Towards Obtaining Affect-Based Proxies for Attentional Behaviour in TEL

Kshitij Sharma, Norwegian University of Science and Technology, kshitij.sharma@ntnu.no
Shitanshu Mishra, Vanderbilt University, shitanshu.mishra@vanderbilt.edu
Zacharoula Papamitsiou, Norwegian University of Science and Technology, zacharoula.papamitsiou@ntnu.no
Anabil Munshi, Vanderbilt University, anabil.munshi@vanderbilt.edu
Bikram Kumar De, Vanderbilt University, bikram.k.de@vanderbilt.edu
Gautam Biswas, Vanderbilt University, gautam.biswas@vanderbilt.edu
Michail Giannakos, Norwegian University of Science and Technology, michailg@ntnu.no

Abstract: Current multimodal studies have a common limitation of not being able to scale up the implications since the apparatus used is not scalable. In this paper, we propose a simple method to find measurements from scalable data modes such as facial data and examine the measures in richer and more granular data modes like eye-tracking that they correspond most closely to. In other words, we find pervasive proxies to the measurements that have been reported to be obtrusive. We exemplify this approach using eye-tracking and facial data from two different studies.

Keywords: Eye-tracking, Facial Expressions, Multimodal Learning Analytics.

Introduction and background

The motivation for this research is twofold. First, sensing students' cognitive and affective states in online learning environments needs to be achieved through ubiquitous devices such as webcams or Fitbits, since current off-the-shelf solutions for eye-tracking and EEG measures are not scalable in terms of cost-effectiveness. Secondly, ubiquitous sensing of student behaviors during learning has potential for classroom orchestration research, with the aim of informing the teacher about students' cognitive and/or affective states at any moment during interaction with the learning environment (Prieto et al., 2018). In classroom settings, especially with a large student population, current methods for reliably sensing students' cognitive and affective states do not support principles of ecological validity. A solution would be using multimodal learning analytics (MMLA) techniques to fuse data sources that identify cognitive-affective states - such as eye-tracking, electroencephalography (EEG), heart rate, facial data, to mention a few.

Current research in MMLA (Giannakos et al., 2019; Sharma et al., 2019) is focused on fusion of different data modalities to model attentive and cognitive processes, affective states along with arousal trends, to explain the learning outcomes and behaviors. Since all the different data sources used in current MMLA research are not ecologically valid, fusing these sources together can not only help us achieve a higher level of understanding about student behaviors but also enable improvement of our sensing methods to move towards a ubiquitous sensing of learning processes. In this contribution, we present an exemplar method to achieve this by finding interconnections between gaze and facial data streams using data from two different classroom studies.

Eye-tracking has been used to understand the learning processes (Sharma et al, 2015) and provide students with adaptive feedback on their learning processes (D'Angelo & Begel, 2017). Within affective states, eye-tracking has been used to gauge student boredom, curiosity, and attention or mind-wandering (Bixler & D'Mello, 2016); identifying these states helps design interventions to regulate student affect during learning. Bearing in mind the innate connection between emotions and facial expressions, a sizeable body of research is dedicated to assessing learning and performance through emotions inferred from facial expressions (Graesser et al, 2006; Baker et al, 2010;). D'Mello et al. (2012) collected facial expressions of students while interacting with AutoTutor; and modeled transition likelihoods among affective states of boredom, flow (engagement), and confusion during learning (D'Mello et al., 2012).

Methodology

Study 1 This study used Betty's Brain, an open-ended learning environment (OELE) that helps middle school students learn scientific processes (Leelawong & Biswas, 2008). During our study in Dec 2018 with 60 sixth-grade participants of a public middle school in USA, students worked on Betty's Brain to build causal models of climate change, using individual laptops equipped with webcams and Tobii-4c eye-tracking devices, thereby enabling the collection of facial landmarks data from webcam video frames and gaze data from eye-tracker logs. After initial processing, the facial and gaze data logs were synchronized by time for analysis purposes.

Study 2 The second study took place in controlled settings. Students' responses and system usage logs were collected with LAERS (Papamitsiou et al, 2013), a web-based implementation of a layered architecture for testing systems. Thirty-two undergraduate students at a European university enrolled in an online adaptive self-assessment procedure for a Web Technologies course (related to front-end development) in Oct 2018. Students' gaze was recorded using the Tobii X3-120 eye-tracking device at 120 Hz sampling rate using 5-point calibration. The device was mounted at the bottom of the screen, facial video recording was done with 640x480 video resolution at 10 FPS using a Logitech Webcam that was zoomed 150% onto participant faces.

Finding facial proxies to gaze data First, we scale gaze and facial measurements to be in the same range of zero to one. Next, we conduct rotated Principal Component Analysis (PCA) on the combined data (gaze + face). We then select the appropriate number of Principal Components (PC) that explain the relation between gaze and facial measurements, applying a PC threshold of 75%. For each gaze measure that has smallest angle with each PC, we rotate the vector corresponding vector to be parallel to the PC, and recompute all the angles for facial measures with the selected PC. We then identify the top five closest facial vectors in each of the rotated dimensions to be the proxies to the gaze measurements.

Measurements From facial data we extracted the intensity of 16 Action Units (AUs) such as ChinRaiser, JawDrop, LipPressor, LipStretcher, and LipPucke using iMotions Affectiva (for Betty's Brain) and OpenFace (for adaptive assessment system). From gaze data we computed the following gaze variables (for definitions and relation between these measurements and cognitive/behavioral processes, cf., Holmqvist et al, 2011): Mean Saccade Angle (MSA), Scan-path Velocity (SPV), Saccade Velocity Skewness (SVS), Local/Global Information processing (IP), Saccade Amplitude (SA), Saccade Velocity (SV) .

Results and discussion

For Betty's Brain, PCA resulted in 12 components explaining a total of 75% variance from the original data. For adaptive assessment, PCA resulted in 7 components explaining a total of 79% variance from the original data. Striking similarities are observed in terms of the relationship between gaze and facial measurements across the two studies, with an overlap of 86.16% in terms of affinity (correlation irrespective of being positive or negative) between the individual gaze measures and AUs, and an overlap of 66.67% (when considering affinity direction i.e., positive or negative correlation). Only in 16.84% cases the two studies are completely different, which could be attributed to the diversity in the two contexts, one being a complex Intelligent Tutoring system (Betty's Brain) and the other an adaptive self-assessment testing system. The similarities in gaze-face relations across the two studies could be exploited to design interventions solely based on the intensity of AUs. For example, low intensities of ChinRaiser, JawDrop, LipPressor, LipStretcher, and LipPucker might suggest global processing. Similarly, high intensities of LipPressor, LipStretcher, LitTightener, NoseWrinkler and UpperLipRaiser might suggest high saccade velocity skewness, which depicts strong anticipatory patterns. Such inferences can enable the provision of appropriate feedback to students or adaptation of the learning system.

References

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *Intl. Journal of Human-Computer Studies*, *68*(4), 223-241.

Bixler, R., & D'Mello, S. (2016). Automatic gaze-based user-independent detection of mind wandering during computerized reading. User Modeling and User-Adapted Interaction, 26(1), 33-68.

D'Angelo, S., & Begel, A. (2017, May). Improving communication between pair programmers using shared gaze awareness. In *Procs. of the 2017 CHI Conf. on Human Factors in Computing Systems* (pp. 6245-6290). ACM.

D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, *22*(2), 145-157.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). FACS investigator's guide. A human face, 96.

Graesser, A. C., McDaniel, B., Chipman, P., Witherspoon, A., D'Mello, S., & Gholson, B. (2006). Detection of emotions during learning with AutoTutor. In *Proceedings of the 28th annual meetings of the cognitive science society* (pp. 285-290).

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, *48*, 108-119.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Sharma, K., Papamitsiou, Z., & Giannakos, M. (2019). Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach. *British Journal of Educational Technology*, *50*(6), 3004-3031.

Sharma, K., Caballero, D., Verma, H., Jermann, P., & Dillenbourg, P. (2015). Looking AT versus looking THROUGH: A dual eye-tracking study in MOOC context. International Society of the Learning Sciences, Inc.[ISLS]..