Computational Semantics for Natural Language Processing (FS2022)

Combined Transformers for Conversational Question Answering

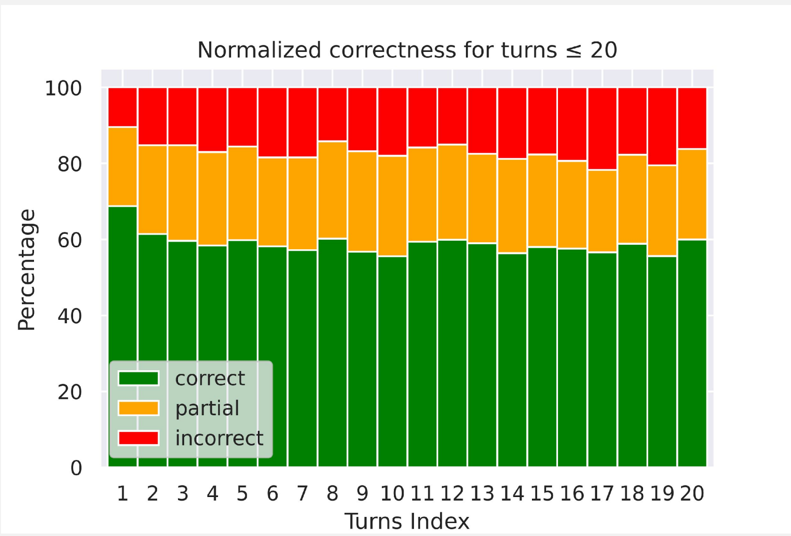
Yuanzhi Zhu[1], Bartosz Dzionek[1,2]
[1]ETH Zürich, [2]University of Edinburgh

1 Introduction

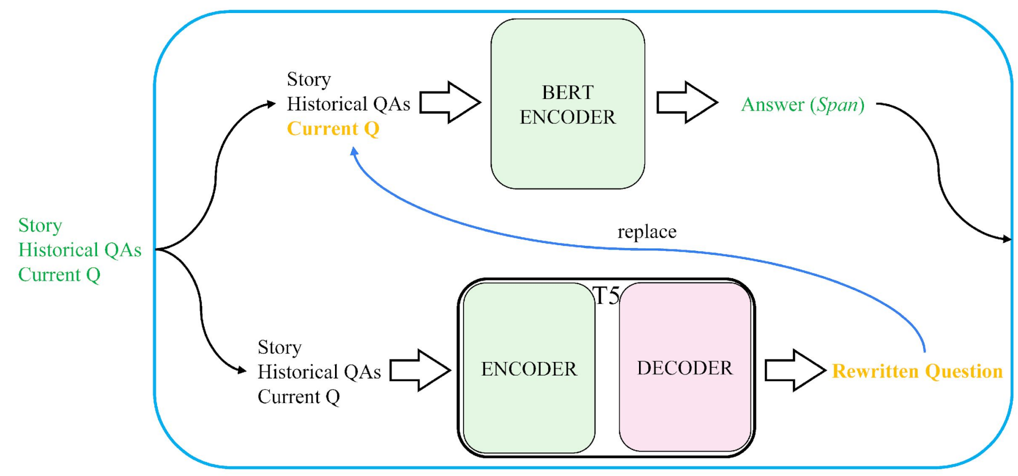
In the era of voice assistants, it is very important to enable machines to answer sequential questions in a given context in a conversational manner[1]. In this project, we investigate the possibility of combining a generative model (eg. T5[2]) with a baseline BERT[3] for Conversational Question Answering.

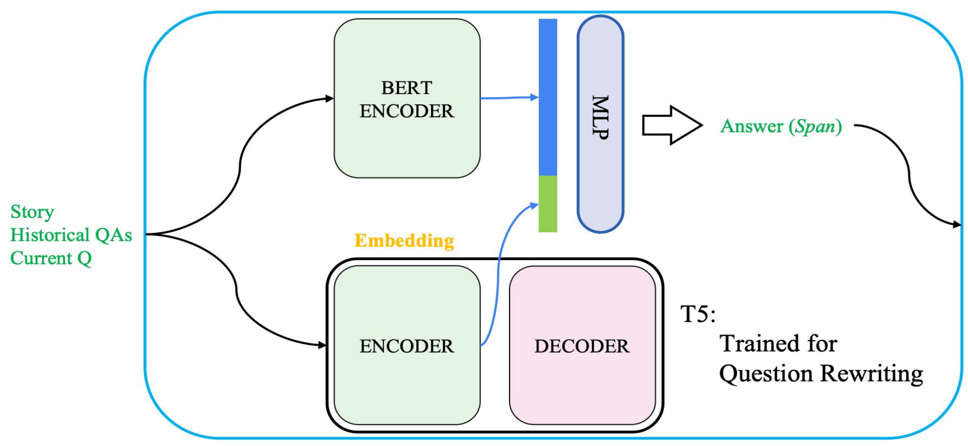
2 Baseline Analysis

- BERT has better performance in the first round.
- In the next turns, its performance is lower and stable.



3 Our Methods





First idea: Get a rewritten question from T5 and feed it as an additional input for BERT to get a more accurate answer.

Second idea: Use just the encoder of T5. Append the T5 embedding to the BERT embedding and fine tune with MLP.

4 Results and Discussion

Inputs	EM	F1
Baseline (2 epochs)	67.2	77.1
Replaced with Rewritten Q (4 epochs)	64.8	74.8
Replaced all (2 epochs)	62.4	71.8
Concatenated w/o history (2 epochs)	63.8	73.5
Concatenated with Rewritten Q (4 epochs)	66.5	76.7
Concatenated with Rewritten Q (2 epochs)	**67.3**	**77.2**
With T5 embedding (4 epochs)	**67.4**	**77.2**
Pooled T5 embedding (3 epochs)	66.9	76.8

EM means the generated answer exactly matches one of the ground truths..
F1 means the F1 similarity score to the ground truths..
For more results, please refer to our final report in our Github repository[4] under folder *reports*

5 Conclusion

- We get results that are comparable with the baseline. It was challenging to make an improvement without modifying the BERT model itself.
- We believe that many more experiments can be done to fill the gap between first and the rest turns' performance.
- The power of T5 for Question Rewriting is limited and the output from T5 is not always helpful.
- In the future, we should pay more attention to the alignment of the two models, such that they can understand each other better (T5 sentence embedding should be better)

References

1. M. Zaib, W. E. Zhang, Q. Z. Sheng, A. Mahmood, and Y. Zhang, *"Conversational question answering: A survey,"* arXiv preprint arXiv:2106.00874, 2021.
2. C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," Journal of Machine Learning Research, vol. 21, no. 140, pp. 1–67, 2020.
3. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, *"BERT: Pre-training of deep bidirectional transformers for language understanding,"* in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), (Minneapolis, Minnesota), pp. 4171–4186, Association for Computational Linguistics, June 2019
4. yuanzhi-zhu/CSNLP-Project-ETH: This is the repo for our computational semantics natural language processing course project (github.com)