THUIR_SS at the NTCIR-17 Session Search (SS) Task

Xinyan Han¹, Yiteng Tu², Haitao Li¹, Qingyao Ai¹, Yiqun Liu¹

¹ Department of Computer Science and Technology, Tsinghua University,
   Zhongguancun Laboratory, Beijing 100084, China
2 Renmin University of China, Beijing 100872, China

yitengtu16@gmail.com
Introduction

• Session Search holds significant importance in the field of information retrieval and user experience.

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

• Our methods significantly outperform other competitors.
Methods

Search Context Definition

**FOSS:**
\[ C = [q_1, d_1^+, \ldots, q_{k-1}, d_{k-1}^+, q_k] \]

**POSS:**
\[ C = [q_1, d_1^+, \ldots, q_n, d_n^+, q_{n+1}, \ldots, q_j], \quad (n + 1 \leq j \leq k) \]

- \( q_i \) is the i-th query in a session
- \( d_i^+ \) is the first clicked document title of \( q_i \) (may be empty)
Methods

Feature Extraction

11 features: both term level and semantic level, both single-turn level and session level

- 2 classic but effective sparse retrieval methods, BM25 and QLD
- A dense retrieval model trained on TianGong-ST using contrastive learning
- A context-aware ranking model, DCL, trained with a dual curriculum learning framework
Methods

Feature Extraction

Sparse Retrieval

- Model Selection: BM25 & QLD
- Retrieval Strategy: w/ & w/o RM3 pseudo-relevance feedback
- Query Text: Single-turn Query $q_k$ & Session Query $C$
Methods

Feature Extraction

Dense Retrieval

• Model Structure: Dual-encoder

\[ r_q = BERT_{CLS}(q) \]

\[ r_d = BERT_{CLS}(d) \]

\[ s(q, d) = r_q \cdot r_d \]

• Training Data: Tiangong-ST

• Training Technique: Contrastive Learning (InfoNCE Loss)

\[ \mathcal{L} = -\log \frac{\exp (s(q, d^+))}{\exp (s(q, d^+)) + \sum_j \exp (s(q, d^-))} \]
Methods

Feature Extraction

Context-aware Ranking Model: DCL

• Model Structure: BERT + MLP

• Training Data: Tiangong-ST

• Training Technique: A Dual Curriculum Learning Framework*

\[ X = [CLS]q_1[EOS]d_1[EOS]...q[EOS][SEP]d[EOS][SEP] \]
\[ r = BERT(X)[CLS] \]
\[ OUTPUT = MLP(r) \]

Methods

Feature Extraction

Context-aware Ranking Model

- The Curriculum Learning Framework


Figure 2: The training process of our framework. For the curriculum of positive pairs, only easy samples are used at the beginning ($t_1$). Along the training process, the positive sampling space is gradually extended to the whole positive pairs ($t_3$). For the curriculum of negative pairs, the sampling space is shrinking from all samples ($t_1$) to only hard samples ($t_3$).
Methods

Fusion

- RRF
  \[ RRF(d \in D) = \sum_{r \in R} \frac{1}{k + r(d)} \]
- Linear Combination
  - 1 sparse score + 1 DCL score
- Learning-to-rank
  - LightGBM
  - LambdaMART

Table 1: features of learning-to-rank model

<table>
<thead>
<tr>
<th>feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ad-hoc score of BM25</td>
</tr>
<tr>
<td>2 ad-hoc score of QLD</td>
</tr>
<tr>
<td>3 ad-hoc score of BM25 with RM3</td>
</tr>
<tr>
<td>4 ad-hoc score of QLD with RM3</td>
</tr>
<tr>
<td>5 session score of BM25</td>
</tr>
<tr>
<td>6 session score of QLD</td>
</tr>
<tr>
<td>7 session score of BM25 with RM3</td>
</tr>
<tr>
<td>8 session score of QLD with RM3</td>
</tr>
<tr>
<td>9 ad-hoc score of DR model</td>
</tr>
<tr>
<td>10 session score of DR model</td>
</tr>
<tr>
<td>11 session score of DCL model</td>
</tr>
</tbody>
</table>
Results and Future Work

• Linear combination achieves the best performance

• Room for improvements:
  • More features (user interaction, query/document length ……)
  • Semantic inconsistency in session
Conclusion

- Our team (THUIR_SS) participate in the FOSS and POSS subtask of the NTCIR-17 Session Search (SS) Task.
- We try several retrieval and ranking approaches on different levels and explore different fusion methods.
- The submission using linear combination achieves the best performance in both FOSS and POSS subtasks.
Thanks!

Q&A