AI for human assessment: What do professional assessors need?

Riku Arakawa (CMU, ACES, Inc.), Hiromu Yakura (UTsukuba, AIST) †equal contribution
https://rikky0611.github.io/
https://yumetaro.info/

Domain: Human Assessment

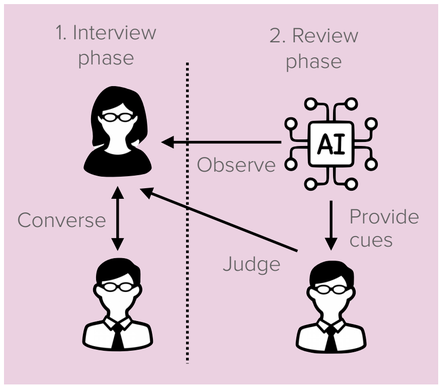
- Assess employees' suitability (e.g., as a manager)
- Professional assessors make decisions through conversation with the employees
  - **Check behaviors in interviews, especially nonverbal cues**

- **Time-consuming** ⏳
- **Wrong decision due to assessors' subjectivity** 🧐

Q: How can AI systems support professional assessors' decision-making?

🏢 Japanese assessment company

conducted a workshop with two professional assessors ✏️



Lessons Learned

- It is neither recommended nor feasible to train an AI model that replicates assessors' decision-making process
  - Inevitable inconsistency among their processes

- Our design of **separating observation and judgment** is a promising approach in such highly contextual domains
  - Importantly, our goal is not replacing human decision

now conducting further case studies

We appreciate your comments on how we can gain new knowledge in human-AI trust from this project 🙌

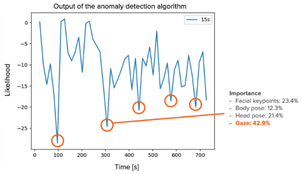
System Requirements

- The assessors were **skeptical about AI-based end-to-end decision making** because human assessment should consider various factors specific to each assessee
  - They are highly human-contextual and difficult to be captured by computers

- The assessors **expected AI systems to help them not miss important behavior cues due to their subjectivity or mental demands**
  - Then, the assessors can revise their judgment by taking the contextual meaning of such AI-detected cues into consideration

Hypothesis: Separating observation (by AI) and judgment (by professional)

conducted a feasibility evaluation using the videos of actual assessments



Findings

- We developed an **unsupervised anomaly-detection-based observation algorithm** to detect behavior cues that are highly contextual and specific to each assessee
  - Evaluated the agreement between the algorithm and the assessors

- The assessors found that the algorithm would facilitate their assessment, even though the algorithm does not completely replicate their annotation
- **The interpretable output of the anomaly-detection-based algorithm guided them to infer the reason behind the detection**
  - Crucial to maintain the assessors' trust in the case of false-positives 🤝

A: The separation contributed to the trust in this highly contextual domain