RUCIR at the NTCIR-16 Session Search (SS) Task
Haonan Chen and Zhicheng Dou

Haonan Chen hnchen@ruc.edu.cn
Gaoling School of Artificial Intelligence, Renmin University of China
Background

• To complete a complex search task, a user often has to issue multiple queries to find the information they need.

• Studies have shown that utilizing the contextual information of the search session can help understand the user's current search intent.
Backbone -- COCA

• Using data augmentation strategies to generate possible variations of user behavior sequences from a search log.

• Based on the augmented data, COCA uses contrastive learning to optimize the BERT encoder to train a more robust model to deal with new behavior sequences.

Data Augmentation Strategies

- Term Mask
- Query/Document Deletion
- Behavior Reordering
Model Architecture

Augmented Sequence $H'_n$

Augmented Sequence $H''_n$

Original User Behavior Sequence $H_n$

(1) Sequence Representation Optimization

(2) Document Ranking

\[ l(i, j) = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{k \neq i} \exp(\text{sim}(z_i, z_k)/\tau)}, \]

InfoNCE Loss

\[ \mathcal{L}_{\text{rank}} = -\frac{1}{N} \sum_{i=1}^{N} y_i \log z_i + (1 - y_i) \log(1 - z_i), \]

CrossEntropy Loss
Some Tricks

• Using BM25 algorithm as regularization.

\[
Score(\text{candidate}) = \alpha \times Score_{COCA} + (1 - \alpha) \times Score_{BM25}
\]

• Keeping the document that has the largest usefulness value per historical query.
## Results

<table>
<thead>
<tr>
<th>FOSS</th>
<th>NDCG@3</th>
<th>NDCG@5</th>
<th>NDCG@10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>COCA+BM25</td>
<td>0.4783</td>
<td>0.4785</td>
<td>0.4939</td>
<td></td>
</tr>
<tr>
<td>COCA+U</td>
<td>0.5365</td>
<td>0.5406</td>
<td>0.5570</td>
<td></td>
</tr>
<tr>
<td>COCA+BM25+U</td>
<td>0.5525</td>
<td>0.5623</td>
<td>0.5693</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>POSS</th>
<th>RsDCG</th>
<th>RsRBP</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>COCA</td>
<td>0.4355</td>
<td>0.5640</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COCA+BM25</td>
<td>0.4738</td>
<td>0.6281</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COCA+BM25+U</td>
<td>0.5439</td>
<td>0.7466</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Thank You for Listening!