MULTIFOCUS: MULTImodal learning analytics FOr Co-located collaboration Understanding and Support

Sambit Praharaj, Maren Scheffel, Hendrik Drachsler and Marcus Specht

Open Universiteit, Valkenburgerweg 177, 6419AT Heerlen, The Netherlands
firstname.lastname@ou.nl

Abstract. This PhD research project has multiple focus points. Using the help of Multimodal Learning Analytics with the help of sensors to understand how co-located collaboration takes place, identifying the indicators of collaboration (such as pointing at peer, looking at peer, making constructive interruptions, etc.) and designing a collaboration framework model which defines the aspects of successful collaboration. These insights will help us to build the support framework to enable efficient real-time feedback during a group activity to facilitate co-located collaboration.

Keywords: multimodal indicators, multimodal interaction, multimodal learning analytics, co-located collaboration, feedback, cscl

1 Introduction

Collaboration is an important skill in the 21st century [4]. It occurs when two or more persons work towards a common goal [9]; and when this goal is associated with learning then it is known as collaborative learning (CL) [6]. CL can occur remotely or in co-located (i.e. face-to-face) settings. Our primary focus is on co-located collaboration (CC). CC can take place in different *contexts* like: collaborative brainstorming [18], collaborative programming [7], collaborative meeting [8] and collaborative problem solving [2]. Based on past research in CC, a list of multiple *dimensions* of collaboration (i.e. mutual understanding, dialogue management, reaching consensus, task division, etc.) has been outlined in [10]. These dimensions can be a proxy to measure the collaboration levels. Most of the studies [3, 11] in CC employ human observers which is quite laborious and not effective. With the ubiquity of sensors, the use of semi-automated mechanisms like sensor-based tracking in collaborative learning have taken up pace [2, 7]. These sensors can be anything from a simple microphone or a camera to a more complex integrated type of sensor like Kinect which can function simultaneously as an infrared, depth, audio and video sensor. So, Kinect can track gestures, postures, facial expressions and audio characteristics simultaneously [7, 17].

The data traces collected with the help of these sensors can be useful to analyze characteristics of individual group members and gain meaningful insights.

As this data is obtained from multiple modalities (like audio, video and depth), this work has recently been linked to the term “Multimodal Learning Analytics” (MMLA) [5]. Out of the different settings in which MMLA has been used to evaluate successful CC, it is difficult to determine the indicators like synchrony in participation [2] and joint visual attention [13]. Moreover, the challenge is the identification and interpretation of these multimodal indicators in real-time. Given these circumstances, these indicators could be used to facilitate CC with the help of indirect or direct feedback which is supported by MMLA.

Some previous works used MMLA in CC. Schneider and Blikstein [12] used Tangible User Interface (TUI) for pairs of students to predict learning gains by analyzing data from multimodal learning environments. They tracked the gesture and posture using a Kinect Sensor (Version 1) which can track the posture and gesture of a maximum of four students at a time based on their skeletal movements. They found that hand movements and posture movements (i.e. coded as active, semi-active and passive) are correlated with learning gains. Even the number of transitions between these three phases was a strong predictor of learning. Students who used both hands showed higher learning gains. The logs obtained from the TUI activities (like the frequency of opening the information box in the TUI) was associated with learning gain.

Besides, eye gaze can be a good indicator of collaboration as found by Schneider and Pea [13]; they found that (JVA) Joint Visual Attention (i.e. the proportion of times gazes of individuals are aligned by focusing on the same area in the shared object or screen) is a good predictor of the quality of collaboration of a group which is reflected by the group's performance. Schneider et al. [15] got the same results by replicating the experiment in a co-located setting. The work by Schneider and Pea [14] used JVA, network analysis and machine learning to determine different dimensions of a good collaboration (like mutual understanding, dialogue management, division of task, signs of coordination).

Moving on to the different contexts in which CC has been studied, Spikol et al. [17] studied CL in the context of *collaborative problem solving* (CPS) using MMLA. They used a combination of hand movements, head direction and physical engagement (by coding 0 for passive, 1 for semi-active and 2 for active) to detect synchrony. Another work by Grover et al. [7] studied collaborative problem solving in a *pair programming* context based on a pilot study. They captured data from different modalities (i.e. video, audio, clickstream and screen capture) unobtrusively using Kinect and other tools. For initial training of the classifiers using machine learning, experts coded the video recordings with three annotations (i.e. High, Medium and Low) when they found evidences of collaboration (i.e. pointing to the screen, grabbing the mouse from the partner and synchrony in body position) between the dyads.

Most of the studies using MMLA in Co-located Collaboration (CC) either analyzed the collected data as a post-processing measure to determine the multiple dimensions of collaboration; or used a mere reflection mechanism to act as a feedback for the group members during collaboration. For instance, Bachour et al. [1] designed the “Reflect” table to address the problem of unequal audio participation; they made every group member aware about their total speaking time with the help of a LED light feedback display on the table. Besides, there have been works where feedback in a group setting was managed by human moderators. Groupgarden [18] was one such example of a metaphorical feedback system which supported co-located brainstorming in a group.

To sum up, CC can be composed of multiple dimensions. Based on our study, synchrony, engagement, participation of students, visual attention of students in a collaborative learning scenario have been detected before using different multimodal cues like finger pointing, head movement, sitting posture, hand movement, eye gaze, etc. Most experiments used a post-hoc feedback mechanism while some used a reflection mechanism for the group. We can outline some of the drawbacks in the previous studies as: 1) There is a *large gap* between the theory surrounding the dimensions of collaboration and the dimensions of collaboration detected by sensors. Although theoretically there are multiple dimensions of collaboration, only a few of them have been detected so far with the help of sensors; 2) Most of the feedback systems built to facilitate collaboration provide a *post-hoc* feedback or a real-time *reflective* feedback; this does not promote active involvement of collaborating members rather assumes that delayed-reflection or self-reflection will facilitate collaboration; 3) There has been a *dearth* of studies on automated multimodal analysis in non-computer supported environments [19].

To this end, we seek answers to the following research questions:

RQ1 What *multimodal indicators* (MI) can give evidences of *quality* of collaboration in a CC setting?
1a What are the *dimensions* (or indexes) of co-located collaboration?
1b How can we define the *mapping* between the multimodal indicators and the dimensions of collaboration?
RQ2 How can we *measure* MI in CC with sensor technology and build efficient multimodal data models to enable real-time data aggregation and analysis?
2a How can we create a *framework* (or vocabulary) to annotate the collaboration indicators based on different multimodal channels?
RQ3 How can we enable efficient *real-time feedback* supported by MMLA to facilitate CC?
3a How can we understand different *levels* of collaboration in co-located settings using a combination of different modalities in a continuous fashion?
3b How do we decide on the *level* (i.e. individual or group) and *type* (i.e. private, public or mixed type display) of real-time feedback?

The remainder of the paper is structured as follows: in the *challenges* (Sect. 2) section we outline the main challenges of the research project; it is followed by an explanation of our proposed *methodology* (Sect. 3); finally, *conclusions* (Sect. 4) are drawn.

2 Challenges

We have outlined the main research challenges as follows:

C 1 – *Designing the collaboration task and its outcomes* – Define the complexity and nature of the task and the possible outcomes. Before narrowing down the indicators, we need to narrow down the context of collaboration.

C 2 – *Annotating the multimodal data set* – We will need human annotators hired by crowdsourcing to annotate the large data set. We will also need to use an annotation tool or interface to annotate the data sets. For training these large data sets we need to use semi-supervised learning later.

C 3 – *Architectural design* – Designing an architecture which collects, processes and predicts by taking help of different modalities in real-time. For this, we will need to use a Deep Neural Network (DNN) architecture.

C 4 – *Efficient feedback system* – Design a real-time feedback or intervention system which is efficient and works with minimum latency. A decision also needs to be made on the level of display which can range from a basic private mobile phone display to a public display.

3 Methodology

We have sub-divided the planned methodology into multiple tasks as:

Task 1: *Extensive systematic literature study* – First, to answer RQ1 we are conducting an extensive literature study to determine the multimodal indicators and dimensions of collaboration quality. Then, we try to determine the suitable mappings for each collaboration indicators to different dimensions of collaboration. For the systematic literature study, we have come up with this search term after multiple iterations: *'multimodal indicators' + 'multimodal learning analytics' + 'collaborative' + 'quality of collaboration'*.

Task 2: *Formative study with the prototype and collection of the training data* – We will conduct a pilot study (to answer RQ2) in a small room in co-located settings with around 3-6 members in a group performing some collaborative task like a group meeting where we use Kinect(s) and microphones to detect various multimodal cues. Later we will conduct design-based workshops to look into different situations of collaboration and the different indicators and feedback mechanisms that can be associated with those situations. Then we use these indicators to answer RQ3. In the meantime, we break down the prototype design into different simple use-cases where we track each multimodal indicator alongwith a simple feedback mechanism. For instance, in one use-case we can track the audio characteristics (i.e. total speaking time and turn taking while speaking) using only a microphone.

In later stages, we need to design a Deep Neural Network (DNN) classifier which can work in different domains and predict the level of collaboration. We need to train this DNN classifier with large number of data sets collected from these multiple case studies. Before training, we need to use feature engineering to extract the important features which are fed as input to the DNN classifier. For example, to process an audio stream, it is fed as an input to the classifier; next step is feature extraction where we extract different features like pitch, amplitude, number of pauses, etc.; then the classifier is trained based on these extracted features and later makes predictions on the level of collaboration as the output.

Task 3: *Study for accuracy of the DNN classifiers* – We need to compare the predictions of the classifier to the ground truth data (which is made from the annotations given by the human observers for the video recordings). From that we can determine the precision and recall to determine the accuracy of the classifier. This is essential to predict the level of collaboration as efficiently as possible in real-time or near to real-time. This will help us to answer RQ3a.

Task 4: *Summative study on shaping behaviour of group members* – We need to determine if there is any effect of real-time feedback on the level of collaboration based on the multimodal cues observed during collaboration. To enrich this, we need to gather feedback with the help of feedback questionnaires (on the nature, type and effect of real-time feedback during collaboration). This can act as an indicator level on the possible effects of real-time feedback during collaboration. It can help us to modify the real-time feedback accordingly. This will help us to answer RQ3b.

4 Conclusions

Collaboration being an important skill and ubiquitously present in our day to day activities, we plan to build a real-time feedback mechanism (or support framework) for the PhD research project to facilitate collaboration in real-time. Contrary to past research, we plan to implement a real-time feedback to guide the collaborators. Thus, we want to move from a *mirroring feedback* mechanism to a *guiding feedback* mechanism in a co-located collaboration setting [16]. Besides, our aim is to bridge the gap between the theory and practical aspects encompassing CC. In future, our plan is to offer the system as a *Collaboration Coach* service in classroom or other group settings.

References

1. Bachour, K., Kaplan, F., Dillenbourg, P.: An interactive table for supporting participation balance in face-to-face collaborative learning. IEEE Transactions on Learning Technologies 3(3), 203–213 (2010)
2. Cukurova, M., Luckin, R., Mavrikis, M., Millán, E.: Machine and human observable differences in groups' collaborative problem-solving behaviours. In: European Conference on Technology Enhanced Learning. pp. 17–29. Springer (2017)
3. Davidsen, J., Ryberg, T.: "this is the size of one meter": Children's bodily-material collaboration. International Journal of Computer-Supported Collaborative Learning 12(1), 65–90 (2017)
4. Dede, C.: Comparing frameworks for 21st century skills. 21st century skills: Rethinking how students learn 20, 51–76 (2010)
5. Di Mitri, D., Klemke, R., Drachsler, H., Specht, M.: Towards a real-time feedback system based on analysis of multimodal data (2017)

6. Dillenbourg, P.: What do you mean by collaborative learning? (1999)
7. Grover, S., Bienkowski, M., Tamrakar, A., Siddiquie, B., Salter, D., Divakaran, A.: Multimodal analytics to study collaborative problem solving in pair programming. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge. pp. 516–517. ACM (2016)
8. Kim, T., Chang, A., Holland, L., Pentland, A.S.: Meeting mediator: enhancing group collaboration using sociometric feedback. In: Proceedings of the 2008 ACM conference on Computer supported cooperative work. pp. 457–466. ACM (2008)
9. Martinez-Moyano, I.: Exploring the dynamics of collaboration in interorganizational settings. Creating a culture of collaboration: The International Association of Facilitators handbook 4, 69 (2006)
10. Meier, A., Spada, H., Rummel, N.: A rating scheme for assessing the quality of computer-supported collaboration processes. International Journal of Computer-Supported Collaborative Learning 2(1), 63–86 (2007)
11. Scherr, R.E., Hammer, D.: Student behavior and epistemological framing: Examples from collaborative active-learning activities in physics. Cognition and Instruction 27(2), 147–174 (2009)
12. Schneider, B., Blikstein, P.: Unraveling students' interaction around a tangible interface using multimodal learning analytics. Journal of Educational Data Mining 7(3), 89–116 (2015)
13. Schneider, B., Pea, R.: Real-time mutual gaze perception enhances collaborative learning and collaboration quality. International Journal of Computer-supported collaborative learning 8(4), 375–397 (2013)
14. Schneider, B., Pea, R.: Toward collaboration sensing. International Journal of Computer-Supported Collaborative Learning 9(4), 371–395 (2014)
15. Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., Pea, R.D.: 3d tangibles facilitate joint visual attention in dyads. In: Proceedings of 11th International Conference of Computer Supported Collaborative Learning. vol. 1, pp. 156–165 (2015)
16. Soller, A., Martínez, A., Jermann, P., Muehlenbrock, M.: From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. International Journal of Artificial Intelligence in Education 15(4), 261–290 (2005)
17. Spikol, D., Ruffaldi, E., Cukurova, M.: Using multimodal learning analytics to identify aspects of collaboration in project-based learning. Philadelphia, PA: International Society of the Learning Sciences. (2017)
18. Tausch, S., Hausen, D., Kosan, I., Raltchev, A., Hussmann, H.: Groupgarden: supporting brainstorming through a metaphorical group mirror on table or wall. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. pp. 541–550. ACM (2014)
19. Worsley, M., Blikstein, P.: Leveraging multimodal learning analytics to differentiate student learning strategies. In: Proceedings of the Fifth International Conference on Learning Analytics And Knowledge. pp. 360–367. ACM (2015)