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

Is Internal State Feedback in an E-Learning Environment Acceptable to People?

Atsushi ASHIDA[a*], Ryosuke KAWAMURA[b], Shizuka SHIRAI[a], Noriko TAKEMURA[c], Mehrasa ALIZADEH[d], Hideaki HAYASHI[a] & Hajime NAGAHARA[a]
[a]*D3 Center, Osaka University, Japan*
[b]*Fujitsu Research of America, Inc., United States of America*
[c]*Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, Japan*
[d]*Institute for General Education, Otemon Gakuin University, Japan*
*ashida@ids.osaka-u.ac.jp

Abstract: In on-demand e-learning environments, the lack of direct intervention can lead to a decline in learners' engagement. To address this issue, systems that estimate the learners' attitudes and provide feedback have been proposed. However, the acceptability of such systems has not been sufficiently researched. In this study, we investigated the acceptability by people to an e-learning system with internal state feedback, for future personalized learning support. To this end, we developed a system that estimates and visualizes the learner's internal state in real-time. The system was exhibited in a public space for free use, and users' impressions were analyzed. To estimate the learners' internal state, we developed a machine-learning model that recognizes learners' alertness from facial videos. The system was deployed in an exhibition space, and 131 responses were collected. These responses were coded and analyzed using a co-occurrence network. The result indicated that learners tend to dislike the system due to feelings of being observed by supervisors. In contrast, instructors expressed favorable options toward the introduction of the system.

Keywords: Video-based-learning, e-learning, engagement, alertness, feedback

1. Introduction

In e-learning environments, maintaining and enhancing engagement is both important and challenging. Engagement, defined as the degree of active participation by learners, is a critical determinant of academic outcomes. In video-based e-learning environments, instructors are unable to observe the learners' attitudes, making it difficult to intervene when engagement declines. Research has indicated that e-learning environments are associated with higher dropout rates compared to traditional face-to-face classes (Tello, 2008).

Several efforts have been made to measure engagement in e-learning environments as an alternative to direct observation. One approach is to use log data of interactions in an e-learning system. Beck et al. (2005) estimated engagement using quizzes and applying item response theory, considering factors such as problem difficulty, response time, and response accuracy. Various studies have employed text activities to measure learners' engagement (Toti et al. 2021; D'mello et al. 2008). However, such methods are not applicable in scenarios where learners primarily watch instructional videos without interactive components.

Other approaches involve the use of biometric data. Nomura et al. (2018) measured temporal changes in seating pressure as an indicator of engagement. Arakawa and Yakura (2019) proposed a method to detect unconscious behaviors using multimodal inputs combined with anomaly detection techniques. Shinohara et al. (2023) estimate learners' emotional concealment during e-learning using biometric data and facial features. Although these technologies can measure engagement effectively, they require specialized equipment. As an alternative, deep learning and computer vision techniques have been proposed to estimate learner engagement using only facial images and videos (Murshed et al. 2019). However, it remains unclear whether these methods are truly accepted by learners and educators. These systems involve estimation and feedback of learner's internal states, such as alertness and emotions, to measure and enhance engagement. This could be perceived as an intrusive form of machine surveillance, potentially causing aversion among some users.

This paper explores user acceptance of internal state feedback within a video-based e-learning system. To achieve this objective, we developed a system that estimates the learners' internal states from facial video images captured during video lectures and provides this information as feedback to the learner. The system was deployed in an exhibition space for public use. We analyzed feedback comments collected from the public after the deployment to assess the level of acceptance for such technology.

2. Alertness Estimation System

2.1 System Overview

We developed an alertness estimation system that can visualize learners' alertness in real-time while they are watching video classes. This system aims to make learners aware of their own decline in alertness and encourage them to improve their engagement.

An overview of the developed system is shown in Figure 1. The system is implemented as a web application and consists of client and server computers. The client is a laptop computer that performs facial video capturing, presents learning contents, and visualizes alertness. The server computer runs as a back end in a location not directly accessible to the learner and estimates the alertness by processing captured facial videos. The learner accesses the system page on the client to watch the learning content. While the learner is watching a video lecture, the client captures the learner's facial video and transmits it to the server. The server estimates the learner's alertness levels from the facial videos. The estimated alertness levels are sent back to the client. The client generates real-time visualizations based on the received alertness data to provide feedback to the learner.

2.2 Alertness Estimation Model

The proposed model focuses on behavioral engagement in learning (Fredricks et al., 2004) and recognizes the learner's alertness. Let $S_{i:i+n}$ denote video segments from the i-th to $(i+n)$-th frames and $y_{i:i+n}$be the alertness state during the period from the i-th t $(i+n)$-th frames. In the recognition task, the alertness state during a period of the input sequence is the target of prediction. Therefore, we build a model that takes $S_{i:i+n}$ as input and predicts $y_{i:i+n}$. In our method, we set n to 10, corresponding to one second of the sequence. The overview of our model is shown in Figure 2. The model consists of a convolutional neural network (CNN) for image feature extraction, a bidirectional gated recurrent unit (BiGRU) for learning temporal dependencies, and a fully connected (FC) layer for alertness recognition. The features extracted by the CNN-BiGRU are input into two FC layers to classify the alertness state.

The CNN-BiGRU module extracts video features with one-second windows to recognize the alertness state for each second. We employed Efficientnet-B3 whose weights are initialized by ImageNet for the CNN-BiGRU module (Tan & Le, 2019). Finally, our model outputs the probability of the alertness state. The information from facial images is effective for recognizing the current alertness state. In our work, we focus on eye regions for alertness state prediction. The squared region around the right eye is cropped based on five facial landmarks detected by Dlib.

We collected learners' facial images during e-learning to evaluate the accuracy of our model and confirm the effectiveness of alertness recognition. We used data from 53 undergraduate students, obtained with informed consent, while they were watching video lectures (Kawamura et al., 2018). After the experiment, students' alertness levels were coded

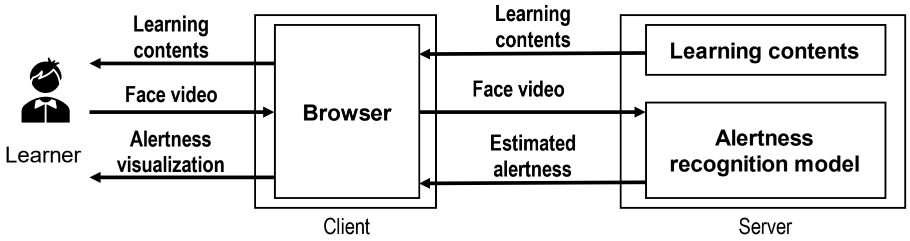

Figure 1. The overview of the alertness estimation system.

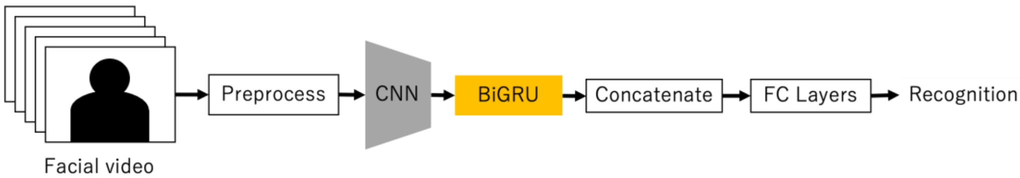

Figure 2. Overview of the alertness estimation model.

as binary values. The students who did not fall asleep were excluded our experiment. As a result, out of the 53 students, the data from 35 students who fell asleep during the experiment at least once were selected for inclusion in our database. We selected four students from our database who exhibited consistent sleeping patterns, using data from one as the test dataset and data from the remaining 34 students as the training data. This process was repeated until all four selected students had been evaluated. To address the influence of class imbalance on model performance, we randomly sampled 1,000 instances from each class in the training dataset. Our model achieved an average F1 macro score of 0.89.

2.3 Alertness Visualization

The user interface of the client system is shown in Figure 3. In the class video section, the e-learning class video is displayed. In the facial video section, the video captured by the built-in camera of the client computer is displayed, allowing the learner to watch the facial video used to estimate alertness. In the alertness visualization section, alertness is visualized with a graph where the horizontal axis represents the number of estimations, and the vertical axis represents alertness levels. An alertness estimation is attempted every second, and upon a successful estimation, the graph is updated by adding a new alertness level on the right side. If an estimation fails, the previous prediction is carried over. Using this interface, learners can confirm their estimated alertness levels, allowing them to objectively observe their own learning attitude during the lecture.

3. Methodology

3.1 Data Collection

The alertness estimation system has been deployed in *The Lab*, an exhibition facility located within Grand Front Osaka, a commercial complex in Osaka, Japan. The purpose of this facility is to allow the public to experience the system and to raise awareness of university research. Given the diverse types of visitors, five different class video contents are prepared, each lasting about 10 minutes. Visitors can select one of these videos and freely experience the system. The facility conducts a free-response survey. This research analyzes the responses from the

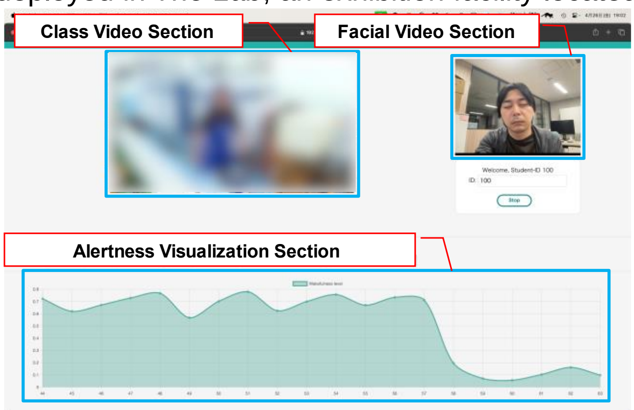

Figure 3. The system interface.

surveys to investigate the acceptability of the developed system in educational settings.

The data collection period spanned from August 2022 to January 2024, and 131 responses were obtained. The collected data includes age, gender, and comments containing impressions and opinions on the system. The distribution of the respondents is shown in Table 1. The age range of respondents spans from under 10 years old to those in their 70s. Notably, people in their 20s, 30s, and 40s were the majority, accounting for 70% of the respondents. While the demographic data indicates a skewed number of respondents in each age group, it also shows that responses were obtained from a wide range of demographics.

3.2 Data Analysis

To analyze the comments collected in the survey, codes were individually assigned to each free-text comment. The list of codes, along with their corresponding quantities and meanings, is presented in Table 2. This table includes only the codes with more than five responses. Open coding was used, where the types of codes are not predetermined. Multiple codes were allowed for a single response. Responses not related to the system, e.g., comments about the experimental environment and computer specifications, were excluded. The assignment of the codes was performed by the first author of this paper based on the careful review of the content. After the initial coding was completed, the assigned codes were reviewed, merged if they represented the same meaning, or rewritten to adjust granularity.

4. Results and Discussion

4.1 Co-occurrence Network

As most responses have multiple codes, we analyzed the co-occurrence of codes in a response using a network to clarify relationships between codes assigned to each opinion and determine prevailing topics in the obtained opinions. Figure 4 shows the created co-occurrence network. In the figure, the nodes represent codes. The size of the nodes

Table 1. The distribution of ages of respondents.

Age group	Total (%)	Males (%)	Females (%)
Less than 20	10 (7.63)	5 (3.82)	5 (3.82)
20-29	35 (26.72)	24 (18.32)	11 (8.40)
30-39	33 (25.19)	18 (13.74)	15 (11.45)
40-49	29 (22.14)	16 (12.21)	13 (9.92)
50 and older	24 (18.32)	15 (11.44)	9 (6.87)
Total (%)	131 (100)	78 (59.54)	53 (40.46)

Table 2. The codes assigned to the survey responses.

Code	Number of responses	Code meanings
User	50	Opinions from learners' perspective
Negative	38	Negative opinions about the system
Positive	35	Positive opinions about the system
Supervisor	31	Opinions from educators' perspective
Application	27	Opinions on where to apply the technology
Algorithm	14	Opinions about methods of the feedback
Usability	11	Opinions about the usability of the system
Feedback	10	Opinions about the alertness visualization
Parent	9	Opinions from parents' perspective
Experience	8	Opinions about one's experience
Interface	7	Opinions about user interface
Mismatch	5	Opinions about discomfort with feedback

corresponds to the frequency of occurrence of the codes, as indicated by the number of responses in Table 2. The thickness of the links represents the number of co-occurrences of codes on the connected edges, i.e., the strength of the co-occurrence relationships. In Figure 4, links are drawn between pairs of codes that co-occurred two or more times.

From Figure 4, it is evident that there is a strong relationship between "User" and "Negative," as well as between "Supervisor" and "Positive." This indicates that learners who use the system tend to have negative opinions, whereas educators tend to have positive opinions.

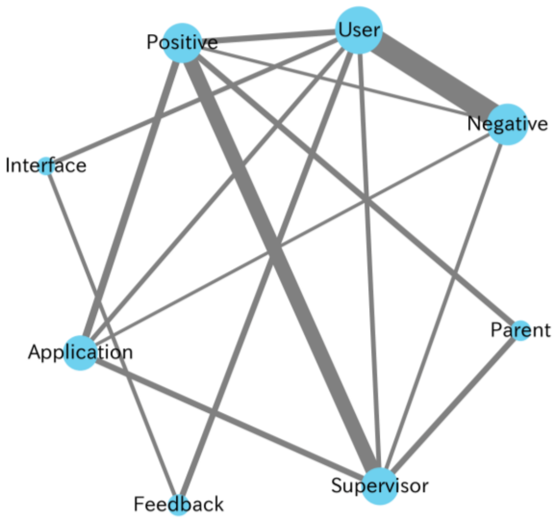

Figure 4. Created co-occurrence network.

4.2 Specific Comments that Coded "User-Negative" and "Supervisor-Positive"

To clarify the acceptance of the developed system, we also analyzed which aspects of the system are seen negatively from the learners' perspective and positively from the supervisors' perspective, based on the comments to which each code was assigned. There were 35 responses coded as "User-Negative." Representative comments included: "I feel scared to be monitored all the time" and "I would like to measure my own level of concentration personally, but I don't like the idea of comparing my concentration with others." These comments indicate a sense of dislike towards the implementation of the system. There were also more specific comments such as "If this were implemented at a company and if the estimated alertness levels were significantly low, would it lead to a pay cut? If that's the case, I would dislike it." These comments suggested that they felt as if their learning attitudes are being monitored or judged by someone else, leading to their aversion. However, it has also been reported that being observed or receiving attention can improve performance (Cottrell et al., 1968). From this, the feeling of being monitored is not necessarily disadvantageous. Negative comments from learners indicate potential concerns about their learning attitude. In the field of learning support systems, excessive intervention can lead to a lack of agency for learners (Koedinger et al., 2007). In the developed system, constant real-time feedback might have caused annoyance or a feeling of being always monitored due to excessive feedback. Moving forward, it will be necessary to clarify the appropriate feedback methods that maintain high learner engagement. This could involve comparing the current system with alternatives that provide feedback only when concentration decreases.

The "Supervisor-Positive" code was assigned to 22 responses. A common trend in the comments was the desire to use information about learners' alertness as a guide for instructional improvement. For example, comments like "It's good for teachers to know if learners are sleeping or bored," and "It would be good if the teacher could look at this data and explain again in the next class the parts where everyone was not able to concentrate," were received. Additionally, there were comments from a parent's perspective, such as "It's good. If a child's concentration drops, you can tell them to stop." These comments showed a favorable view of knowing the alertness levels.

5. Conclusion

In this study, to investigate whether internal state feedback in an e-learning environment is acceptable to people, we developed a system that can visualize alertness in real-time during on-demand lectures. The system was deployed in a public space, and opinions were collected from people who experienced it. We analyzed the collected responses. From the perspective of learners, opinions were obtained expressing a dislike for the system, particularly the feeling of being monitored by supervisors. This suggests that compared to situations without feedback, it is possible to induce changes in learners' attitudes toward learning, but the system's feedback may have been excessive. From the instructors' perspective, many favorable opinions were received about implementing this system in their teaching. Particularly, it was found that the alertness levels estimated by this system could serve as a valuable source of information for improving teaching.

In future work, opinions such as feeling monitored or being distracted by the changing graph were also obtained from the learners' perspective. It is necessary to clarify the most effective feedback methods by varying the frequency and approach of feedback to enhance learning outcomes. Since it was suggested that the alertness levels of learners might be able to be an element of improvements in teaching, developing a dashboard that can display the alertness information of learners is desired to facilitate teaching improvements.

Acknowledgements

This paper is partially supported by Innovation Platform for Society 5.0 from Japan Ministry of Education, Culture, Sports, Science and Technology, and JSPS KAKENHI Grant Number 24K16753.

References

Tello, S. F. (2008). An analysis of student persistence in online education. In *Information communication technologies: Concepts, methodologies, tools, and applications* (pp. 1163–1178). IGI Global.

Beck, E. J. (2005). Engagement tracing: using response times to model student disengagement. *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology*, *125*, 88–95.

Toti, D., Capuano, N., Campos, F., Dantas, M., Neves, F., & Caballé, S. (2021). Detection of student engagement in e-learning systems based on semantic analysis and machine learning. In *Advances on P2P, Parallel, Grid, Cloud and Internet Computing: Proceedings of the 15th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing, 15,* 211–223.

D'mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., & Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User modeling and user-adapted interaction*, *18*, 45-80.

Nomura, K., Iwata, M., Augereau, O., & Kise, K. (2018). Estimation of Student's Engagement Using a Smart Chair. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, 186–189.

Arakawa, R., & Yakura, H. (2019). REsCUE: A framework for REal-time feedback on behavioral CUEs using multimodal anomaly detection. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 572–585.

Shinohara, K., Muramatsu, K., & Matsui, T. (2023). Development of Estimation Method for Learner's Emotional Concealment During Learning Using Biometric Information and Feedback Model. *Proceedings of the 31st International Conference on Computers in Education,* 132–134.

Murshed, M., Dewan, M. A. A., Lin, F., & Wen, D. (2019). Engagement detection in e-learning environments using convolutional neural networks. In *2019 IEEE International Conference on Dependable, Autonomic and Secure Computing, International Conference on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, International Conference on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, 80–86.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74(1), 59–109.

Kawamura, R., Shirai, S., Takemura, N., Alizadeh, M., Cukurova, M., Takemura, H., & Nagahara, H. (2021). Detecting drowsy learners at the wheel of e-learning platforms with multimodal learning analytics. *IEEE Access*, *9*, 115165–115174.

Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning, 6105–6114.

Knowledge capital, The Lab., https://kc-i.jp/en/facilities/the-lab/ (Last accessed: 2024/04/13)

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. (1968). Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of personality and social psychology*, 9(3), 245.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*, 239–264.