Final Report: Using Natural Language Processing to Study the Role of Argument Moves, Specificity and Evidence Type in Text-Based Classroom Discussions

Diane Litman and Amanda Godley

July 2016 - June 2018

Goals

Collaborative argumentation - the building of evidence-based, reasoned knowledge and solutions through dialogue - is essential to individual learning as well as group problem-solving in STEM and other disciplines and a defining characteristic of 21st century workplaces. The goal of our project was to leverage Natural Language Processing (NLP) approaches to study collaborative argumentation in student-centered, text-based classroom discussions in English language Arts.

Activities and Accomplishments

Generate research-based coding scheme for important features of student talk. We proposed a new annotation scheme for student talk during ELA "text-based" discussions - that is, discussions that center on a text or piece of literature. The annotation scheme was developed to capture four features of classroom talk theorized in the literature as important to discussion quality and learning opportunities: 1) argumentation (the process of systematically reasoning in support of an idea; coded as *claims, evidence, reasoning*), 2) specificity (the quality of belonging or relating uniquely to a particular subject; coded as *low, medium, high*), 3) knowledge domain (area of expertise represented in the content of the talk, coded as *textual* or *experiential)*, and 4) collaboration (how students build on other students' ideas, using a coding scheme that synthesized findings in both the classroom discourse and computer-supported collaborative learning literature).

Manually code secondary data and analyze inter-rater reliability. We demonstrated the reliability of our coding scheme via an annotation study of a secondary dataset of discussion excerpts that we created from 13 peer-reviewed articles and dissertations. The dataset contained 70 discussions of 32 different literary texts, where the discussions involved 19 teachers and their students in diverse U.S. schools. All discussion excerpts were double-coded by trained graduate and undergraduate students. Our results demonstrated the reliability of our coding scheme, with Kappas ranging from .65-.8 across the four features of coded student talk. In addition, our results provided a snapshot of high school ELA discussions: argumentation involved many claims but fewer descriptions of supporting evidence and explanations; specificity of the argumentation was rarely high, and collaboration involved building on each other's ideas but had few challenges or synthesis.

Analyze relationship between coded features of student talk and discussion quality. We had experts rate the overall quality of discussions in order to identify the elements of our coding scheme that best predicted discussion quality. 41 discussion excerpts from our dataset were rated for quality by 2 English Education faculty with expertise in classroom discourse, yielding Quadratic Weighted Kappa > .6. Using these quality ratings as dependent variables, we then performed stepwise regression with independent variables created from our coded features of student talk. The most predictive features of discussion quality were lower ratio of teacher to student talk, frequency of challenges/rebuttals, high specificity, and focus on text domain. Our results demonstrated that all four of our student talk features had significant relationship with discussion quality, yielding

validity evidence for our coding scheme.

Develop NLP methods to automate the coding the features of student talk. When predicting specificity, we showed that an existing general-purpose system achieved significantly better performance when its predictive features were used for retraining on educational data. We also showed that performance could be further improved by using additional features from the NLP literature, especially when combined with new neural network machine learning methods and new features tailored to text-based classroom discussion. When predicting argument moves, we showed that an existing system developed for another educationally-oriented domain performed poorly on our dataset. We then showed that feature sets from prior work on argument mining for student essays and online dialogues could be used to improve performance considerably. We also again explored the use of neural network machine learning approaches, and found that while these models were not always able to outperform traditional models (given our small dataset), we were able to gain some insights on the utility of different neural approaches. We also developed models for predicting knowledge domain, which had not previously been studied in NLP.

Contributions. With respect to literacy research, our work is a step towards enabling the exploration of hypotheses concerning discussion quality over large datasets, spanning multiple classes and including a large number of students, which would otherwise require a prohibitive amount of work for manually annotating data. In addition, we provided empirical support for existing theories regarding important features of student talk. With respect to NLP, our work showed that by tailoring our predictive models for automatically coding argumentation and specificity to dialogue in an educational context, we could outperform the state of the art.

Successful funding applications for which the LRDC monies served as a seed grant

- EAGER: Discussion Tracker: Development of Human Language Technologies to Improve the Teaching of Collaborative Argumentation in High School English Classrooms, NSF, 2018 - 2019, $149,825
- Discussion Tracker: Development of Human Language Technologies to Improve the Teaching of Collaborative Argumentation in High School, NSF, 2019 – 2022, $749,581

Graduate students supported by the grant

- Luca Lugini, School of Computing and Information
- Christopher Olshefski, School of Education

Undergraduate/FER students participating in the grant via academic credit

- Kaya Bartolomeo, Sonia Cromp, Clare Miller, Luke Profy, Anika Roy, Annika Swallen, Devaughn Wilkerson, Zinan Zhuang

Publications and Presentations

- **Amanda Godley** and **Christopher Olshefski**. 2017. The role of argument moves, specificity and evidence type in meaningful literary discussions across diverse secondary classrooms. Unpublished paper presented at *Literacy Research Association 67th Annual Conference*
- **Luca Lugini** and **Diane Litman**. 2017. Predicting specificity in classroom discussion. *Proceedings of the12th Workshop on Innovative Use of NLP for Building Educational Applications.*

- **Luca Lugini**, **Diane Litman**, **Amanda Godley**, and **Christopher Olshefski**. 2018. Annotating Student Talk 0n Text-based Classroom Discussions, *Proceedings of the 13th Workshop on Innovative Use of NLP for Building Educational Applications*
- **Luca Lugini** and **Diane Litman.** 2018**.** Argument Component Classification for Classroom Discussions, *Proceedings of the 5th Workshop on Argument Mining*
- **Amanda Godley** and **Christopher Olshefski** (2018). Characteristics of student talk in high-quality text discussions across secondary ELA classrooms. Poster presented at the 2018 Annual Meeting of the American Educational Research Association, New York, NY.
- **Amanda Godley** and **Christopher Olshefski (**2019). Promises and Limitations of Applying NLP to Classroom Discourse Analysis. Paper presented at the 2019 Annual Meeting of the American Educational Research Association, Toronto, Ontario.