Using AI to Promote Equitable Classroom Discussions: The TalkMoves Application*

Abhijit Suresh[1], Jennifer Jacobs[1], Charis Clevenger[1], Vivian Lai[1] , Chenhao Tan[1], James H. Martin[1], and Tamara Sumner[1]

[1] [1] Institute of Cognitive Science,
[2] Department of Computer Science
University of Colorado Boulder
{FirstName.Lastname}@colorado.edu

Abstract. Inclusion in mathematics education is strongly tied to discourse rich classrooms, where students ideas play a central role. Talk moves are specific discursive practices that promote inclusive and equitable participation in classroom discussions. This paper describes the development of the TalkMoves application, which provides teachers with detailed feedback on their usage of talk moves based on accountable talk theory. Building on our recent work to automate the classification of teacher talk moves, we have expanded the application to also include feedback on a set of student talk moves. We present results from several deep learning models trained to classify student sentences into student talk moves with performance up to 73% F1. The classroom data used for training these models were collected from multiple sources that were pre-processed and annotated by highly reliable experts. We validated the performance of the model on an out-of-sample dataset which included 166 classroom transcripts collected from teachers piloting the application.

Keywords: Accountable talk · deep learning· BERT· equity-focused instruction· AI application· mathematics · teacher learning· classroom discourse

1 Introduction

1.1 Overview of the TalkMoves application

The TalkMoves application[1] serves as an exemplar for a new type of translational activity enabled by big data: the reification of existing, well-researched theoretical frameworks in deep learning models. In particular, this application draws on accountable talk theory [6, 9] as a framework for providing fully automated feedback to mathematics teachers on specific discourse moves they and

* Supported by National Science Foundation under Grant No.1837986 : The TalkBack Application: Automating Analysis and Feedback to Improve Mathematics Teachers' Classroom Discourse

their students used during classroom instruction. The application is an innovative AI-driven platform that builds on advances in deep learning for natural language processing and speech recognition to automatically analyze teaching episodes and offer users near real-time feedback.

Using talk moves is an equity-focused endeavor [8]. Forming and sustaining a learning environment based on accountable talk theory can be particularly beneficial for girls and students from home backgrounds, inculturating them into the norms of democratic discourse that can later be realized in wider civic spheres [6]. Shifting away from traditional discourse patterns towards accountable talk makes space for students' contributions, especially for English Language Learners, through a focus on communicating mathematically and presenting arguments rather than acquiring vocabulary and other low-level linguistic skills [4, 7]. Furthermore, increased participation by students from non-dominant groups can foster dispositions in which attention is given to competencies and resources rather than deficits and obstacles [3]. The current version of the TalkMoves application includes feedback on a set of six teacher (keeping everyone together, getting students to relate to another's ideas, restating, pressing for accuracy, revoicing and pressing for reasoning) and four student talk moves (relating to another student, asking for more info, making a claim, and providing evidence/reasoning).

2 Approach

2.1 Data

We collected 461 written transcripts of mathematics teaching episodes from elementary and secondary (K-12) math lessons drawing on multiple sources. In addition, 166 transcripts were available from 21 teachers who piloted the TalkMoves application during the 2019-2020 academic year. Written transcripts were segmented into sentences using an automated script. Each sentence in the transcript was manually coded by two annotators for teacher and student talk moves; these coded sentences served as the "ground-truth" training dataset for our models. The annotators established reliability prior to applying the codes and again when they were approximately halfway through coding to ensure that their annotations remained accurate and consistent. Reliability, calculated using Cohen's kappa, was high for each talk move at both time periods and ranged from 0.88-1.0.

2.2 Model development and performance

The set of 461 transcripts was used for training while the set of 166 transcripts was used as an out-of-sample testing set. The dataset of 461 transcripts is made up of 176,757 sentences including 115,418 teacher utterances and 61,339 student utterances. The data was split into training and validation set according to a 90/10 % split. The validation set was stratified to mimic the distribution of the labels in the training set. Similarly, the out of sample test set of 166 transcripts had 49,048 teacher utterances and 13,968 student utterances collected from the pilot study. In this initial study, we chose to train two independent deep learning models to automatically identify the teacher and student talk moves. For the teacher model, starting with LSTM neural networks, we explored different model architectures including recent transformer models such as BERT and RoBERTa [10, 11]. The inputs to the model were student-teacher "sentence pairs", which refers to a combination of a teacher sentence concatenated with the immediately prior student sentence. For example, a sentence pair can include a student utterance "I'm pretty sure the numerator would be four" followed by a teacher utterance "Okay, why do you think the numerator would be four?". This sentence pair is a good example of the teacher encouraging a student to reason (pressing for reasoning). The output of the deep learning model is a 7-way sequence classification (softmax) over the six teacher talk moves and "None".

The inputs to the student model were student-student "sentence pairs", which refer to a combination of a student sentence concatenated with the immediately prior student sentence. The output was a 5-way sequence classification (softmax) over the four student talk moves and "None". After hypertuning parameters of the model, we found that RoBERTa base performed the best among the BASE models for both teacher and student sequence classification (see Table 1). We did not find evidence for a significant change in performance on other variants of BERT including XLNet-base and AIBERT-base.

Table 1. Performance of teacher and student models on the out-of-sample test set

	F1 score (in %)	MCC
Teacher Model		
- BERT-base [2]	76.05	0.7519
- RoBERTa- base [5]	**77.29**	**0.7627**
Student Model		
- BERT-base [2]	71.96	0.6585
- RoBERTa-base[5]	**73.04**	**0.6727**

2.3 Confusion matrix

The confusion matrix for the student talk moves model (see Table 2) indicates that the talk move "relating to another student" performed worse relative to the other student talk moves. We intend to conduct additional experiments to determine whether the performance can be improved by extending sentence pairs to include multiple previous student sentences as context.

Table 2. Confusion matrix of the student talk moves model

RoBERTa-base : 73.04% F1	0	1	2	3	4	Precision	Recall	F1
0 - None	5284	221	139	415	61	0.82	0.86	0.84
1 - Relating to another student	209	407	8	149	84	0.46	0.47	0.47
2 - Asking for more information	59	6	245	17	6	0.56	0.71	0.63
3 - Making a claim	798	209	34	3611	313	0.82	0.73	0.77
4 - Providing Evidence	59	39	11	210	1364	0.74	0.81	0.77

3 Conclusion

The TalkMoves application provides teachers with detailed feedback on their use of research-based discourse practices, in the form of specific instructional talk moves that prior research suggests promote inclusion and equity [6]. The strong performance of both the student and teacher talk moves models illustrates the reliability and robustness of artificial intelligence algorithms applied to noisy real-world classroom data. The student utterances extracted from classroom transcripts are especially challenging to interpret relative to teacher speech. For example, there are numerous instances where the utterances lacked well-formed syntax, including missing words. Despite the limitations of this noisy dataset, BASE models achieved good results in classifying the student talk moves, particularly Roberta BASE which was used for both the teacher and student models. As a next step, we plan to experiment with strategies such that the student model performs on par with the teacher model. In addition, future research is needed to continually improve the model performance of individual talk moves, better understand teachers' perceptions and use of the application, and consider how it can be incorporated into structured professional learning opportunities that promote discourse-rich pedagogy.

References

1. http://talkmoves.com/
2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
3. Hand, V.: Seeing culture and power in mathematical learning: Toward a model of equitable instruction. Educational Studies in Mathematics **80**(1), 233–247 (2012)
4. Khisty, L.L., Chval, K.B.: Pedagogic discourse and equity in mathematics: When teachers' talk matters. Mathematics education research journal **14**(3), 154–168 (2002)
5. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692 (2019)
6. Michaels, S., O'Connor, C., Resnick, L.B.: Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. Studies in philosophy and education **27**(4), 283–297 (2008)

7. Moschkovich, J.: Principles and guidelines for equitable mathematics teaching practices and materials for english language learners. Journal of Urban Mathematics Education **6**(1), 45–57 (2013)
8. O'Connor, C., Michaels, S.: Supporting teachers in taking up productive talk moves: The long road to professional learning at scale. International Journal of Educational Research **97**, 166–175 (2019)
9. O'Connor, C., Michaels, S., Chapin, S.: Scaling down" to explore the role of talk in learning: From district intervention to controlled classroom study. Socializing intelligence through academic talk and dialogue pp. 111–126 (2015)
10. Suresh, A., Sumner, T., Huang, I., Jacobs, J., Foland, B., Ward, W.: Using deep learning to automatically detect talk moves in teachers' mathematics lessons. In: 2018 IEEE International Conference on Big Data (Big Data). pp. 5445–5447. IEEE (2018)
11. Suresh, A., Sumner, T., Jacobs, J., Foland, B., Ward, W.: Automating analysis and feedback to improve mathematics teachers' classroom discourse. In: Proceedings of the AAAI conference on artificial intelligence. vol. 33, pp. 9721–9728 (2019)