Detection of Collaboration: Relationship Between Log and Speech-Based Classification

Sree Aurovindh Viswanathan(✉) and Kurt Vanlehn(✉)

Arizona State University, Arizona, USA
{sviswal0,kvanlehn}@asu.edu

Abstract. Research in the field of collaboration shows that students do not spontaneously collaborate with each other. A system that can measure collaboration in real time could be useful by, for example, helping the teacher locate a group requiring guidance. To address this challenge, my research focuses on building and comparing collaboration detectors for different types of classroom problem solving activities, such as card sorting and hand writing. I am also studying transfer: how collaboration detectors for one task can be used with a new task. Finally, we attempt to build a teachers dashboard that can describe reasoning behind the triggered alerts thereby helping the teachers with insights to aid the collaborative activity. Data for building such detectors were collected in the form of verbal interaction and user action logs from students' tablets. Three qualitative levels of interactivity was distinguished: Collaboration, Cooperation and Asymmetric Contribution. Machine learning was used to induce a classifier that can assign a code for every episode based on the set of features. Our preliminary results indicate that machine learned classifiers were reliable.

Keywords: Collaborative learning · Machine learning · Learning analytics

1 Introduction and Problem Statement

Collaboration is a 21st century skill as well as an effective method for learning [1, 2]. However, collaboration between students is not spontaneous and acquiring collaboration skills is not straightforward. Several theoretical frameworks of collaboration [3–5] connect variations of social interactions to effectiveness of learning. Various dimensions of effective collaboration have been identified in the literature [6, 7]. Transactivity has been identified as one of the important characteristics of collaboration grounded in frameworks of Piaget [8] and Vygotsky [9], and it has shown to facilitate acquiring domain knowledge [1]. Chi's ICAP framework [5] includes transactive process in it category Interactive. Of the four categories of overt behavior, Interactive process foster the most learning.

Many projects have worked on the challenge of automating the analysis of interaction among group members. These antecedents will be briefly reviewed by defining two dimensions, *purpose* and *input*, then describing the few systems whose position along these two dimension match the position of the project reported here. The two dimensions are excerpted from several similar multi-dimensional reviews [10, 11].

S. Isotani et al. (Eds.): AIED 2019, LNAI 11626, pp. 327–331, 2019.
https://doi.org/10.1007/978-3-030-23207-8_60

When a large number of projects could be cited as illustrations of a dimension, only those published most recently will be cited.

The first dimension concerns the purpose or function of the collaboration measure. That is, what does the system do with the output of the collaboration detector? This dimension has the following categories: Clustering, Classification, Mirroring, Meta-cognitive, Guiding, Orchestration and Restructuring. Our project fits into two of the categories: *Classification* and *Orchestration*. Projects in classification category [12, 13] used human judges to code group interactions into a variety of collaboration categories, then used supervised machine learning methods to induce classifiers (also called detectors). The main research questions is: how accurate is the induced detector? The projects in Orchestration categories [14] display the amount of collaboration per group on a dashboard held by the teacher. This allows the teacher to visit groups that need help collaborating. The main research question is whether such collaboration detection is useful to the teacher and effective at increasing collaboration in the classroom.

The second dimension classifies prior work by input to the detector. All collaboration detection projects so far have students work in a shared workspace, so their detectors take the users' interactions (log data) as one input. Most projects also analyzed some form of communication among group members. The communication input can be classified as:

- Group members communicated in a formal language [15].
- Group members used a small set of buttons to express agreement/disagreement [16].
- Group members communicated by typing natural language and classifying their contribution using a menu of sentence openers or speech acts. Some systems ignored the text and used *only* the students' classifications of their text [17].
- Group members communicated via typing (chat), with or without sentence openers. The text was analyzed by human "wizards" [18], keywords [19, 20] or machine-learned text classifiers [21].
- Group members conversed in unconstrained speech, recorded by individual microphones [12, 20, 22, 23].

The design of our classification codes matches that of Chi's ICAP Framework. My thesis project falls into the *Classification* category of the purpose dimension and *unconstrained speech* in the input dimension. In addition, I am developing collaboration detectors that can generalize across different tasks. Finally, I use the collaboration codes generated by the system and the underlying data to populate a dashboard that not only shows teachers which groups are not collaborating but also explains what evidence supports its assessment.

2 Methodology and Progress

The overall dissertation work focuses on building collaboration detectors that measure the quality of collaboration in real time. Laboratory studies were conducted with more than sixty pairs of students working on two different types of tasks. In order to create and evaluate collaboration detectors, the judgments of human coders were used as the 'gold standard' classification of the group's interactions. The coders had both high quality audio and several videos to aid their judgment. Collaboration detectors were then machine-learned from the human judgments. Their accuracies were measured using 10 fold cross validation.

My thesis project is divided into the tasks briefly described below:

1. The first task involved students collaboratively working on a card-moving task which required interpreting time-distance graphs. Machine learned detectors were built by using speech and log data to measure collaboration [24]. The results were promising with a high level of agreement. However, it has to be noted that the particular task made it relatively easy to measure collaboration. (Complete)
2. The second task involved students working on a collaborative task where they were required to analyze solutions of four hypothetical students. They had to write paragraph long explanations. An in-depth analysis of video tapes and logs of tablets were performed to understand how students write on the surface of the tablets. It also highlighted the fact that superficial measures of collaboration may not be adequately useful for detection of collaboration in hand writing settings. (Complete)
3. The third task involved determining whether collaboration detection could be accurate when student voices are converted to a privacy-preserving binary signal (1 = speaking, 0 = silence) before being transmitted and stored. Data were collected as students wrote paragraphs together and solved problems. A speech signal was processed at a microphone by voice activity detector to produce the binary signal. The results indicate that binary based collaboration detectors yielded only slightly less accuracy than detectors that took the high quality audio signal as input. (Complete)
4. Whereas task 1 above showed that a log-based collaboration detector was just as accurate as speech-based collaboration detection, the card-moving task made such detection easy. This fourth task investigated log-based collaboration detection with a more common task, collaborative writing. Data came from students who analyzed mistaken problem solutions done by four hypothetical students. The students then wrote an analysis of each solution. The results indicate that log-based collaboration detection accuracy was low to moderate for this collaborative writing task. Comparing the features of the collaborative writing task to the card-moving task allows speculations on what task properties facilitate log-based collaboration detection. (Complete)
5. The fifth task will involve creating a general collaboration detector that will function well with multiple collaborative tasks. Features would be extracted from acoustic and prosodic characteristics of audio signal along with its time series characterization. If a generalized collaboration detector is reliable, then it could be used in various tasks to measure collaboration. This would help the researchers avoid the laborious work of annotating the video/audio files manually to understand the process of collaboration. (In progress)
6. Finally, in collaboration with a larger group of students, I am attempting to create a visualization dashboard that will provide insights to teachers about the collaboration based on speech and actions in collaborative group activity. It will also provide the teacher with suggestions for improving the collaboration of specific groups. (in progress).

3 Contributions and Impact

The thesis explores methods to automatically measure the types collaboration exhibited by students working together on learning activities. Collaboration detectors are based on building machine learning models of log and/or speech data. Firstly, this work complements research in collaborative learning environments with a goal to classify collaborative activity in MOOCs and other environments where students communicate in text. Secondly, if task-general classification of spoken collaboration is successful, it would reduce the laborious process of human coding required to establish reliability and would potentially allow the researchers to build various systems that utilize the underlying categories of collaboration. Finally, the proposed dashboard would provide insights into student's speech and actions with a goal of reducing teachers' cognitive overload and provides teachers with information to facilitate the classroom.

Acknowledgements. This research was funded by the Diane and Gary Tooker chair for effective education in Science Technology Engineering and Math, by NSF grant IIS-1628782, and by the Bill and Melinda Gates Foundation under Grant OP1061281.

References

1. Chen, J., Wang, M., Kirschner, P.A., Tsai, C.-C.: The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: a meta-analysis. Rev. Educ. Res. **88**(6), 799–843 (2018)
2. Vogel, F., Wecker, C., Kollar, I., Fischer, F.: Socio-cognitive scaffolding with computer-supported collaboration scripts: a meta-analysis. Educ. Psychol. Rev. **29**(3), 477–511 (2017)
3. Stahl, G.: Theories of cognition in collaborative learning. In: Hmelo-Silver, C., O'Donnell, A., Chan, C., Chinn, C. (eds.) The International Handbook of Collaborative Learning, pp. 74–90. Taylor & Francis, New York (2013)
4. Hartmann, C., Angersbach, J., Rummel, N.: Social Interaction, Constructivism and their Application within (CS) CL Theories. International Society of the Learning Sciences, Inc. [ISLS] (2015)
5. Chi, M.T., Wylie, R.: ICAP: a hypothesis of differentiated learning effectiveness for four modes of engagement activities. Educ. Psychol. **49**(4), 219–243 (2014)
6. Kahrimanis, G., et al.: Assessing collaboration quality in synchronous CSCL problem-solving activities: adaptation and empirical evaluation of a rating scheme. In: Cress, U., Dimitrova, V., Specht, M. (eds.) EC-TEL 2009. LNCS, vol. 5794, pp. 267–272. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04636-0_25
7. Meier, A., Spada, H., Rummel, N.: A rating scheme for assessing the quality of computer-supported collaboration processes. Int. J. Comput. Support. Collab. Learn. **2**(1), 63 (2007)
8. Berkowitz, M.W., Gibbs, J.C.: Measuring the developmental features of moral discussion (1982)
9. Schwartz, D.L.: The productive agency that drives collaborative learning. In: Dillenbourg, P. (ed.) Collaborative Learning: Cognitive and Computational Approaches, pp. 197–218. Elsevier Science, New York (1999)
10. Soller, A., Martinez, A., Jermann, P., Muehlenbrock, M.: From mirroring to guiding: a review of state of the art technology for supporting collaborative learning. Int. J. Artif. Intell. Educ. **15**, 261–290 (2005)

11. VanLehn, K.: Regulative loops, step loops and task loops. Int. J. Artif. Intell. Educ. **26**(1), 107–112 (2016)
12. Martinez-Maldonado, R., Kay, J., Yacef, K.: An automatic approach for mining patterns of collaboration around an interactive tabletop. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 101–110. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_11
13. Gweon, G., Agarawal, P., Raj, B., Rose, C.P.: The automatic assessment of knowledge integration processes in project teams. In: Proceedings of Computer Supported Collaborative Learning (2011)
14. Martinez-Maldonado, R., Clayphan, A., Yacef, K., Kay, J.: MTFeedback: providing notifications to enhance teacher awareness of small group work in the classroom. IEEE Trans. Learn. Technol. **8**, 187–200 (2014)
15. Tedesco, P.A.: MArCo: building an artificial conflict mediator to support group planning interactions. Int. J. Artif. Intell. Educ. **13**(1), 117–155 (2003)
16. de los Angeles Constantino-Gonzalez, M., Suthers, D.D., de los Santos, J.G.E.: Coaching web-based collaborative learning based on problem solution differences and participation. Int. J. Artif. Intell. Educ. **13**(2–4), 263–299 (2003)
17. Baghaei, N., Mitrovic, A., Irwin, W.: Supporting collaborative learning and problem-solving in a constraint-based CSCL environment for UML class diagrams. Comput. Support. Collab. Learn. **2**, 159–190 (2007)
18. Tsovaltzi, D., Rummel, N., McLaren, B., Pinkwart, N., Scheuer, O., Harrer, A., Braun, I.: Extending a virtual chemistry laboratory with a collaboration script to promote conceptual learning. Int. J. Technol. Enhanc. Learn. **2**(1/2), 91–110 (2010)
19. Dragon, T., Floryan, M., Woolf, B., Murray, T.: Recognizing dialogue content in student collaborative conversation. In: Aleven, V., Kay, J., Mostow, J. (eds.) ITS 2010. LNCS, vol. 6095, pp. 113–122. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13437-1_12
20. Martinez-Maldonado, R., Yacef, K., Kay, J.: TSCL: a conceptual model to inform understanding of collaborative learning processes at interactive tabletops. Int. J. Hum Comput Stud. **83**, 62–82 (2015)
21. Walker, E., Rummel, N., Koedinger, K.R.: Adaptive intelligent support to improve peer tutoring in algebra. Int. J. Artif. Intell. Educ. **24**(1), 33–61 (2014)
22. Gweon, G., Jain, M., McDonough, J., Raj, B., Rose, C.P.: Measuring prevalence of other-oriented transactive contributions using an automated measure of speech style accommodation. Int. J. Comput. Support. Collab. Learn. **8**(2), 245–265 (2013)
23. Bassiou, N., et al.: Privacy-preserving speech analytics for automatic assessment of student collaboration. In: Privacy-Preserving Speech Analytics for Automatic Assessment of Student Collaboration, pp. 888–892 (2016)
24. Viswanathan, S.A., VanLehn, K.: Using the tablet gestures and speech of pairs of students to classify their collaboration. IEEE Trans. Learn. Technol. **11**, 230–242 (2018)