What does shared understanding in students' face-to-face collaborative learning gaze behaviours "look like"?

Qi Zhou[1], Wannapon Suraworachet[1], Oya Celiktutan[2], Mutlu Cukurova[1]

[1] University College London, London, UK
qtnvqz3@ucl.ac.uk
[2] King's College London, London, UK
oya.celiktutan@kcl.ac.uk

Abstract. Several studies have shown a positive relationship between measures of gaze behaviours and the quality of student group collaboration over the past decade. Gaze behaviours, however, are frequently employed to investigate i) students' online interactions and ii) calculated as cumulative measures of collaboration, rarely providing insights into the actual *process* of collaborative learning in *real-world settings*. To address these two limitations, we explored *the sequences of students' gaze behaviours* as a process and its relationship to *collaborative learning in a face-to-face environment*. Twenty-five collaborative learning session videos were included from five groups in a 10-week post-graduate module. Four types of gaze behaviours (i.e., gazing at peers, their laptops, tutors, and undefined objects) were used to label student gaze behaviours and the resulting sequences were analyzed using the Optimal Matching (OM) algorithm and Ward's Clustering. Two distinct types of gaze patterns with different levels of shared understanding and collaboration satisfaction were identified, i) peer-interaction focused (PIF), which prioritise social interaction dimensions of collaboration and ii) resource-interaction focused (RIF) which prioritise resource management and task execution. The implications of the findings for automated detection of students' gaze behaviours with computer vision and adaptive support are discussed.

Keywords: Learning Analytics, Face-to-face Collaborative Learning, Gaze Behaviours, Process Mining, Computer Vision

1 Introduction

In recent years, multiple data sources and analytics techniques have been applied to extract insights from collaborative learning settings. However, the majority of existing research focuses on log data of student interactions in digital settings, followed by questionnaires and verbal documentation which are then analysed with descriptive and inferential statistics [1]. As presented in a recent systematic review on social learning, the dominant analytical approach researchers use is social network analysis, followed by inferential statistics and the dominant data source used is students' online traces [2] while almost completely ignoring what is happening outside of the digital space. Nevertheless, the overreliance on digital traces from a single platform arguably provides insufficient information, overlooks learning as an ecosystem [1], and undervalues many real-world social context complexities that are crucial for social learning [2]. Investigation of students' real-world nonverbal behaviours from video data and computer vision techniques is an understudied area for AIED. Here, we investigated the sequences of students' gaze behaviours from a real-world face-to-face collaborative learning activity from videos and analysed their relationship to perceived shared understanding.

2 Background Research on Gaze Behaviours in Collaboration

Gaze behaviours are considered to be a crucial element for the building of shared understanding in collaborative learning. Learners use gaze to streamline speech, co-present, and disambiguate and direct others' attention. In eye-tracking research, gaze behaviours have been shown to have good potential for understanding and predicting the quality of collaboration through different measures such as joint visual attention (JVA) [3], gaze overlap [4], and attention similarity [5]. However, these features, which measure the cumulative frequency of whether learners are looking at the same object, can hardly be used to represent the complex process of disambiguating and directing attention [6]. As Fan and colleagues [7] argued, the establishment of "shared attention" in social contexts through gaze, consists of a sequence of gaze behaviours from involved agents rather than being a single act. It usually requires initial mutual attention in time, referring to the point of attention, following the reference, and shared attention. Considerations of gaze behaviours as a process might provide better insights into students' collaborative learning but are rarely considered in educational research studies.

Most existing gaze behaviour investigations in collaborative learning research come from eye-tracking studies. However, existing studies on gaze behaviours in collaborative learning are limited due to various inherent challenges. Firstly, limited by equipment and technology, most of the studies looked at collaboration in digital learning environments [8]. Previous work has used eye trackers [9] or markers in the real world [10] to capture learners' attentive region. These studies illustrated the close relationship between learners' visual attention and their collaborative learning outcomes. However, they focused more on the visual attention in the collaborative working space rather than the attention among peers, which also has been considered an important gaze behaviour during collaborative learning [11]. Secondly, nearly all published studies were conducted in a laboratory context rather than investigating natural real-world learning environments [10]. The effectiveness of the identified proxies has not been studied in an ecological setting which may have more interference and may be longer in duration than in studied experimental conditions.

Yet, understanding gaze communication dynamics *in face-to-face collaborative learning settings and interpreting students' gaze behaviours from video data with computer vision are understudied*. In this paper, we present a novel representation of gaze communication dynamics specific to real-world collaborative learning environments, which has significant implications for developing novel computer vision algorithms and AIED tools to provide timely and useful interventions and feedback.

3 Methodology

3.1 Context of study, Data Collection, and Pre-processing

The data was collected from a 10-week postgraduate module. Students were assigned into groups of 4 or 5 students with interdisciplinary backgrounds, mixed-gender and varied first languages. Within each week, students were requested to attend a 1-hour face-to-face session to discuss and complete a weekly task collaboratively on Miro (miro.com) which they accessed through their laptops/tablets.

During the sessions, students were seated as a group around a T-shaped table, facing a camera. Twenty-three sessions, lasting from about 33 minutes to about 67 minutes, have been used as the final dataset in this study. The first frames of each second from a particular session were extracted to generate a new video for the labelling of gaze behaviours in the analysis. After each session, students were asked to fill in a post-survey with 5-points Likert scale questions about their shared understanding. Ethics approval was received from the institution and individual consents were given by students before the start of the study.

3.2 Coding the Gaze Behaviours

We categorized the learners' gaze behaviours into four main categories: looking at a student (S), looking at a laptop (L), looking at a tutor (T), and looking at other objects (O). To be more specific, code S refers to the gaze behaviours of a student looking at another student in the same group. The learners in the group were labelled from 1 to n, where n is the number of students within the group. By using the code S1 to Sn, the actual learner who has been gazing at can be identified. Code L represents situations when the learner was looking at the laptop on the desk. L1 is used when the learner was looking at his/her own laptop while L2 is used when looking at another member's laptop. Code T refers to a situation when the learner was looking at the tutor who appeared in the video. Code O is used when the learner was looking at other objects which have not been defined above. For example, learners who were looking at their own gestures while speaking, or looking at food/cups on the table would be coded as O.

Computer Vision Annotation Tool (CVAT) (cvat.org) tool was used for video annotation. The coding scheme was implemented by two researchers. A sample video of 1000 frames was coded by both to achieve the consensus of coding with high reliability (Cohen's Kappa = 0.98).

3.3 Feature Engineering from Labelled Gaze Behaviours and Analysis

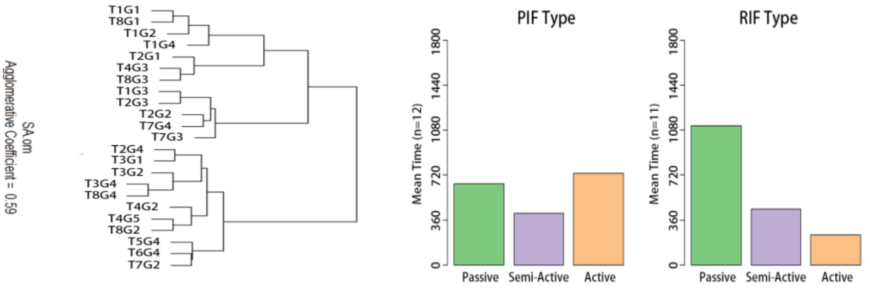
The shaping of shared understanding does not happen in a single gaze moment and requires to be analysed as a process. Here, we engineered a process feature named Shared Attention (SA) as a proxy to measure whether learners shared gaze attention in a specific time period. Ten-frame windows (representing ten seconds in original videos) were used to generate the process-based feature. In a specific window, the students who have been gazed at by over half of the students and the students who gazed at them were marked as “1”, which means they might participate in building shared attention. To increase the accuracy of the processing, the overlapping window method was used. The window size was chosen as ten frames and the window was moved two frames further for each time. The output SA sequence for the whole group is consist of the ratio of shaping shared gaze attention for each frame in this session.

The Optimal Matching Algorithm (OMA) was applied to explore the gaze behaviour sequences. Based on the distance matrix obtained from OMA, further cluster analysis was applied. Before implementing OMA, numerical values in SA sequences were converted into codes. The numerical value “0” in the original SA sequences was labelled as “passive (P)” since learners showed no shaping shared gaze attention when this ratio is 0. On the contrary, the value “1” was coded as “active (A)”. The values between 0 and 1 were coded as “Semi-active (S)”. Since students had to follow the same set of activities regardless of their sessions and the length of activities varied, to avoid value loss, the first thirty minutes of each sequence were used. In total, 23 sequences with 1800 frames were included in the analysis.

A 23X23 matrix was the output of OMA at the session-level. Each cell in the matrix represented the “distance” between the following sequences. Then, Ward’s Clustering was applied to hierarchically cluster the sequences with similar patterns across sessions. The agglomerative coefficient, which reflects the tightness of clustering, was 0.59.

4 Results and Discussion

Fig. 1(a) shows the clusters of gaze behaviour sequences in sessions. According to this tree graph, we divided 23 input sessions into two types.



(a) (b)

Fig. 1. (a) A hierarchical tree represents the result from Ward’s clustering in which T represents a task number and G represents a group number. (b) A relative frequency of codes (1=Passive, 2=Semi-active, 3=Active) in each cluster (Type 1 and Type 2).

The first type contains the top 12 sessions and the second type contains the bottom 11 sessions on the tree. Fig.1(b) shows the frequency of 3 codes in these two types. The green, purple and orange bars represent “Passive”, “Semi-active”, and “Active” shared gaze periods respectively. According to Fig. 1(b), these two types have a similar frequency of “semi-active” states. Meanwhile, type 1 presents more frequency of being “active” than type 2. The “active” status in type 1 appeared more frequently which means a longer period of active state was achieved compared to type 2. It can also be inferred that the shared gaze attention lasted longer in type 1 sessions. In other words, students from type 1 exhibited patterns of longer shared gaze periods. On the contrary, students from type 2 tended to focus more on completing the task on their Miro boards and gazing at their laptops. Type 1 sequences of gaze behaviours might be better associated with interactive and socio-emotional dimensions [12] of collaborative learning. These gaze sequences are more likely to occur when students are interacting with peers, actively listening to others, encouraging participation and inclusion of peers etc. On the other hand, Type 2 gaze behaviours may be better associated with the behavioural and regulative dimensions [13]. These gaze sequences are more likely to occur while students are doing resource management, taking actions on their laptops and during task execution phases. Therefore, we named type 1 sequences as the peer interaction focused (PIF) type and type 2 as the resource interaction focused (RIF) type. It is worth noting that, the types of tasks and groups did not show significantly different distribution between PIF and RIF patterns (Fig.1 (a)). This illustrates the potential of these sequences to be task and group size-independent features.

The PIF type (m = 3.54, SD = 0.41) and the RIF type (m = 3.54, SD = 0.31) did not show statistically significant difference in terms of their perceived shared understanding (SU). It means that a higher frequency of shared gaze attention with peers may not always lead students to perceive a better shared understanding in collaboration. Rather, groups that lack shared understanding might spend long periods of PIF sequences of gaze behaviours, trying to establish a shared understanding. Meanwhile, the shared understanding values of the PIF type are distributed wider than in the RIF type. Previous research illustrated that JVA (measured as overlapping gaze areas) had a significantly positive relationship with shared understanding in collaboration [6]. However, this result may mainly reflect that students who already have established a shared understanding are more likely to overlap in their gaze areas in collaborative learning tasks. On the other hand, if students are initially trying to build such shared understanding this might require extended periods of peer-interaction focused sequences of gaze behaviours.

5 Conclusion

In this paper, we identified two distinct types of gaze behaviour patterns of students from twenty-three face-to-face collaborative learning sessions. Peer-interaction focused (PIF) patterns, which prioritise social interaction dimensions of collaboration, might lead to a more shared understanding and higher satisfaction for students compared to resource-interaction focused (RIF) patterns, which prioritise resource management and task execution. This work has significant implications for developing novel computer vision algorithms and hence designing fully automatic behavioural analytics tools to provide intervention and feedback in real-world learning environments.

References

1. Mangaroska, K., Giannakos, M.: Learning Analytics for Learning Design: A Systematic Literature Review of Analytics-Driven Design to Enhance Learning. IEEE Trans. Learn. Technol. 12, 516–534 (2019). https://doi.org/10.1109/TLT.2018.2868673.
2. Kaliisa, R., Rienties, B., Mørch, A.I., Kluge, A.: Social learning analytics in computer-supported collaborative learning environments: A systematic review of empirical studies. Comput. Educ. Open. 3, 100073 (2022). https://doi.org/10.1016/j.caeo.2022.100073.
3. Schlösser, C., Schlieker-Steens, P., Kienle, A., Harrer, A.: Using Real-Time Gaze Based Awareness Methods to Enhance Collaboration. In: Baloian, N., Zorian, Y., Taslakian, P., and Shoukouryan, S. (eds.) Collaboration and Technology. pp. 19–27. Springer International Publishing, Cham (2015).
4. D'Angelo, S., Gergle, D.: An Eye For Design: Gaze Visualizations for Remote Collaborative Work. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–12. ACM, Montreal QC Canada (2018). https://doi.org/10.1145/3173574.3173923.
5. Papavlasopoulou, S., Sharma, K., Giannakos, M., Jaccheri, L.: Using Eye-Tracking to Unveil Differences Between Kids and Teens in Coding Activities. In: Proceedings of the 2017 Conference on Interaction Design and Children. pp. 171–181. ACM, Stanford California USA (2017). https://doi.org/10.1145/3078072.3079740.
6. Sharma, K., Olsen, J.K., Verma, H., Caballero, D., Jermann, P.: Challenging Joint Visual Attention as a Proxy for Collaborative Performance. 8 (2021).
7. Fan, L., Wang, W., Zhu, S.-C., Tang, X., Huang, S.: Understanding Human Gaze Communication by Spatio-Temporal Graph Reasoning. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 5723–5732. IEEE, Seoul, Korea (South) (2019). https://doi.org/10.1109/ICCV.2019.00582.
8. D'Angelo, S., Schneider, B.: Shared Gaze Visualizations in Collaborative Interactions: Past, Present and Future. Interact. Comput. 33, 115–133 (2021). https://doi.org/10.1093/iwcomp/iwab015.
9. Yang, C.-W., Cukurova, M., Porayska-Pomsta, K.: Dyadic joint visual attention interaction in face-to-face collaborative problem-solving at K-12 Maths Education: A Multimodal Approach. 10.
10. Sung, G., Feng, T., Schneider, B.: Learners Learn More and Instructors Track Better with Real-time Gaze Sharing. Proc. ACM Hum.-Comput. Interact. 5, 1–23 (2021). https://doi.org/10.1145/3449208.
11. Joiner, R., Scanlon, E., O'Shea, T., Smith, R.B., Blake, C.: Evidence from a series of experiments on video-mediated collaboration: does eye contact matter? In: Proceedings of CSCL. p. 371. https://doi.org/10.3115/1658616.1658669.
12. Rogat, T.K., Adams-Wiggins, K.R.: Interrelation between regulatory and socioemotional processes within collaborative groups characterized by facilitative and directive other-regulation. Comput. Hum. Behav. 52, 589–600 (2015). https://doi.org/10.1016/j.chb.2015.01.026.
13. Malmberg, J., Järvelä, S., Järvenoja, H.: Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. Contemp. Educ. Psychol. 49, 160–174 (2017). https://doi.org/10.1016/j.cedpsych.2017.01.009.