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 when fine-tuning the BERT re-ranker.

Method

• Cross-Encoder BERT Ranker:

\[ s(q, d) = \text{BERT}([\text{CLS}]q[\text{SEP}]d[\text{SEP}])_{\text{CLS}} \times W_p \]

• Representation Learning with BERT:

\[ r_q = \text{BERT}([\text{CLS}]q[\text{SEP}])_{\text{CLS}} \quad \text{and} \quad r_d = \text{BERT}([\text{CLS}]d[\text{SEP}])_{\text{CLS}} \]

• Pairwise Cross-entropy Softmax Loss:

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

• Triplet Loss:

\[ l_{\text{representation}} = \max (f(r_q, r_d) - f(r_q, r_d) + \text{margin}, 0) \]

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

\[ l_{\text{aggregate}} = l_{\text{ranking}} + \lambda l_{\text{representation}} \]

Multi-Task Fine-Tuning of BERT Ranker

The evaluation results of MTFT-BERT with various \( \lambda \) for COLIEE 2021

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

Results