Identifying Teacher Questions Using Automatic Speech Recognition in Classrooms

Nathaniel Blanchard[1], Patrick J. Donnelly[1], Andrew M. Olney[2], Xiaoyi Sun[3], Brooke Ward[3], Borhan Samei[2], Sean Kelly[4], Martin Nystrand[3], and Sidney K. D'Mello[1]

[1] University of Notre Dame, [2] University of Memphis, [3] University of Wisconsin-Madison, [4] University of Pittsburgh

Abstract

We investigate automatic question detection from recordings of teacher speech collected in live classrooms. Our corpus contains audio recordings of 37 class sessions taught by 11 teachers. We automatically segment teacher speech into utterances using an amplitude envelope thresholding approach followed by filtering non-speech via automatic speech recognition (ASR). We manually code the segmented utterances as containing a teacher question or not based on an empirically-validated scheme for coding classroom discourse. We compute domain-independent natural language processing (NLP) features from transcripts generated by three ASR engines (AT&T, Bing Speech, and Azure Speech). Our teacher-independent supervised machine learning model detects questions with an overall weighted F_1 score of 0.59, a 51% improvement over chance. Furthermore, the proportion of automatically-detected questions per class session strongly correlates (Pearson's r = 0.85) with human-coded question rates. We consider our results to reflect a substantial (37%) improvement over the state-of-the-art in automatic question detection from naturalistic audio. We conclude by discussing applications of our work for teachers, researchers, and other stakeholders.

Overview

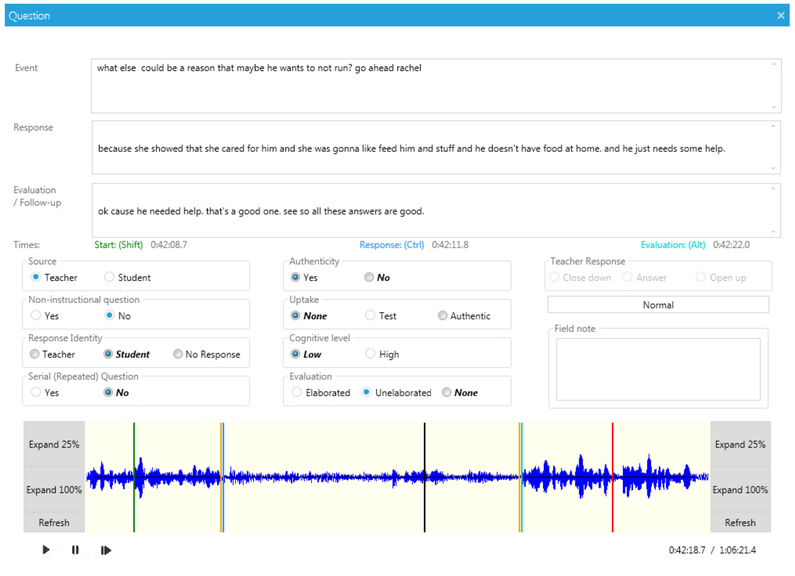
- Certain types of questions are powerful predictors of student engagement and achievement.
- The overall goal of the project is to provide feedback to teachers.
- A processing pipeline must be fully automated deploy in the field.
- This work presents an automated process to identify teacher questions.

Data Collection

- Teachers wore a headset microphone to capture high fidelity speech
- Sessions averaged 60±30 minutes
- 37 class sessions collected
- 32 hours of audio
- 11 teachers



Teacher's Microphone

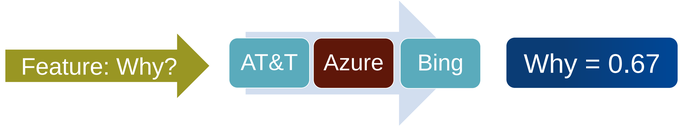


Question Coding

- Audio automatically divided into *utterances* based on silence
- Utterances are automatically, albeit imperfectly. segmented
- Seven research assistants tagged each utterance as either *containing* or *not containing* a question
- 10080 utterances in which 3586 contain questions

Feature Extraction

Three ASR engines transcribed utterances. 37 NLP features were extracted from each ASR transcript. We also created a fourth set of NLP features that combined, at the feature level, the features from the individual ASRs .



ASR	Word Error Rate	Simple Word Overlap
Bing Speech	0.45 (0.10)	0.55 (0.06)
AT&T Watson	0.63 (0.11)	0.42 (0.11)
Azure Speech	0.49 (0.07)	0.64 (0.16)

Question Detection Results

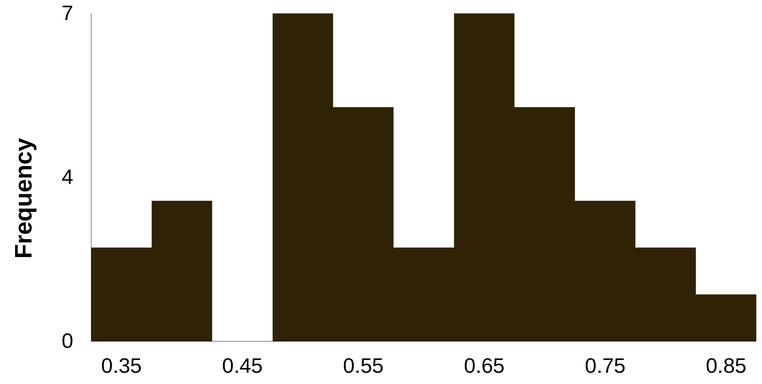
Machine learning models are trained to identify important components of dialogic instruction and are validated independently of the teacher. Combined ASR model trained boosting J48 trees with a MetaCost weight two.

ASR Model	F_1 Question	F_1 Not-Question	F_1 Overall
AT&T	0.52	0.68	0.63
Azure	0.53	0.67	0.63
Bing	0.54	0.67	0.63
Combined	0.59	0.74	0.69

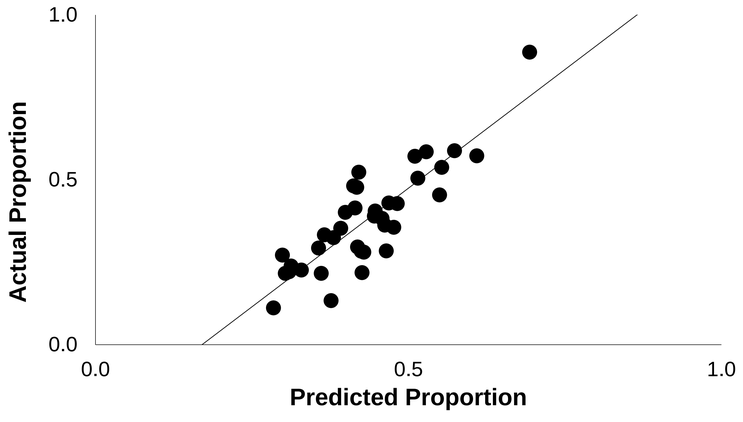
Actual		Predicted	
Prior %		**Question**	**Non-Question**
0.36	**Question**	0.63	0.37
0.64	**Non-Question**	0.30	0.70

Consistency by Class Session

Models are analyzed for consistency across the 37 class sessions.



Above: Histogram of F_1 scores for 37 class sessions.
Below: Predicted and actual proportions of questions per class session.



Future Work

- Inclusion of acoustic features
- Classification of question properties (authenticity and uptake)
- Assessment of the ability to predict student achievement using automatically coded dialogic instruction, compared to manual coding
- Dissemination of software/hardware to teachers, researchers, and professional development specialists

This research was supported by the Institute of Education Sciences (R305A130030)

UNIVERSITY OF NOTRE DAME
EMOTIVE COMPUTING LAB