IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol. 003, LIIR064
DOI: https://doi.org/10.52731/liir.v003.064

A Real-time Engagement Assessment for Learner in Asynchronous Distance Learning

Shofiyati Nur Karimah [*] , Shinobu Hasegawa [†]

Abstract

Early notice of a learner's disengagement during learning is an essential signal for an educator to either change the pedagogy or give personal support. Hence, assessing learner's engagement is important to avoid dropout. However, assessing learner's engagement in distance learning is a challenge due to the learner-educator interaction limit. To address the challenge, in this paper we proposed a real-time automatic engagement estimation system to assess learner's engagement from facial landmarks and body pose during his/her learning activity in asynchronous distance learning, where there is no direct interaction between learner and educator. A web-based application has been developed as an early stage implementation in a real education setting. The prototype has been successfully recognizing learner's engagement in three-level: Very Engaged, Normal Engaged, and Not Engaged.

Keywords: Engagement estimation, face detection, asynchronous distance learning, web-based application.

1 Introduction

Engagement is a complex human behavior state which includes affective, cognitive, and behavioral factors [1]. In educational point of view, engagement as an inner state was classified into three types, i.e, emotional, behavioral, and cognitive engagements [2]. Emotional engagement is related to affective revealing emotion during learning, e.g., boredom, interest, happiness, and sadness. Whereas behavioral engagement is shown by learners' active participation in a class or tasks, and cognitive engagement refers to self-regulation and psychological investment in learning.

Regardless the learning settings, measuring learner's engagement is essential in a learning process to increase productivity and learning outcome, as well as provoke insight for personalized support, and reduce dropout [3, 4]. In traditional classroom settings, measuring engagement can be done directly by the educator during the class or using evaluation

[*] Graduate School of Advanced Science, Japan Advanced Institute of Science and Technology (JAIST)
[†] The Center for Innovative Distance Education and Research, Japan Advanced Institute of Science and Technology (JAIST)

check with rating scales. Meanwhile in distance learning setting, measuring engagement is more challenging since there is a limitation in leaner-educator interaction.

Based on learner-educator interaction, distance learning can be distinguished into two categories, i.e., synchronous and asynchronous learning. Synchronous learning enables learners and educators to communicate directly through online classroom or personal tutoring through video conference applications such as Zoom, Cisco Webex, Google Meet, etc. Meanwhile, in asynchronous learning, learners mostly interact with learning content in virtual learning environments instead of with the educator, for example, in learning through massive open online courses (MOOCs) or learning management systems (LMS) such as Moodle.

In this paper, we address the problem of assessing learner engagement in an asynchronous distance learning setting. We focus on automatically estimating emotional engagement because we use appearance-based cues, i.e., facial landmarks and body pose, to measure engagement. We proposed a web-based application for a real-time engagement assessment, where face detection and engagement model were utilized, to estimate the learner's engagement state in three-level: very engaged, normal engaged, and not engaged. Real-time estimation is important to understand learners' emotional engagement during their interaction with the learning materials in MOOCs or LMS. By recognizing learners' engagement state real-time with their activity with the learning material, an educator can give feedback to the learners who showed any sign of disengagement by giving personalized support or modifying the learning-content delivery.

2 Related Work

Several appearance-based automatic engagement estimations have been proposed [6, 7, 8, 9, 10, 11]. Appearance-based automatic engagement estimation means extracting visual traits (e.g., facial expression, eye gaze, and body pose) captured in a video and analysed using computer vision analysis, particularly machine-learning algorithms.

Shen et al. [6] assessed learning engagement based on facial expression recognition in a MOOC environment. They utilized the images from publicly available facial expression datasets such as JAFFE, CK+, and RAFDB to be estimated using a lightweight attentional convolutional network. Before the fed to the network for prediction, Kernel Maximum Mean Discrepancies (MK-MMD) was used to calculate the distribution between the extracted features. The Four-class engagement was predicted with 56% of accuracy.

Similarly, Sumer et al. [7] were also using face features and head poses shown in videos and estimated using several machine learning algorithms including support vector machine (SVM), random forest (RF), multi-layer perceptron (MLP), and long short-term memory (LSTM). Before the estimation, RetinaFace was used for face detection, and fine-tuning with AffectNet and Attention-Net (300W-LP) was done to enhance the prediction of three-class engagement classification.

In human-robot interaction, Youssef et al. [8] analyzed the image captured from a robot's camera and extract the information of the head, gaze, and face stream. Logistic regression (LR), linear discriminant analysis (LDA), RF, and MLP were used for binary classification.

Also using publicly available datasets, Liao et al. [10] used DAiSEE [12] and EmotiW [13] datasets to develop an automatic engagement dataset. MTCNN was used for face detection and to estimate the engagement, they proposed a deep facial spatiotemporal network

(DFSTN) which is a combination network: pre-trained SE-ResNet-50 for extracting facial spatial features, and LSTM Network with global attention (GALN).

Despite the massive development of automatic engagement estimation, there is an implementation gap between computer science studies and distance learning practices. The recent automatic engagement estimation module cannot give immediate impact in distance learning process, especially with the lack of technologically savvy environment for educators and education managers to interpret the report. Therefore, this study proposes a system for real-time automatic engagement estimation implementation, especially, in asynchronous distance learning.

Unlike in synchronous distance learning, where the educator can visually observe learner engagement, learner is mostly alone in asynchronous distance learning setting. Therefore, real-time automatic engagement assessment not only benefits the educators to adjust their teaching strategy the way they do in a traditional classroom (e.g., by suggesting some useful reading materials or changing the course contents [14]), but also for self-monitoring by the learner her/himself.

3 Methodology

In this section, the problem definition of assessing learners' engagement in an asynchronous distance learning setting, such as learners majorly using LMS, was described. The architecture of the proposed system to overcome the problem will also be explained.

3.1 Problem definition

A learning management system (LMS), such as Moodle[1], is one practical example of distance learning, both synchronous and asynchronous. The LMS shaped the face of e-learning nowadays since it facilitates many essential educational activities including managing enrollments, creating learning plans, delivering learning content, and grading works in one platform.

While synchronous LMS may offer flipped classroom, in fully asynchronous LMS, educators normally cannot see the learners' faces during their interaction with learning material, and thus measuring their engagement is difficult other than by checking their cognitive activity, e.g., from certain tasks, assignments, or exam score. However, the cognitive result cannot guarantee if the learners are actually engaged. Besides, due to lack of visibility, the educator cannot check if the learner did the assignment or exam in the LMS by themselves, or whether the learners are struggling to understand the learning material. Therefore, we believe that additional visual information about learners would be beneficial in recognising learners' emotional engagement in asynchronous learning.

Visual analysis through facial recognition is suitable to assess non-verbal behaviors without interrupting the learning process. However, unlike in synchronous distance learning settings, the learners' visuals in asynchronous distance learning might not be accessible. Not to mention the limited bandwidth can be an additional problem for multimedia streaming. Alternatively, a log file that contains log information of learners during their interaction with the learning materials in an LMS can be an option as shown in Figure 1, where the problems should be solved with the following system architecture.

[1] `https://moodle.org/`

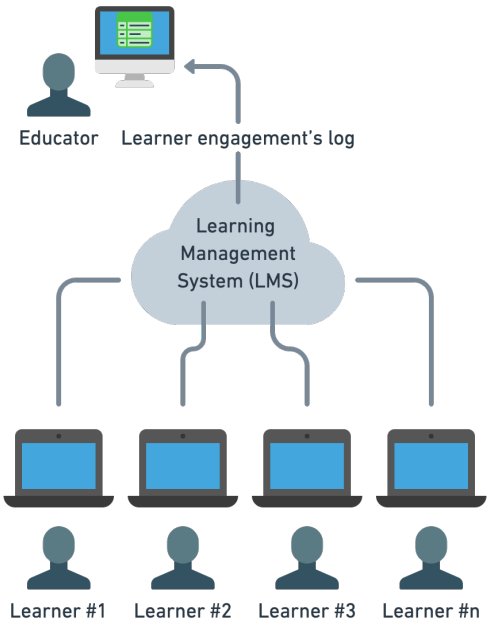


Figure 1: Asynchronous learning scenario.

3.2 System Architecture

From the aforementioned problem definition, we designed a real-time automatic engagement estimation system (Figure 2). The system was incorporated with three main steps: data collection, model training, and online implementation.

The data were collected from three poses representing three engagement states. In this work, we only focus on building the architecture and developing the prototype for real implementation. Therefore, we did not focus on the data collection or estimation performance in depth.

We simply defined the poses that represent the three engagement levels by obviously visible gestures, which are mainly based on the distance between the learner's face from the monitor. Very engaged means that the learner continuously faces the monitor closely. Similarly, normal engaged also shows the learner fully attention to the monitor only with more distance than the very engaged. Meanwhile, not engaged is to classify the learners when their faces were directing away from the monitor. In this work, we did not take into account more situations such as note-taking, in which the learner is concentrating by taking a note, thus the face is perceived as as not engaged or even cannot be captured by the camera.

The data were collected by using a face detector to extract the face, hand, and body landmarks, which were defined into three class labels for supervised learning. The collected features and classes were packed in an engagement dataset, namely called Engagement.csv. To build an automatic engagement estimation module, the dataset is trained in some classic machine learning models, such as logistic regression (LR), random forest (RF), and gradient

boosting (GB). The trained model aimed to classify three engagement levels: very engaged, normal engaged, and not engaged.

In this work, we experimented with three classical machine learning models, i.e., LR, RF, and GB. Due to the limited amount of dataset and fixed poses, the classification performance among the three models are the same, with the accuracy 1. However, we found that the accuracy of the RF was more stable than the other two to train our current data. Therefore, we used the RF trained model for the implementation in the system.

The trained model was saved and implemented in a web-based application that learners can access at the same time during their visit to an LMS. Therefore, their engagement was estimated in real-time. The engagement states were automatically recorded in a log file when the estimation process start to run analysing learners' face and body features. The log files were accessible for the educator to understand the engagement of the learners and give further feedback.

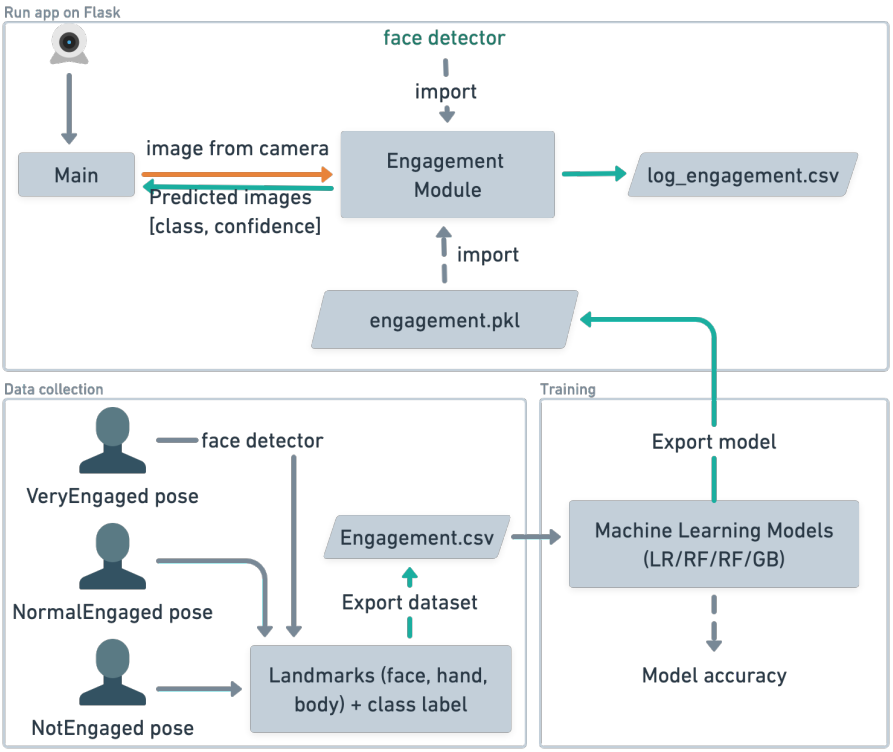


Figure 2: The proposed system architecture.

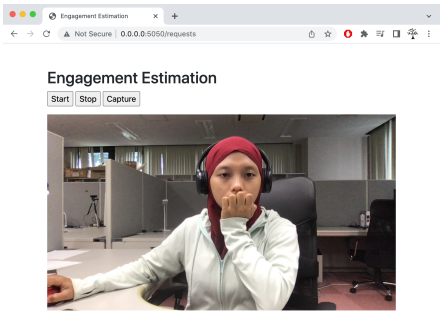
4 Prototype Development

The system was implemented in a web-based application with Python run on Flask. We used Mediapipe[2] for the face detector used in both data collection as well as the online running. Before building the web application for the real-time estimation, we also first built

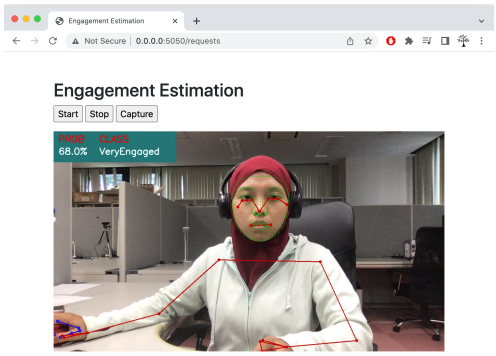
[2]https://google.github.io/mediapipe/

an application for the data collection. We used Mediapipe because it is a community-based open source work that offers several machine learning solutions including face detection, face mesh, iris, hands, pose, and holistic. Most importantly, it offers cross-platform, and customisable for live and streaming media, which is suitable for our current work.

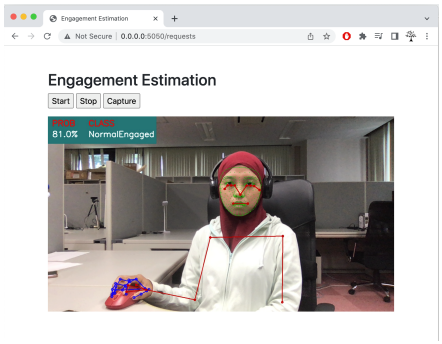
Figure 3 shows the running application in four states, i.e., the idle state, in which the estimation is not running, and the three engagement states. For an ethical reason, the estimation was not running immediately when the application first runs. Instead, we provided three buttons: start, stop, and capture, which gives the student free will to activate (start button) and deactivate (stop button) the engagement prediction. The capture button is an additional button in case images were needed for further analysis. As a reference, Figure 4 shows the screenshot of the engagement log file.



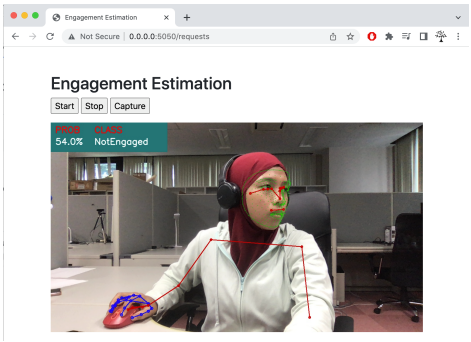
(a) Screen view without engagement estimation.



(b) A learner is in a very engaged state.



(c) A learner is in a normally engaged state.



(d) A learner is in not engaged state.

Figure 3: Web-based application of engagement estimation. (3a) is the screen in the first load or when the stop button was pressed, whereas (3b-3d) are the screen views showing the three engagement stages when the start button was pressed.

5 Conclusions and Discussion

In asynchronous distance learning, there is no direct interaction between learners and educators. Instead, the learners majorly interact with learning materials in a learning management system (LMS). To help educators understand their learners' engagement, we proposed a system that enables them to obtain a record of their learners' engagement with their consent. This paper presented an application of real-time automatic engagement estimation to assess learner's engagement in an asynchronous distance learning setting.

log_engagement-20220825_10

10:51:45	VeryEngaged	70.0%
10:51:46	VeryEngaged	68.0%
10:51:46	VeryEngaged	70.0%
10:51:46	VeryEngaged	70.0%
10:51:46	VeryEngaged	64.0%
10:51:47	VeryEngaged	70.0%
10:51:47	VeryEngaged	66.0%
10:51:47	VeryEngaged	68.0%
10:51:48	VeryEngaged	68.0%
10:51:48	VeryEngaged	75.0%
10:51:48	VeryEngaged	76.0%
10:51:49	VeryEngaged	75.0%
10:51:49	VeryEngaged	77.0%
10:51:49	VeryEngaged	67.0%
10:51:49	VeryEngaged	66.0%
10:51:50	NormalEngaged	82.0%
10:51:50	NormalEngaged	82.0%
10:51:50	NormalEngaged	87.0%
10:51:51	NormalEngaged	90.0%
10:51:51	NormalEngaged	84.0%
10:51:51	NormalEngaged	89.0%
10:51:52	NormalEngaged	86.0%
10:51:52	NormalEngaged	88.0%
10:51:52	NotEngaged	76.0%
10:51:52	NotEngaged	91.0%

Figure 4: A snapshot of the automatic engagement log in a CSV file.

5.1 Contributions and Findings

The prior automatic engagement estimation focuses on method and development from the computer science point of view, which cannot be immediately benefited in the education studies, especially distance learning settings. This study addresses the implementation gap between the automatic engagement estimation of computer science studies and distance learning studies.

We presented the problem definition (Figure 1) to give an overview of the scenario where learners, which mainly interact with the learning materials in LMS, have no direct interaction with the educator. Our proposed solution (Figure 2) enabled the educator to obtain engagement state logs of their student in a less-bandwidth-demand form.

Although the images and videos of learners will not be recorded, our proposed system analysed emotional engagement, where visual information of face and body were extracted. Therefore, we provided the application with start and stop buttons so they are aware when their faces will be analyzed (Figure 3).

5.2 Limitations and Future Works

The current dataset was limited in terms of number and collection environment (i.e., posed instead of spontaneous) since, in this work, we only focused on building the architecture and developing the prototype for real implementation. Likewise, the estimation performance was not in-depth discussed in this work.

For future works, the current prototype can be modified so that possible for online data collection. For this purpose, the engagement module can be deactivated, while activating the landmark recording when face and body landmarks information is intended. With more module modifications, the face capture also could be recorded for image data collection. The newly collected data can be further used to update/re-train the engagement model for better accuracy of estimation.

Our proposed system is useful to educators to know when their learner losing their engagement in general, whereas the reason for the disengagement was not be analysed. A more detailed report is a challenge for future works, which not only provide individual feedback but also provides feedback to the educators by putting them together. Furthermore, implementing the prototype in real LMS is suggested for real implementation in an educational learning process.

References

[1] J. Reeve and C.-M. Tseng, "Agency as a fourth aspect of students' engagement during learning activities," Contemporary Educational Psychology, 36(4):257–267, 10 2011. ISSN 0361476X. doi: 10.1016/j.cedpsych.2011.05.002. URL https: //linkinghub.elsevier.com/retrieve/pii/S0361476X11000191.

[2] J. A. Fredricks, P. C. Blumenfeld, and A. H. Paris, "School Engagement: Potential of the Concept, State of the Evidence," Review of Educational Research, 74(1):59–109, 3 2004. ISSN 0034-6543. doi: 10.3102/00346543074001059.

[3] K. L. Alexander, D. R. Entwisle, and C. S. Horsey, "From First Grade Forward: Early Foundations of High School Dropout," Sociology of Education, 70(2):87, 4 1997. ISSN 00380407. doi: 10.2307/2673158. URL https://www.jstor.org/stable/2673158?origin=crossref.

[4] H. Monkaresi, N. Bosch, R. A. Calvo, and S. K. D'Mello, "Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate," IEEE Transactions on Affective Computing, 8(1):15–28, 1 2017. ISSN 1949-3045. doi: 10.1109/TAFFC.2016.2515084. URL http://ieeexplore.ieee.org/document/7373578/.

[5] J. Whitehill, Z. Serpell, Y. C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagement from facial expressions," IEEE Transactions on Affective Computing, 5(1):86–98, 2014. ISSN 19493045. doi: 10.1109/TAFFC.2014.2316163.

[6] J. Shen, H. Yang, J. Li, and Z. Cheng, "Assessing learning engagement based on facial expression recognition in MOOC's scenario," Multimedia Systems, 28(2):469–478, 4 2022. ISSN 0942-4962. doi: 10.1007/s00530-021-00854-x.

[7] O. Sumer, P. Goldberg, S. D'Mello, P. Gerjets, U. Trautwein, and E. Kasneci, "Multimodal Engagement Analysis from Facial Videos in the Classroom," IEEE Transactions on Affective Computing, 2021. ISSN 19493045. doi: 10.1109/TAFFC.2021.3127692. URL https://ieeexplore.ieee.org/document/9613750.

[8] A. Ben-Youssef, C. Clavel, and S. Essid, "Early Detection of User Engagement Breakdown in Spontaneous Human-Humanoid Interaction," IEEE Transactions on Affective Computing, 12(3):776–787, 7 2021. ISSN 1949-3045. doi: 10.1109/TAFFC.2019.2898399. URL https://ieeexplore.ieee.org/document/8637822.

[9] P. Goldberg, Sumer, K. Sturmer, W. Wagner, R. Gollner, P. Gerjets, E. Kasneci, and U. Trautwein, "Attentive or Not? Toward a Machine Learning Approach to Assessing Students' Visible Engagement in Classroom Instruction," Educational Psychology Review, 33(1):27–49, 3 2021. ISSN 1040-726X. doi: 10.1007/s10648-019-09514-z. URL https://link.springer.com/article/10.1007/s10648-019-09514-z#citeas.

[10] J. Liao, Y. Liang, and J. Pan, "Deep facial spatiotemporal network for engagement prediction in online learning," Applied Intelligence, 51(10):6609–6621, 10 2021. ISSN 0924-669X. doi: 10.1007/s10489-020-02139-8. URL https://link.springer.com/10.1007/s10489-020-02139-8.

[11] W. H. Yun, D. Lee, C. Park, J. Kim, and J. Kim, "Automatic Recognition of Children Engagement from Facial Video Using Convolutional Neural Networks," IEEE Transactions on Affective Computing, 11(4):696–707, 10 2020. ISSN 19493045. doi: 10.1109/TAFFC.2018.2834350.

[12] A. Gupta, A. D'Cunha, K. Awasthi, and V. Balasubramanian, "DAiSEE: Towards User Engagement Recognition in the Wild," 14(8):1–12, 9 2016. doi: https://doi.org/10.48550/arXiv.1609.01885. URL https://arxiv.org/abs/1609.01885.

[13] A. Dhall, A. Kaur, R. Goecke, and T. Gedeon, "EmotiW 2018: Audio-Video, Student Engagement and Group-Level Affect Prediction," In Proceedings of the 2018 on International Conference on Multimodal Interaction - ICMI '18, number October, pages 653–656, New York, New York, USA, 2018. ACM Press. ISBN 9781450356923. doi: 10.1145/3242969.3264993.

[14] B. Woolf, W. Burleson, I. Arroyo, T. Dragon, D. Cooper, and R. Picard. Affect-aware tutors: recognising and responding to student affect. International Journal of Learning Technology, 4(3/4):129, 2009