Improving BERT-based Query-by-Document Retrieval with Multi-Task Optimization

Universiteit Leiden

Amin Abolghasemi[1], Suzan Verberne[1], Leif Azzopardi[2]
[1] Leiden University, [2] University of Strathclyde
{m.a.abolghasemi, s.verberne}@liacs.leidenuniv.nl leif.azzopardi@strath.ac.uk

SCAN ME

ECIR 2022

Introduction

- Query-by-document (QBD) retrieval is an Information Retrieval task in which a seed document acts as the query and the goal is to retrieve related documents.
- Transformer-based ranking models have proven to be highly effective at taking advantage of context. However, recent work showed that transformer-based models which handle longer input sequences are not necessarily more effective when being used in retrieval tasks on long texts.
- We improve the retrieval effectiveness of the BERT re-ranker, proposing an extension to the fine-tuning step to better exploit the context of queries.
- We use an additional document-level representation learning objective besides the ranking objective when fine-tuning the BERT re-ranker.

Method

- Cross-Encoder BERT Ranker:

$$s(q,d) = BERT([CLS]q[SEP]d[SEP])_{[CLS]} * W_p$$

- Representation Learning with BERT:

$$r_q = BERT([CLS]q[SEP])_{[CLS]} \qquad r_d = BERT([CLS]d[SEP])_{[CLS]}$$

- Pairwise Cross-entropy Softmax Loss:

$$l_{ranking} = -\log \frac{e^{score(q,d^+)}}{e^{score(q,d^+)} + e^{score(q,d^-)}}$$

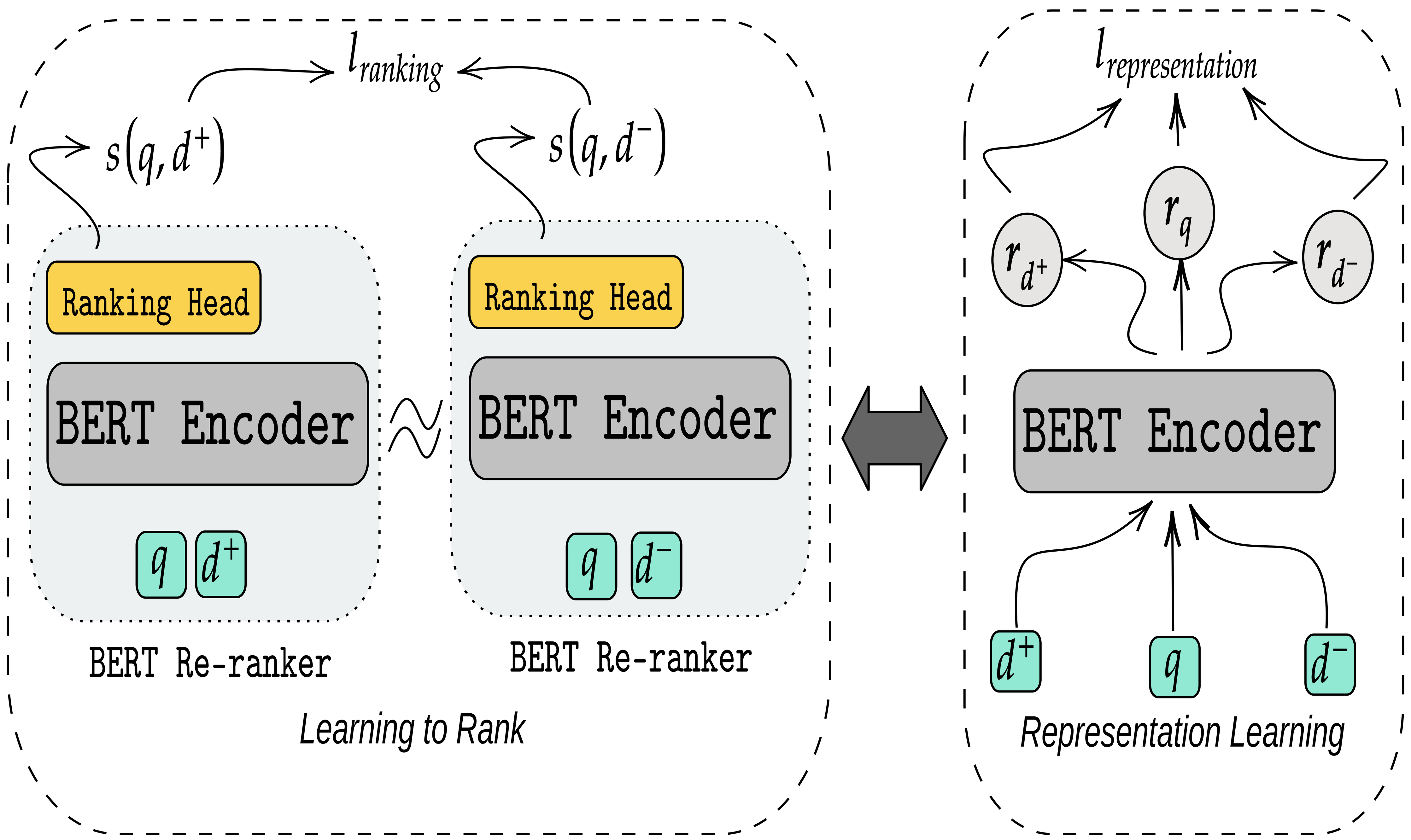
- Triplet Loss:

$$l_{representation} = max\{(f(r_q, r_{d^+}) - f(r_q, r_{d^-}) + margin), 0\}$$

- In the fine-tuning step, we jointly optimize both the ranking loss and representation learning loss:

$$l_{aggregated} = l_{ranking} + \lambda\, l_{representation}$$

Multi-Task Fine-Tuning of BERT Ranker



- The Multi Task Fine Tuning (MTFT) is achieved by providing training instances consisting of triples.
- The same training triples (q, d^+, d^-) are used in each step. The BERT re-rankers are the same, and the BERT encoder is shared between the ranking and representation learning tasks.
- The resulting model MTFT-BERT re-ranker obtains consistently better retrieval quality than the original BERT re-ranker using the same neural ranking architecture.

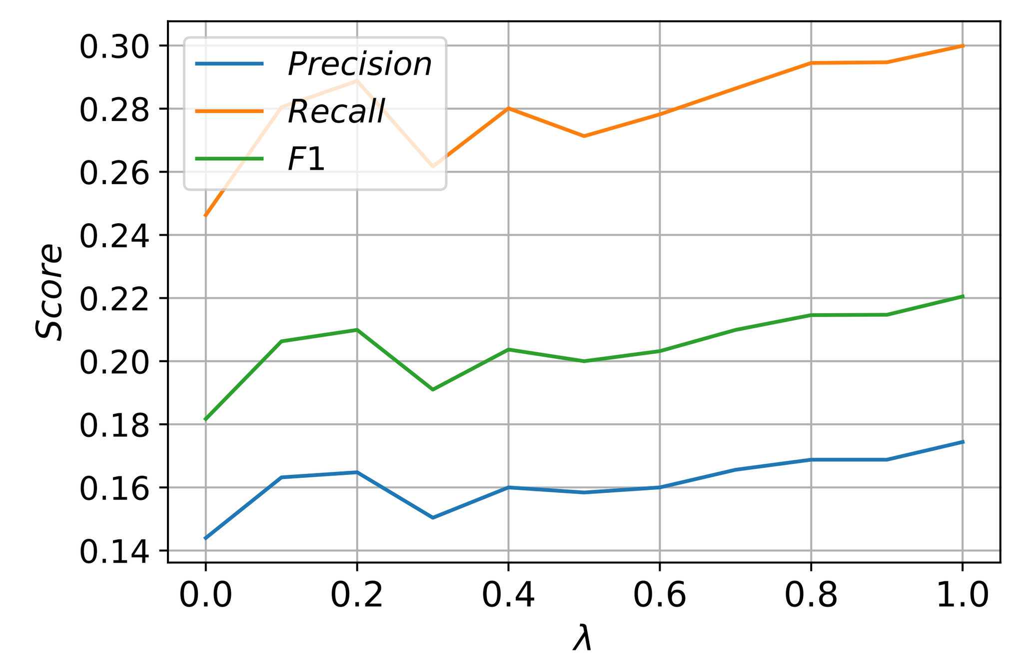
Results

The ranking results with BM25 and optimized BM25 as initial rankers for COLIEE 2021

Model	Initial Ranker	Precision %	Recall %	F1 %
BM25	-	8.80	16.51	11.48
TLIR	-	15.33	25.56	19.17
Optimized BM25	-	17.00	25.36	20.35
BERT	BM25	10.48	18.80	13.46
MTFT-BERT	BM25	12.08	21.59	15.49
BERT	Optimized BM25	14.40	24.63	18.17
MTFT-BERT	Optimized BM25	17.44	29.99	22.05

The ranking results on the SciDocs benchmark

	Co-View		Co-Read		Cite		Co-Cite	
Model	MAP	nDCG	MAP	nDCG	MAP	nDCG	MAP	nDCG
SPECTER	83.6%	0.915	84.5%	0.924	88.3%	0.949	88.1%	0.948
SPECTER w/ HF	83.4%	0.914	85.1%	0.927	92.0%	0.966	88.0%	0.947
BM25	75.4%	0.874	75.6%	0.881	73.5%	0.876	76.3%	0.890
Optimized BM25	76.3%	0.877	76.1%	0.881	75.3%	0.884	77.4%	0.896
BERT	85.2%	0.925	87.5%	0.940	94.0%	0.975	89.7%	0.955
MTFT-BERT	86.2%	0.930	87.7%	0.940	94.2%	0.976	91.0%	0.961



The evaluation results of MTFT-BERT with various λ for COLIEE 2021