EMNLP 2021

Joint Passage Ranking for Diverse Multi-Answer Retrieval

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

University of Washington, Google Research

† Work done while interning at Google Research

Agenda

01 Problem: Multi-answer retrieval

02 Model: JPR (Joint Passage Retrieval)

03 Experiments

04 Summary

01

Problem: Multi-answer retrieval

Question What was Eli Whitney's job?

Answers
Inventor
Farm laborer
School teacher

passage

passage

passage

UWNLP Google Research

Question What was Eli Whitney's job?

Answers
Inventor
Farm laborer
School teacher

WIKIPEDIA

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

Whitney worked as a **farm laborer** and **school teacher** ...

... was created by American **inventor** Eli Whitney.

Multi-answer retrieval: find passages with the maximum coverage of **all distinct answers** to the question

UWNLP Google Research

Question What was Eli Whitney's job?

Answers Inventor

Farm laborer

School teacher

Existing retrieval

$P(p_i|q)$

for single-answer retrieval

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

Whitney worked as a **farm laborer** and **school teacher** ...

... was created by American **inventor** Eli Whitney.

1) valid passages competing with each other

2) may repeatedly retrieve the same answer

Need new formulation

Question What was Eli Whitney's job?

Answers Inventor

Farm laborer

School teacher

Our new formulation

Joint retrieval

$$P(p_1, p_2 \ldots p_k | q)$$

for multi-answer retrieval

Question What was Eli Whitney's job?

Answers Inventor

Farm laborer

School teacher

Our new formulation

Joint retrieval

$P(p_1, p_2 \ldots p_K | q)$

for multi-answer retrieval

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

Whitney worked as a **farm laborer** and **school teacher** ...

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

... was created by American **inventor** Eli Whitney.

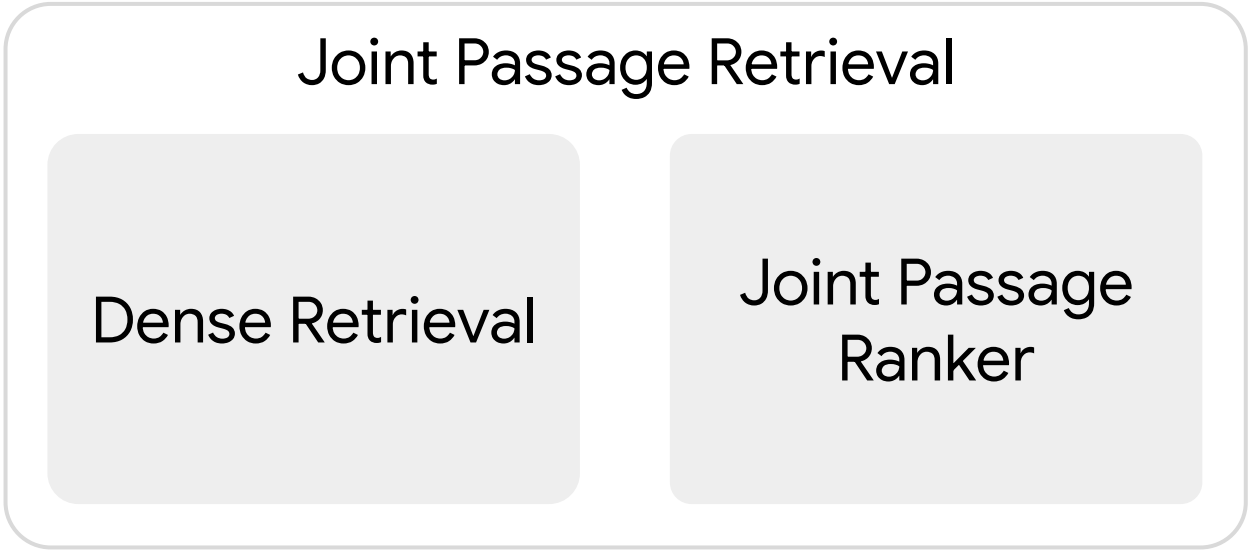
We study the **multi-answer retrieval problem** that is an underexplored problem

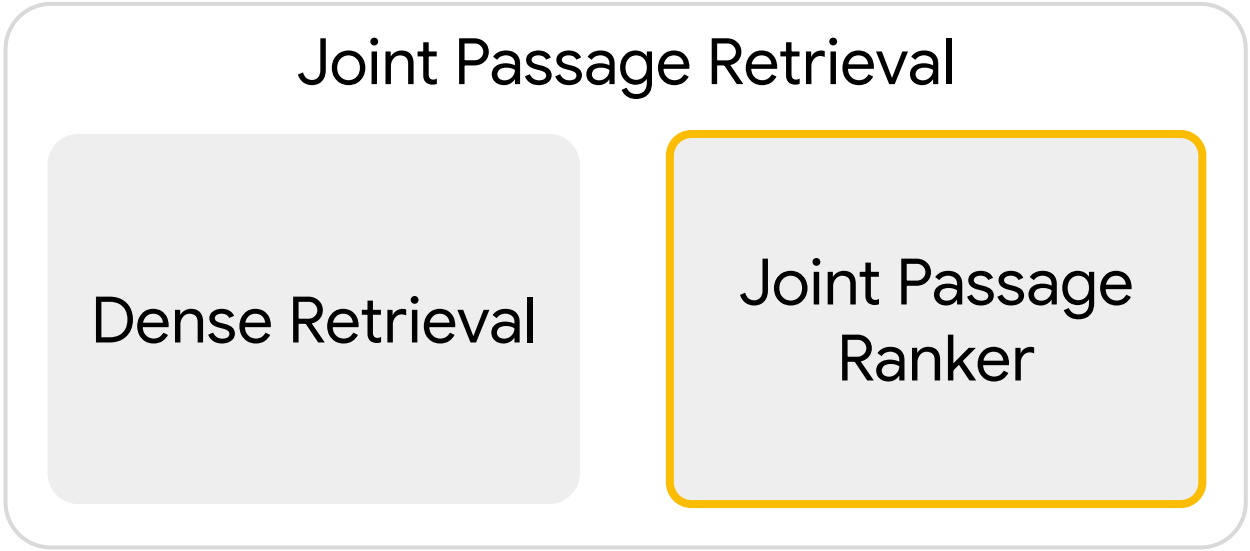
We claim that existing retrieval is not designed to maximize the answer coverage, and propose a new formulation that computes the **joint probability of $P(p_1, p_2 \ldots p_K | q)$**

We introduce **JPR (Joint Passage Retrieval)** that is an instance of our proposed formulation

02

Model: JPR - Joint Passage Retrieval


Joint Passage Retrieval
Dense Retrieval
Joint Passage Ranker

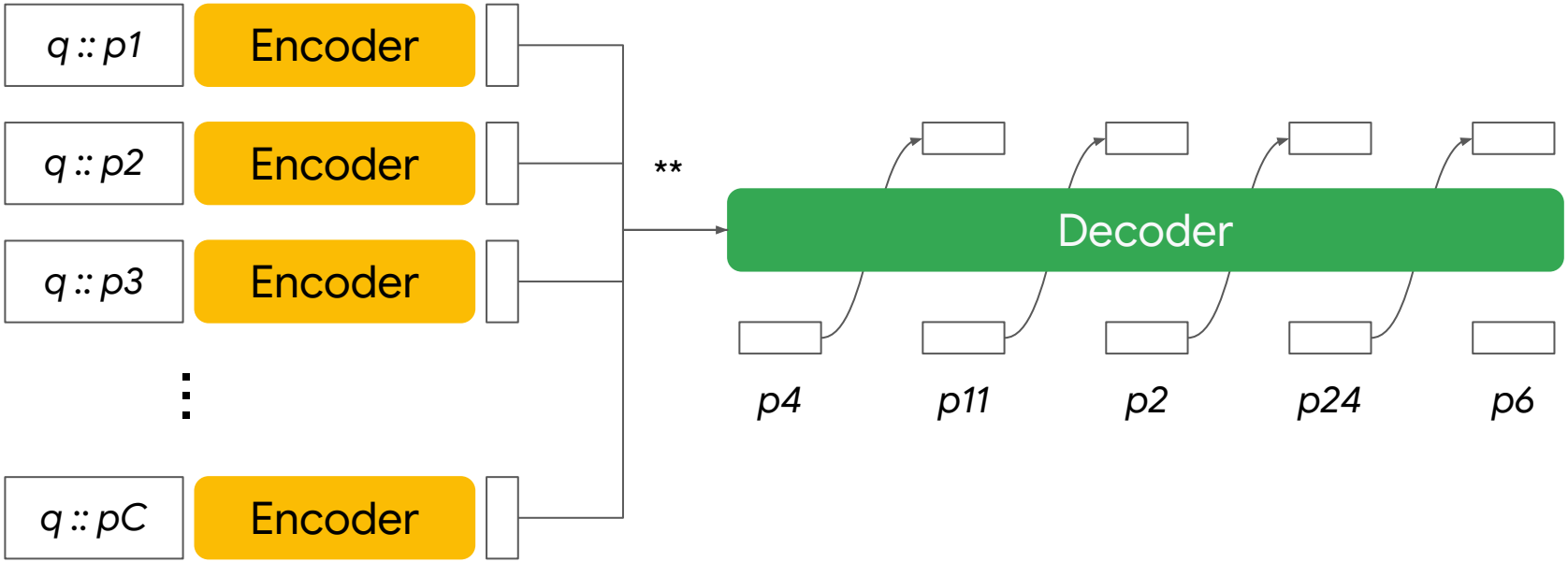


Rank ***C*** candidate passages (***C***>=100) to retrieve ***k*** passages (***k***=5, 10)

Usually the most expressive part compared to dense retrieval
Allow us to add joint ranking formulation in a more flexible way

Architecture

Exploit autoregressive architecture*

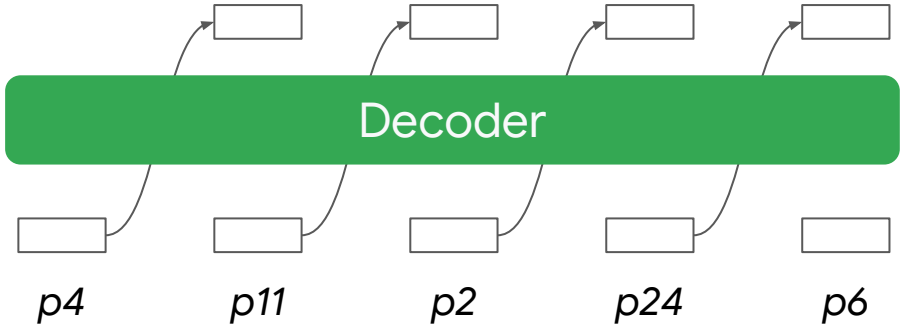


* We use T5 (Raffel et al., 2020)

** Taken from FiD (Izacard & Grave, 2021) as a way of operating with many passages

Modeling Challenges: Ordering Problem

- For training, groundtruth ordering is unknown
- Discrepancy between decoding a sequence and decoding a set



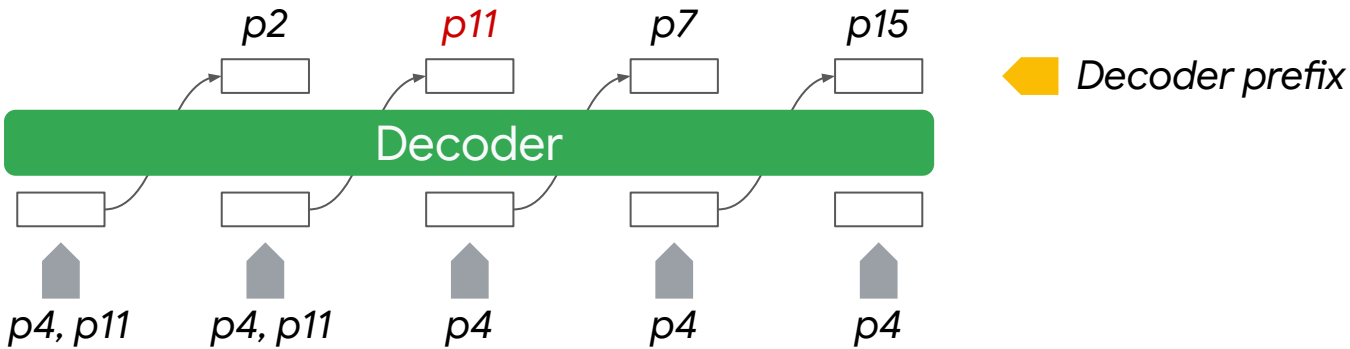
Need new training & decoding methods

Dynamic Oracle Training

Gold passages: p4, p11

: supervised through cross entropy

[p2, p11, p7, p15, p4]



learn to predict any gold that are *not covered yet*

(Less dependent to predefined ordering between gold passages)

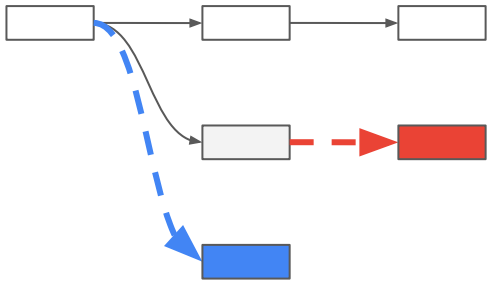
Tree Decoding

Sequence Decode

Moving on to the next step == Exploring new answers

In practice, the number of distinct answers is usually less than *k* and more predictions from the same step can be better (to recover mistakes from earlier steps)

Tree Decode



See the paper for the exact algorithm!

Move on to the next step vs. More prediction from the same step

03

Experiments

Datasets

	# questions			Avg. # answers
	Train	Dev	Test	
WebQSP (Yih et al. 2016)	2,756	241	1,582	12.4
AmbigQA (Min et al. 2020)	10,036	2,002	2,004	2.2
TREC (Baudis & Sedivy, 2015)	1,250	119	654	4.1

Baselines

- Dense retrieval only

Baselines

- Dense retrieval only
- Dense retrieval + Nogueira et al. 2020 (SOTA reranker)

Baselines

- Dense retrieval only
- Dense retrieval + Nogueira et al. 2020 (SOTA reranker)
- IndepPR (Strict non-autoregressive version of JPR)

Baselines

- Dense retrieval only
- Dense retrieval + Nogueira et al. 2020 (SOTA reranker)
- IndepPR (Strict non-autoregressive version of JPR)
  - Very strong baseline; SOTA on competitive single-answer benchmark (NQ)

(All independently ranking passages)

Evaluation

1. Retrieval recall (Intrinsic evaluation)
   - MRecall* on *k* retrieved passages (*k*=5, 10; in this talk, *k*=10)

* Considered as “Hit” when all answers are covered by k passages; Exact definition in the paper

Evaluation

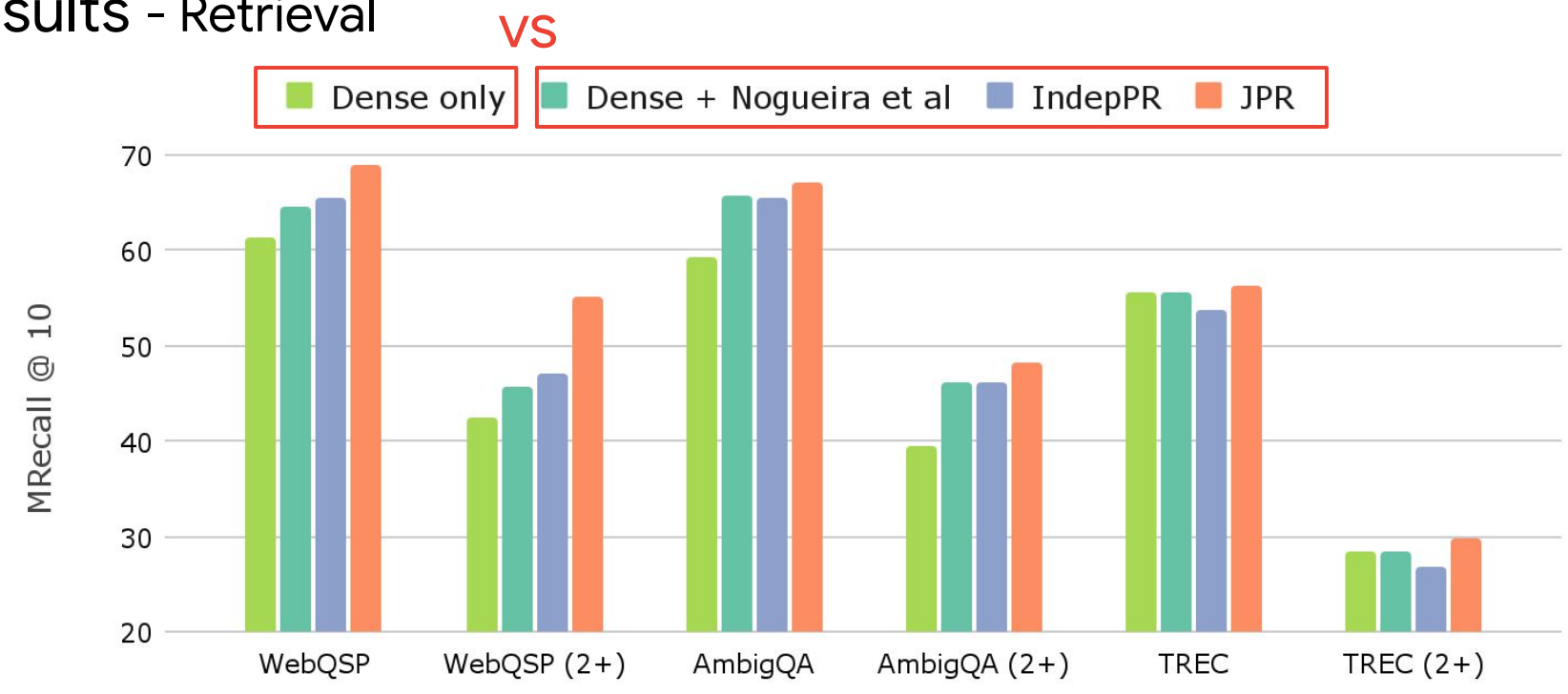
1. Retrieval recall (Intrinsic evaluation)
   - MRecall* on *k* retrieved passages (*k*=5, 10; in this talk, *k*=10)

2. End QA accuracy (Extrinsic evaluation)
   - F1 on short answers generated by a subsequence answer generation model**

Better retrieval → Better answer generation

* Considered as “Hit” when all answers are covered by k passages; Exact definition in the paper
** We use Fusion-in-Decoder (Izacard & Grave, 2021)

Results - Retrieval

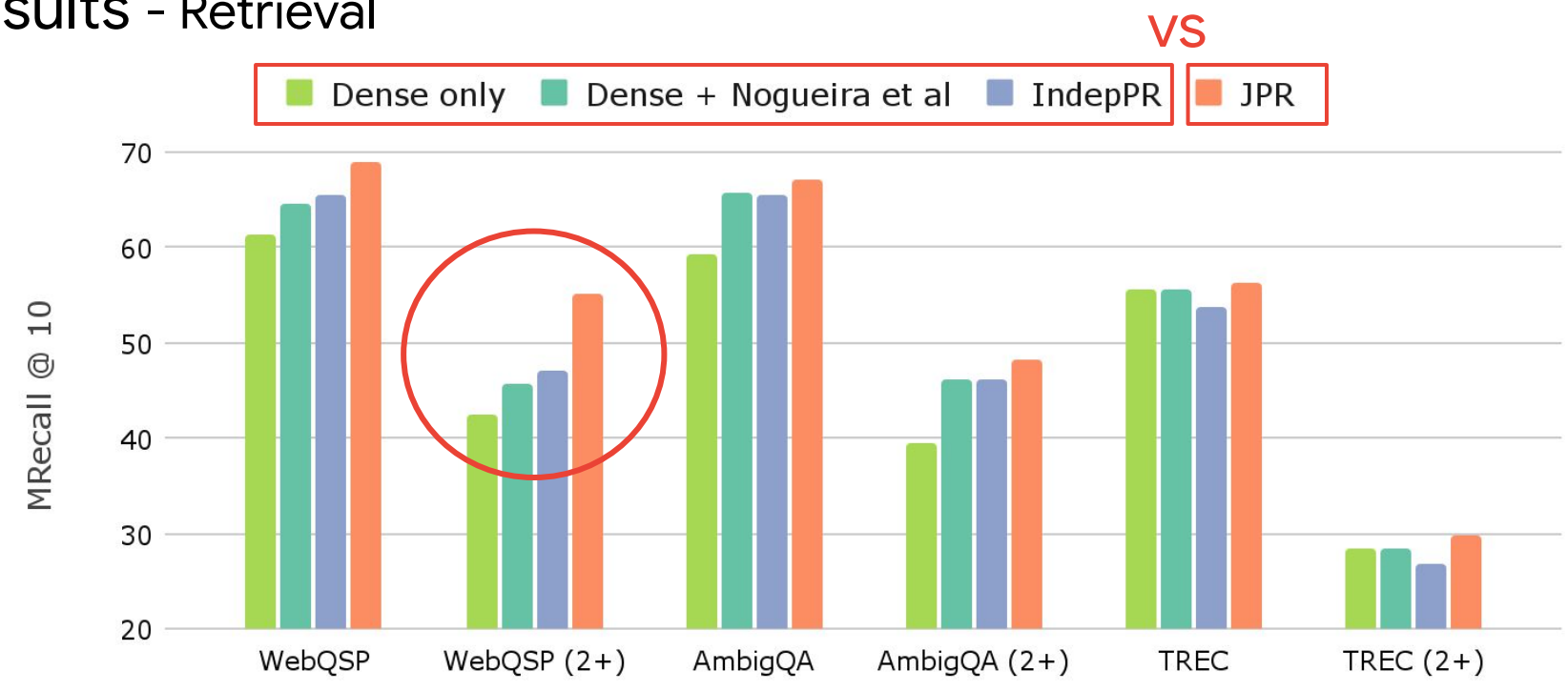


1. Reranking is important

* result on dev

UWNLP Google Research

Results - Retrieval



1. Reranking is important

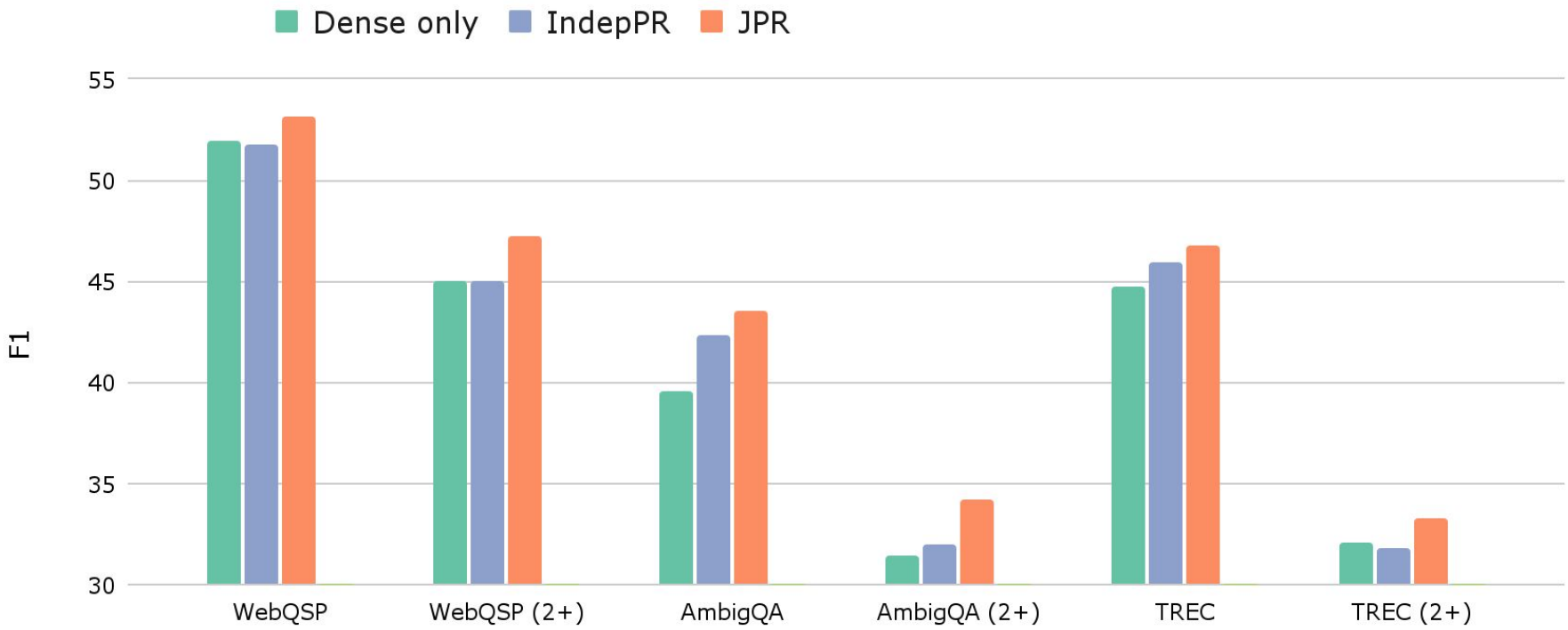
2. JPR is consistently better than independent ranking baselines

* result on dev

UWNLP Google Research

Results - Question Answering

k=10, answer generation based on T5-3B

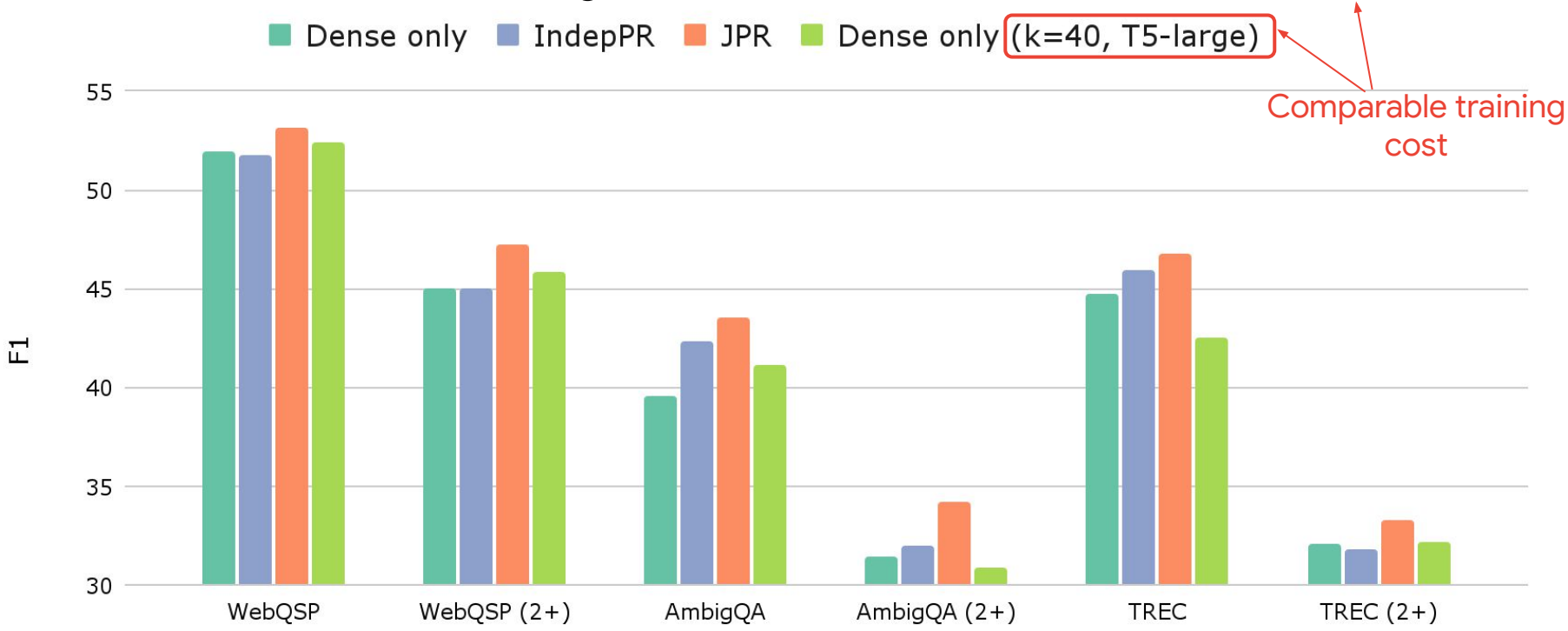


1. No reranking < independent ranking < joint ranking

What if we feed more passages instead of using a reranker?

* result on test

Results - Question Answering



1. No reranking < independent ranking < joint ranking

2. Better reranking allows you to use fewer passages & larger answer generation model

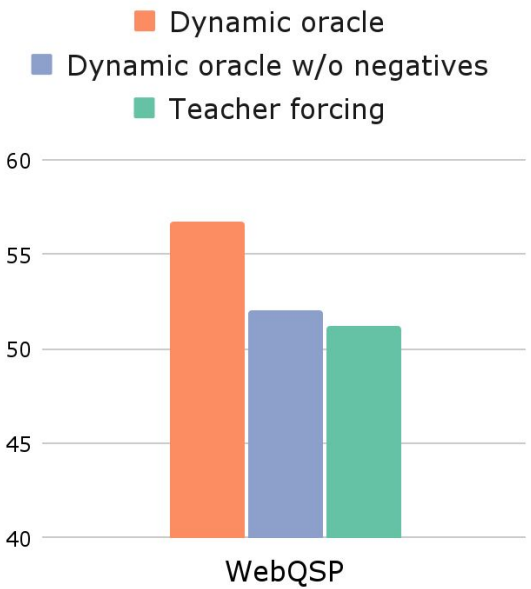
* result on test

Google Research

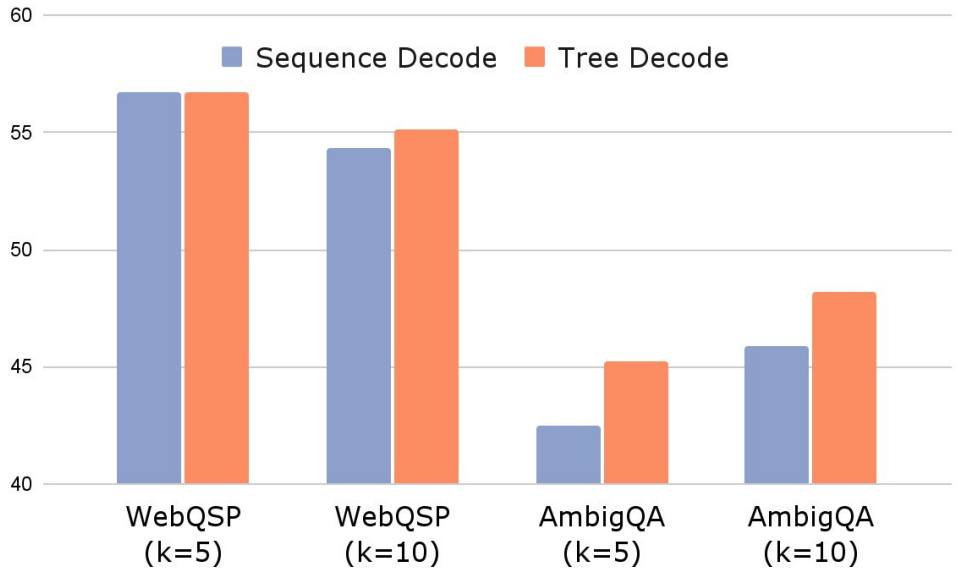
Ablations

Take a look at our paper for more details

Ablation in training methods



Ablation in decoding methods



Dynamic oracle is important for training

TreeDecode boosts the performance
(esp. on datasets with smaller number of distinct answers)

* result on dev

(Metric: MRecall on questions w/ 2+ answers)

UWNLP Google Research

04

Summary

We propose a formulation for **jointly ranking passages** to maximize the answer coverage in multi-answer retrieval

We introduce **JPR (Joint Passage Retrieval)** based on an autoregressive architecture and better training & decoding

Experiments on three multi-answer datasets show JPR achieves **better answer coverage** and leads to a **new SOTA result in QA**

Future work could extend the scope beyond QA

UWNLP Google Research

Thank You

Come to the poster session!

Paper: arxiv.org/abs/2104.08445
Code/data will be available soon!

UWNLP Google Research