Kashihara, A. et al. (Eds.) (2024). Proceedings of the 32nd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Developing a Multimodal Learning Analytics Approach to Examine Students' Cognitive Presence and Metacognition in a Metaverse Environment

Yanjie SONG*, Lei TAO, Hao DENG & Jiachen FU
Department of Mathematics and Information Technology,
The Education University of Hong Kong, Hong Kong SAR, China
*ysong@eduhk.hk

Abstract: The rise of the metaverse as an educational platform has introduced new opportunities and challenges in understanding students' cognitive processes. Traditional learning analytics methods often fail to fully capture the dynamic patterns of cognitive presence and metacognitive activities in immersive virtual environments, particularly when students interact with an artificial intelligence (AI)-powered digital human. To address this gap, this study aims to develop a multimodal learning analytics (MMLA) approach to analyse cognitive presence and metacognition in Learningverse - a metaverse platform. To address the limitations of traditional methods, this research integrated eye-tracking data with dialogue text from the AI-powered digital human to provide a more comprehensive understanding of these patterns. A pilot study involving undergraduate students was conducted to collect data, revealing specific patterns of cognitive presence and metacognition. The findings suggest that the MMLA approach could offer deeper insights into students' learning behaviours, providing valuable implications for the design of educational tools and the development of more effective learning strategies in virtual environments.

Keywords: Metaverse, cognitive presence, metacognition, multimodal learning analytics, AI-powered digital human

1. Background

1.1 Learning in the metaverse and the role of AI-powered digital humans

The metaverse, as a new digital learning environment, has gained increasing attention in recent years within the field of education. Platforms like Learningverse (Song et al., 2023) enable students to act as avatars and collaborate in learning activities, offering an immersive and interactive learning experience without the need for a head-mounted display. In Learningverse, students interact not only with the virtual learning environment but also with AI-powered digital humans. These digital humans, driven by advanced natural language processing (NLP) technology, simulate human interactions by providing real-time feedback and guidance. This interaction could deeply engages students by allowing them to ask questions, receive tailored explanations, and engage in reflective dialogue (Lynch et al., 2023). These digital humans are not just tools for delivering information; they are designed to promote active learning by encouraging students to think critically, reflect on their understanding, and construct knowledge through interactive dialogues (Kong & Yang, 2024; Lin & Chang, 2023). Recent research underscores the growing importance of digital humans in education, showing that these virtual assistants can effectively promote students' cognitive development and increase learner engagement through personalised dialogues (Gan et al., 2023). Understanding how these interactions can be optimised to support deeper learning outcomes is crucial. This forms the foundation for exploring advanced analytics approaches, such as

multimodal learning analytics, which can capture and analyse the rich data generated by these interactions within the metaverse.

1.2 Current state and challenges of multimodal learning analytics

With the advancement of educational technology, researchers have increasingly focused on the application of multimodal learning analytics (MMLA) in complex learning environments. MMLA integrates data from multiple sources, including textual, behavioural, and physiological data, to capture a more comprehensive view of students' learning processes (Mu et al., 2020). However, recent studies have pointed out that existing learning analytics methods still face challenges when applied to complex environments like the Metaverse.

Traditionally, learning analytics have predominantly relied on text data, such as discussion posts and assignment submissions, to assess students' cognitive activities. While useful, these data types may not fully capture the rich cognitive processes occurring within immersive environments, where interactions are often more dynamic and complex (Du et al., 2023; Gan et al., 2023). Among the various forms of data, eye-tracking technology stands out as an essential tool in MMLA. Eye-tracking provides detailed insights into students' attention distribution and information processing during learning, making it particularly valuable in dynamic and immersive learning environments (Van Gog & Jarodzka, 2013). The potential for eye-tracking to complement textual analysis has been recognized as a promising avenue for future research. For instance, Ouyang and Zhang (2024) proposed that multimodal data, including eye-tracking and emotional data, could be used to simultaneously track cognitive and non-cognitive processes. However, it is crucial to note that this integration remains a proposed direction. To address this limitation, the current study aims to explore the potential of integrating eye-tracking data with text data to develop a robust MMLA approach.

1.3 Multimodal learning analytics of cognitive presence and metacognition

Building on the identified need for more comprehensive analytical frameworks, this study applies the proposed MMLA approach to the analysis of cognitive presence and metacognition within the Community of Inquiry (CoI) framework, particularly in the context of the metaverse. Cognitive presence is a core concept in the CoI framework, referring to the extent to which learners can construct and confirm personal meaning through reflection and discourse (Garrison, 2000). In recent years, researchers have increasingly recognised that relying solely on text data to analyse cognitive presence may be insufficient in some complex learning environments (Garrison, 2016). In environments such as the metaverse, students' behavioural data, such as eye-tracking data, can provide richer insights into cognitive processes.

Shea et al. (2022) proposed the concept of shared metacognition, suggesting the addition of a fourth dimension—learning presence—related to metacognition and self-regulation in the CoI framework. While some researchers support this idea, issues exist. For instance, Garrison (2022) argues that introducing learning presence as a new dimension might disrupt the original constructivist premise of the CoI framework. Instead, Garrison suggests that metacognitive processes should be captured by supplementing the existing framework rather than adding a new dimension.

In Learningverse, combining eye-tracking data with students' dialogue text with the AI-powered digital human allows researchers to conduct more in-depth analysis of students' cognitive and metacognitive activities. This multimodal analysis approach can capture the complex cognitive processes of students in immersive learning environments and offer new perspectives for coding cognitive presence(Ouyang et al., 2023). However, the challenge remains in systematically integrating and analysing these multimodal data sources to fully capture the intricate patterns of cognitive presence that emerge in interactions with the digital human in such complex environments. In view of this, this proposed study aims to address these challenges by developing a multimodal learning analytics (MMLA) approach to understand students' cognitive presence and metacognition in the metaverse platform – Learningverse. This research will address the following questions:

RQ1: How can a multimodal learning analytics approach be designed to analyse cognitive presence and metacognition in students interacting with the digital human in the metaverse environment?

RQ2: What patterns of students' cognitive presence and metacognition can be identified using the integrated MMLA approach in the metaverse environment?

2. Methodology

2.1 Data collection

In the metaverse environment, "Learningverse", the digital human serves as a scaffold, offering personalised learning experiences (See Figure 1 in Learningverse for users and digital human). Communication with the digital human (dialog text data) and eye-tracking data are logged as multimodal data for further analysis. Therefore, we conducted a pilot study using inquiry-based learning instructional design on the topic of "How to Use Atomic Theory to Mitigate Climate Change". The inquiry-based learning instructional design consisted of five stages: Engage, Explore, Analyze, Explain, and Reflect. The digital human was involved in all the inquiry stages to provide support in their learning process. Seven participants from the Education University of Hong Kong were involved in the study, all aged over 18 and lacking prior knowledge of atomic theory.

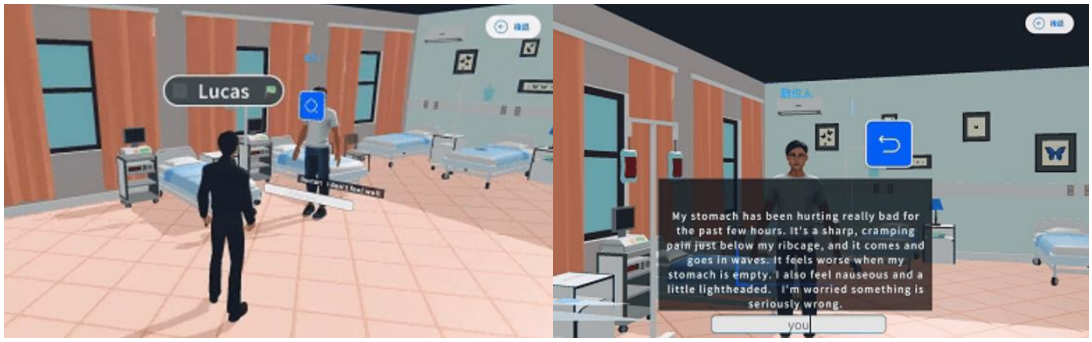

Figure 1. Users and digital human in Learningverse

2.2 Data analysis

The collected data underwent preprocessing based on a coding scheme for cognitive presence, which includes both text and behavioural components. This coding scheme integrates approaches from Shea et al. (2010) for metacognition and (Garrison, 2016) for cognitive presence (refer to Figure 2). This preprocessing facilitates subsequent analyses in MMLA.

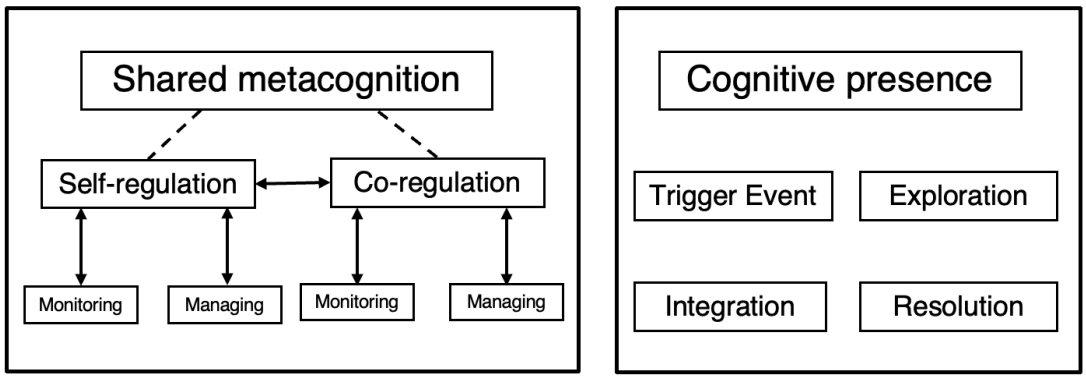

Figure 2. Integrated coding scheme for metacognition and cognitive presence from Shea et al. (2010) and Garrison (2016) to analyse text and behavioural data

2.3 Three-layer multimodal learning analytic approach

The MMLA approach was adapted from Ouyang et al. (2023) (refer to Figure 3). This approach integrates dual-channel sequence analysis, cluster analysis, and hidden Markov models to comprehensively analyse the multimode data including dialog data from participants and digital humans, as well as eye-tracking data.

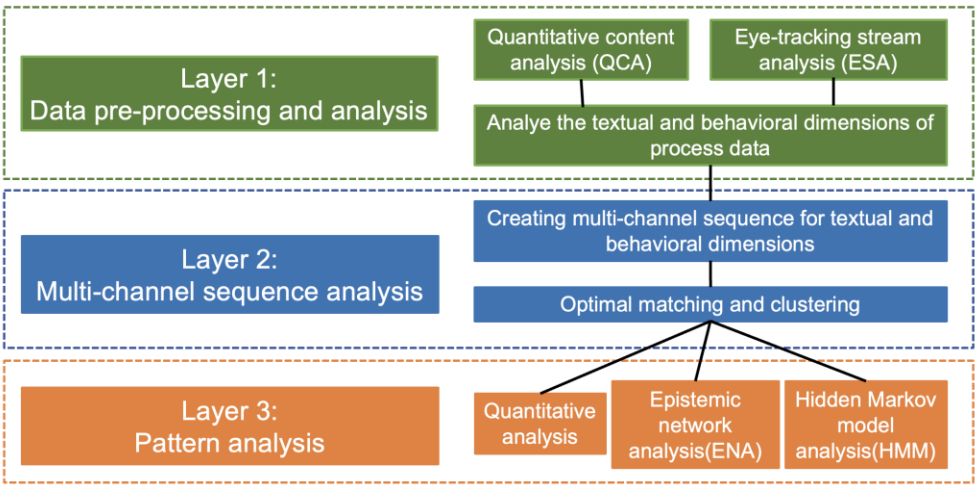

Figure 3. Three-layer multimodal learning analytic approach

In the first layer, two researchers code the time-stamped text and behavioural data based on the coding schemes from Shea et al. (2010) and Garrison (2016). The coding scheme included CR (co-regulation), SR (self-regulation), TE (triggering event), E (exploration), I (integration), and R (resolution). Initially, Rater 1 codes 30% of the dataset. Subsequently, Rater 2 recoded this portion to identify and resolve any discrepancies, aiming for a Krippendorff's alpha reliability of over 0.80 to ensure inter-rater reliability. After this reconciliation process, Rater 1 completed the coding of the remaining dataset, with Rater 2 reviewing and finalising the codings to ensure accuracy and consistency across the analysis.

In the second layer, multi-channel sequence analysis (MCSA) was used to examine the similarities of the cognitive presence of individuals' learning activities in the metaverse to detect different types of cognitive presence. MCSA is a sequence analysis method derived from the field of bioinformatics that is used to simultaneously analyse multiple parallel trajectories (e.g., dimensions, states) of a time series (Eisenberg-Guyot et al., 2020; Gauthier et al., 2010). MCSA examined similarities in cognitive presence across learning activities within the metaverse. It involved converting data into two-channel sequences for each student, aligning these using the optimal matching algorithm, and clustering them into types with similar patterns using ward clustering. Cluster selection was based on goodness-of-fit, dendrograms, and interpretability.

In the third layer, we demonstrated three perspectives of different cluster groups, namely, quantitative perspective, structural perspective and transitional perspective. In the quantitative perspective, we conducted descriptive analysis (MEAN and SD) on each coding dimension of different cluster groups. Epistemic Network Analysis (ENA) demonstrated the structure of cognitive existence of different cluster groups. The structural characteristics of ENA of different groups were shown by the location of the centroid of the network. The structure of the transition process indicated that the cognitive presence of students in different clusters changed during the learning process. This transition state could be revealed by the probabilistic method of Hidden Markov Model (HMM). This method was used to describe Markov chains with implicit parameters, detect the underlying process with a certain number of hidden states, and identify the expected transition pattern between hidden states (Eddy, 1996). The required number of states was determined by the best fit determined by the Bayesian Information Criterion (BIC).

3. Result and discussion

In the second layer, following preprocessing, encoding, and analysis in the first layer, the results of optimal Ward clustering based on similarity are shown in Figure 4. The optimal clustering results revealed patterns of cognitive presence and shared metacognition among two clusters within Learningverse. The first cluster is represented by orange and consists of the coded sequences of four students; the second cluster is represented by green and comprises the coded sequences of three students.

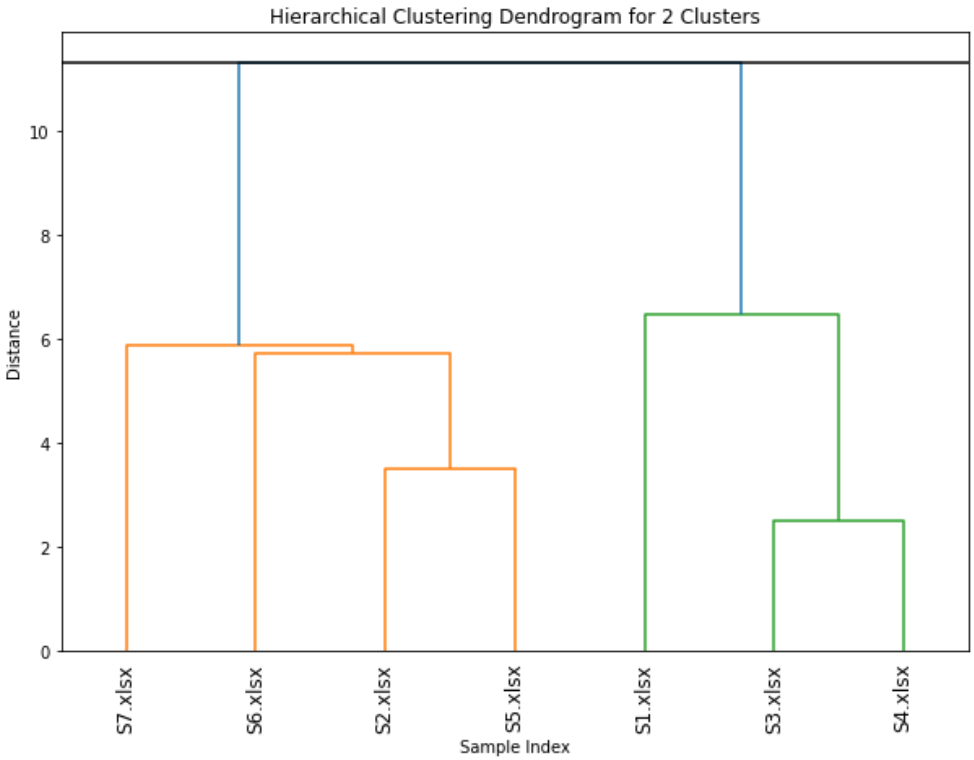

Figure 4. Hierarchy clustering result for 2 cluster

Figure 5 visualises the results of the dual-channel sequences for the two categories based on the optimal clustering results. In Figure 4, each channel represents the changes in coded events within the dimension over time. Specifically, the first channel is green, representing the text sequence. The second channel is blue, representing the behavioral sequence. The blank part indicates that no coded event occurred. The first channel describes the cognitive presence of students in learning according to the CoI framework, while the second channel serves as a complement to cognitive presence within the same framework. Together, these two channels describe the cognitive presence and shared metacognition of students in Learningverse. From Figure 5, two distinctly different learning patterns could be observed in the optimal clusters. Students in the first cluster tended to engage in self-regulation before collaborating with digital beings in subsequent learning processes. Most of the coded events of cognitive presence in this cluster are TE and E.

In contrast, students in the second cluster are more inclined to initially communicate with digital beings, generating events of shared metacognition. Following this, they engage less with digital beings and fewer events of cognitive presence are identified. Moreover, in later stages, events of shared metacognition and self-regulation become sparser and of shorter duration.

Shared metacognition significantly facilitates the construction of knowledge in students (Zheng et al., 2021). Additionally, students become more aware of their progress in the learning process through increased shared metacognition (Hadwin et al., 2017). The first category of students exhibits more self-regulation and shared regulation, while the second category exhibits less. It can be inferred that the first category of students may achieve better

knowledge construction and monitoring of their learning processes. Järvelä et al. (2013) noted that ideal learning involves a timely transition between self-regulation and co-regulation, which influences the learning outcomes. The first cluster closely aligns with this ideal state, whereas the second cluster does not conform as closely to this ideal state.

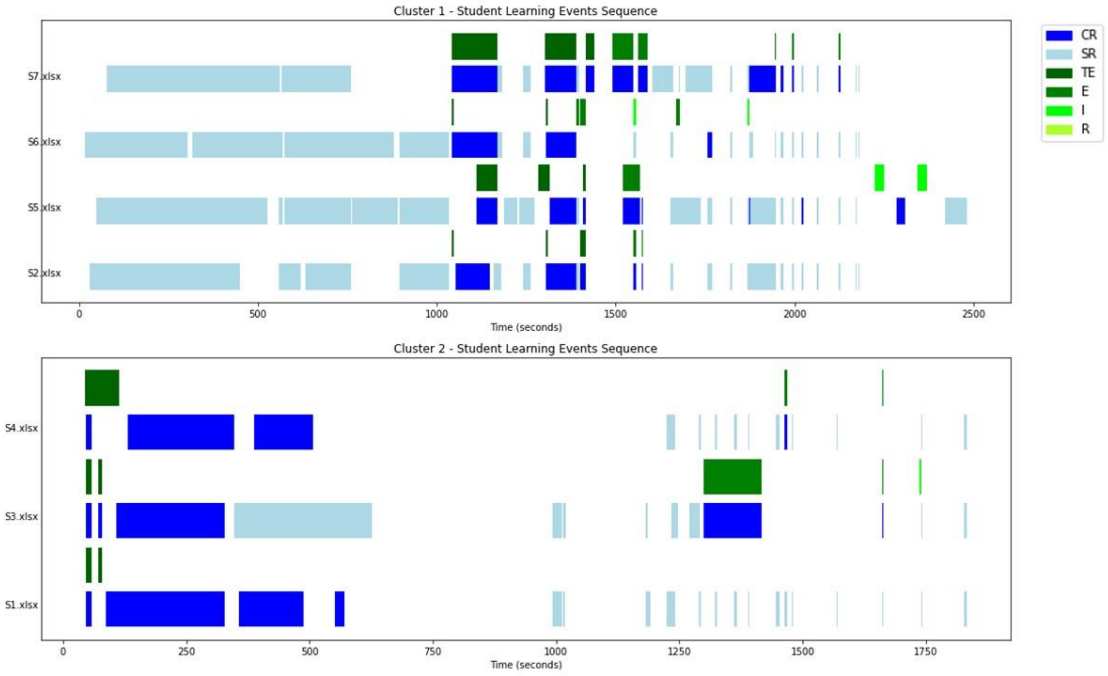

Figure 5. Different types of clusters' MCSA visualisation

From the perspective of cognitive presence, R and I encoding events are of a higher order compared to TE and E encoding events Sadaf and Olesova (2017). The first category of students exhibited a higher frequency of I events. From this perspective, it can be argued that the learning of the first category of students involves a higher order of cognitive presence. This structural difference aligns with the two distinct patterns exhibited in Figure 6 of the ENA analysis and Table 1.

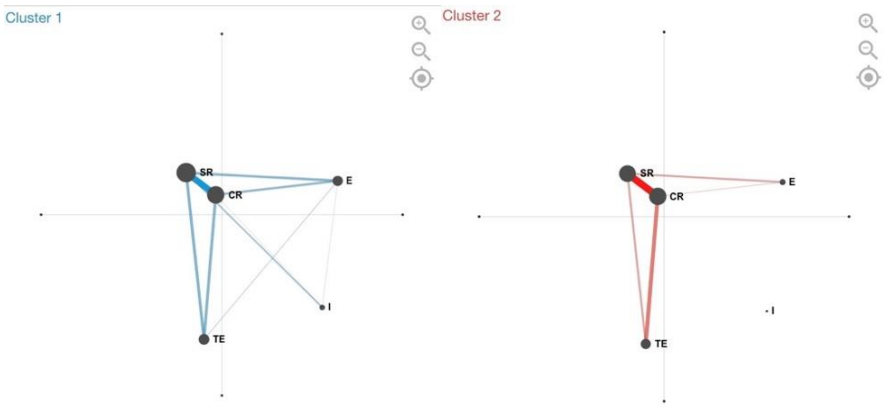

Figure 6. ENA network of two types of patterns

Table 1. *Descriptive Analysis of Coded Events Across Different Clusters*

	Cluster1 (n=4)		Cluster2 (n=3)	
Code	Mean	SD	Mean	SD
CR	6.00	2.449	4.33	1.528
E	3.00	0.816	1.50	0.707

I	2	0.000		
SR	19.50	2.887	13.67	4.726
TE	3.00	0.816	1.67	0.577

From a status perspective, we have identified two distinct learning patterns. In the first cluster, the first state has a close distribution between TE, SR and CR, while the second state displays a larger amount of CR and I. In the second cluster, the dominant observation states in the first hidden state are E, I and CR. The second state shows a more I event and less CR happen.

In terms of state transitions, the first cluster demonstrates frequent transitions between states. The second cluster, on the other hand, exhibits a higher likelihood of transitioning from the first to the second state, while the reverse transition from the second to the first state is comparatively less probable. Cluster 1 shows a higher frequency of state transitions, indicating more active and frequent dynamic changes between hidden states. This transitional behavior is consistent with earlier discussions focused on sequence visualization and structural analysis. The asymmetry in state transitions in cluster 2 suggests the presence of a "primary" state.

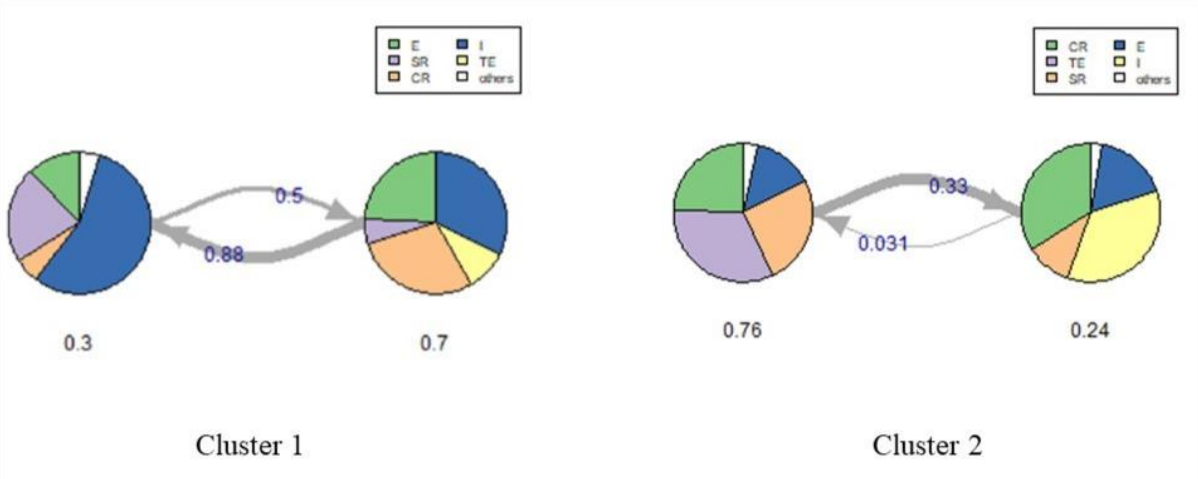

Figure 7. Hidden state description of two clusters' HMMs

4. Conclusion

In the current phase, our research developed a three-layer MMLA approach that combines eye-tracking with interactions involving LLM-supported digital humans for dual-channel analysis. Through a pilot study, we collected a limited dataset and have preliminarily identified two distinct patterns of students' cognitive presence.

Future work will focus on (1) integrating functional near-infrared spectroscopy (fNIRS) data to deepen the analysis of cognitive presence, (2) refining instructional designs based on the patterns identified in our study, (3) conducting a comprehensive experimental study to assess these enhancements, and (4) expanding data analysis techniques by incorporating additional data types into MMLA, such as student artefact assessments, and fine-tuning a LLM for automated scoring.

References

Du, X., Dai, M., Tang, H., Hung, J.-L., Li, H., & Zheng, J. (2023). A multimodal analysis of college students' collaborative problem solving in virtual experimentation activities: A perspective of cognitive load. Journal of Computing in Higher Education, 35(2), 272-295.

Gan, W., Qi, Z., Wu, J., & Lin, J. C.-W. (2023). Large language models in education: Vision and opportunities. 2023 IEEE international conference on big data (BigData),
Garrison, D. R. (2016). E-learning in the 21st century: A community of inquiry framework for research and practice. Routledge.
Garrison, R. (2000). Theoretical challenges for distance education in the 21st century: A shift from structural to transactional issues. International Review of Research in Open and Distributed Learning, 1(1), 1-17.
Hadwin, A., Järvelä, S., & Miller, M. (2017). Self-regulation, co-regulation, and shared regulation in collaborative learning environments. In Handbook of self-regulation of learning and performance (pp. 83-106). Routledge.
Järvelä, S., Järvenoja, H., Malmberg, J., & Hadwin, A. F. (2013). Exploring socially shared regulation in the context of collaboration. Journal of Cognitive Education and Psychology, 12(3), 267-286.
Kong, S.-C., & Yang, Y. (2024). A Human-Centred Learning and Teaching Framework Using Generative Artificial Intelligence for Self-Regulated Learning Development through Domain Knowledge Learning in K–12 Settings. IEEE Transactions on Learning Technologies.
Lin, M. P.-C., & Chang, D. (2023). CHAT-ACTS: A pedagogical framework for personalized chatbot to enhance active learning and self-regulated learning. Computers and Education: Artificial Intelligence, 5, 100167.
Lynch, C., Wahid, A., Tompson, J., Ding, T., Betker, J., Baruch, R., Armstrong, T., & Florence, P. (2023). Interactive language: Talking to robots in real time. IEEE Robotics and Automation Letters.
Mu, S., Cui, M., & Huang, X. (2020). Multimodal data fusion in learning analytics: A systematic review. Sensors, 20(23), 6856.
Ouyang, F., Xu, W., & Cukurova, M. (2023). An artificial intelligence-driven learning analytics method to examine the collaborative problem-solving process from the complex adaptive systems perspective. International Journal of Computer-Supported Collaborative Learning, 18(1), 39-66.
Ouyang, F., & Zhang, L. (2024). AI-driven learning analytics applications and tools in computer-supported collaborative learning: A systematic review. Educational Research Review, 44, 100616.
Sadaf, A., & Olesova, L. (2017). Enhancing cognitive presence in online case discussions with questions based on the practical inquiry model. American Journal of Distance Education, 31(1), 56-69.
Shea, P., Hayes, S., Vickers, J., Gozza-Cohen, M., Uzuner, S., Mehta, R., Valchova, A., & Rangan, P. (2010). A re-examination of the community of inquiry framework: Social network and content analysis. The Internet and Higher Education, 13(1-2), 10-21.
Shea, P., Richardson, J., & Swan, K. (2022). Building bridges to advance the community of inquiry framework for online learning. Educational Psychologist, 57(3), 148-161.
Song, Y., Jiaxin, C., Lei, T., & Gašević, D. (2023). A holistic visualisation solution to understanding multimodal data in an educational metaverse platform–Learningverse. International Conference on Computers in Education 2023,
Van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In International handbook of metacognition and learning technologies (pp. 143-156). Springer.
Zheng, L., Zhong, L., Niu, J., Long, M., & Zhao, J. (2021). Effects of personalized intervention on collaborative knowledge building, group performance, socially shared metacognitive regulation, and cognitive load in computer-supported collaborative learning. Educational Technology & Society, 24(3), 174-193.