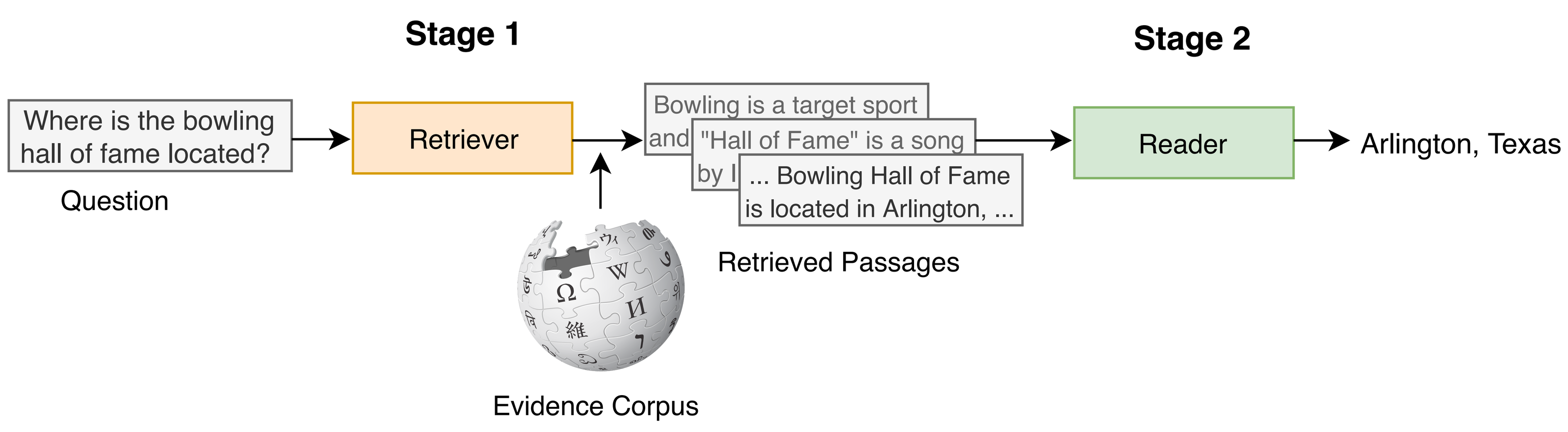
Improving Passage Retrieval with Zero-Shot Question Generation

Devendra Singh Sachan[1,2], Mike Lewis[3], Mandar Joshi[4], Wen-tau Yih[3], Joelle Pineau[1,2,3], Luke Zettlemoyer[3,4]
[1]McGill University, [2]Mila, [3]Meta AI, [4]University of Washington
sachande@mila.quebec, mikelewis@meta.com

Introduction

Open-Domain Retrieval Pipeline



- **Stage 1**: Given an information-seeking question, the first-stage retriever obtains a ***set of relevant passages*** from the evidence.
- **Stage 2**: These passages are attended to by a reader network to generate an answer for the question.

Research Question

How to Improve the First-Stage Passage Retrieval Accuracy?

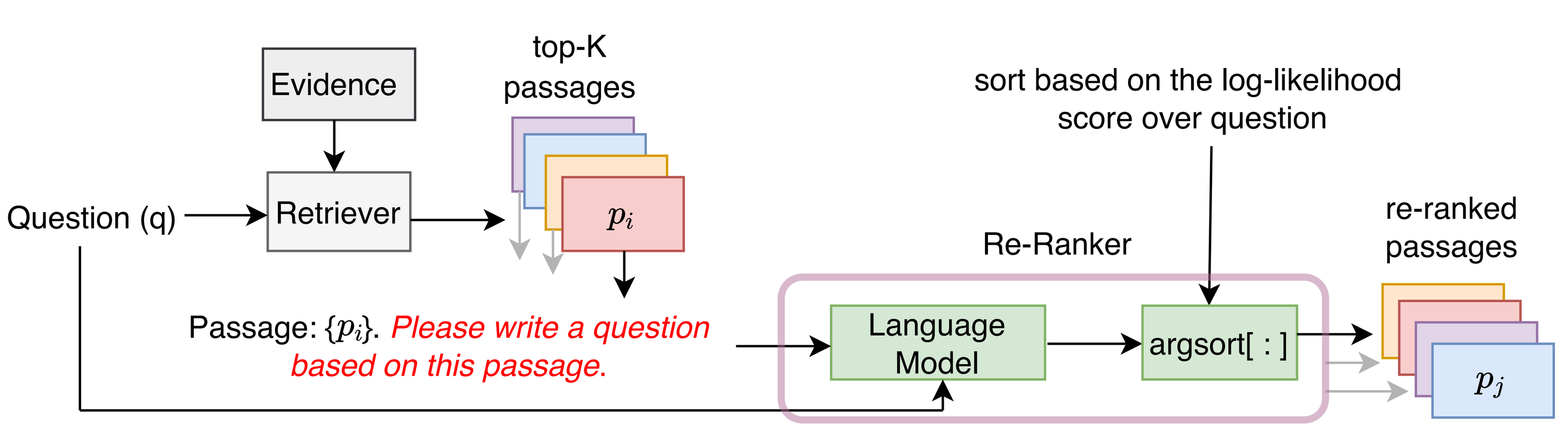
Significance

- *An improvement in retrieved passage rankings directly leads to more accurate answers.*

Method

UPR: Unsupervised Passage Re-ranking

We propose an unsupervised re-ranker to improve the ranking of a first-stage retriever.



1. Input: (i) Question (**q**)
   (ii) Evidence passages, such as segmented Wikipedia $\mathcal{D} = \{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_M\}$
2. Conditioned on the question, retrieve a large set of matching passages from evidence
$$\mathcal{Z} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K\}, \ \ \boldsymbol{z}_i \in \mathcal{D}$$
3. Compute the log-likelihood of the question tokens conditioned on a passage with teacher-forcing using a large pre-trained language model.
$$\log p(\boldsymbol{q} \mid \boldsymbol{z}_i) = \frac{1}{|\boldsymbol{q}|} \sum_t \log p(q_t \mid \boldsymbol{q}_{<t}, \boldsymbol{z}_i; \Theta), \ \ \forall \boldsymbol{z}_i \in \mathcal{Z}$$
4. Sort the passage ordering based on the log-likelihood score p(**q** | **z**$_i$) and select the top 100 passages for final QA task.

Experimental Settings

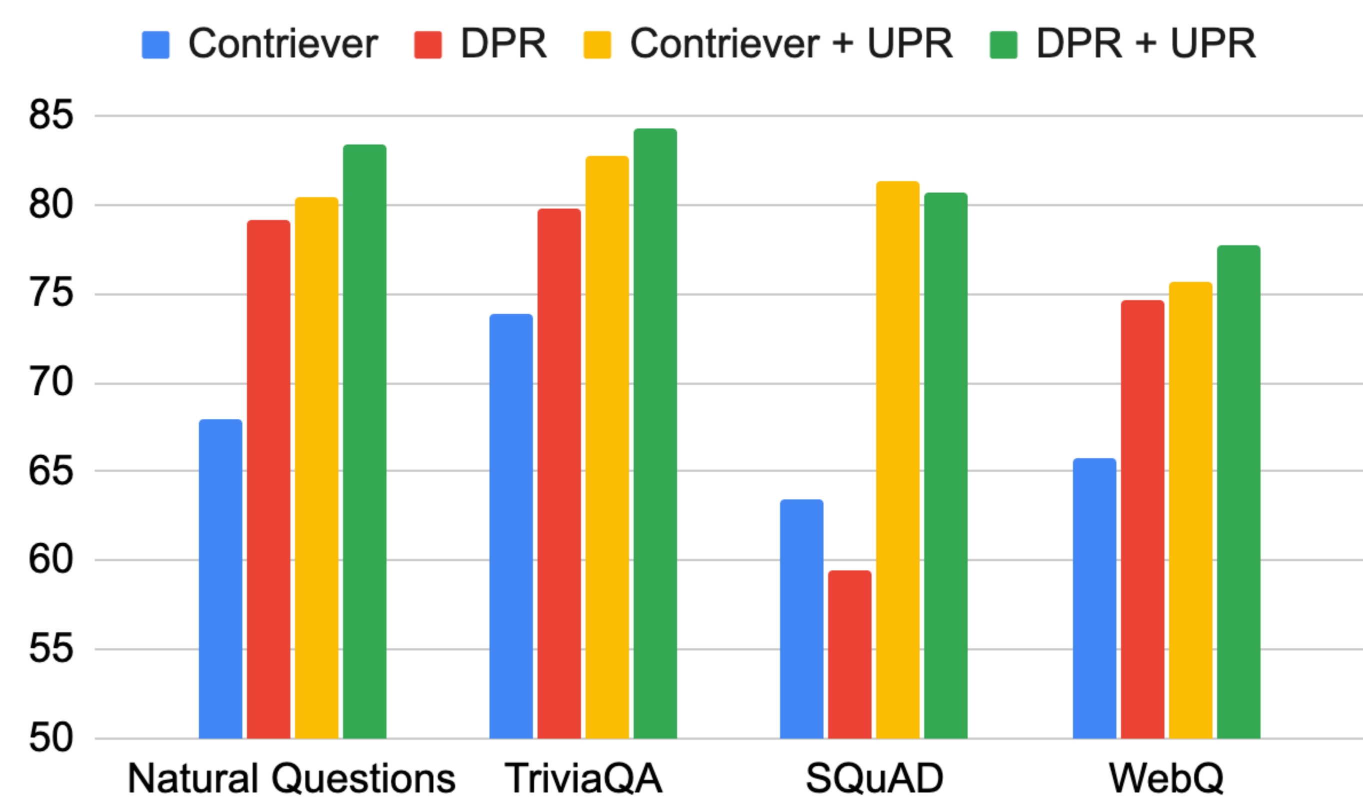
- We select the top-K = 1000 passages for re-ranking.
- We use the instruction-tuned T0 (3B) pre-trained language model in UPR.

UPR: Strengths

- ✓ No training data is needed.
- ✓ Uses off-the-shelf pre-trained language models without finetuning.
- ✓ Can be applied to both sparse retrievers (such as BM25) and dense retrievers (such as DPR).
- ✓ Leverages rich cross-attention between question and passage tokens resulting in improved passage rankings.

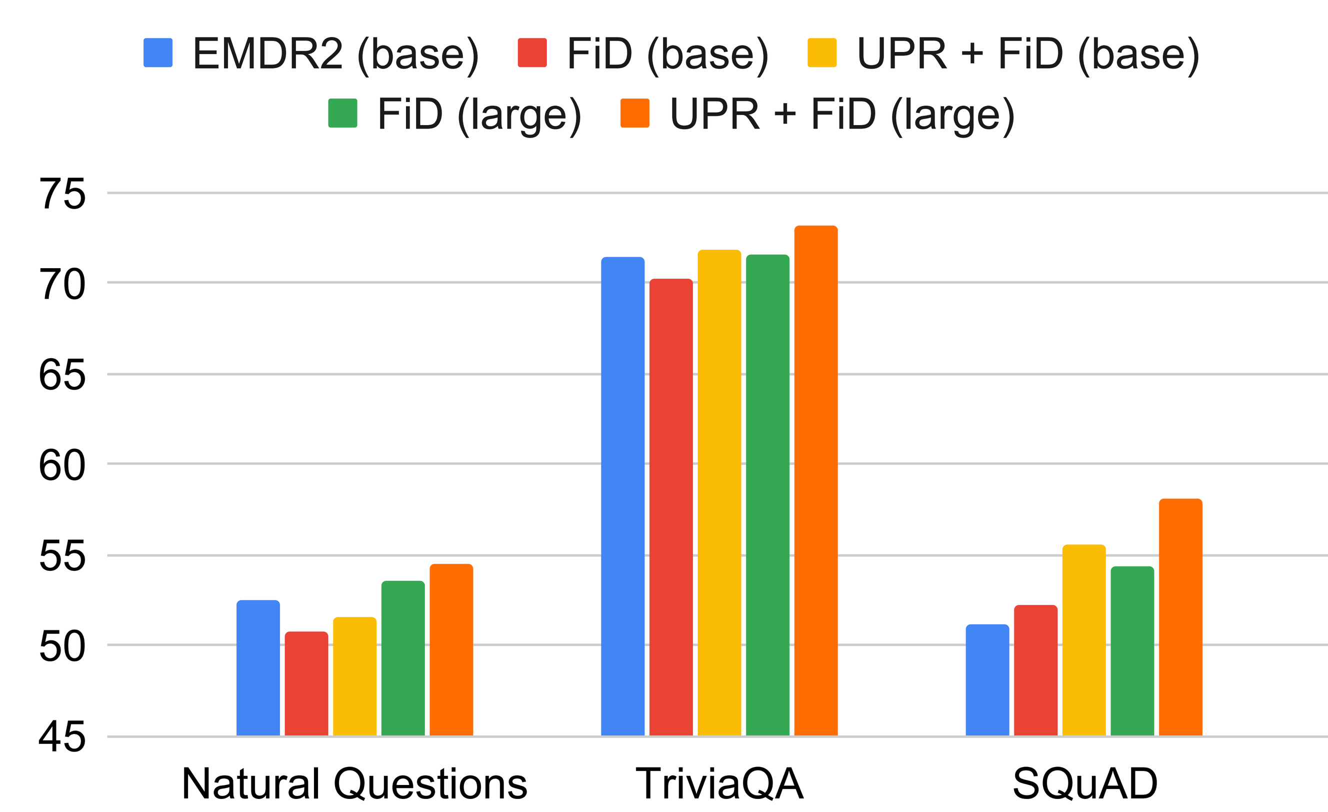
Results

Top-20 Passage Retrieval Accuracy



- UPR improves accuracy of both supervised and unsupervised retrievers by 6%-18%.
- An **unsupervised pipeline** consisting of Contriever and UPR outperforms supervised models like DPR.

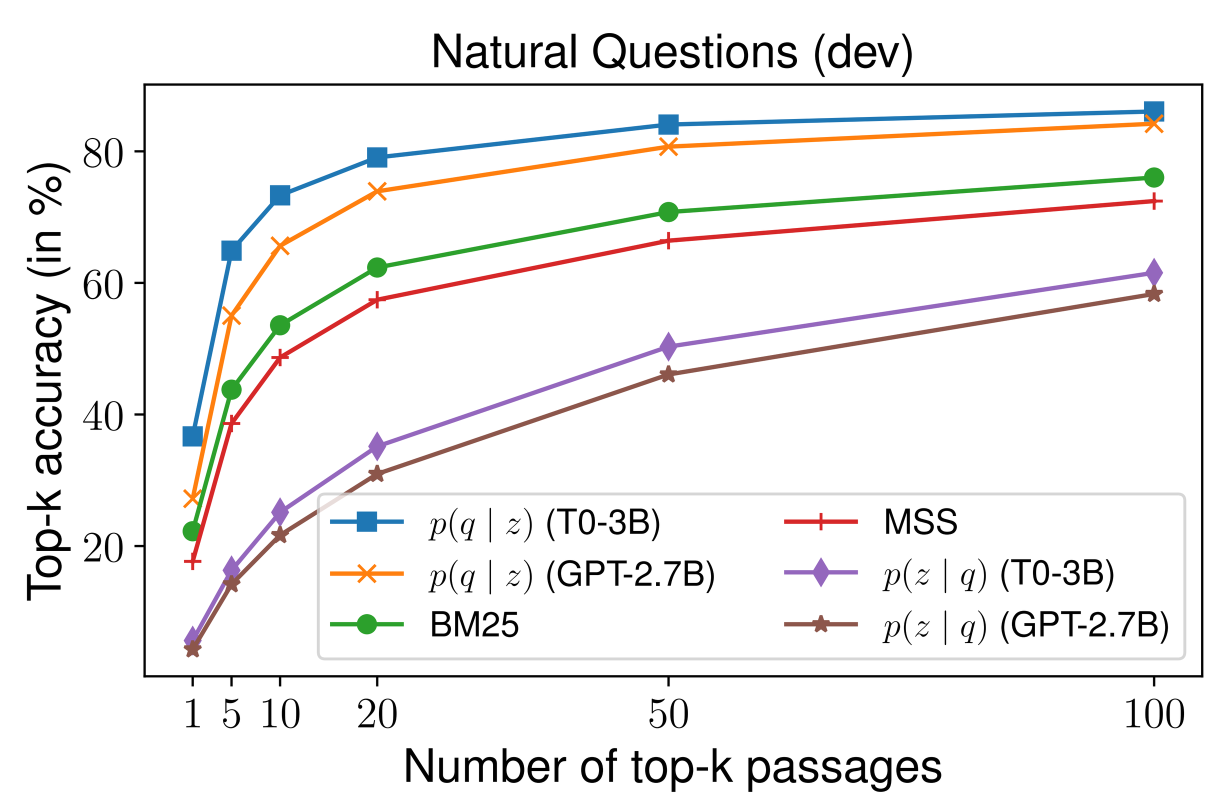
Open-Domain Question Answering



- Performance gains up to 3 points by doing inference with re-ranked passages.
- UPR + FiD pipeline is a more scalable alternative for working with large models.

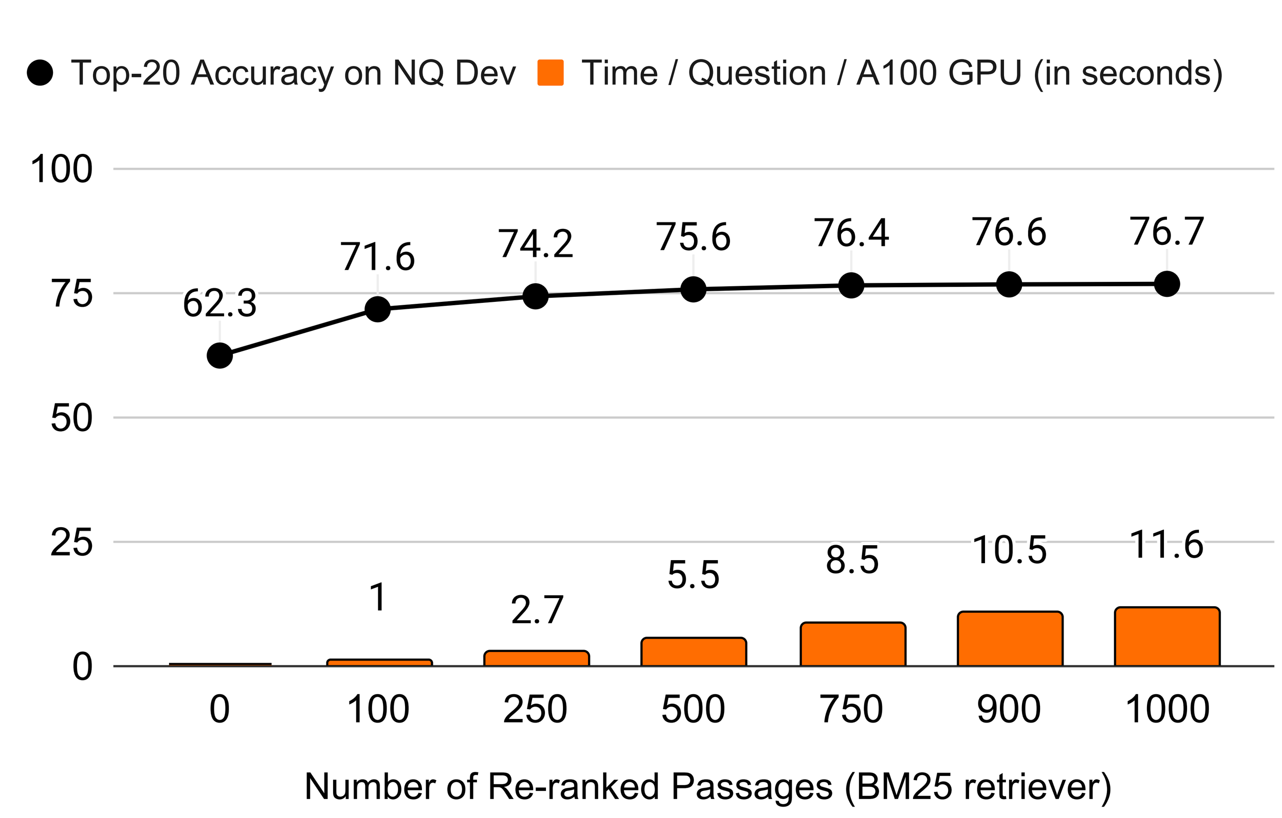
Analysis

Question Generation or Passage Generation



- Re-ranking based on passage generation p(**z**|**q**) is sub-optimal.

Passage Candidate Size and Latency



- **Pros**: Retrieval accuracy improves with a larger pool of candidate passages.
- **Cons**: Latency increases linearly with the number of passages.

EMNLP 2022

Code: *https://github.com/DevSinghSachan/unsupervised-passage-reranking*