What Did I Miss? Assisting User-adaptive Missed Content Reviewing in MOOC Learning

Qian Zhu[1,2] and Shuai Ma[1,2] *
[1] HCI Lab, ISCAS. [2] School of Computer Science and Technology, UCAS.
{zhuqian172, mashuai171}@mails.ucas.ac.cn   *: Corresponding author

ABSTRACT
In Massive Open Online Courses (MOOCs), learners face a lot of distractions which will cause divided attention (DA). However, it is not easy for learners to realize that they are distracted and to find out which part of the course they have missed. In this paper, we present *Reminder*, a system for detecting divided attention and reminding learners what they just missed on both PC and mobile devices with a camera capturing their status. To get learners' attention level, we build a regression model to predict attention score from an integrated feature vector. Meanwhile, we design an interactively updating method to make the model adaptive to a specific user. We also propose a visualization method to help learners review missed content easily. User study shows that *Reminder* detects learners' divided attention and assists them to review course contents effectively.

Author Keywords
Attention monitoring; MOOC learning; Content reviewing;

CCS Concepts
• Applied computing → E-learning;

INTRODUCTION
Massive Open Online Courses (MOOCs), as a scalable paradigm in online education, had attracted over 101 million learners by December 2018 [1]. However, learners are easily distracted when taking MOOC lessons compared with offline classroom education as there is no teacher monitoring their disengagement [12, 13]. The most direct way to address this challenge is to monitor learners' studying status. Various learning systems have been built to detect and respond to learners' affective and cognitive states [3, 4, 7, 8] via physiological signals, such as heart rates [6, 11], eye gaze [7], and EEG signals [3, 4]. To implicitly detect learners' attention, OneMind [13] and AttentiveLearner [6, 10] introduced implicit physiological signal sensing on unmodified mobile phones by letting users put their finger on the phone's back camera. However, these models are trained only based on overall users' data, it is inapplicable to an individual user. Another problem is that since there is no way to perfectly eliminate distractions, most attention detection systems don't provide learners with a review method for distraction. Although review can effectively solve the problem of missing knowledge points caused by distraction, it is sometimes difficult for users to realize that they are distracted. Even if they realize that they are distracted, they have missed a long period of learning content. So, it is difficult for users to know clearly how much learning content they have missed, which makes it difficult for users to locate the missed content.

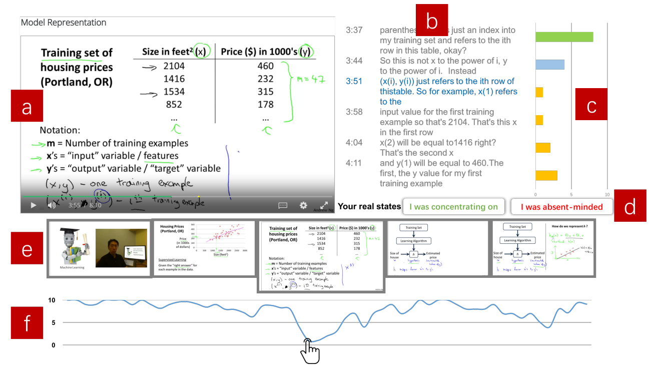

Figure 1. Overview of *Reminder*. (a) is a normal player; (b) is a script area changing with lecture and the current script is highlighted; (c) is the attention score of every script in a slide; (d) is a feedback area; (e) is a thumbnail list of slides in this lecture; (f) is an attention score curve in the whole lecture.

To solve these problems, we propose an implicit burden-free attention detection system, *Reminder*, which can integrate learners' multi-modal information to predict their attention level and improve its performance by interactive learning with individual user's feedback. Meanwhile, *Reminder* provides a multi-scale distraction search function to help users remind and review the missed lectures.

In the user study, we first collected 40 participants' attention data to train a model. Then we deployed the model to build *Reminder* and let participants watch MOOC lecture videos using *Reminder* and a normal player. Participants finished watching and reviewing tasks in less time using *Reminder* and commented that the review assistance is quite helpful.

SYSTEM DESIGN

Interface
Reminder has three main features provided to help users monitor their attention and review missed knowledge.

Real-time attention monitoring: When users watching the lecture, the front-camera on their PC, pad, phone will capture their shot every second. At the same time, an

CHI 2020, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04...$15.00.
DOI: https://doi.org/10.1145/3313831.XXXXXXX

attention curve (figure 1 (f)) will be formed based on users' attention score and displayed in real-time.

Attention-based review: We provide a two-scale attention visualization modes. Slide-scale (figure 1 (f)) helps users explore their attention changing over slides. Script-scale (figure 1 (c)) supports users to explore their attention changing over scripts.

Attention feedback: Because of the different manifestation of distraction among different users, the attention score displayed to users is not accurate sometimes. For example, some learners are accustomed to watching videos with their heads hold in one hand, but we can't say that they are distracted even though their eyes are slanted. As a result, a more accurate and personalized model should be introduced. We allow users to correct their attention in a script with two buttons which are "I was concentrating on" and "I was absent-minded" shown in figure 1 (d). Users will get a more and more accurate prediction of their attention with the increasing use of this function.

Algorithm

The algorithm of *Reminder* contains two parts, attention prediction and model updating (overview in figure 2).

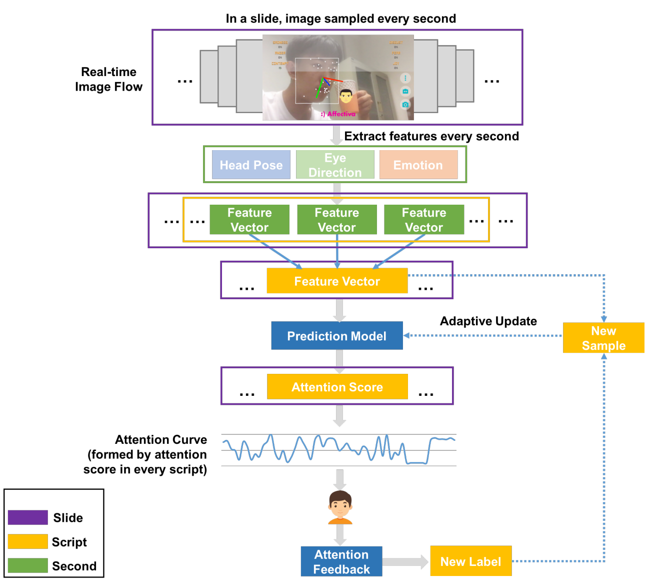

Figure 2. Overview of the user-adaptive attention monitoring algorithm.

Attention prediction: The input of the model is users' shots captured by the front-camera. We deployed three modal of recognition methods to extract features from the captured shots. Partly followed [2], we used OpenFace to track head pose [9], gaze direction [5] and used Affectiva[1] to recognize users' emotion. We set the camera capture rate to 1 FPS to reduce computational cost. From each shot, we extracted a 20-dimensional feature set, which contains head pose (6-dimension), eye gaze (8-dimension) and facial emotion (6-dimension) features. Then we trained a logistic regression model to predict a scalar probabilistic value and mapped it to 0 to 10 to get the attention score.

Model updating: As the buttons ("I was concentrating on" and "I was absent-minded") clicked, the former will give a label 0 to a script which is regarded as a negative sample, and the latter will give a label 1 to a script which is regarded as a positive sample. They will serve as a new personalized training date to interactively train users' personal attention prediction models. As our model is essentially a logistic regression, the computation is efficient.

USER STUDY

We first invited 40 participants to watch some short-time lecture videos containing 200 scripts with either engagement or disengagement, and we trained and evaluated the model with 8000 samples by cross-validation. Then, we let 40 participants watch 4 different lecture videos, 20 watched lecture 1&2 with a normal player and lecture 3&4 with *Reminder*, and others were the opposite. During the viewing, we limited participants to look at the progress bar time and gave some interference to them in a random time, such as noise, new mobile phone notification and so on. We requested that participants respond positively to the disturbances we send. After that, we asked every participant to find the exact time of the disturbance to simulate the reviewing process. The review included two parts, one was reviewing after the slide is finished and another was after the lecture is finished. In the end, participants were required to fill a post-task questionnaire on a 5-point Likert scale.

Result

In the 10-fold cross-validation, we got an average result of accuracy 0.92, recall 0.89, and precision 0.94. In the distraction and review study, we got the average review time of 40 participants using a normal player or *Reminder* in two conditions, review after the slide and after the lecture. Figure 3 (a) shows that *Reminder* outperformed the baseline player in review time by a large margin both in slide-scale and lecture-scale. Based on pair-wise t-tests we found that both the differences are significant ($p<0.001$). Figure 3 (b) showed that users were very satisfied with *Reminder*.

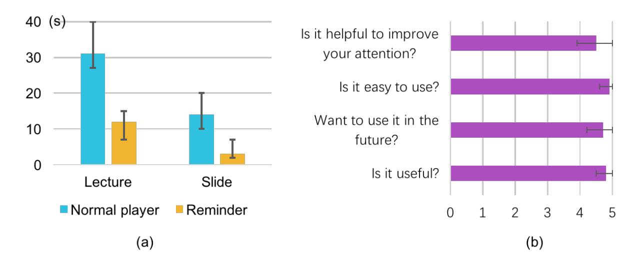

Figure 3. Results of user study. (a) shows the time spent with two systems; (b) shows the result in the questionnaire.

FUTURE WORK

In this paper, we propose a review method and a user-adaptive model updating method. However, the latter hasn't been verified yet. In the future, we will design a long-term user study to investigate whether our methods can effectively suit different users gradually.

[1] https://www.affectiva.com/

REFERENCES

1. 2019. By The Numbers: MOOCS in 2018. https://www.classcentral.com/report/mooc-stats-2018/.
2. Chang, C., Zhang, C., Chen, L., & Liu, Y. 2018. An ensemble model using face and body tracking for engagement detection. In *Proceedings of the 2018 on International Conference on Multimodal Interaction*, 616-622.
3. Szafir, D., & Mutlu, B. 2013. ARTFul: adaptive review technology for flipped learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 1001-1010.
4. Szafir, D., & Mutlu, B. 2012. Pay attention!: designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI conference on human factors in computing systems,* 11-20.
5. Wood, E., Baltrusaitis, T., Zhang, X., Sugano, Y., Robinson, P., & Bulling, A. 2015. Rendering of eyes for eye-shape registration and gaze estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, 3756-3764.
6. Pham, P., & Wang, J. 2015. AttentiveLearner: improving mobile MOOC learning via implicit heart rate tracking. In *International Conference on Artificial Intelligence in Education*, 367-376.
7. D'Mello, S., Olney, A., Williams, C., & Hays, P. 2012. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies*, *70*(5), 377-398.
8. D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., & Graesser, A. 2010. A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In *International conference on intelligent tutoring systems*, 245-254
9. Baltrusaitis, T., Robinson, P., & Morency, L. P. 2013. Constrained local neural fields for robust facial landmark detection in the wild. In *Proceedings of the IEEE international conference on computer vision workshops*, 354-361.
10. Xiao, X., Han, T., & Wang, J. 2013. LensGesture: augmenting mobile interactions with back-of-device finger gestures. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, 287-294.
11. Xiao, X., & Wang, J. 2015. Towards attentive, bi-directional MOOC learning on mobile devices. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 163-170.
12. Xiao, X., & Wang, J. 2016. Context and cognitive state triggered interventions for mobile MOOC learning. *In Proceedings of the 18th ACM International Conference on Multimodal Interaction*, 378-385.
13. Xiao, X., & Wang, J. 2017. Undertanding and detecting divided attention in mobile mooc learning. *In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 411-2415.