Introduction

➢ We participated in FOSS and POSS subtasks in NTCIR17 Session Search task.

➢ In both subtasks, we tried different approaches for feature fusion, including Learning-to-Rank and linearly combination.

➢ The final report of the SS-2 task demonstrate the effectiveness of our method, significantly outperforming other competitors.

FOSS Subtask

➢ In FOSS subtask, we concatenate all the queries and the first clicked document title of each query except the last one in a session as session context.

Learning-to-Rank

➢ In our approaches, we incorporate a total of 11 features which include both term-level features from traditional sparse retrieval methods and semantic-level features obtained through deep neural networks.

➢ We choose two classic but effective sparse retrieval methods, BM25 and QLD.

➢ We train a dense retrieval model on TianGong-ST, using the InfoNCE Loss as the loss function.

➢ We also train a fine-grained context-aware ranking model, DCL, with a curriculum learning framework.

➢ We feed the 11 features (Table 1) into two widely-used learning-to-rank models, LightGBM and LambdaMART.

Linear combination

➢ We use the linear combination of two ad-hoc scores to generate LS score.

➢ In THUIR_SS-FOSS-NEW-3 and THUIR_SS-POSS-NEW-3, we choose the ad-hoc score of BM25 as S1 and the score computed by DCL model as S2.

➢ In submission THUIR_SS-FOSS-NEW-6 and THUIR_SS-POSS-NEW-6, we replace S1 with QLD ad-hoc score.

RRF

➢ We sort documents according to the score of 11 features. Then we use all 11 rankings to calculate RRF scores.

POSS Subtask

➢ In POSS subtask, the user interaction information for the last k-n queries is not provided, we just skip the clicked document and concatenate query.

➢ We use the same method of FOSS subtask to rank documents in POSS subtask.

Submitted Runs and Evaluation

➢ FOSS subtask.

➢ The preliminary evaluation of our runs in FOSS subtask are shown in Table 2.

➢ The linear combination method using the ad-hoc score of BM25 and the score computed by DCL model achieved the best performance.

➢ POSS subtask.

➢ The preliminary evaluation of our runs in POSS subtask are shown in Table 3.

➢ The linear combination method using the ad-hoc score of QLD and the score computed by DCL model achieved the best performance.

➢ Our learning-to-rank method still has room for improvement.

➢ First, more features can be selected to feed into the model, such as user interaction information (clicks and timestamp)

➢ Second, we find that sometimes the query at the beginning of a session and the query at the end of a session are not very semantically related. These unrelated queries bring noise to our result.

Conclusion & Future work

➢ Our team (THUIR_SS) participates in the FOSS and POSS subtask of the NTCIR-17 Session Search (SS) Task.

➢ We try learning to rank model, RRF method, and linear combination method.

➢ The submission using linear combination achieves the best performance in both FOSS and POSS subtasks.

Email: yitengtu16@gmail.com