Learning About the Norms of Teaching Practice: How Can Machine Learning Help Analyze Teachers' Reactions to Scenarios?

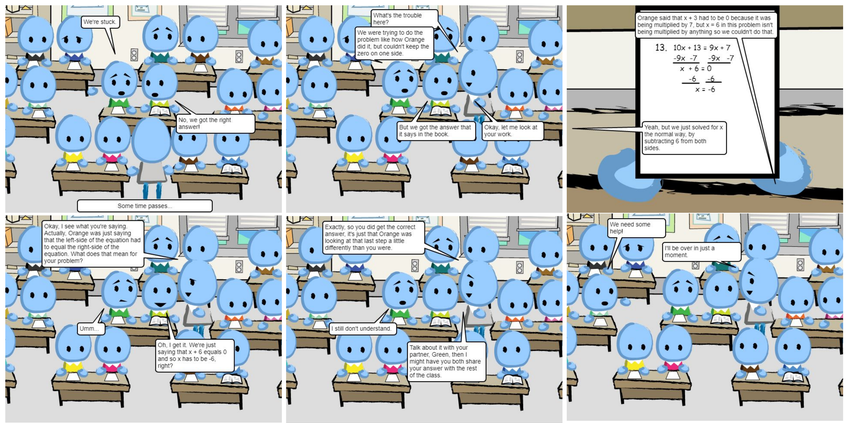
Mike Ion Emanuele Bardelli Patricio Herbst
University of Michigan — School of Education

Abstract

The study of teachers' perspectives on the work of teaching, particularly of its norms, has benefitted from teachers' responses to multimodal scenarios where hypothesized norms are at stake. The analysis of open-ended responses to those scenarios by hand, however, is time- consuming and achieving interrater reliability for linguistics-informed coding is challenging. Using open-ended responses from a national sample of teachers, we first develop a custom word embedding, representative of teacher discussions of classroom events. A word embedding is a mapping words into a continuous (and fairly low-dimensional) vector space, where 'semantically-similar' words are mapped to nearby points. While other popular pre-trained word embeddings exist (e.g., Word2Vec and Glove), our custom model optimizes the embeddings in a way that is sensitive to the subject-specificity of classroom situations. We then use a convolutional neural network (CNN) to classify teachers' responses based on their appraisal of classroom practice. Using Cohen's Kappa, we find high inter-rater reliability between the computer model and human coders, which shows promise that machine learning methods can improve and enhance our current understanding and research of teaching.

Data Collection

We collected data from a national sample of high school teachers. We contacted 1000 teachers and asked them to complete a series of scenario-based surveys. Between 395 and 456 teachers completed the scenarios that we use in these analyses. These scenarios presented teaching situations depicting a typical mathematics lesson. At the end of each scenario, the storyboard presented an unexpected move on the part of the teacher (we call these breaches of the instructional norms). We analyse participants' open responses to the question "what did you see happening" in the scenario.



Human Coding

- Two researchers on our team coded the 5366 geometry responses for the presence of positive and negative appraisal.
- Examples of positive appraisal:
  - "It just seemed like that is how **a good lesson** starts."
  - "Teacher is **listening to input** from the class and **facilitating the process**.... not just telling them."
- Examples of negative appraisal
  - "He was **very unclear**."
  - "Without a visual representation the students were **primed for misunderstanding**."

Regression Models

- Bag of words representation for open-ended responses
  - Remove common stop words and stem using snowball stemmer
  - Keep 9000 features, up to 3-grams and n-grams that were used at least 3 times across the corpus
- Use the human coders classification on the geometry items as training and test sets
  - Training set included 4829 responses (90%)
  - Test set included 537 responses (10%)

Two major results emerge: (1) We find strong agreement between the human and machine codings of both the positive and negative appraisals of teachers' actions; and (2) we find promise in the future use of the same model for out of sample coding of other algebra-related items.

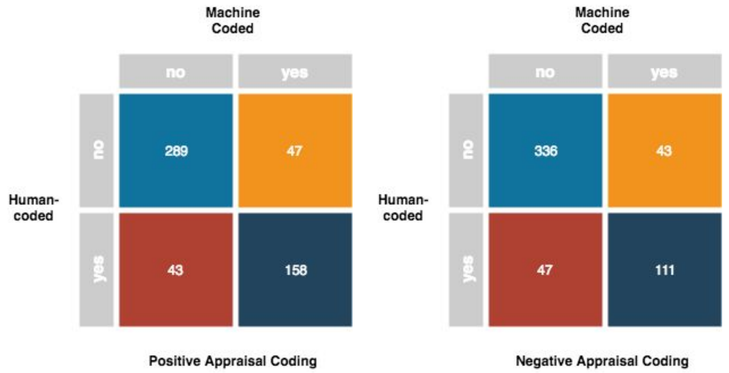


Figure 1. Confusion matrices for positive and negative appraisal codes.

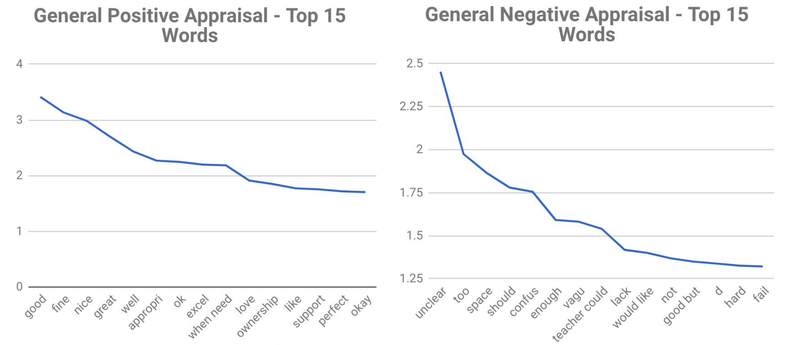
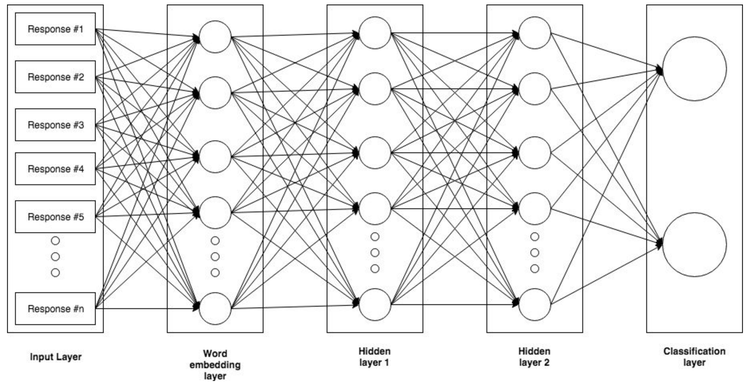


Figure 2. Word weights for positive and negative appraisal codes.

Next Steps: Neural Network

Using teachers' responses to multimodal classroom scenarios, we created a word embedding that we believe can improve our future models analyze text that is contextually dependent on mathematics classroom dialogue. Next, we hope to use this constructed word embedding as a first layer of a convolutional neural network, in hopes to improve upon generalizability and accuracy of these initial results. We also hope that this method can help us expand our appraisal coding to other situations of qualitative coding.



Acknowledgements

The work presented here has been done with the support of NSF Grant DRL- 0918425 to P. Herbst. All opinions are those of the authors and do not necessarily represent the views of NSF.