Joint Passage Ranking for Diverse Multi-Answer Retrieval

Sewon Min†, Kenton Lee, Ming-Wei Chang, Kristina Toutanova, Hannaneh Hajishirzi

†Work done while interning at Google Research   contact: sewon@cs.washington.edu / @sewon__min

UWNLP

Google Research

1. Multi-answer retrieval

Q: What was Eli Whitney's job?

P1: Eli Whitney was an American **inventor**, widely known for ...

P2: Whitney worked as a **farm laborer** and **school teacher** to ...

P3: ... cotton gin was created by American **inventor** Eli Whitney.

Goal: maximize the coverage of multiple answers

3. Datasets & Metrics

	Avg. # answers
WebQSP	12.4
AmbigQA	2.2
TREC	4.1

✔ **MRecall**: Credit if all or 10+ answers are covered (*Intrinsic eval*)
✔ **F1**: End QA accuracy (*Extrinsic eval*)

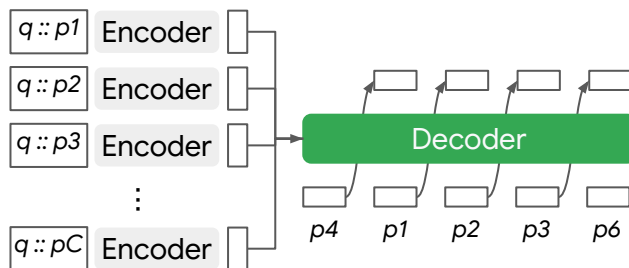## 2. JPR: Joint Passage Retrieval

Prior work
Independent ranking

$$P(p_i|q)$$

might choose P1 and P3 over P2

Our new formulation
***Joint** ranking*

$$P(p_1,p_2...p_K|q)$$

should choose {P1, P2} or {P2, P3} over {P1, P3}

$q :: p1$ → Encoder
$q :: p2$ → Encoder
$q :: p3$ → Encoder
⋮
$q :: pC$ → Encoder
→ Decoder →
p4 p1 p2 p3 p6

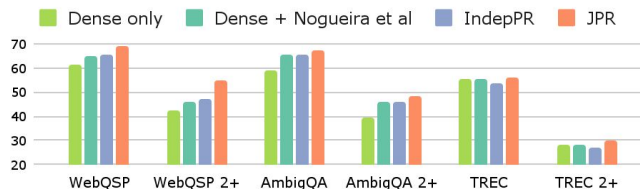1. Encoder-decoder architecture (T5)

2. Dynamic oracle training
   ✔ Each step is optimized with gold passages not already included in decoder prefix

3. Tree decoding algorithm
   ✔ Each step chooses between moving on to the next step vs. predicting more from the current step
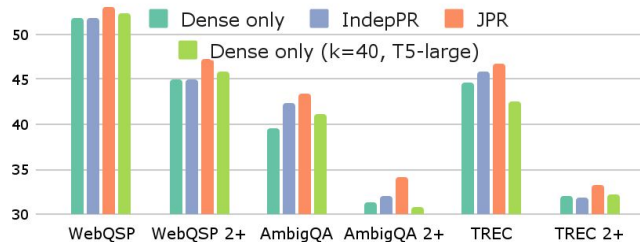
4. Experiments

Retrieval Recall (MRecall)

Legend: Dense only | Dense + Nogueira et al | IndepPR | JPR

WebQSP, WebQSP 2+, AmbigQA, AmbigQA 2+, TREC, TREC 2+

✔ Reranking is important
✔ JPR is better than independent ranking

Question Answering Accuracy (F1)

Feeding k=10 passages to Fusion-in-Decoder based on T5-3B

Legend: Dense only | IndepPR | JPR | Dense only (k=40, T5-large)

WebQSP, WebQSP 2+, AmbigQA, AmbigQA 2+, TREC, TREC 2+

✔ Reranking is important
(allows using fewer passages and bigger answer generation model with the same training cost)
✔ JPR is better than independent ranking