UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

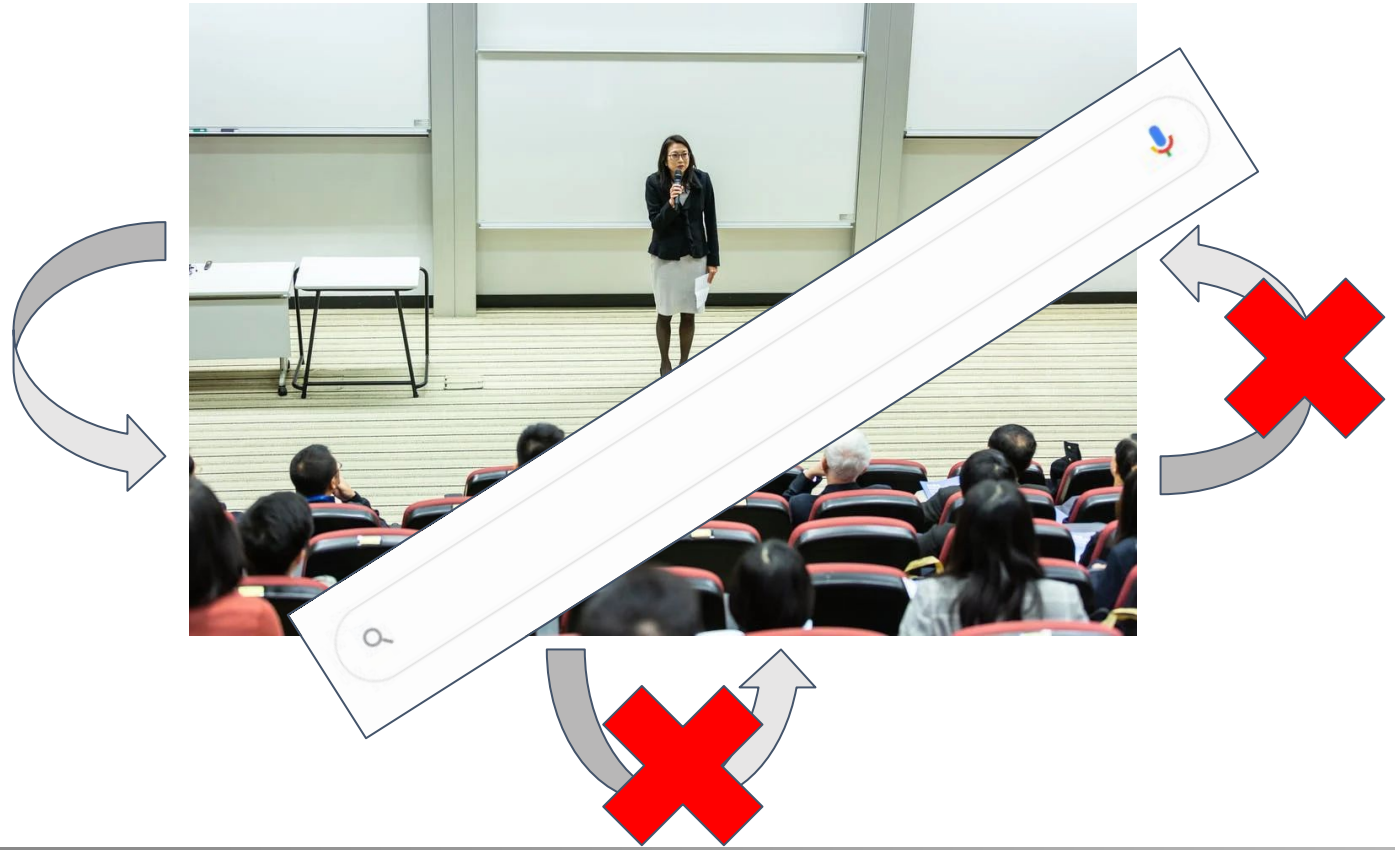
Generation of Student Questions for Inquiry-based Learning

Kevin Ros, Maxwell Jong, Chak Ho Chan, ChengXiang Zhai

Presented by Kevin Ros, kjros2@illinois.edu

In a traditional, in-person classroom…

In an online, asynchronous classroom…



How can this be addressed?

By learning how to generate likely questions given a lecture context!

Collect questions coupled to the time in the lectures when they were asked

Use lecture context windows to generate questions

The Lecture-Question Data Set

- Asynchronous course where primary lectures were from two MOOCs
  - *Text Retrieval and Search Engines* and *Text Mining and Analytics*
  - Also have lecture transcripts with timestamps
- Asked students, if student had a question while watching a lecture, to anonymously submit the question and corresponding lecture context
  - <Lecture name, start time, end time, question>
- Total resulting data set: 536 questions across 90 lectures
- **Questions are *about* lecture content, often not answerable by content**

Unigrams	Trigrams
is (264)	what is the (65)
how (188)	how do we (30)
what (181)	the meaning of (13)
does (107)	the difference between (12)
why (104)	why do we (11)
are (103)	is it possible (9)

Table 1: A list of some of the most frequent unigrams and trigrams in student questions. The number in parentheses indicates the occurrence frequency.

Question Examples
What is the point of compression? Will the access times really be that impactful to the overall indexing?
Are the doc-ids sorted with the term-ids in the "local" sort?
Can we get more examples of using gamma-code?
How does the gamma-code intergar compression method work? I did not understand the example from the video
I'm still very confused how integer compression actually reduces size of storage since some of the examples make it seem like you're using more bits than before on some inputs

Table 2: A few example questions from the lecture-question data set.

Using the Lecture-Question Data Set to Generate Questions

- Goal: Given a lecture window, generate the question associated with that window
- Low amount of data, so tried two low-data techniques with T5 - a pre-trained generative language model based on the Transformer [1]

- **Research Question #1**: How does pre-training on search engine query generation affect student question generation performance?
  - Used docTTTTTquery [2], a version of T5 fine-tuned to generate search engine queries given the ground truth passage

- **Research Question #2**: How does continuous prefix tuning affect student question generation performance?
  - Implemented prefix tuning on top of T5 and docTTTTTquery
    - Fine-tune generative language model by freezing the original parameters and adding a learnable prefix that is prepended to every input [3]

Findings

Main Takeaways

- **RQ1**: Pre-training on search engine query generation appears to offer clear benefit in increasing the precision, though the benefit appears to be more for traditional fine-tuning.
- **RQ2**: There seems to be marginal benefit for using continuous prefix tuning in a low-data setting to generate student questions.

Model	R	P	F_1
	ROUGE-1 (%)		
t5-base FT	20.06	14.47	14.82
t5-base Prefix	20.13	21.56	18.63
docTTTTTquery	14.41	25.17	16.83
docTTTTTquery FT	15.70	23.34	17.45
docTTTTTquery Prefix	17.19	24.00	18.74
	ROUGE-2 (%)		
t5-base FT	1.697	1.656	1.502
t5-base Prefix	3.267	3.391	3.043
docTTTTTquery	3.237	4.596	3.358
docTTTTTquery FT	4.011	4.730	3.903
docTTTTTquery Prefix	4.790	6.247	5.010
	ROUGE-L (%)		
t5-base FT	15.82	11.57	11.77
t5-base Prefix	16.89	17.65	15.47
docTTTTTquery	13.17	22.32	15.18
docTTTTTquery FT	14.34	20.64	15.76
docTTTTTquery Prefix	15.47	21.00	16.73

Some Examples of Generated Questions

Model	Question
Ground Truth	Does is the delta-code use gamma-code twice recursively?
t5-base FT	What is the difference between delta coding and delta coding? Is it possible to use delta coding for inverted index distribution?
t5-base Prefix	What is the difference between delta and gamma?
docTTTTTquery	what is gamma coding
docTTTTTquery FT	what is the difference between delta and delta coding?
docTTTTTquery Prefix	what is the difference between delta and gamma coding?
Ground Truth	What is the reasoning for making the first (1+logx) unary and the x-2(logx) uniform? The method for encoding seem random to me.
t5-base FT	What is the problem? Is there a way to use binary code to calculate the remaining part of the value of x?
t5-base Prefix	what is the difference between unary code and binary code?
docTTTTTquery	what is the difference between unary and binary code
docTTTTTquery FT	what is the difference between unary and binary code?
docTTTTTquery Prefix	what is the difference between unary and binary code?
Ground Truth	What is the risk associated with discarding documents that are potentially relevant?
t5-base FT	is there any way to determine the most likely relevance in the documents?
t5-base Prefix	What is the purpose of having a diverse set of ranking methods?
docTTTTTquery	what is diverse set of ranking methods
docTTTTTquery FT	what is the difference between a diverse set of ranking methods and an unjudged pool of documents?
docTTTTTquery Prefix	what is meant by having a pool of relevant documents that aren't being ranked?

Table 5: A few examples of the ground truth question compared to each model's generated question.

Conclusion and Future Work

- This is an exciting, new area!
- We are collecting more questions (Spring 2022, planned for Fall 2022)
- Some future directions:
  - Types of questions that students and instructors find helpful
  - How background knowledge affects the types of questions asked
  - More robust evaluation and analysis
  - Practical applications (integrating into classrooms)

Thank you for your time!

Corresponding author: Kevin Ros, kjros2@illinois.edu

References

[1] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv:1910.10683.

[2] Rodrigo Nogueira, Jimmy Lin, and AI Epistemic. 2019a. From doc2query to docTTTTTquery. Online preprint.

[3] Xiang Lisa Li and Percy Liang. 2021. Prefix-tuning: Optimizing continuous prompts for generation. arXiv preprint arXiv:2101.00190.