Investigating New Methods for Automated Teacher Discourse Analysis

Emily Jensen, Frederick Thayer, Mohammed Ali, Chloe Bruce

1 Problem and Motivation

Education is a cornerstone of modern society: engaged students become innovators and key economic drivers of the future. The worsening trend of student engagement is thus an existential danger to society and worthy of concentrated research. Research shows that 53% of grade 5-12 students are disengaged from education, resulting in a projected $4.75 trillion per year aggregate social burden (Hodges, 2018; Belfield et al., 2012). Extensive research has documented that the best ways to improve student engagement are with dialogic and exploratory teaching (Taylor and Parsons, 2011; Kelly, 2007; Lai and McNaughton, 2013). Unfortunately, without regular and objective feedback, teachers are unable to implement improvements to their classroom practices. Most evaluation methods are costly in time and money (Archer et al., 2016) and do not provide actionable feedback teachers need to grow over time.

2 Data and Previous Work

Prior research (Jensen et al., 2020; Stone et al., 2019) provides an approach for real-time, automated speech processing to identify the presence of discourse variables and provide teachers with feedback. Teachers used a wireless headset to record 127 classroom lessons. This yielded around 64,000 utterances after transcription. Approximately 17,000 utterances were coded by human experts for 7 discourse variables: instructional talk, questions, authentic questions, elaborated evaluation, cognitive level, goal specificity, and subject-specific terms.

Researchers represented utterances as bags of n-grams and used a random forest classifier to predict the presence of discourse variables. They report an average AUROC of 0.77 (compared to a 0.5 baseline) and an average accuracy of 0.71 (Jensen et al., 2020).

We can improve this classification task through deep learning models and word embeddings, due to the sequential nature of classroom instruction. Thus, our project aims to use the preexisting data set with several models, features, and architecture variations to determine an optimal system for teacher discourse analysis.

3 Proposed Methods

The previous work (Jensen et al., 2020; Stone et al., 2019) in teaching discourse classification utilized feature-based modeling (random forests). We will expand this work by comparing feature-based methods and deep learning methods. We will explore the predictive accuracy of transfer learning and various neural network architectures.

We will first use transfer learning (Pratt, 1993), which uses pre-trained networks and tunes specific layers. In addition, we will consider feed-forward and recurrent (LSTM) (Hochreiter and Schmidhuber, 1997) neural networks. These are very applicable to this task since discourse is contextual and sequential in nature. We will utilize F1 scores and accuracy to evaluate model performance.

For feature engineering, we will experiment with several word embedding methods. Previous research done on word embeddings shows the Skip Gram model (Huang et al., 1993) can represent rare words with small amounts of data. We will also consider Common Bag Of Words (Mikolov et al., 2013), which is faster and has better representations for more frequent words.

For transfer learning, we can consider using state-of-the-art pre-trained models such as ULMFiT (Howard and Ruder, 2018), Transformer (Vaswani et al., 2017), Google's BERT (Devlin et al., 2018), Transformer-XL (Dai et al., 2019), OpenAI's GPT-2 (Radford et al., 2019). Each has unique strengths that may help improve classification performance.

References

Jeff Archer, Steve Cantrell, Steven L. Holtzman, Jilliam N. Joe, Cynthia M. Tocci, and Jess Wood. 2016. *Better Feedback for Better Teaching: A Practical Guide to Improving Classroom Observations*. Bill and Melinda Gates Foundation.

Clive R. Belfield, Henry M. Levin, and Rachel Rosen. 2012. *The Economic Value of Opportunity Youth*. Corporation for National and Community Service, Washington, DC.

Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc V. Le, and Ruslan Salakhutdinov. 2019. Transformer-XL: Attentive Language Models Beyond a Fixed-Length Context.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Tim Hodges. 2018. Not Just a Buzzword: Using survey data to foster engagement among students, parents, and teachers. *Principal*, pages 10–13.

Jeremy Howard and Sebastian Ruder. 2018. Universal Language Model Fine-tuning for Text Classification.

Xuedong Huang, Fileno Alleva, Hsiao-Wuen Hon, Mei-Yuh Hwang, Kai-Fu Lee, and Ronald Rosenfeld. 1993. The SPHINX-II speech recognition system: an overview. *Computer Speech & Language*, 7(2):137–148.

Emily Jensen, Meghan Dale, Patrick J. Donnelly, Cathlyn Stone, Sean Kelly, Amanda Godley, and Sidney K. D'Mello. 2020. Toward Automated Feedback on Teacher Discourse to Enhance Teacher Learning. In *2020 CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2020)*. ACM Press.

Sean Kelly. 2007. Classroom Discourse and the Distribution of Student Engagement. *Social Psychology of Education*, 10(3):331–352.

Mei Kuin Lai and Stuart McNaughton. 2013. Analysis and Discussion of Classroom and Achievement Data to Raise Student Achievement. In *Data-based Decision Making in Education*, pages 23–47. Springer Netherlands, Dordrecht.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space.

Lorien Y Pratt. 1993. Discriminability-based transfer between neural networks. In *Advances in neural information processing systems*, pages 204–211.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8):9.

Cathlyn Stone, Patrick J. Donnelly, Meghan Dale, Sarah Capello, Sean Kelly, Amanda Godley, and Sidney K. D'Mello. 2019. Utterance-level Modeling of Indicators of Engaging Classroom Discourse. In *The 12th International Conference on Educational Data Mining*, pages 420–425, Montreal, Canada.

Leah Taylor and Jim Parsons. 2011. Improving Student Engagement. *Current Issues in Education*, 14(1).

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need.