Master thesis report 1

University of Tartu

February 13, 2023

1 Papers

- Challenges in Generalization in Open Domain Question Answering([Liu et al., 2021])
- Undersensitivity in Neural Reading Comprehension([Welbl et al., 2020])
- Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks([Lewis et al., 2020a])
- Prefix-Tuning: Optimizing Continuous Prompts for Generation([Li and Liang, 2021])
- Parameter-Efficient Transfer Learning for NLP([Houlsby et al., 2019])
- Evidentiality-guided Generation for Knowledge-Intensive NLP Tasks([Asai et al., 2021])
- Leveraging Passage Retrieval with Generative Models for Open Domain Question Answering([Izacard and Grave, 2020])
- Semantic Sentence Embeddings for Paraphrasing and Text Summarization([Zhang et al., 201
- REALM: Retrieval-Augmented Language Model Pre-Training([Guu et al., 2020])
- Simple and Effective Multi-Paragraph Reading Comprehension([Clark and Gardner, 2017])
- Dense Passage Retrieval for Open-Domain Question Answering([Karpukhin et al., 2020])
- Question and Answer Test-Train Overlap in Open-Domain Question Answering Datasets([Lewis et al., 2020b])

2 Proposal 1

2.1 Problem

In [Liu et al., 2021] one of identified problem with bad performance for comp-gen questions related to the retriever that provided passages are not always co-locate with the anchor words of a question.

For example:

Q: Who played Mary in Christmas with the Kranks?

—

P_w: Julie Gonzalo Julieta [...] is an [...] actress. [She] is also knows for her roles ”Christmas with the Kranks”.

A_w: Julie Gonzalo

—

P_c: She also starred in [...] ”Christmas with the Kranks”.

A_c: Felicity Huffman

—

Since “Mary” is not mentioned in either passage, it is impossible to infer that the correct answer is Felicity Huffman.
The second problem also related to the retriever: for high frequency wiki entities the retriever provides many relevant passages which are distracting the model.

2.2 Solution

In that case the only improvement we can do is to improved retriever. Also, I can conduct similar experiment, and try to answer those questions myself by reading all the passages. Then:

1. If it would be solvable by human, in that case there would be something that model could also use to infer correct answer beside anchor words.

2. If it not solvable by human, in that case we need to improve the retriever.

The second problem is actually quite challenging for human as well. In order to solve that as humans we read thoroughly around the anchor words.

3 Proposal 2

The language is very complicated, thus maybe some of the initial experiment can be conducted with scan dataset, since its language is small, and therefore more managable. We could even extend it to account for our possible needs.

In scan 'length' split represent comp-gen, 'jump'/'run' commands represent novel-gen, and definitely there are a lot of overlaps.

- The cons - easy to reason about.
- The pros - addition work for extension and maintenance of such a dataset.

4 Proposal 3

Let's consider following question "Panda is national animal of which country?" from [Liu et al., 2021], that we want our model to find an answer for. However, the word "panda" wasn't present in similar context in training set, but it was in different: "When did the first panda come to America?". Thus it is comp-gen question, in that case the model might fail to find an answer.

To solve that we could find similar example in the training set that are similar to the given question, like "Cow is a national animal of which country?". And provide question+passage+answer, as an example for the model how to find the answer to the first question, in similar fashion like in-context learning.

Remark: Is it possible to train a model with in-context learning?

References

[Asai et al., 2021] Asai, A., Gardner, M., and Hajishirzi, H. (2021). Evidentiality-guided generation for knowledge-intensive nlp tasks.

[Clark and Gardner, 2017] Clark, C. and Gardner, M. (2017). Simple and effective multi-paragraph reading comprehension.

[Guu et al., 2020] Guu, K., Lee, K., Tung, Z., Pasupat, P., and Chang, M.-W. (2020). Realm: Retrieval-augmented language model pre-training.

[Houlsby et al., 2019] Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., de Laroussilhe, Q., Gesmundo, A., Attariyan, M., and Gelly, S. (2019). Parameter-efficient transfer learning for nlp.

[Izacard and Grave, 2020] Izacard, G. and Grave, E. (2020). Leveraging passage retrieval with generative models for open domain question answering.

[Karpukhin et al., 2020] Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., and Yih, W.-t. (2020). Dense passage retrieval for open-domain question answering.

[Lewis et al., 2020a] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W.-t., Rocktäschel, T., Riedel, S., and Kiela, D. (2020a). Retrieval-augmented generation for knowledge-intensive nlp tasks.

[Lewis et al., 2020b] Lewis, P., Stenetorp, P., and Riedel, S. (2020b). Question and answer test-train overlap in open-domain question answering datasets.

[Li and Liang, 2021] Li, X. L. and Liang, P. (2021). Prefix-tuning: Optimizing continuous prompts for generation.

[Liu et al., 2021] Liu, L., Lewis, P. S. H., Riedel, S., and Stenetorp, P. (2021). Challenges in generalization in open domain question answering. *CoRR*, abs/2109.01156.

[Welbl et al., 2020] Welbl, J., Minervini, P., Bartolo, M., Stenetorp, P., and Riedel, S. (2020). Undersensitivity in neural reading comprehension.

[Zhang et al., 2018] Zhang, C., Sah, S., Nguyen, T., Peri, D., Loui, A., Salvaggio, C., and Ptucha, R. (2018). Semantic sentence embeddings for paraphrasing and text summarization.