SIGHT: A Large Annotated Dataset on Student Insights Gathered from Higher Education Transcripts

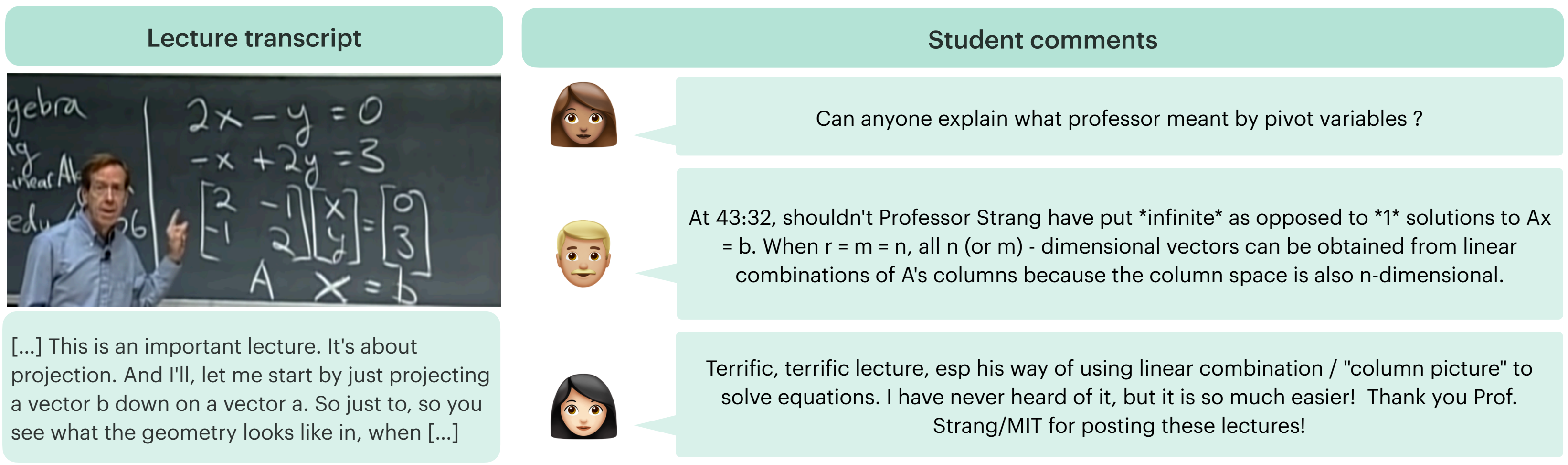
Rose E. Wang* Pawan Wirawarn* Noah Goodman Dorottya Demszky

*Equal contributions rewang@cs.stanford.edu

Introduction

- **Motivation:** Lectures are a learning experience for both students and teachers. Students learn from teachers about the subject material, while teachers learn from students about how to refine their instruction.
- **Challenge:** However, online student feedback is unstructured and abundant, making it challenging for teachers to learn and improve.
- **Prior works:** While prior works have used qualitative analysis methods to uncovering insightful feedback, applying these methods to large sources of data is challenging as it requires manual annotation [4, 2, 1, 5, 6].
- **Research question:** How can we automatically synthesize insights and learn from abundant, unstructured student feedback in online educational settings?

Contributions



Lecture transcript

[...] This is an important lecture. It's about projection. And I'll, let me start by just projecting a vector b down on a vector a. So just to, so you see what the geometry looks like in, when [...]

Student comments

Can anyone explain what professor meant by pivot variables ?

At 43:32, shouldn't Professor Strang have put *infinite* as opposed to *1* solutions to Ax = b. When r = m = n, all n (or m) - dimensional vectors can be obtained from linear combinations of A's columns because the column space is also n-dimensional.

Terrific, terrific lecture, esp his way of using linear combination / "column picture" to solve equations. I have never heard of it, but it is so much easier! Thank you Prof. Strang/MIT for posting these lectures!

Figure 1. A peek into SIGHT.

1. We create SIGHT, a dataset of 288 lecture transcripts from MIT OpenCourseWare (OCW) mathematics courses and of 15,784 annotated user comments.
2. We develop an annotation rubric of feedback types found in YouTube comments using a qualitative analysis approach.
3. We release a set of best practices for using large language models (LLMs) with qualitative coding rubrics for scaling annotation.
4. We analyze the quality of the annotation and the diverse types of student feedback uncovered via our automated annotation procedure.

Feedback Rubric

Our process for developing the rubric follows the procedure outlined in [6, 7, 3]:

- Two authors read a sample of comments and develop the initial categories collaboratively.
- The categories are adapted to be specific and iterated until authors agree that the categories sufficiently cover the comments.

General: The comment expresses a general opinion about the video's content and/or about the teaching/professional characteristics of the instructor.

Confusion: The comment asks a math-related question, expresses math-related confusion, and/or points out a math-related mistake in the video.

Pedagogy: The comment mentions an instructional method. Instructional methods include the use of examples, applications, worked out problems, proofs, visualizations, elaboration, and analogies.

Teaching setup: The comment describes or mentions the lecture's teaching setup. Teaching setup includes the chalk, chalkboard, microphone or audio-related aspects, and camera or camera-related aspects (e.g., angle).

Personal experience: The comment mentions the user's personal experience or context with respect to the lecture. Personal experience or context includes the user's own math learning or teaching experiences.

Clarification: The comment clarifies someone's math-related misunderstanding or elaborates content from the video, and includes an '@' that is immediately followed by a username.

Gratitude: The comment contains the word "thanks" or "thank".

Non-English comment: The comment is not in English.

N/A: The comment expresses a joke or is a troll comment, and/or the comment says something that is hard to connect to the video content, and/or the comment does not fall into any of the categories above.

References

[1] Martin W Bauer and George Gaskell. *Qualitative researching with text, image and sound: A practical handbook for social research*. Sage, 2000.

[2] Juliet Corbin et al. Basics of qualitative research grounded theory procedures and techniques. 1990.

[3] Juliet M Corbin and Anselm Strauss. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, 13(1):3–21, 1990.

[4] Frederick Erickson et al. *Qualitative methods in research on teaching*. Institute for Research on Teaching, 1985.

[5] Colin D Harrison, Tiffy A Nguyen, Shannon B Seidel, Alycia M Escobedo, Courtney Hartman, Katie Lam, Kristen S Liang, Miranda Martens, Gigi N Acker, Susan F Akana, et al. Investigating instructor talk in novel contexts: Widespread use, unexpected categories, and an emergent sampling strategy. *CBE—Life Sciences Education*, 18(3):ar47, 2019.

[6] Cliodhna O'Connor and Helene Joffe. Intercoder reliability in qualitative research: debates and practical guidelines. *International journal of qualitative methods*, 19:1609406919899220, 2020.

[7] Shannon B Seidel, Amanda L Reggi, Jeffrey N Schinske, Laura W Burrus, and Kimberly D Tanner. Beyond the biology: A systematic investigation of noncontent instructor talk in an introductory biology course. *CBE—Life Sciences Education*, 14(4):ar43, 2015.

SIGHT dataset

- The videos are from all the MIT OCW math lecture playlists.
- There are 288 lectures, 15,784 comments, each labelled with 9 labels.
- Topics range from general mathematics courses on calculus and linear algebra to more advanced topics like graph theory and functional analysis.

Annotation examples

Category	Example comment	%
general	Best video I have watched so far, I was with him all the way and my concentration never dipped.	28.37%
confusion	34:43 why "directional second derivative" would not give us a clue of whether it is a min or max? I thought it is a promising way. hmmm.	20.76%
pedagogy	From this lecture, I really understand Positive Definite Matrices and Minima thanks to Dr. Gilbert Strang. The examples really help me to fully comprehend this important subject.	7.27%
setup	Oh.. my god.. the board and chalk are phenomenal..!	3.81%
personal	sweet, did this like a term and a half ago in higshcool. aced the test for it too :D gosh calculus is awesome!	9.00%
clarification	@[USERNAME] Actually, if a constant k=1/1m is used, then in the final formula for V you will end up with subtracting m^1 from m^2 which is apparently not correct.	2.42%
gratitude	Thank you very much! Amazing lectures!	13.49%
nonenglish	Tłumaczenie na polski wymiata	6.57%
na	sounds drunk on 0.5 speed	42.21%

Table 1. Example comments for each comment annotation category. The category percentage of the sample dataset is reported in the column %. Note, a comment can be labeled with multiple categories so the percentages do not add up to 100%.

- Each comment is annotated as a binary classification task per category, i.e., does this category apply to this comment?
- **Example:** "His teaching style seems casual and intuitive. I go to a small public college and the course is much more formal and proof driven. These lectures are a great addition to (as well as a nice break from) formal proofs. Thanks MIT!"
- This comment is labeled as pedagogy, personal, and gratitude categories.

Scaling Annotation and Results

Takeaways

- The model's annotations reflect the variability observed in human opinions (Figure 2).
- The model misses subtle references to categories (Table 2).
- The effect of auxiliary information varies across categories (Table 3).
- The annotations allow for a diversity of feedback types to be discovered (e.g., Table 4 within the confusion category).

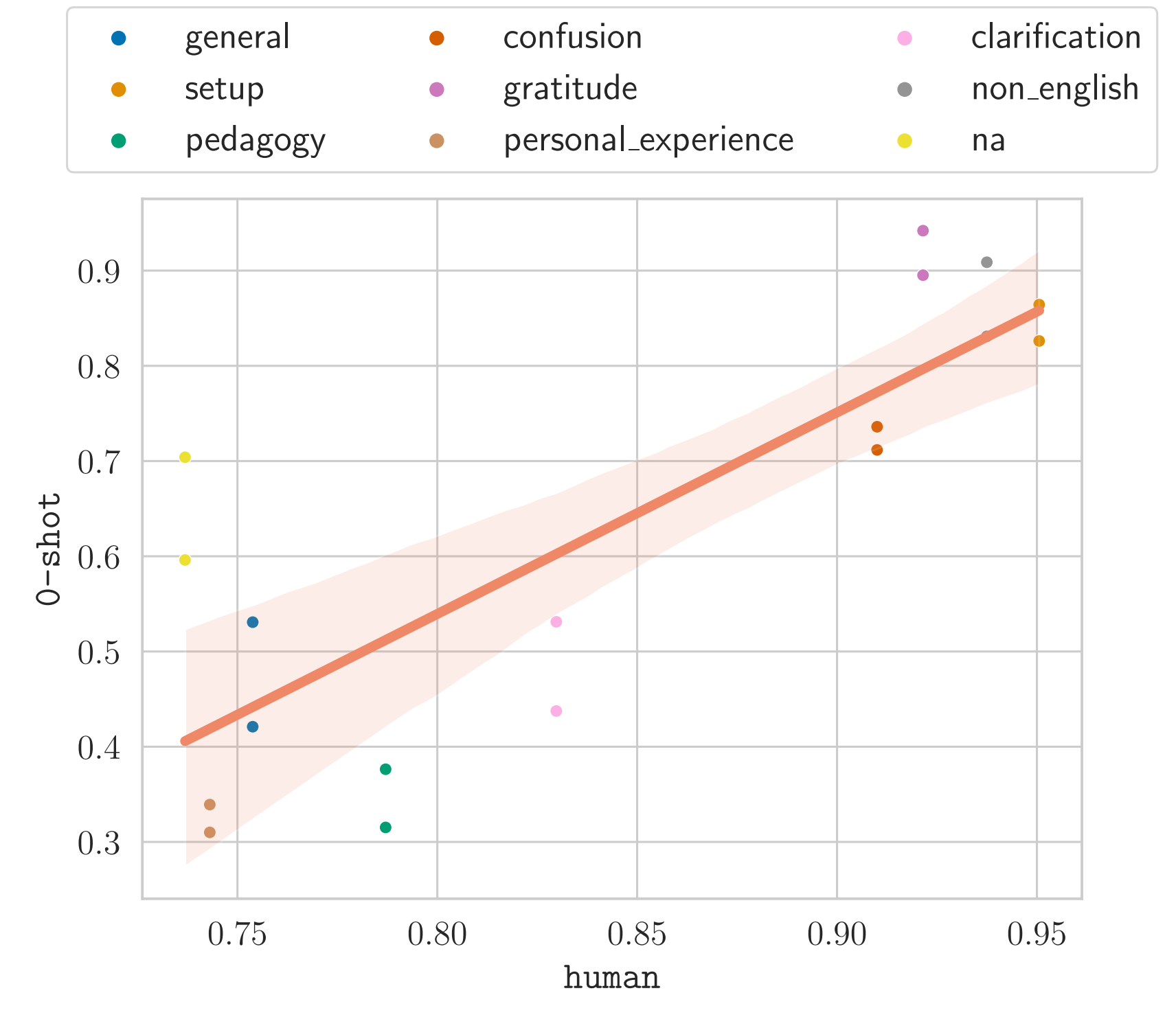


Figure 2. Human inter-annotator agreement (human) vs. human-model inter-annotator agreement (0-shot).

#	Category	Comment	H	M
A	pedagogy	This guy is great. I studied engineering at a university less prestigious than MIT, and I remember professors refusing to explain their algebra steps. They were like "you should know this already".	1	0
B	personal	Wish this guy taught me Math 293 and 294 at Cornell. My guy could barely speak English, let alone explain what we were trying to accomplish. I understood that if we wanted eigenvectors perpendicular to x we'd get lift relative to flow...but this guy would have made the math a bit simpler.	1	0
C	pedagogy	41:53 These are questions that should be asked in recitation, not in lecture.	0	1
D	personal	why is iteration in newtons done..i cant understand the logic behind this	0	1

Table 2. Error analysis on pedagogy and personal, the two lowest agreement categories on the zero shot setting (0-shot). The **H** column is the category label that both humans assigned the comment to, and the **M** column is the label that the model assigned the comment to. 1 indicates that the annotator believes the category *does* apply to the comment, whereas 0 is where the category is presumed *not* to apply.

IRR	gen.	conf.	peda.	set.	pers.	clar.	gra.	noneng.	na
human	0.75	0.91	0.79	0.95	0.74	0.83	0.92	0.94	0.74
0-shot	0.48	0.72	0.35	**0.85**	0.32	**0.48**	0.92	**0.87**	**0.65**
3-shot	0.50	0.69	0.52	0.75	**0.57**	0.16	0.85	0.64	0.50
3-shot-R	**0.52**	**0.76**	**0.57**	**0.85**	0.37	0.32	**0.93**	0.50	0.47

Table 3. Cohen's kappa scores for measuring inter-rater reliability (IRR) within humans (human) and within human-model pairs across the rubric categories when the model is prompted 0-shot (0-shot), 3-shot (3-shot), or 3-shot with reasoning (3-shot-R).

Subcategories	Comments labeled as confusion
Conceptual	Can anyone explain what professor meant by pivot variables ?
Conceptual	Can anyone help me understand, why the professor keep saying at 19:01 that we can't solve 4 equation with 3 unknowns?
Potential mistake	i think the explanation of the first queston was a little bit wrong it seems. because he wrote the equation to diagonalize the matxix P even though it does not have 3 independent eigen vectors.
Potential mistake	Anyone understand the equation at 32:15? I think x_free should be above x_pivot?
Resources	What is good homework to test if we clearly understand this lecture? Is there such corresponding homework?
Resources	Does anyone know which lecture he derive the general equation for a determinant? Would be a massive help thanks!

Table 4. Example comments in the confusion category.