Monitoring Cyber Peer-Led Team Learning: A Multimodal Human-in-the-loop Approach

Karen DSouza[1], Pratibha Varma-Nelson[2], Shiaofen Fang[1] and Snehasis Mukhopadhyay[1]

[1]*Department of Computer Science, Indiana University Purdue University Indianapolis, Indiana, USA*
[2]*Department of Chemistry, Indiana University Purdue University Indianapolis, Indiana, USA*

Abstract

The recent adoption of generative artificial intelligence (AI) tools in education has transformed education and AI-assisted learning. However, researchers embracing applied machine learning (ML) in pedagogical settings continue to face challenges. Lack of publicly available multimodal datasets involving gesture and emotion recognition specifically in education is a bottleneck for building AI-enabled e-learning platforms. In the paper, we take a constructive stance at monitoring group behavior in cyber peer-led team learning (cPLTL) classes in Organic Chemistry using ML. Although past studies have attempted to quantify student engagement in e-learning, their use in the cPLTL use case for AI modeling has not yet been established. The hypothesis underlying our proposed framework is that online peer group behavior can be characterized by a human-in-the-loop model that relies on multiple input modalities. Thus, the aim is to identify behavioral patterns in head and facial movements that are augmented by lexical based sentiment and audio feature extraction. To combat the small data challenge, we propose a framework for the human-in-the-loop (HITL) system that actively learns the past group modalities. HITL strategies enable the algorithm to learn more efficiently from less data iteratively. The model will be implemented using active learning, measures of uncertainty, random sampling and entropy which are key in the design of the study. A qualitative comparison of sentiment modality with ChatGPT's participant performance evaluation has been discussed. The study will increase the use of AI in tools that support educators in universities using pedagogies of active engagement in science, technology, engineering, and math.

Keywords

Multimodal AI, Active Learning, Human-in-the-loop, ChatGPT, Generative AI

1. Introduction

Generative large language models such as ChatGPT and AI tools such as chatbots in education have shifted the perception of educational outcomes viewed by educators, students, and policy makers. The sentiment towards using generative AI tools in education is viewed as positive and transformational [1]. Specifically, the shift is due to the potential of generative AI models to enhance the learning experience in classrooms and online settings. However, the widespread use of large language models in education raises issues in addressing bias and privacy that continue to grow.

HITLAML'23: Human-in-the-Loop Applied Machine Learning, September 04–06, 2023, Belval, Luxembourg
kmdsouza@purdue.edu (K. DSouza); pvn@iupui.edu (P. Varma-Nelson); shfang@iupui.edu (S. Fang); smukhopa@cs.iupui.edu (S. Mukhopadhyay)

Integrating AI into contemporary pedagogical tools such as Peer-Led Team Learning (PLTL) and Process Oriented Guided Inquiry Learning (POGIL) requires large volumes of data from classrooms [2] [3]. The scarcity of publicly trained datasets in education limits the scope of researchers engaged in improving the quality of education through AI-enabled tools. The problem is challenging because the lack of labeled multimodal data that incorporates the behavior of students poses additional limitations on finances. Labeling datasets has a high cost due to the manual labor involved in labeling data.

An adaptation of PLTL known as Cyber Peer-Led Team Learning (cPLTL) has moved group learning from a face-to-face setting to a synchronous online learning environment [4]. During the Covid-19 pandemic, educators migrated to the online learning modality using cPLTL to continue fostering science retention in universities [2]. cPLTL benefits such as high GPA and academic success have been evidenced qualitatively through statistical measures [5] [6]. cPLTL workshops are labeled good when the level of active peer participation and engagement is collaborative leading to problem solving [7]. By contrast, the communication, debate, and discussion between the students is low in a poor cPLTL workshop and these videos are given a low score.

Currently, instructors must rely on peer leaders for weekly peer group progress. The study aims to support educators by supplementing a method of feedback through AI. Two research questions are posed (i) We seek to develop a machine learning model to predict the quality of a cPLTL workshop using small scale data from video recordings. (ii) We attempt to use AI to identify patterns in multiple modalities – lexical, audio, head movements and facial expressions from online peer group learning.

The contributions of the paper are as follows. (i) A framework for the multimodal system with human feedback using the student engagement index in cPLTL has been developed. To the best of the authors' knowledge, no existing work has studied the integration of AI for peer feedback. (ii) A qualitative comparison of sentiment polarity from machine learning and ChatGPT has been presented. (iii) An active learning machine learning technique that will work well with limited quantity of data to fit the educational case study has been discussed.

2. Related Work

Student engagement can be studied by measuring physiological, behavioral, and cognitive factors. A recent engagement analysis using DAiSEE- an existing multi-label video dataset based on user affective states has achieved only 63.8% accuracy in predicting boredom levels from labeled engagement features [8][9]. The low accuracy reflects the complexity of predicting human behavior from videos. Another study performed real time simulation of learner engagement and classified it with 85% accuracy using physiological measurements such as respiration and heart rate [10]. Invasive methods involving EEG and eye tracking sensors have been used to determine student's motivation [11]. Other methods to monitor the emotional and psychological state of students include but are not limited to monitoring skin temperature, keyboard, and mouse response times.

On the other hand, non-invasive methods that track student engagement include wearable devices on the wrist or head that are powered by AI. Emotion recognition from physiological

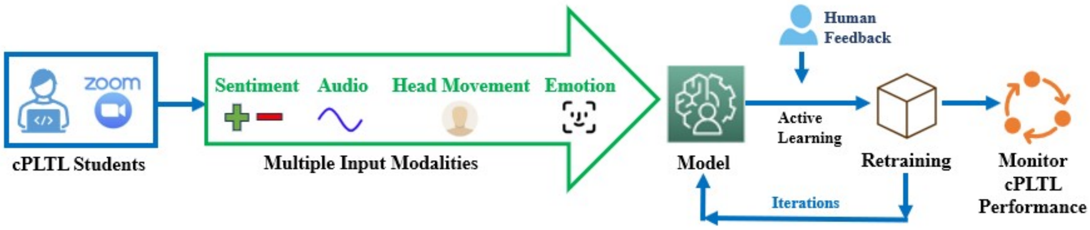

Figure 1: Monitoring cPLTL performance using multiple input modalities and continuous human feedback through active learning methodology.

and behavioral factors is an effective predictor of student engagement [12]. Another area of focus is on intelligent tutoring systems (ITS). These ITS systems characterize engagement tracing using an affective model that validates psychometric properties to quantify individual student response times to problem solving [13]. Researchers achieved a 92.58% accuracy from a real time e-learning environment using individual student's appearance and geometric-based cues [14]. In comparison, our use case focuses on online group learning, so individual student's cues in an individual learning environment cannot be considered. cPLTL is modeled around successful group interactions among peers.

Lastly, we use an active learning algorithm to implement HITL. Active learning algorithms are highly successful because they permit the learner to choose training samples. Thereby, the algorithm learns faster using less data [15]. In this paper, we extract modalities using a non-invasive technique that models the interplay between various modalities while preserving the relationships between them. In addition, we use active learning to model the HITL algorithm.

3. Method

To capture the student's engagement activity, we propose a metric called the student engagement index (SEI). The SEI value is defined as the aggregation of sentiment polarity, head gestures, facial expressions, and audio features. To calculate the SEI value, each modality is first extracted into a vector format. The complete system is displayed in Fig. 1. The privacy of the participants is maintained by removing all participant identifiers. Since the weekly videos are over two hours long, the original video is split into a shorter 15-minute window. This makes processing more appropriate to capture finer details in the sentiment variance and audio tones.

The first modality is the sentiment of the peer group conversation. We compared a sampling of peer group sentiment using Python's Sentiwordnet package with ChatGPT's qualitative output for the transcript. This is a two-step process. In the first step, for every 15-minute peer group transcript, a numerical sentiment score and polarity is populated from Sentiwordnet. Polarity is neutral if the generated sentiment score is between -0.5 to 0.5; positive for values greater than 0.5 and negative for sentiment less than -0.5. In the second step, the same 15-minute transcript is submitted to ChatGPT to evaluate the overall sentiment, tone of the conversation and engagement styles.

Table 1 shows the qualitative result of both steps for transcripts extracted from 7 groups.

Table 1
A Comparison of Sentiment Polarity from 7 groups in cPLTL Workshops Extracted from Machine Learning and ChatGPT's outputs relating to overall sentiment, participant engagement and tone.

Group	Sentiment	ChatGPT Sentiment	ChatGPT Participant Engagement Tone
1	Positive	Positive	Informative, friendly, collaborative
2	Negative	Confusion	Unsure seeking clarification, collaborative
3	Positive	Neutral	Collaborative, educational, constructive
4	Neutral	Neutral	Problem solving, informative, focused
5	Positive	Neutral	Discussing concepts, informative, analytical
6	Positive	Neutral	Discussing concepts, informative, explanatory

The results from ChatGPT's overall sentiment score, participant engagement and tone of the conversation produced more insights into the peer group learning. Also, the participant engagement and tone from ChatGPT together with the sentiment extracted from the ML algorithm is useful in determining the engagement value. In future, large language models such as ChatGPT could prove valuable in an education setting to support educators by providing feedback.

The second modality is the audio signal from the videos. Each 15-minute audio clip is synthesized into a spectral waveform using Python's standard Pyaudio and Librosa packages. The audio features are then extracted into individual numerical vectors that can be input into a multimodal neural network.

For the third modality, we designated capturing the student's raised head position to indicate attentiveness. When the student lowers his head, this action is captured as a disengagement. Computer vision algorithms are still not successful in identifying bent humans on video. So, the head movement modality is challenging to extract correctly. The small breakout room window on zoom does not sufficiently encapsulate the images that can be used appropriately in a neural network.

The fourth modality relies on dominant facial emotions. Seven standard facial characteristics – angry, happy, disgusted, fearful, neutral, sad, and surprised are extracted from each video. This is achieved by training the video dataset on a publicly available Facial Expression Recognition 2013 Dataset (FER2013) through a shallow convolutional neural network (CNN) to categorize the dominant emotion.

After processing all four modalities, the vector is input into an AI model that is powered by an active learning algorithm. Active learning is used to relabel the incorrect session samples in each iteration using three separate sampling methods. These are random sampling, lowest confidence, and maximum entropy. For random sampling, the training algorithm chooses a sample randomly with no prior history or likelihood of being the best sample [16]. However, in the lowest confidence and maximum entropy sampling techniques, the training algorithm queries samples to promote the samples with highest uncertainty based on their metrics of confidence or entropy [16]. The model is then retrained and the prediction accuracy of cPLTL scores is calculated.

During the HITL iterative training, qualitative feedback is input from the key decision makers. In our use case, the decision makers are the educators. They will improve the labeled score performance of cPLTL sessions. We propose the use of an average labeling score in each iteration. The average labeling score is the average of any two scores provided by the educators. Collecting new labels from at least two educators who are subject matter experts will reduce bias towards either end of the session score which ranges from 1 to 5. The feedback will be collected through each iteration of the active learning method in a simulation run. The multimodal HITL system will improve the score predictions and thereby the performance of cPLTL sessions.

4. Conclusion

The HITL system incorporates educator's feedback in a cPLTL workshop through ML. Guided human feedback improves cPLTL quality and thereby enhances the effectiveness of future cPLTL workshops. The unique contributions of the paper include the HITL framework for incorporating multiple modalities in a cPLTL education setting while using a student engagement index. We presented a qualitative comparison of sentiment modality with ChatGPT's outcomes from participant engagement tones and sentiment. The productive insight from the comparison opens research focused on enhancing education tools using large language models. Future work will incorporate generative AI techniques such as summarization of transcripts to produce an improved lexical modality. Lastly, the use of active learning in cPLTL datasets has not been attempted before. Our method discusses the adoption of randomness, uncertainty, and entropy metrics in the iterative modeling process. The proposed AI-backed model will be valuable in domains such as healthcare and training that rely on peer-to-peer motivated learning.

References

[1] Y. K. Dwivedi, N. Kshetri, L. Hughes, E. L. Slade, A. Jeyaraj, A. K. Kar, A. M. Baabdullah, A. Koohang, V. Raghavan, M. Ahuja, et al., "so what if chatgpt wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational ai for research, practice and policy, International Journal of Information Management 71 (2023) 102642.

[2] C. Servin, M. Pagel, E. Webb, An authentic peer-led team learning program for community colleges: A recruitment, retention, and completion instrument for face-to-face and online modality, in: Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 1, 2023, pp. 736–742.

[3] R. S. Moog, J. N. Spencer, Pogil: An overview, in: Process Oriented Guided Inquiry Learning (POGIL), ACS Publications, 2008, pp. 1–13.

[4] K. Mauser, J. Sours, J. Banks, J. Newbrough, T. Janke, L. Shuck, L. Zhu, G. Ammerman, P. Varma-Nelson, Cyber peer-led team learning (cpltl): Development and implementation, EDUCAUSE Review Online 34 (2011) 1–17.

[5] J. Y. Chan, C. F. Bauer, Effect of peer-led team learning (pltl) on student achievement, attitude, and self-concept in college general chemistry in randomized and quasi experimental designs, Journal of Research in Science Teaching 52 (2015) 319–346.

[6] S. P. Shields, M. C. Hogrebe, W. M. Spees, L. B. Handlin, G. P. Noelken, J. M. Riley, R. F.

Frey, A transition program for underprepared students in general chemistry: Diagnosis, implementation, and evaluation, Journal of Chemical Education 89 (2012) 995–1000.
[7] J. Smith, S. B. Wilson, J. Banks, L. Zhu, P. Varma-Nelson, Replicating peer-led team learning in cyberspace: Research, opportunities, and challenges, Journal of Research in Science Teaching 51 (2014) 714–740.
[8] A. Gupta, A. D'Cunha, K. Awasthi, V. Balasubramanian, Daisee: Towards user engagement recognition in the wild, arXiv preprint arXiv:1609.01885 (2016).
[9] N. Solanki, S. Mandal, Engagement analysis using daisee dataset, in: 2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV), IEEE, 2022, pp. 223–228.
[10] M. Carroll, M. Ruble, M. Dranias, S. Rebensky, M. Chaparro, J. Chiang, B. Winslow, Automatic detection of learner engagement using machine learning and wearable sensors, Journal of Behavioral and Brain Science 10 (2020) 165–178.
[11] R. Wang, L. Chen, A. Ayesh, Multimodal motivation modelling and computing towards motivationally intelligent e-learning systems, CCF Transactions on Pervasive Computing and Interaction 5 (2023) 64–81.
[12] M. Bustos-López, N. Cruz-Ramírez, A. Guerra-Hernández, L. N. Sánchez-Morales, N. A. Cruz-Ramos, G. Alor-Hernández, Wearables for engagement detection in learning environments: A review, Biosensors 12 (2022) 509.
[13] E. Joseph, Engagement tracing: using response times to model student disengagement, Artificial intelligence in education: Supporting learning through intelligent and socially informed technology 125 (2005) 88.
[14] S. Gupta, P. Kumar, R. Tekchandani, A multimodal facial cues based engagement detection system in e-learning context using deep learning approach, Multimedia Tools and Applications (2023) 1–27.
[15] B. Settles, Active learning literature survey, Computer Sciences Technical Report 1648 (2009).
[16] Y. Fu, X. Zhu, B. Li, A survey on instance selection for active learning, Knowledge and information systems 35 (2013) 249–283.