Speech analytics on individual and group audio data to understand collaboration

Robin Jephthah Rajarathinam, Cynthia M. D'Angelo, Emma Mercier
rjrthnm2@illinois.edu, cdangelo@illinios.edu, mercier@illinois.edu
University of Illinois Urbana-Champaign

Abstract: Collaborative learning in classrooms require instructors to monitor student groups to ensure they make progress with the tasks. One way learning analytics has helped facilitating such classrooms is by providing speech-based solutions to help instructors monitor. In this poster, we investigate two different ways of collecting audio data from group work namely, group audio data and individual audio data and how voice activity detection (VAD) can be used to predict student collaboration. Both types of audio data were collected from classes focused on collaborative problem solving that were part of an introductory undergraduate engineering course. Preliminary analysis of 8 groups of audio data using VAD indicate that individual audio data could provide information regarding turn ending, turn overlap, and turn duration of individual students which can be critical in understanding the quality of collaboration of a group which cannot be obtained consistently using group audio data.

Introduction and theoretical background

Facilitating collaborative learning tasks in typical classrooms is difficult when teachers are required to monitor multiple groups with 2-4 students in each group. Ideally, teachers would be able to monitor student interactions and talk in a group long enough to understand whether the group needs help and what form that help should take. Effective intervention cannot be provided by instructors for classrooms with more than few groups (Kaendler et al., 2014). Students talking to one another in face-to-face environments is the natural setting for classroom collaboration and speech-based solutions will provide necessary information to facilitate such large classrooms (D'Angelo et al., 2019). In this work, we investigate the feasibility and use of students' speech during collaborative activity collected both as a group, and as individual streams of data, to determine the quality of group collaboration. We discuss the tradeoffs and utility of both the individual and group audio streams and how these differences relate to understanding and measuring collaboration in a classroom setting.

Methods

Audio data was collected from a 16-week introductory undergraduate engineering course's discussion sections. Three instructors, comprised of a graduate assistant and two undergraduates, facilitated each discussion session. Students worked in groups of 3 or 4 on ill-structured collaborative tasks which were developed by members of the research team to support collaboration (Shehab et al., 2017). The students worked on synced tablets which had a shared workspace to collaborate. A total of 68 students in 18 groups participated in the audio recording study over 4 weeks. Each consented student wore a head-mounted noise-cancelling microphone. For this analysis, audio data from a single week was selected because it had the most groups with 3 or 4 students who wore the head-mounted microphones properly.

Voice Activity Detection (VAD) from the openSMILE toolkit (Eyben et al. 2010) was used to discriminate human speech and non-speech sounds. The background speech from other groups is moderated using the thresholds (parameters) in VAD. The threshold was adjusted such that the turns detected by the VAD captured all the student speech for both the group and individual audio streams. The thresholds were set differently for each stream as the individual audio streams captured less background audio than the group audio stream. The accuracy of the VAD settings was determined by manually coding for speech and non-speech on (randomly selected) 10% of the audio data and then comparing it with the output produced by the VAD in the openSMILE toolkit.

Findings

Data from the group audio stream included lots of human speech from the surrounding groups in the classroom and the facilitating instructors, as well as actual background noise (e.g., chairs moving). This made the VAD pick up background noise as turns when the student group being recorded is quieter than the neighboring groups or when the instructor is intervening with a nearby group. This made selection of an appropriate threshold for the VAD complicated as the decision had to consider the tradeoff of noise being captured versus capturing all the student speech. VAD was unable to clearly segment turns from group audio data when there was overlap of speech between students.

Data from the individual audio stream included noise from within the group, surrounding groups in the classroom, and the facilitating instructors. VAD consistently identified individual turns of the speaker once the threshold for the VAD was appropriately set up. This ensured that from the individual audio stream, turn overlap, an important indicator of collaboration could be determined. VAD, as expected, picked up instructor speech when the student was interacting with an instructor who was physically located very close to the student.

Discussion

VAD provides information on the individual student turns and who is speaking during a collaborative activity. Given that the *sine qua non* of collaborative leaning is peer interaction, analysis on the speech patterns including turn overlap and turn taking have shown to be useful in predicting speech patterns (Levinson & Torreira, 2015).

Group audio streams could be used along with speaker identification software to detect individual turns. But this performs only when the students do not have overlap of turns. Overlap of turns prevents group audio data from reliably capturing turn taking. Individual audio stream allows for capturing turn taking and turn overlap which are key indicators of collaboration. Combining individual audio streams with instructor audio streams could allow for clear distinction of turns by both the individual students as well as the instructors. This will lead to a deeper understanding of collaboration.

It should be noted that collecting individual audio data can be difficult based on the tool being used. This implementation used head-mounted microphones which the students sometimes removed, possibly because they were uncomfortable wearing them. There were several instances where the students decide to stop wearing the microphones and removed them midway through the discussion section.

Conclusion and implication

Use of audio data to facilitate collaborative learning tasks in classrooms holds great promise as it allows for informed support from instructors to help the groups collaborate. Using VAD on group audio stream is useful only when there is very little ambient noise. VAD on individual audio stream allows for identification of individual student turns and overlap of turns. Measures should be taken to isolate instructor interaction with the student group. Collecting individual audio stream can be perceived as intrusive which can lead to some students declining use of these microphones. But the overall information obtained from individual audio stream could outweigh the drawbacks. Future directions involve generating a prediction model from the audio data on the state of collaboration of the student group, combined with prior work that incorporates the synchronized tablet logfile data. The work reported in the poster is the first step towards building a complex collaboration prediction model.

References

D'Angelo, C. M., Smith, J., Alozie, N., Tsiartas, A., Richey, C., & Bratt, H. (2019). Mapping individual to group level collaboration indicators using speech data. In Lund, K., Niccolai, G., Lavoué, E., Hmelo-Silver, C., Gweon, G., & Baker, M. (Eds). A Wide Lens: Combining Embodied, Enactive, Extended, and Embedded Learning in Collaborative Settings: *13th International Conference on Computer Supported Collaborative Learning*. Lyon, France: International Society of the Learning Sciences.

Eyben, F., Wöllmer, M., & Schuller, B. (2010). OpenSMILE - The Munich versatile and fast open-source audio feature extractor. MM'10 - *Proceedings of the ACM Multimedia 2010 International Conference*, 6(4), 1459–1462. https://doi.org/10.1145/1873951.1874246

Kaendler, C., Wiedmann, M., Rummel, N., & Spada, H. (2014). Teacher Competencies for the Implementation of Collaborative Learning in the Classroom: a Framework and Research Review. *Educational Psychology Review*, *27,* 505-536.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in psychology*, *6*, 731.

Shehab, S., Mercier, E., Kersh, M., Juarez, G., & Zhao, H. (2017). Designing engineering tasks for collaborative problem solving. In *Making a Difference—Prioritizing Equity and Access in CSCL: The 12th International Conference on Computer Supported Collaborative Learning*. Philadelphia: The International Society of the Learning Sciences.

Acknowledgments

This material is based upon work supported by the Campus Research Board of the University of Illinois Urbana-Champaign and the National Science Foundation under Grant No.#1628976. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the University of Illinois Urbana-Champaign or the National Science Foundation.