Real-Time Feedback on Reader's Engagement and Emotion Estimated by Eye-Tracking and Physiological Sensing

Akshay Palimar Pai
University of Kaiserslautern & DFKI GmbH
Kaiserslautern, Germany
Akshay_Palimar.Pai@dfki.de

Jayasankar Santhosh
University of Kaiserslautern & DFKI GmbH
Kaiserslautern, Germany
Jayasankar.Santhosh@dfki.de

Shoya Ishimaru
University of Kaiserslautern & DFKI GmbH
Kaiserslautern, Germany
Shoya.Ishimaru@dfki.de

ABSTRACT

The primary goal of this study is to estimate engagement and emotion in a reading task using machine learning techniques and then to utilize the data to design a visualization tool that depicts the differences in engagement and emotion of various readers at regular intervals. A reading experiment with 20 participants and 14 documents was designed which was followed by a questionnaire to rate engagement, arousal, and valence after reading each document on a scale of 1-5. Tobii 4C eye tracker was used along with Empatica E4 wristband to collect data from participants. Different machine learning models were employed to estimate the engagement, arousal, and valence as rated by participants. A 1D-Convolutional Neural Network achieved the highest mean accuracy of 73% for engagement detection, and a Fully Convolution Network network achieved the highest mean accuracy of 66% and 64% for prediction of arousal and valence in a leave-one-participant-out cross-validation. From the evaluation results, a working prototype with an engagement gauge and emoji was developed to visualize the variations in engagement and emotion at timely intervals for each user.

CCS CONCEPTS

• **Human-centered computing** → *Interactive systems and tools.*

KEYWORDS

Eye-Tracking, Physiological Sensing, Deep Neural Network, Affective Computing

ACM Reference Format:
Akshay Palimar Pai, Jayasankar Santhosh, and Shoya Ishimaru. 2022. Real-Time Feedback on Reader's Engagement and Emotion Estimated by Eye-Tracking and Physiological Sensing. In *Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC '22 Adjunct), September 11–15, 2022, Cambridge, United Kingdom.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3544793.3560329

1 INTRODUCTION

The prediction of a reader's internal states, such as engagement, could facilitate in making the reading material more interactive [1]. To recognize the user behavior, an in-depth understanding of his/her engagement or involvement in a particular task is required, which can be leveraged to foster learning and motivate users. Jacob *et al.* presented how physiological signals acquired from an Empatica E4 wristband could be used for detecting interest in a user for a reading task [2]. This study aims to detect the engagement and emotion of readers in a reading task and to develop a visualization tool that depicts the variations in engagement and emotion at timely intervals for different readers. Tobii 4C eye tracker was used along with Empatica E4 wristband to collect the data. The main contributions of this work include (1) Design and Implementation of a data set with 20 participants, rating the responses for engagement, arousal, and valence in a reading task, (2) Comparison of machine learning models for the prediction of engagement, valence, and arousal using sensor data collected from eye tracker and a wristband, and (3) A feedback based visualization tool depicting the variations in engagement and emotions using a gauge meter and emojis.

UbiComp/ISWC '22 Adjunct, September 11–15, 2022, Cambridge, United Kingdom

ACM ISBN 978-1-4503-9423-9/22/09.
https://doi.org/10.1145/3544793.3560329

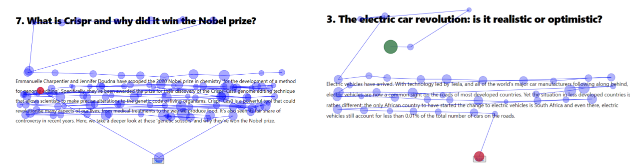

Figure 1: Scanpaths of gaze data for most engaging document vs least engaging document as rated by one participant

2 APPROACH

Our system records eye gaze and physiological signals while reading to estimate affective states including engagement, arousal, and valence. We utilize Tobii 4C remote eye-tracker for recording 90 Hz gaze coordinates and pupil diameters. Physiological data collected from E4 wristband sensor includes 3-axis Acceleration (ACC; 32Hz), Blood Volume Pulse (BVP; 64Hz), Electrodermal activity (EDA; 4Hz), Skin Temperature (TEMP; 4Hz), and Heart Rate (HR; 1Hz). Samples for the classification are created as sliding windows with a size of 30 sec and with a 0.50-second step. We implement the following two types of approaches and chose the one with the best performance.

Feature Extraction Based Affective State Estimation – Fixations and saccades are pre-processed from raw gaze coordinates, and the mean, std, min, and max of the fixation duration, saccade length, saccade velocity, saccade angle, and pupil size are computed. In addition, the number of fixations is also computed. Empatica E4 is comprised of physiological signals like EDA, ACC, BVP, TEMP, and HR. In addition to the statistical features, the phasic (SCR) and tonic (SCL) components were extracted from the EDA signal. The

Table 1: Summary of classification results using machine learning models including deep learning.

Model	Sensor	Engagement		Arousal		Valence	
		Accuracy	F1-score	Accuracy	F1-score	Accuracy	F1-score
SVM	E4 Wristband	49.32	47.09	48.80	48.56	37.59	37.38
	Eye-Tracker	62.76	59.43	55.12	54.20	51.80	51.05
	Combined	66.87	63.80	56.38	55.26	48.65	46.20
FCN	E4 Wristband	70.56	68.17	53.35	50.39	54.47	51.06
	Eye-Tracker	70.12	70.66	**66.15**	**61.65**	**64.17**	**63.97**
CNN	E4 Wristband	71.51	67.58	50.64	46.23	54.56	50.89
	Eye-Tracker	**72.87**	**72.25**	63.27	62.48	63.55	62.57

frequency and time domain features were computed from the HR data as well. After calculating features, we utilize a Support Vector Machine (SVM) to classify each sample.

Deep Learning Based Affective State Estimation – In this approach, a deep neural network receives raw signals from an eye-tracker and/or physiological signals and classifies each sample into affective states. We implement a Fully Convolution Network (FCN) [3] in addition to 1-Dimensional Convolutional Neural Network (CNN).

3 EVALUATION

We conducted an experiment involving 20 master-pursuing students reading 14 documents with two sessions of 25 minutes each with 10 minutes of break between them. All participants were asked to rate documents based on the levels of engagement, arousal, and valence on a scale of 1-5, where one was the least and five the most. Figure 1 shows the variation in gaze patterns for documents rated the most engaging and least engaging. We evaluated the performances as binary classification tasks (negative for three or fewer). By assuming the development of a user-independent system, we separated data into training and testing in a leave-one-participant-out cross-validation manner.

Results – Table 1 shows the comparison of sensing modalities and machine learning models. It was observed that 1D-CNN and FCN were able to achieve better accuracy and F1-scores specifically for engagement classification compared with SVM. Although omitted from the table, we made a performance comparison with other classifiers (e.g., Random Forests, k-Nearest Neighbor) and SVM recorded the highest performance in the category of feature extraction based approaches. Note that we calculated a decision-level fusion of E4 and Eye-Tracker [4] but could not observe better results than individual performances.

4 APPLICATION

Figure 2 shows an application using the trained model. The dashboard contains an engagement gauge that shows the engagement level of a reader based on the predicted class probabilities and also contains an emotion emoji that shows the emotion of a reader. The 1D-CNN model was used to predict the engagement of the user. The variations in the engagement gauge were visualized from 0-100 by predicting the class probabilities of engagement. The best-performing model in terms of accuracy was used to predict the valence and arousal to show the varying emotions of the user.

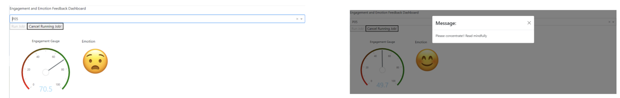

Figure 2: Screenshots of the demo with engagement gauge and alert message when engagement level falls below a threshold value

The dashboard was designed with alert messages to notify the user when their engagement level falls behind a threshold value and also a motivation message to encourage the user when their engagement levels were consistently above a specific value. For future applications, we are planning for a user study with integrated real-time feedback to the user based on sensor data received.

5 CONCLUSION

In this work, we demonstrated a real-time dashboard that monitors affective states including engagement and emotions while reading that are estimated from eye movements and physiological signals. The estimation model was evaluated with an experiment involving 20 readers and a leave-one-participant-out cross-validation. Engagement, arousal, and valence were classified into two classes with accuracies of 73%, 66%, and 64%, respectively.

REFERENCES

[1] Shoya Ishimaru, Syed Saqib Bukhari, Carina Heisel, Nicolas Großmann, Pascal Klein, Jochen Kuhn, and Andreas Dengel. 2018. Augmented learning on anticipating textbooks with eye tracking. In *Positive Learning in the Age of Information.* Springer, 387–398.

[2] Soumy Jacob, Shoya Ishimaru, and Andreas Dengel. 2018. Interest detection while reading newspaper articles by utilizing a physiological sensing wristband. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers.* 78–81.

[3] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 3431–3440.

[4] Hamed Monkaresi, Nigel Bosch, Rafael A Calvo, and Sidney K D'Mello. 2016. Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing* 8, 1 (2016), 15–28.