Interaction Analysis Practice Within Computational Trends: Promises and Challenges

Mengxi Zhou, Indiana University Bloomington,mz13@iu.edu
Joshua Danish, Indiana University Bloomington, jdanish@iu.edu

Joyce Fonteles, Eduardo Davalos, Efrat Ayalon, Alicia C. Lane, Ashwin T.S., Namrata Srivastava
joyce.h.fonteles@vanderbilt.edu, eduardo.davalos.anaya@vanderbilt.edu, efrat.ayalon@vanderbilt.edu, alicia.c.lane@vanderbilt.edu, ashwindixit9@gmail.com, namrata.srivastava@vanderbilt.edu,
Vanderbilt University

Selena Steinberg, ssteinb@iu.edu, Indiana University Bloomington

Gabriella Anton, Noel Enyedy, Gautam Biswas
gabriella.anton@vanderbilt.edu, noel.d.enyedy@vanderbilt.edu, gautam.biswas@vanderbilt.edu
Vanderbilt University

Abstract: Artificial intelligence (AI) approaches to analyzing multimodal data often rely on cognitive learning theories. More research is needed to explore AI's role in analyzing and identifying learning from a sociocultural perspective (Giannakos & Cukurova, 2023). This poster demonstrates an exploration of integrating computational methods into Interaction Analysis (Jordan & Henderson, 1995) and reports observed promises and challenges.

Introduction

Since the inception of Learning Sciences (LS), video analysis has been central to the field (Hall & Nemirovsky, 2012). Yet, analyses often overemphasize spoken words (Mondada, 2013) and stationary participants (Leander et al., 2010), and background other essential modalities in moving bodies (Hall & Nemirovsky, 2012). This lens has unnecessarily narrowed our understanding of learners' diverse forms of participation. Video analysis can elicit emic/participants' perspectives by inviting them to the analysis or by examining the organization of turn-taking, though memory loss and varied turn-taking forms (e.g., turn with bodies/artifacts) pose challenges (Jordan & Henderson, 1995; Sacks et al., 1974). Thus, attending to more multimodal resources of participants enriches analysis and aligns closer with emic perspectives. The rise of sensor technology and artificial intelligence (AI) holds potential to capture a broader range of multimodal behaviors (Spikol & Cukurova, 2019). We focus on one way of analyzing video records, Interaction Analysis (IA; Jordan & Henderson, 1995) and explore using computational methods to complement it. Our IA approach follows relational epistemology and views learning as socially emergent. Also, this exploration features an interdisciplinary collaboration between LS researchers with sociocultural views and Computer Science (CS) researchers.

Method

A target learning context for computationally supported interaction analysis

Our co-design explores using computational methods to complement IA of students' learning in a mixed-reality environment GEM-STEP (Danish et al., 2022), where several elementary students explore science phenomena through self-initiated and spontaneous whole-body movements and actions. While all learning is embodied, here students' embodiments are explicitly encouraged and extensively used. They physically move in a tracking space while observing their virtual agent's movements and interactions with other agents on a shared projection. This poster presents data collected from a fourth-grade class in a Southeastern city. Students role-played as molecules to explore the photosynthesis process. The simulation alternates between day and nighttime, and students act as molecules (oxygen, water, sugar, carbon dioxide) to move between locations (chloroplast, mouse, stem, root) to explore molecular transformation. During the day, a student playing water and another playing carbon dioxide meet at the chloroplasts to complete photosynthesis and transform themselves into oxygen and sugar. Next, the oxygen student moves to the mouse to become carbon dioxide, and the sugar student moves to the stem and then root to become water. This process sets a new stage for the next photosynthesis cycle. We collected video recordings of students' participation, screen recordings and software-generated logs of students' interactions.

Organizing computational analysis into a visualization tool

This multi-party, whole-body movement context represents a complex learning environment that can challenge IA. Researchers need to consider the simulation, the physical space, and multiple participants. Although students' movement and their agents are synchronized, multiple mediators (e.g., social norms) shape how they move, interact, and learn. LS researchers led the focus on the analysis of ongoing interactions in the video and incorporated observations during data collections. CS researchers processed logs and videos, and then created a tool that integrated video data with computational visualizations of system logs and AI analysis of gaze and affect to chronologically represent those multimodal data (Fonteles et al., 2024). This visualization has gone through iterative refinements to enhance its utility in IA by LS researchers, including displaying data of multiple students, switching camera angles, and transcription (Figure 1). Those refinements emerged from iterative testing and discussions where the LS team shared IA from a sociocultural perspective, and the CS team demystified the AI black-box (e.g., confidence value, training models limitation, data training process). Through this iterative process of refining design and results, we hope to ensure our understanding of students' views and participation in a dignity-affirming manner by recognizing and honoring their perspectives (Espinoza et al., 2020). We recorded weekly discussions and gathered reflective journals in Spring 2024. Then, we surveyed additional LS researchers (n=11) to capture their perceptions of the visualization as a supplement for IA.

Figure 1
The Visualization Tool Interface

Findings

Identifying interactional hotspots and formulating hypotheses for zoomed-in analysis

LS researchers found the camera-switching and transcription features useful, though improved accuracy and diarization are needed. The visualization of logs was considered particularly valuable in identifying productive moments in the video record, or what we called 'hotspots' for focused analysis. For instance, the visualization of multiple students' log data helped pinpoint instances of successful photosynthesis between two students. Those initial observations supported formulating hypotheses for IA examination. When conducting IA in contexts of multi-party interactions, human analysts can focus on certain participants but cannot "turn people off," but rows within the visualization can be turned on/off to help analysts focus. Emergent hypotheses identified for IA with select rows in the visualization showed the value of turning this tool on/off in supporting focused analyses. Our goal is not to suggest this is always better than directly analyzing video. Rather, combining methods/tools can facilitate a more fluid and confident analysis and reduce the feeling of "too much" to observe.

Discrepant views between AI and IA on students' affect

CS researchers included AI-identified gaze and affect in the visualization. Yet, LS researchers noted significant discrepancies between these results and those developed through IA, leading to concerns. They argued that interpreting affect using AI may lead to shallow/inaccurate understanding and expressed concerns about gendered and ableist aspects of facial/embodied performance. This distrust of AI is well-grounded, as AI reflects given data and training models, which often represents normative behaviors and inadvertently frames diverse learners in a deficit manner (Maddaio et al., 2022). Also, relying solely on facial expressions to infer affect repeated the narrow single-modality lens. We are now testing AI to identify multiple patterns in lower-inference content (e.g., speed, movement path, volume) to aid human interpretations. While improving AI transparency and explainability, we acknowledge human biases in IA and suggest AI-detected descriptions could challenge our biases.

Discussion

AI holds the potential to support IA (Davidsen & Steier, 2024), and the interpretation of multimodal data should consider contextual meanings (Giannakos et al., 2022; Worsley et al., 2021). Our co-design effort informs us that when using AI with IA, a closer collaboration throughout the ongoing analytical process is needed to ensure the benefits of AI technologies without losing the humanistic lens of IA (Kubsch et al., 2023).

References

Danish, J. A., Anton, G., Mathayas, N., Jen, T., Vickery, M., Lee, S., ... & Ryan, Z. (2022). Designing for shifting learning activities. Journal of Applied Instructional Design.

Davidsen, J., & Steier, R. (2024). Near future practices of interaction analysis: technology-mediated trends in educational research. *International Journal of Research & Method in Education*, 1-17.

Fonteles, J., Davalos, E., Ashwin, T. S., Zhang, Y., Zhou, M., Ayalon, E., ... & Biswas, G. (2024, July). A first step in using machine learning methods to enhance interaction analysis for embodied learning environments. In International Conference on Artificial Intelligence in Education (pp. 3-16). Cham: Springer Nature Switzerland.

Giannakos, M., & Cukurova, M. (2023). The role of learning theory in multimodal learning analytics. *British Journal of Educational Technology*.

Giannakos, M., Spikol, D., Di Mitri, D., Sharma, K., Ochoa, X., & Hammad, R. (2022). Introduction to multimodal learning analytics. In D. Di Mitri, M. Giannakos, R. Hammad, X. Ochoa, K. Sharma, & D. Spikol (Eds.), The multimodal learning analytics handbook (pp. 3–28). Springer.

Hall, R., & Nemirovsky, R. (2012). Introduction to the special issue: Modalities of body engagement in mathematical activity and learning. *Journal of the learning sciences*, *21*(2), 207-215.

Kubsch, M., Krist, C., & Rosenberg, J. M. (2023). Distributing epistemic functions and tasks—A framework for augmenting human analytic power with machine learning in science education research. *Journal of Research in Science Teaching*, *60*(2), 423-447.

Leander, K. M., Phillips, N. C., & Taylor, K. H. (2010). The changing social spaces of learning: Mapping new mobilities. *Review of Research in Education, 34*(1), 329394.

Madaio, M., Egede, L., Subramonyam, H., Wortman Vaughan, J., & Wallach, H. (2022). Assessing the fairness of ai systems: Ai practitioners' processes, challenges, and needs for support. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW1), 1-26.

Mondada, L. (2013). The conversation analytic approach to data collection. (In T. Stivers & J. Sidnell (Eds.), The handbook of conversation analysis (pp. 32-56). Oxford: Wiley-Blackwell.

Sacks, H., Schegloff, E., & Jefferson, G. (1974) A simplest systematics for the organization of turn-taking in conversation. *Language, 50*, 696-735.

Spikol, D., & Cukurova, M. (2020). Multimodal learning analytics. In *Encyclopedia of education and information technologies* (pp. 1221-1228). Cham: Springer International Publishing.

Worsley, M., Martinez-Maldonado, R., & D'Angelo, C. (2021). A new era in multimodal learning analytics: Twelve core commitments to ground and grow MMLA. *Journal of Learning Analytics, 8*(3), 10–27.

Acknowledgments

This work is supported by the National Science Foundation under grants 1908632, 1908791, and DRL-2112635. We would like to thank the teacher and students for their wonderful participation in this implementation. We also thank the learning sciences and computer science researchers at Indiana University Bloomington, Vanderbilt University, and North Carolina State University for their valuable feedback and contributions to the project.