Towards Automatic and Pervasive Physiological Sensing of Collaborative Learning

Kshitij Sharma, Ilias O. Pappas, Sofia Papavlasupoulou, and Michail N. Giannakos
kshitij.sharma@ntnu.no; ilpapas@ntnu.no; spapav@ntnu.no; michailg@ntnu.no
Department of Computer Science, Norwegian University of Science and Technology

Abstract: We present a collaborative learning study contextualized within Project based Learning. The main aim of our contribution is to use the physiological data such as heart rate, skin temperature, electrodermal activity and blood volume pressure to quantify the learning experiences of the collaborating teams. We propose an automatic method to extract collaborative measures and study their relationship with the perceived performance, usefulness and satisfaction from the collaborative sessions from various student groups in a university degree course. We aim to contribute towards automatized, pervasive and more generalizable sensing of collaborative learning. Our results show that the synchrony in automatically extracted physiological states correlates positively with perceived performance and satisfaction of teams.

Introduction and background

The proponents of Project Based Learning (PBL) suggest that there are some advantages to it. For example, students in PBL learn better and they engage more actively in learning than normal instruction (Green 1998), students show more reflection on the knowledge gain (Preuss 2002) and there is a gradual decrement in the requirement of a step-by-step instruction (Lenshow, 1998). We present a study from a collaborative PBL course in a university. Our aim is to provide objective measures to capture the self-reported learning experiences of the students in a collaborative scenario such as performance, usefulness and satisfaction. In the present contribution the objective measures come from a wearable wristband that records the heart rate, electro dermal activity, skin temperature and the blood volume pressure of the participants in each student group. In this paper we would collectively refer to these data streams as *wristband data*. With the advancements in the wearable technologies, off-the-shelf devices have made the physiological sensing easier for the researcher than in the past. There have been many studies carried out to explore the relationship between the physiological data and learning outcomes (Pijeira-Díaz et al 2016), collaboration quality (Chenal et al 2013).

We propose three fundamental shifts. First, we move from the invasive devices such as eye-tracking, EKG, EEG to pervasive wristbands. Second, we move from semi-automatic to completely automatic extraction of collaborative segments. Finally, the third shift is from the emotional/semantic sensing of collaborative scenarios to semantic-less sensing of the collaborative scenarios to achieve a greater generalizability of research outcomes. The main contribution of this study is the use of physiological data to define collaborative measures and using them to quantify collaborative learning experiences. Through this contribution, we address the following research question**:** *"how the groups' physiological state affects the learning experiences of the collaborators?"* We define the physiological states using the features extracted from the four collected data streams, i.e., heart rate (HR), blood volume pressure (BVP), electro dermal activity (EDA) and skin temperature (TEMP). We hypothesize that the *synchrony of the physiological state among the collaborators would affect their learning experiences*. To compute the synchrony of the physiological states, we employ previously used methods to compute the synchrony in the collaborative eye-tracking data in various CSCL scenarios called cross-recurrence (Jermann and Nuessli, 2012; Schneider et. al., 2016).

Wearable devices are now widely available and affordable and it has become possible to monitor more subtle phenomena, such as the quality of social interactions, students' mental health and learning engagement (Wang et al., 2018). In terms of measuring collaborative learning gains, Pijeira-Díaz et al (2016) conducted a study with high school students to study the relation between collaboration will, learning product, learning gain and Empatica and eye-tracking data. The main features used by Pijeira-Díaz et al (2016) were the direction of change, direct difference and linear relationship among the different data streams. On the other hand, Chenal et al (2013) used skin temperature and skin conductance with eye-tracking and electrocardiographs (ECG) to assess collaboration quality (measured by grounding, emotion management, conflict resolution, consensus building) using correlation-based features. Blanchard et al (2012) used EEG, GSR and skin temperature to explain the CL process with the focus is on individual analyses. Sottilare et.al. (2012) proposed to measure engagement, workload, motivational level and emotional state using physiological data while interacting with an Intelligent Tutoring System with a special focus on emotions/affective states.

Most of these methods employ eye-tracking, EEG or ECG equipment which hinder the pervasiveness of the devices used (Pijeira-Díaz et al, 2016). Others used human labelling as one of their methods that hinders the automatization of the process (Chenal et. al, 2013). A few other either focused on the individual measurements or they focused on the emotional awareness during collaboration, making the results generalizable in the similar given contexts only (Sollilare et al., 2012). In this contribution, we use only the wristband data, no human labelling, unsupervised segmentation of individual data and a collaborative measure to study the relation between collaborative physiological sensing and learning experiences during collaborative sessions.

Methodology

To capture fine-grained physiological data during a learning experience, we designed an in-the-wild study. The study took place in a regular university course called *Customer Driven Project*. This is a master-level class where groups of 5-7 students work on a software engineering project with a real customer. During the semester, groups have their internal group meetings, meetings with the customer representative and class sessions with their advisor. The role of the advisor is to meet the groups once every week for an approximately 30 minutes to discuss with the group about the progress and objectives of their project.

We recruited a total of 31 university students (12 female and 19 male) aged between 21 and 53 years old (mean = 24.01, S.D. = 5.87). Participants were recruited using convenience sampling from the pool of a major European university. Participants were CS majors and taking the respective course as part of their CS degree. Participants were given a 30 Euro gift card upon completion of the study. In the beginning of each class session, the participant wore the wristband, and attended the class as usual. At the end of the class, each participant completed a questionnaire that was designed (based on the literature) to capture student's learning experience. The class lasted for approximately 30 minutes.

We used surveys (7-point Likert scale) gathered students' learning experience. The survey concerned factors adopted from prior studies, particularly on how students perceived the following three notions: 1) Satisfaction (Gray & Diloreto, 2016), 2) Usefulness (Sánchez & Huero, 2010) and 3) Performance (Kuvaas, 2006). Participants wore the Empatica E4 wristband on their non-dominant hand and four different measurements were recorded: 1) heart rate (HR) at 1 Hz, 2) electro dermal activity (EDA) at 4 Hz, 3) body temperature (TEMP) at 4 Hz, and 4) blood volume pulse (BVP) at 64 Hz.

Wristband data pre-processing

Physiological data, comprising of HR, EDA, BVP, and Temperature, are affected by many biases based on the age, gender, time of the day, physical and medical conditions of the participants. Therefore, it is necessary to remove these biases. We normalize the time series of these data streams for every individual participant using the mean of the first minute of their data. All segments of data will have the same average bias per person, hence when we normalize the data with the first segments of data the biases present in all the segments gets cancelled out. Next, the noise mostly contains the occasional outlier due to the error in the measurement or the repositioning of the wristband by the participants. We use a spline smoothing to remove occasional outliers from the data.

Wristband data feature extraction

We first segmented the data into chunks of 30 seconds each. Then we define the following measures from each of the four data streams (HR, EDA, BVP, and Temperature). *Histogram based*: we compute the basic statistics from the histogram such as: mean, standard deviation, maximum, median, skewness, and kurtosis. *Auto Correlation*: we compute the first N autocorrelation coefficients using the segmented data, the choice of N is based on the AIC value of the autoregressive model. *Frequency domain*: we transform the data from temporal domain to frequency domain using a Fast Fourier Transform, and then take first M coefficients.

Clustering

We use a simple K-means clustering algorithm to obtain 5 clusters. Once again, the number of clusters (5) is determined based on comparing the different number of clusters and the corresponding value of inter-cluster separation. The inter-cluster separation is defined as a ration of the average distance between the point for each cluster and the average distance between the centroids of the clusters. The centroids of the clusters are calculated as the average of all the points in a cluster. We chose the value that maximizes the inter-cluster distance. We can see that the inter-cluster distance is not changing significantly after five clusters. Once we have assigned each segment a cluster number, we then create a time series of cluster values for each participant in every group. Each cluster is indicative of a certain physiological state of the participant. One can use these sequences of states to explain the collaboration processes, outcomes or experiences. The next subsection provides the details of how we used this sequence of states to define collaborative measures to explain the group experiences.

Cross-recurrence (CR)

It is widely used in the theory of dynamical systems to compute the temporal co-occurrence of states of two dynamical systems (Eckmann et. al., 1987). In CSCL this measurement has been used to explain collaborative processes using dual eye-tracking data (Schneider et. al., 2016; Jermann and Nuessli, 2012). In the context of our analysis, each student can be considered as a dynamical system and the physiological data from a given time window represents the state. Thereby, CR analysis can be used to measure how much and when two subjects have similar physiological states. The principle of CR is detailed in Nüssli (2011).

Results

Cluster Analysis: Using the method described in the previous section, we obtained five clusters from the features computed with BVP, HR, EDA, TEMP. Once we obtained the clusters, we chose top 12 features (with largest effect sizes, Figure 1) to explain the clusters in terms of the physiological data.

Cluster 1 (n = 550): This cluster is comprised of the segments with high auto-correlation coefficients for TEMP and BVP. This indicates that TEMP and BVP during such segments have stable TEMP and BVP values, since the auto-correlation coefficients are positive and high for both TEMP and BVP.

Cluster 2 (n = 519): Segments in this cluster are characterized by low auto-correlation coefficients for BVP and high mean and variance for HR. This indicates that the heart rates during these segments are high with a lot of fluctuations. Moreover, since the auto-correlation coefficients for BVP are lower in this case than the other clusters, this shows that there is a considerable variability in the BVP of participants for such durations as well.

Cluster 3 (n = 889): This cluster has the segments with highly skewed EDA and low mean and variance for HR. This indicates that such segments had high EDA levels and low but stable heart rates for the participants.

Cluster 4 (n=133): This cluster has segments which show low auto-correlation coefficients for TEMP but high auto–correlation coefficients for EDA and BVP. This indicates that the EDA and BVP have low variation during such periods, but this is not the case for TEMP.

Cluster 5 (n = 826): Segments in this clusters show low auto-correlation for EDA and low variance for HR. This indicates that the heart rates are stable during these moments, but not the EDA.

CR analysis: We observe a positive and significant correlation between the CR and the perceived performance of the group (Pearson cor. = 0.56, p = .02). We also observe a positive and significant correlation between the CR and the perceived satisfaction of the group (Pearson cor. = 0.56, p = .02). However, we did not observe any significant correlation between the perceived usefulness of the group and the average CR of physiological states.

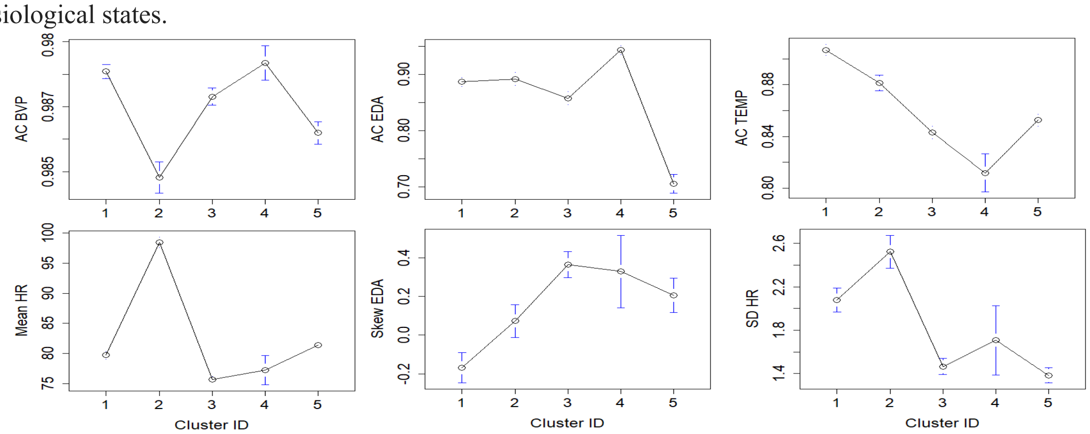

Figure 1. The clusters as defined by the top 12 features (largest effect sizes). These 12 features are: three auto-correlation coefficients for TEMP, EDA, and BVP, mean and variance of HR and the skewness of EDA.

Discussion and conclusion

We present results from a project-based learning study, where we recorded the physiological data (electro dermal activity, heart rate, blood volume pressure, and skin temperature) of the participants. We then defined five different physiological states using the features from the data. Next, we computed the cross-recurrence of these states among the participants. Finally, we show the correlations between the cross-recurrence and the groups' average learning experiences (satisfaction, performance, and usefulness). Each cluster we obtained from the features of the physiological data collected represents a certain physiological state. For example, a state with high variability in HR and EDA might hint towards argumentation during the collaboration (Cluster 2); while the stable episodes with TEMP, EDA and BVP might indicate agreement (Clusters 1 and 4). Cross-recurrence among the participants in a group depicts the physiological synchrony among them. We observe that this synchrony is positively and significantly correlated to the groups' perceived satisfaction and performance. One plausible reason for this might be the fact that having similar physiological states over time indicates similar activities done by every member in the group and hence every one of them had higher ratings for satisfaction and performance. On the other hand, groups with low physiological synchrony might have subgroups within them, and the different subgroups might have different perceptions for performance and satisfaction, this could result in lower average ratings for these two variables. The lack of correlation between the perceived usefulness and the physiological synchrony among the peers could be because of the usefulness could be a construct which is not clearly linked to the physiological data; clearly more experimentation is required to disambiguate these two reasons. The Cross-recurrence can also be used as a temporal measure of synchrony. This provides the possibility to use wearable sensing to automatically capture students' physiological states early in the collaborative sessions. This generates new opportunities for wearable technologies to quantify the learning experiences. Temporal CR of the physiological states might provide early feedback to both students and teachers. Thus, it makes possible to get early warnings of disruptions in collaborations and allows students and teachers to take proactive remedial actions.

Empatica E4 devices are accurate, unobtrusive and require low maintenance; thus, facilitating ubiquity. Eye-tracking glasses are wearable solutions, but there is a cost to accuracy correlation (more accurate off-the-shelf solutions are expensive). Eye-tracking glasses require low maintenance but have certain degree of obtrusiveness. Considering EEG, wearable solutions are available, but they are uncomfortable and obtrusive; EEG is also noise prone and low-cost solutions provide low accuracy and low data resolution. With our results, we show new trends of the shift the three directions (see Introduction). First, we show that unsupervised clustering methods have a potential to be used instead of human labeling of collaborative episodes. Second, we show that synchrony of semantic-less physiological states correlates with the learning experiences. Finally, the third shift towards more pervasiveness, we show that it is possible to quantify collaborative learning experiences using only the wristband data and using similar methods to that from collaborative eye-tracking studies.

References

Blanchard, E., Chalfoun, P., & Frasson, C. (2007). Towards advanced learner modeling: discussions on quasi real-time adaptation with physiological data. In Seventh IEEE International Conference on Advanced Learning Technologies (pp. 809-813).

Chanel, G., Bétrancourt, M., Pun, T., Cereghetti, D., & Molinari, G. (2013). Assessment of Computer-Supported Collaborative Processes Using Interpersonal Physiological and Eye-Movement Coupling. In Proceedings of Humaine Association Conference on Affective Computing and Intelligent Interaction (pp. 116-122).

Eckmann, J. P., Kamphorst, S. O., & Ruelle, D. (1987). Recurrence plots of dynamical systems. EPL, 4(9), 973.

Gray, J. A., & M. Diloreto. (2016). "The Effects of Student Engagement, Student Satisfaction, and Perceived Learning in Online Learning Environments." NCPEA International Journal of Educational Leadership Preparation 11 (1): 98–119.

Green, A. M. (1998). Project-based Learning: Moving Students through the GED with Meaningful Learning.

Jermann, P., & Nüssli, M. A. (2012). Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (pp. 1125-1134). ACM.

Kuvaas, B. (2006). Work performance, affective commitment, and work satisfaction: The roles of pay administration and pay level. Journal of Organizational Behavior, 27, 365–385.

Lenschow, R. J. (1998). From teaching to learning: A paradigm shift in engineering education and lifelong learning. *European Journal of Engineering Education*, *23*(2), 155-161.

Nüssli, M. A. (2011). Dual eye-tracking methods for the study of remote collaborative problem solving.

Pijeira-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2016). Investigating collaborative learning success with physiological coupling indices based on electrodermal activity. In Proceedings of the sixth international conference on learning analytics & knowledge (pp.64-73). ACM.

Preuss, D. A. (2002). Creating a project-based curriculum. *Tech Directions*, *62*(3), 16.

Sánchez, R. A. & Huero, A. D. (2010). Motivational factors that influence the acceptance of Moodle using TAM. Computers in Human Behavior, 26, 6, 1632–1640.

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2016). Using mobile eye-trackers to unpack the perceptual benefits of a tangible user interface for collaborative learning. ACM Transactions on Computer-Human Interaction (TOCHI), 23(6), 39.

Sottilare, R. A., Brawner, K. W., Goldberg, B. S., & Holden, H. K. (2012). The generalized intelligent framework for tutoring. Orlando, FL: US Army Research Laboratory–Human Research & Engineering Directorate.

Wang, R., Wang, W., daSilva, A., Huckins, J. F., Kelley, W. M., Heatherton, T. F., & Campbell, A. T. (2018). Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(1), 43.