Cognitive State Recognition for Developing Anticipating Textbook

Shoya Ishimaru
German Research Center for Artificiall Intelligence (DFKI)
Kaiserslautern, Germany
Email: Shoya.Ishimaru@dfki.de

I. Introduction

Documents traditionally are an important means for learning. The various school subjects are aligned with contents, which are captured in traditional textbooks. Although reading a textbook plays an essential role to obtain new knowledge, students often avert their eyes from reading a textbook because it is static and boring. To solve this problem, I am currently working towards creating an *anticipating textbook*, which can make the materials for learning and instruction adaptive and dynamic on a mobile tablet computer. Figure 1 shows the concept sketch of the anticipating textbook. The system recognizes a student's cognitive states (e.g. attention, interest, comprehension) using several sensors, and changes the content or layout of the textbook dynamically to improve his/her motivation and understanding.

While most of physical activities including walking and sleeping can be recognized by body-mounted motion sensors, recognizing cognitive state is a challenging task. I utilize eye tracking devices for the system becasue the strong relationship between cognitive states and eye movements is well explored in the field of activity recognition and psychology [1]. This paper presents my work in progress to recognize cognitive states while reading a textbook by using two sensing modalities.

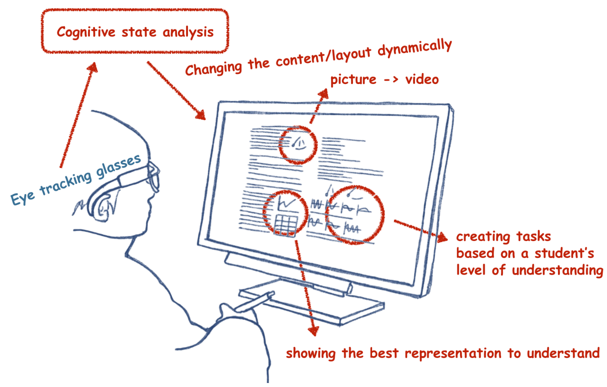

Fig. 1. A concept sketch of an anticipating textbook.

II. Attention Extraction with Eye Tracking Glasses

The first step to develop the anticipating textbook is to visualize students' attentions while reading. On the assumption that the more attention students pay the more time they spend time to read, we have proposed a method to extract attention by using an eye tracking glasses [2].

A. Attention Extracting Method

The method consists of three steps: mapping eye gaze on a document, fixation-saccade detection, and heatmap visualization. Mobile eye tracking glasses are utilized to record students' reading behaviors. Because the outputs of the device are coordinates of eye gaze on a scene camera from the device, we detect the position of the document in a frame of the camera by using SIFT features, and calculate a homograph to map the gaze point to the document. The raw gaze data on the document is classified into fixations (gaze posing in certain position for 200-400 ms) and saccades (jumps of the gaze between two fixations taking 10-20 ms). The system apply the fixation-saccade detection algorithm proposed by Buscher et al. [3]. We divide a textbook beforehand based on the roll and calcualte periods of time a student needs to reads each content to obtain knowledge. Thus, for each area a sum of fixation durations is calculated, which is divided by the size of area to be normalized.

B. Experiment

As an initial experiment, we asked 8 participants (around 12 years old) to wear eye-tracking glasses, to read a physics textbook on a display, and to solve respective exercises. The document presented in the experiment is "Basic Phenomena in Acoustics" in Physics, and it consists of three parts: the introduction, itemized definitions, and applications. Only an explanation of the content was displayed at first. After they understood the content, they could make exercises appear by pressing a button. They could also go buck to read the content to find hints for solving tasks. After the recording, we categorized 8 participants to 3 comprehension levels (novice, intermediate, and expert) according to the score of the exercises.

Figure 2 represents mean attention heatmaps for each comprehension level. It has become obvious that students with high level comprehension do not pay attention to the applications part while both reading and solving tasks compared to other levels. They understand that the applications part is useful for understanding the content, yet there is not much information that can be used as hints for solving tasks. As shown in Figure 3, they preferred to read definitions part

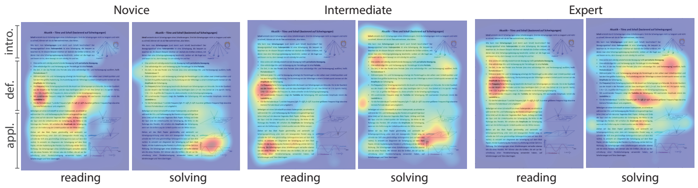

Fig. 2. Attention heatmaps exported by 3 comprehension levels while reading the textbook and solving the exercises.

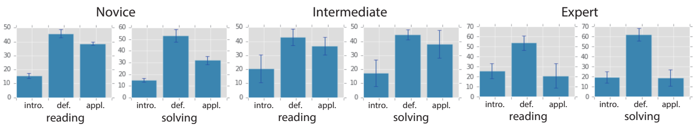

Fig. 3. Percentages of attentions participants paied to introduction, definition, and application part on the textbook.

because there are direct hints (principles, formulas, etc.). Intermediates and novices spend much time to paying attention to the application part while both reading and solving.

III. Tracking Reading Activity with Commercial EOG Glasses

While high spec eye tracking devices are suitable to record human behaviors in controlled experiments, it is still difficult to use them in a real classroom scenario (e.g., the devices are expensive; battery lives are short). Therefore we have also proposed a reading tracking method with commercial device, J!NS MEME [4], [5]. It equips EOG sensors to measure eye movements and motion sensors to measure head movements. I have started from detecting reading activities and estimating the number of words a user read becasue the reading duration and the speed should be related to a user's cognitive states including interest and attention.

A. Word counting Method

The word counting method consists of three processes: obtaining a user's eye movements, detecting forward/backward-saccades, and estimating the number of words he/she read. One reference electrode attached to the forehead and two active electrodes on the nose pad are equipped on J!NS MEME. We calcualte EOG vertical component as an average of two active electrodes and EOG horizontal component as the difference between them. Figure 4 shows an overview of the EOG horizontal component in a one-minute recording that includes reading activity. Negative values represent eye movement right to left, and positive values represent left to right. Saccades towards forward and backward (line breaks) on the reading direction appear as peaks. We detect them and estimated the number of words a user read by support vector regression. Four features are calculated for the regression: the total number of forward-saccades, the mean EOG signal value of forward-saccades, the total number of backward-saccades, and the mean EOG signal value of backward-saccades.

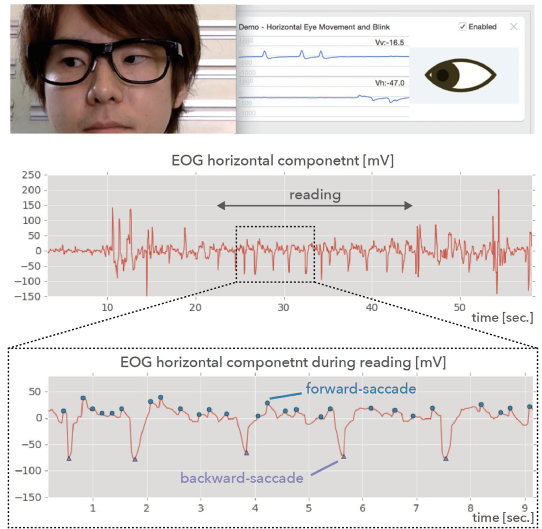

Fig. 4. Overview of J!NS MEME hardware and horizontal EOG sensor signals while reading. Regular patterns of eye movement appear because of line breaks.

B. Experiment

To evaluate the word-counting algorithm, we asked 5 participants to read English essays on an iPad with wearing J!NS MEME. Every participant read 38 paragraphs (minimum: 27 words; maximum: 120 words; average: 60 words in one paragraph). Training and testing were done by leave-one-subject-out as a user-independent approach and 10-fold cross validation as a user-dependent approach.

An average estimation error of five subjects during the readings for each paragraph was 18% on user-independent approach, and decreased to 16% on user-independent approach.

IV. Future Work

In the future, I plan to develop some prototypes of anticipating textbook and find what kind of dynamic changing improves students' learning abitilies. Future work also includes using other sensing modalities (e.g., face temperature, heart rate).

References

[1] R. Kliegl, A. Nuthmann, and R. Engbert, "Tracking the mind during reading: the influence of past, present, and future words on fixation durations." *Journal of Experimental Psychology: General; Journal of Experimental Psychology: General*, vol. 135, no. 1, p. 12, 2006.

[2] S. Ishimaru, S. S. Bukhari, C. Heisel, J. Kuhn, and A. Dengel, "Towards an intelligent textbook: eye gaze based attention extraction on materials for learning and instruction in physics," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 2016, pp. 1041–1045.

[3] G. Buscher, A. Dengel, and L. van Elst, "Eye movements as implicit relevance feedback," in *Proceedings of the 26th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2008, pp. 2991–2996.

[4] S. Ishimaru, K. Kunze, K. Kise, and A. Dengel, "The wordometer 2.0: estimating the number of words you read in real life using commercial eog glasses," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 2016, pp. 293–296.

[5] S. Ishimaru, K. Kunze, K. Tanaka, Y. Uema, K. Kise, and M. Inami, "Smart eyewear for interaction and activity recognition," in *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2015, pp. 307–310.