

# Collective Relevance Labeling for Passage Retrieval

Jihyuk Kim

Yonsei Univ

Minsoo Kim

Seoul National Univ

Seung-won Hwang

Seoul National Univ



YONSEI  
UNIVERSITY



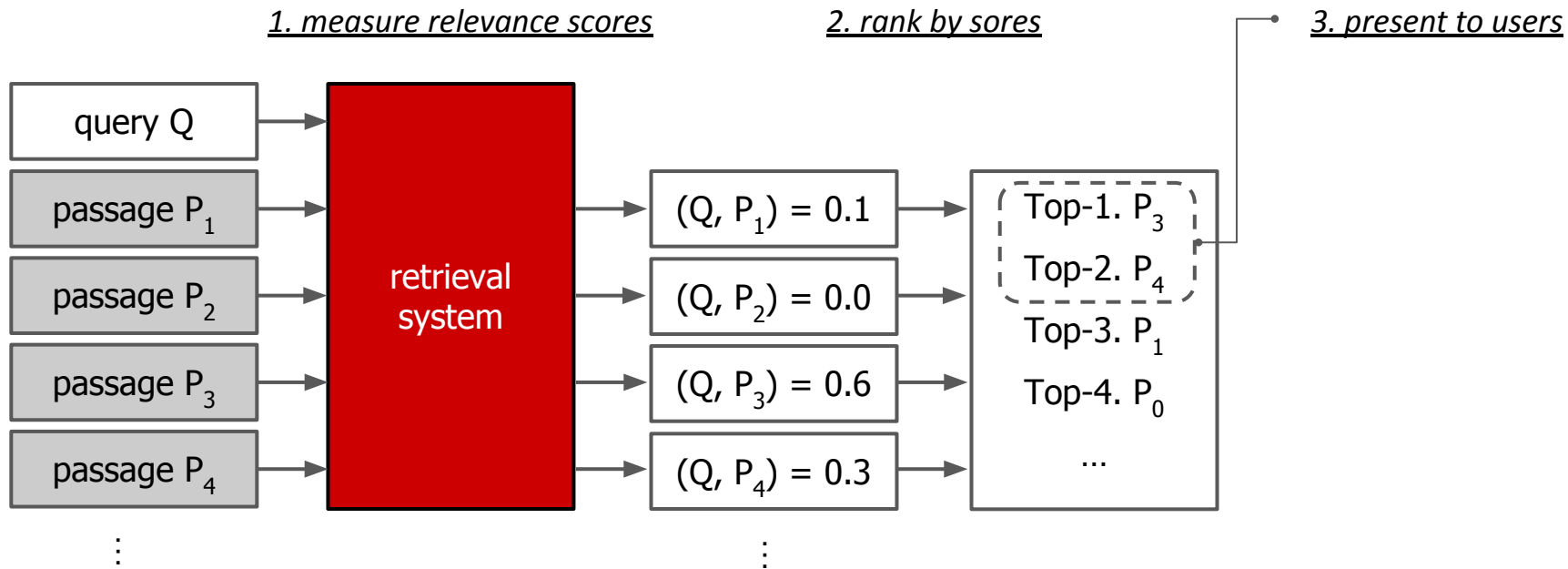
SEOUL  
NATIONAL  
UNIVERSITY



SEOUL  
NATIONAL  
UNIVERSITY

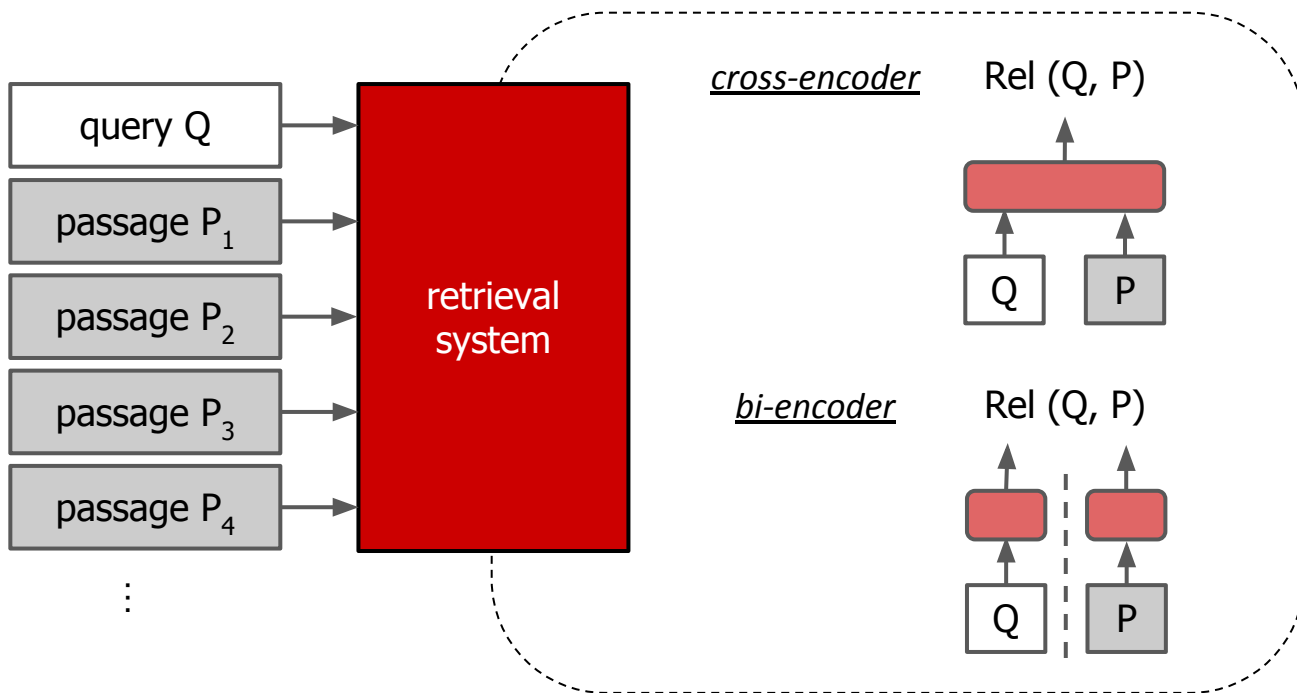
# Task: Passage Retrieval

- We aim to **retrieve passages** that are relevant to the given query.



# Approach: cross- vs bi-encoder

- We adopt the **bi-encoder** approach, for query-document matching,
  - targeting search on large-scale corpus.



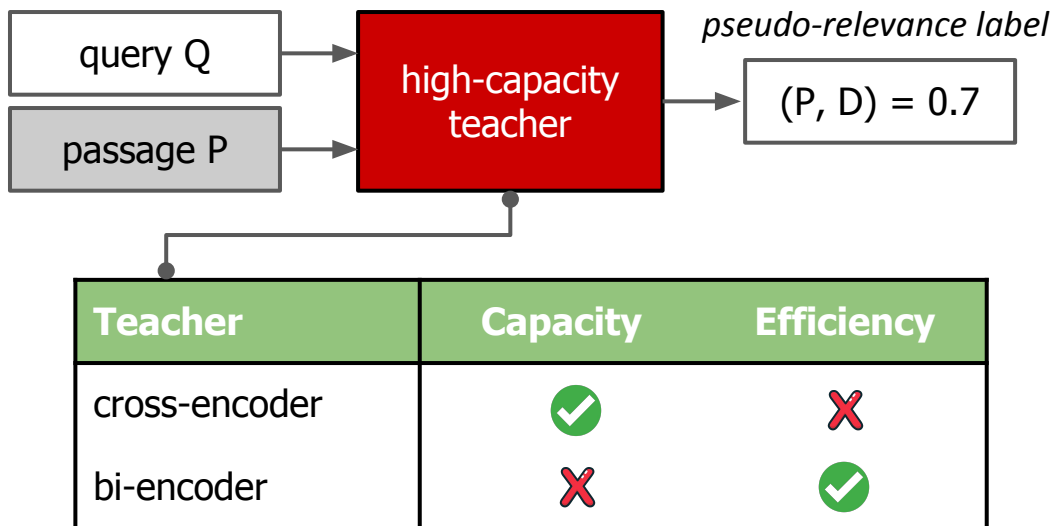
# Challenge

- We tackle the challenge of **incomplete relevance annotation**.
  - Too many passages that should be annotated!

Dataset	# of passages	# of queries	# of relevant passages per query
MSMARO Train set	8.8M	500k	1.1 → Only tiny portions are annotated!
TREC-DL 2019 Test set		43	58.2

# Approach

- We adopt knowledge distillation,
  - to let a high-capacity teacher provide relevance labels.



# Our teacher







- We present a novel teacher, called **collective bi-encoder**,
  - to achieve both high capacity and efficiency.



Teacher	Capacity	Efficiency
cross-encoder	✓	✗
bi-encoder	✗	✓
<b>collective bi-enc</b>	✓	✓










# Motivation

- We leverage **collective knowledge (CK)**, to improve capacity of bi-encoder.

Q/P	Content	Rel Prediction
		Individual
Q	Is <b>caffeine</b> a <b>narcotic</b> ?	
Top-1 P 	... <b>caffeine</b> , and codeine ... <b>treat headaches.</b>	
Top-2 P 	butalbital, and <b>caffeine</b> ... <b>treat tension headaches.</b>	
Top-3 P 	... <b>Caffeine</b> is considered a safe ingredient. ...	

# Motivation

- We leverage **collective knowledge (CK)**, to improve capacity of bi-encoder.

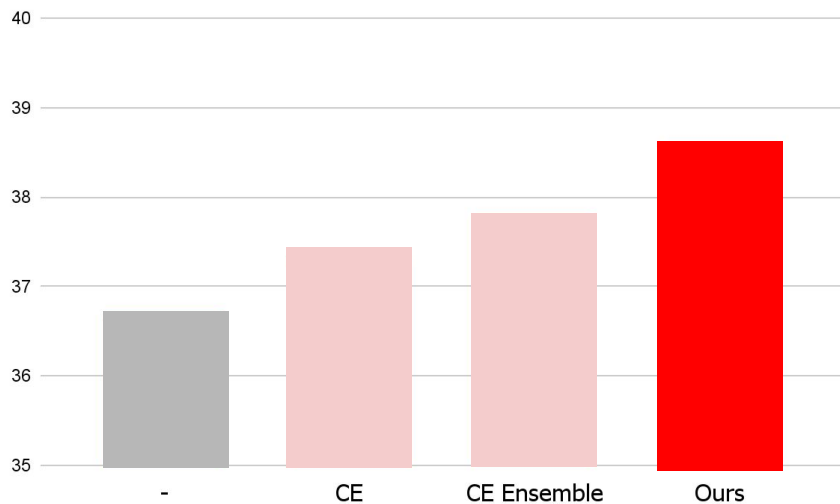
Q/P	Content	Rel Prediction	
		Individual	Collective
Q	Is <b>caffeine</b> a <b>narcotic</b> ?		
Top-1 P 	... <b>caffeine</b> , and codeine ... <b>treat headaches.</b>		
Top-2 P 	butalbital, and <b>caffeine</b> ... <b>treat tension headaches.</b>		
Top-3 P 	... <b>Caffeine</b> is considered a safe ingredient. ...		
CK	<i>treat tension headaches</i>		



# Experimental results

- ✓ **Teachers** provide more informed labels than incomplete labels.
- ✓ **Our collective bi-enc teacher** is the more effective than CE teachers.

MSMARCO (MRR@10)



TREC-DL'19 (NDCG@10)

