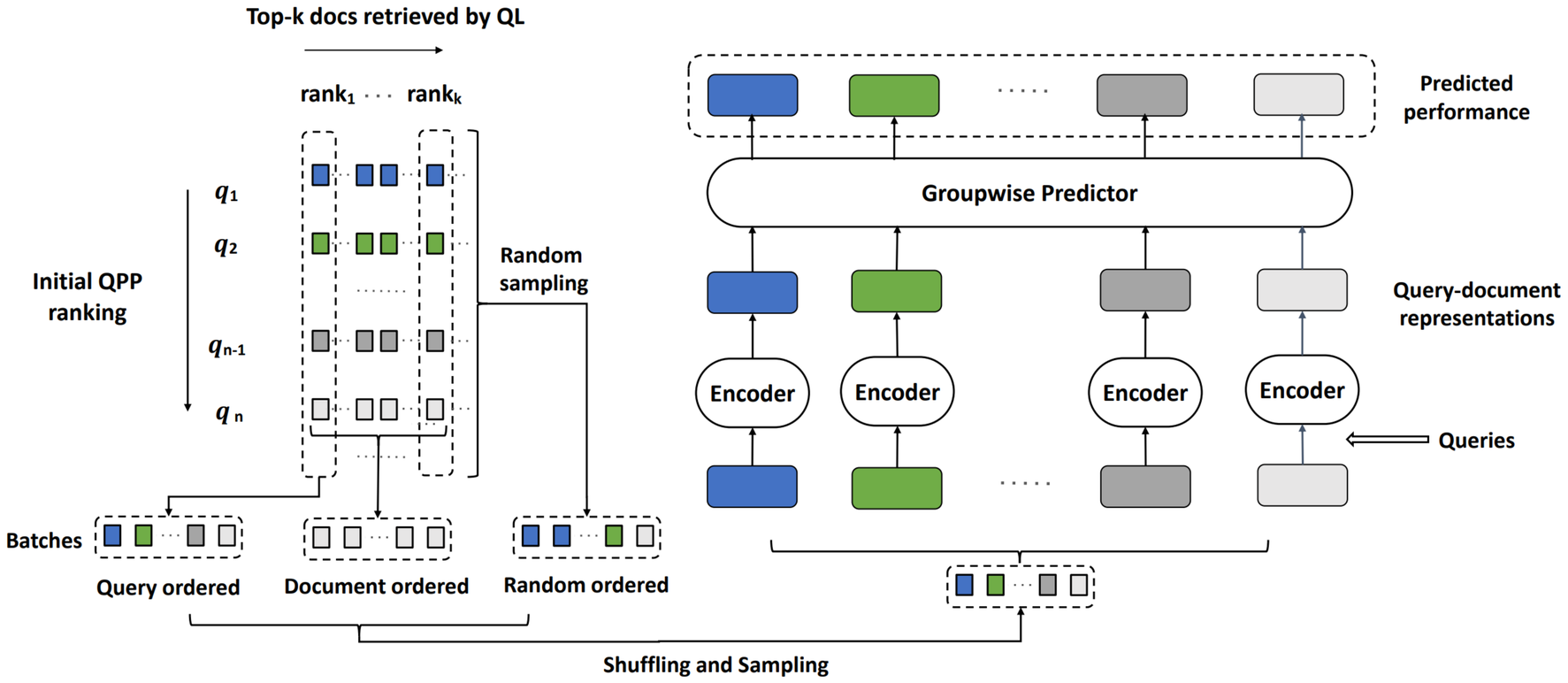
Groupwise Query Performance Prediction with BERT

Xiaoyang Chen[1,2], Ben He[1,2], Le Sun[2]

[1]University of Chinese Academy of Sciences; [2]Chinese Information Processing Laboratory, Institute of Software, Chinese Academy of Sciences



Background

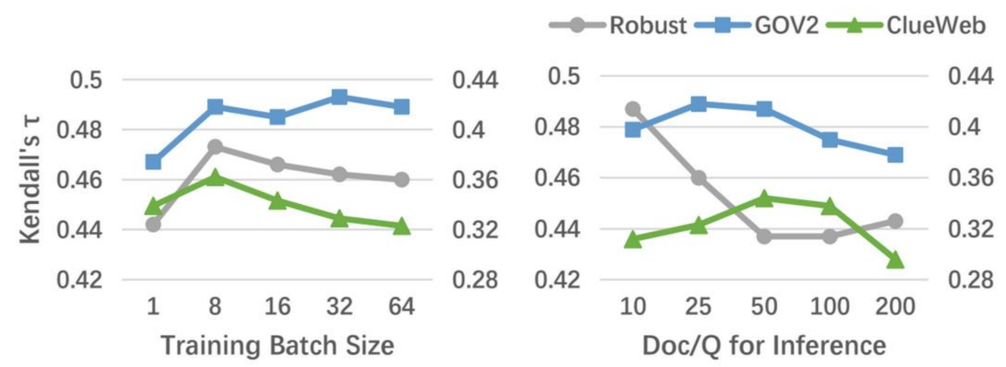
- **Query performance prediction (QPP)** aims to automatically estimate the search results quality of a given query.
- **BERT-QPP[1]** : Recent results demonstrate that BERT effectively improves the performance of **post-retrieval QPP**.
- The **groupwise methods** have achieved superior performance on learning-to-rank [2] and BERT re-ranking benchmarks [3].
- **This paper** proposes a BERT-based groupwise QPP method, which simultaneously incorporates the **cross-query** and **cross-document** information within an end-to-end learning framework.

Method

- **Encoding Query-Document Pairs:**
  - Put *[CLS]Query[SEP]Document[SEP]* into the BERT encoder.
- **Groupwise Prediction:**
  - A four-layer transformer enables the cross attention among the *[CLS]* vectors in each batch.
- **Three Different Ranking Context**
  - **Random Order:** all query-document pairs are shuffled.
  - **Query Order:** position ids are assigned by the initial query order derived by n(σ_X%).
  - **Doc Order:** position ids are assigned by initial document ranking.

Experiments

Method	Robust04		GOV2		ClueWeb09-B	
	P-ρ	K-τ	P-ρ	K-τ	P-ρ	K-τ
Clarity	0.528	0.385	0.428	0.291	0.300	0.213
QF	0.390	0.324	0.447	0.314	0.163	0.072
WIG	0.546	0.379	0.502	0.346	0.316	0.210
NQC	0.516	0.388	0.381	0.323	0.127	0.138
SMV	0.534	0.378	0.352	0.303	0.236	0.183
UEF	0.502	0.402	0.470	0.329	0.301	0.211
σ_k	0.522	0.389	0.381	0.323	0.234	0.177
$n(\sigma_{X\%})$	0.589	0.386	0.556	0.386	0.334	0.247
RSD	0.455	0.352	0.444	0.276	0.193	0.096
WAND[$n(\sigma_{X\%})$]	0.566	0.386	0.580	0.411	0.236	0.142
NeuralQPP	0.611	0.408	0.540	0.357	0.367	0.229
BERT-Small	0.591	0.391	0.615	0.436	0.394	0.278
BERT-Base	0.585	0.423	0.637	0.454	0.447	0.321
BERT-Large	0.579	0.422	0.645	0.461	0.342	0.251
(Random order)-base	0.608*	0.449*	0.665*	0.479*	0.481*	0.353*
(Query order)-base	**0.615***	0.456*	0.676*	0.486*	0.455	0.327
(Doc order)-base	0.563	0.383	0.660*	0.476*	0.365	0.262
(Query+Doc)-base	0.598	0.452*	0.682*	0.496*	0.438	0.317
(R+Q+D)-small	0.590	0.419*	0.680*	0.500*	0.437*	0.305*
(R+Q+D)-base	0.608*	0.460*	0.676*	0.489*	0.449	0.324
(R+Q+D)-large	0.612*	**0.470***	**0.688***	**0.508***	**0.545***	**0.399***



- **Overall effectiveness:**
  - In general, the model incorporating three ranking contexts, i.e. **(R+Q+D)** outperforms the BERT baselines with all three different model sizes.
  - Different ranking contexts leads to different observations on three datasets.
- **Impact of factors:**
  - **Impact of training batch size:** the method works best with a group size of 8.
  - **Impact of documents per query used in inference:** inference with less than 100 documents per query on all three collections yields the best results.

References: [1] N Arabzadeh et al. BERT-QPP: Contextualized Pre-trained Transformers for Query Performance Prediction. In: CIKM. 2021.
[2] L Pang et al. SetRank: Learning a Permutation-Invariant Ranking Model for Information Retrieval. In SIGIR. 2020.
[3] X Chen et al. Co-BERT: A Context-Aware BERT Retrieval Model Incorporating Local and Query-specific Context. In: Arxiv. 2021.