Corrections for the paper

Investigating Passage-level Relevance and Its Role in Document-level Relevance Judgment

To the reader of the SIGIR paper:

We would like to correct one error of the paper accepted by SIGIR'19 titled:

"Investigating Passage-level Relevance and Its Role in Document-level Relevance Judgment"

In Section 5 of the paper, we investigated whether the aggregation results of passage-level relevance signals can promote the performance of existing document retrieval models. We reported the document ranking performances of document-level BM25 scores and aggregated passage-level BM25 scores as Table 1 shows. It showed that the aggregation of fine-grained, passage-level BM25 scores can improve the performance of BM25 in the document ranking task, though the improvements are not significant. However, the inverse document frequencies (IDFs) of terms used in the calculation of document-level and passage-level BM25 scores were calculated based on different corpora. It led to inaccurate results and conclusions. The IDFs we used in the calculation of document-level BM25 scores were calculated on a small corpus. It can't reflect the real importance of a term accurately. Therefore, the document ranking performances of document-level BM25 scores are not so good.

We unify the corpus used for IDF calculation and redo the experiment. The comparison of performances on document ranking between document-level and aggregated passage-level BM25 scores are shown in Table 2. We find that it is hard to promote the document ranking performance with simply aggregated passage-level BM25 scores. The value of learned λ indicates the importance of aggregated passage-level BM25 scores to the document ranking. Compared with document-level BM25 scores, all of the aggregated scores play weaker roles in document ranking ($\lambda < 0.5$). Among methods with the learned λ, aggregated passage-level scores of *maximum* and *length with decay* methods play the most important roles ($\lambda = 0.21$). However, they can not promote the ranking performance. In future work, we would like to try some other methods besides simple aggregation to incorporate passage-level and document-level relevance signals to better rank documents.

Best,

Zhijing Wu, Jiaxin Mao, Yiqun Liu, Min Zhang, Shaoping Ma

Table 1. Comparison of performances on document ranking between document-level and aggregated passage-level BM25 scores (Results in the SIGIR paper, which are inaccurate).

Method	λ	**nDCG@5**	**nDCG@10**	**nDCG@15**
$BM25_{document}$	0	0.522	0.638	0.748
minimum (CRD)	1	0.448	0.561	0.701
maximum (DRD)	1	**0.581**	**0.679**	**0.783**
median (AR)	1	0.481	0.589	0.712
mean (AR)	1	0.491	0.620	0.725
position decay	1	0.518	0.638	0.742
passage length	1	0.497	0.616	0.729
length with decay	1	0.517	0.637	0.743
exact match	1	0.548	0.653	0.764
query similarity	1	0.487	0.618	0.724
minimum (CRD)	0.13	0.530	0.639	0.750
maximum (DRD)	0.39	**0.604**	**0.688**	**0.793**
median (AR)	0.29	0.558	0.650	0.763
mean (AR)	0.37	0.530	0.646	0.754
position decay	0.43	0.537	0.652	0.756
passage length	0.37	0.555	0.655	0.759
length with decay	0.42	0.567	0.664	0.770
exact match	0.47	0.562	0.665	0.769
query similarity	0.38	0.542	0.647	0.757

Table 2. Comparison of performances on document ranking between document-level and aggregated passage-level BM25 scores.

Method	λ	**nDCG@5**	**nDCG@10**	**nDCG@15**
$BM25_{document}$	0	0.620	0.717	0.806
minimum (CRD)	1	0.442	0.557	0.701
maximum (DRD)	1	**0.591**	**0.679**	**0.785**
median (AR)	1	0.479	0.588	0.711
mean (AR)	1	0.489	0.622	0.726
position decay	1	0.518	0.633	0.740
passage length	1	0.505	0.618	0.733
length with decay	1	0.522	0.638	0.745
exact match	1	0.546	0.650	0.763
query similarity	1	0.485	0.616	0.723
minimum (CRD)	0.10	0.621	**0.719**	**0.809**
maximum (DRD)	0.21	0.598	0.692	0.790
median (AR)	0.01	0.622	0.717	0.805
mean (AR)	0.01	0.619	0.716	0.805
position decay	0.03	0.617	0.717	0.805
passage length	0.12	0.613	0.709	0.801
length with decay	0.21	**0.625**	0.714	0.807
exact match	0.05	0.617	0.712	0.802
query similarity	0.01	0.619	0.716	0.805