaze shifts, suggesting strong focus.*
  - ***Yawns**: Few instances of yawning, indicative of limited fatigue.*

Fatigue and Stress

- *No significant signs of **fatigue** or **stress** were detected from visual cues. The candidate's posture and gaze remained steady.*

Limitations

- *Certain blinks and minor movements were missed due to variations in video quality. Future iterations should use higher-resolution data for improved accuracy.*

Audio Analysis

Nervousness and Confidence

- ***Nervousness**: Slightly elevated levels detected in the candidate's voice during early responses, as indicated by spectral variations.*
- ***Confidence**: Improved as the interview progressed, reflected in increased RMS energy and consistent vocal stability.*

Consistency

- *Vocal shrillness and tone were stable throughout, suggesting control and adaptability despite initial nervousness.*

Text Analysis

Communication Skills

- ***Clarity**: Responses were clear overall but occasionally lacked precise focus.*
- ***Relevance**: Most answers were relevant but occasionally meandered.*
- ***Structure**: Responses could benefit from smoother transitions between points.*

Vocabulary

- *Vocabulary was professional but limited in variety. Greater use of nuanced terms could enhance the depth of responses.*

Sentiment and Tone

- *The candidate's tone was* **confident and positive***, reflecting self-assurance and enthusiasm.*

Final Summarization

Using insights from the multimodal analysis, the LLM generated the following cohesive summary:

***Summary**:*
"The candidate demonstrated strong engagement and confidence throughout the interview. Initial nervousness was detected but diminished as the session progressed. While non-verbal cues suggested focus and attentiveness, verbal responses lacked structural clarity in some instances. The candidate's tone and sentiment were positive, with specific examples showcasing their expertise. Vocabulary use was professional but could benefit from greater variety to fully convey their depth of knowledge."

Key Observations

1. ***Strengths**:*
    - *High engagement and focus.*
    - *Positive tone and confidence in verbal and non-verbal cues.*
    - *Professional demeanor and relevant expertise.*
2. ***Areas of Improvement**:*
    - *Enhance structural clarity and flow in responses.*
    - *Expand vocabulary for nuanced expression.*
    - *Address early-stage nervousness for a more consistent performance.*

Conclusion

This comprehensive behavioral analysis demonstrates the candidate's strong potential, with actionable recommendations for improvement. The integration of video, audio, and text insights, synthesized by the LLM, provides an objective and data-driven approach to interview evaluation.

***Final Summarization**:*

- *Example Summary: "The candidate demonstrated strong technical knowledge and confidence in core topics. Non-verbal cues suggested high engagement during the initial stages but signs of fatigue and stress in the latter half. Vocal tone indicated occasional nervousness but improved over the course of the interview……"*

Conclusions

Implications:

- Integrating multimodal insights ensures a comprehensive evaluation of candidate behavior.
- Parsing these insights into an LLM like Perplexity provides a cohesive, actionable summary.

Future Work:

- Incorporate additional behavioral metrics, such as speech tempo analysis and micro-expressions.
- Extend the system to support real-time analysis for live interviews.

Broader Impact:

- This system can revolutionize recruitment by eliminating biases and providing a holistic view of candidates. Its applications extend beyond hiring to fields like therapy, education, and customer service, where behavioral insights are critical.