

Exploring Artificial Intelligence Supported Interaction Analysis

Mengxi Zhou, Indiana University Bloomington, mz13@iu.edu
 Joyce Fonteles, Vanderbilt University, joyce.h.fonteles@Vanderbilt.edu
 Joshua Danish, Indiana University Bloomington, jdanish@indiana.edu
 Eduardo Davalos, Vanderbilt University, eduardo.davalos.anaya@vanderbilt.edu
 Selena Steinberg, Indiana University, ssteinb@iu.edu
 Gautam Biswas, Vanderbilt University, gautam.biswas@Vanderbilt.edu
 Noel Enyedy, Vanderbilt University, noel.d.enyedy@vanderbilt.edu

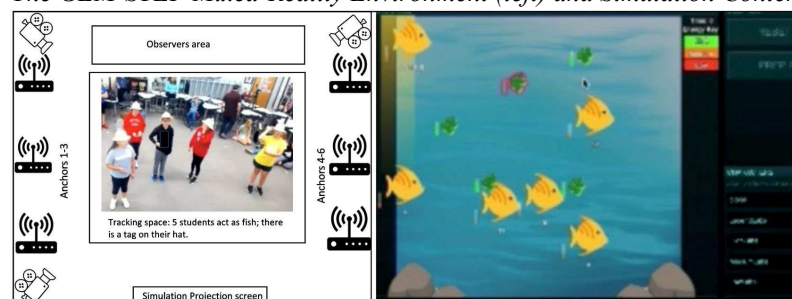
Abstract: Interaction Analysis (IA; Jordan & Henderson, 1995) is a widely used method in the Learning Sciences field. However, it requires working with numerous hours of video to discern patterns, posing challenges in time and effort. We explore a synergy between IA methodologies and Artificial Intelligence algorithms to analyze students' participation in a collective embodied activity and conclude with initial findings and future directions.

Introduction and background

Interaction analysis (IA; Jordan & Henderson, 1995) is a widely used approach to analyzing video data in the Learning Sciences because of its value in understanding participants' emergent interactions in real-world contexts. It begins with content-logging video data, from which researchers may form a hypothesis that will undergo a collective and iterative refinement wherein people from diverse backgrounds co-analyze video data through multiple analysis sessions. Researchers will then work with a significant number of observations of the phenomenon and assess the entire data corpus to check whether the observations indicate a general pattern, or just an idiosyncratic instance. This analysis process is not linear but iterative, which enhances analyses quality. However, there are significant challenges that come with analyzing such rich data because video data situated in learning contexts often contains multiple participants interacting with each other and their environment in varied ways-this makes it challenging for even expert observers to track all that is going on. We are interested in embodied learning contexts, where students explore concepts through the movement through space (e.g. being a fish to learn about energy transfer in aquatic ecosystems). In embodied contexts, IA becomes even harder due to human analysts' limited cognitive resources and the fluid nature of the context, leading to efforts to capture all of complexities in innovative visualizations or transcription methods. Thus, we explore the possibilities of leveraging Artificial Intelligence (AI) techniques to support human analysts in performing IA in an embodied context. This could include but is not limited to a simultaneous focus on multiple actors (which is hard for human analysts); layering AI information over videos or transcriptions; or using AI to identify video segments with particular phenomena. We also argue that aligning the use of AI with learning theories will maximize benefits for Learning Sciences researchers (Worsley & Blikstein, 2015). We adopt the Learning in Embodied Activity Framework (LEAF; Danish et al., 2020) to establish systematic synergistic links between IA and AI. LEAF links sociocultural perspectives and individual perspectives of embodied learning to highlight that students' participation and learning in a collective embodied activity are impacted by various mediators (e.g., rules, tools, object), shaping their movements and attention. As the first step of this synergy exploration, this study utilized AI technology to detect students' gaze patterns to identify key events that represent student behaviors and then conducted IA of the identified events of interest (e.g., gaze shifting, divergent/convergent gaze) to understand their impact on students' participation. We ask: Can AI help interaction analysts identify events in the video that are promising for follow-up analyses in a theoretically grounded manner?

Figure 1

The GEM-STEP Mixed-Reality Environment (left) and Simulation Content (right)



Methods

Context and participants

This study involved 22 fifth graders from a public school in the Midwestern United States in a Mixed-reality (MR) embodied learning environment GEM-STEP (Danish et al., 2022). The study used a motion-tracking computational modeling simulation to support elementary students' exploration of agent-based models. Students act as an agent (e.g., fish) and their movements are tracked and mapped onto a projected simulation of an ecosystem (e.g., pond) (Figure 1). Nine forty-five minutes sessions were recorded by three cameras positioned in three corners of the classrooms and one web camera on top of the projection screen.

Analysis and findings

Currently the most promising results of estimating gaze from videos have been obtained by deriving gaze vectors using L2CS-Net (Abdelrahman et al, 2022), a robust CNN (Convolutional Neural Network)-based model for predicting gaze in unconstrained settings. With the deep learning method, we generated two data outputs of students' gaze patterns: a spreadsheet (indicating timestamps of gaze cohesion and shifting) and video annotations with gaze vectors, from which we identified moments of gaze cohesion or not and then used those moments to focus our human IA efforts. Figure 2 displays two gaze patterns that were also produced in the spreadsheet. Our IA suggested that students' gaze toward/away from the simulation may indicate perspective/role-shifting (as an agent or observer), which supported their noticing of the ecosystem differently and signal the salience of different mediators (e.g., the simulation and then facilitators). In analyzing students' collective gaze toward the simulation, we processed events logs generated in GEM-STEP and produced visualizations (e.g., heat maps) of students' moving paths to support IA. The IA helped us explore how collective gaze and awareness of co-presence were not necessarily related as students focused heavily on different aspects of the simulation. This can then provide the basis for a deeper understanding of the unique patterns of collaboration and attention in this kind of design.

Figure 2

Examples of Students' Gaze Annotations in the Video



Discussion

This goal of this synergistic effort is not to replace human analysts but to leverage AI to support humans in IA. Importantly, we view this as moving away from technology focused human-in-the-loop solutions to what we think of as AI-in-the-loop to highlight the driving role of humans. We seek ways to understand how to refine our AI algorithms to support human analysts, in contexts where there are excessive amounts of information involving multiple actors in a dynamic environment. We hope to create systematic methodologies and visualization schemes where Learning Sciences researchers can benefit from AI to address complex IA problems using multimodal data.

References

- Abdelrahman, A. A., Hempel, T., Khalifa, A., & Al-Hamadi, A. (2022). L2CS-Net: fine-grained gaze estimation in unconstrained environments. *arXiv preprint arXiv:2203.03339*.
- Danish, J., Anton, G., Mathayas, N., Jen, T., Vickery, M., Lee, S., ... & Ryan, Z. (2022). Designing for shifting learning activities. *The Journal of Applied Instructional Design*, 11(4), 169-185.
- Danish, J. A., Enyedy, N., Saleh, A., & Humburg, M. (2020). Learning in embodied activity framework: A sociocultural framework for embodied cognition. *International Journal of Computer-Supported Collaborative Learning*, 15, 49-87.
- Worsley, M., & Blikstein, P. (2015, March). Using learning analytics to study cognitive disequilibrium in a complex learning environment. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 426-427).

Acknowledgement

This work was supported by the National Science Foundation: 1908632, 1908791, and DRL-2112635.