MPII at the NTCIR-15 WWW-3 Task

Canjia Li
University of Chinese Academy of sciences
licanjia17@mails.ucas.ac.cn

Andrew Yates
Max Planck Institute for Informatics
ayates@mpi-inf.mpg.de @andrewyates

Overview

MPII participated in the English subtask of WWW-3 at NTCIR-15 with several variants of our recent PARADE model. Our results support the findings in the PARADE paper [1]: aggregating representations is more effective than aggregating scores, and effectiveness increases with the complexity of the aggregation approach.

Methodology

We evaluated variants of the PARADE model, including PARADE_MAX, PARADE_ATTN, and PARADE.

• PARADE_MAX: Max-pooling over the passage representations.
• PARADE_ATTN: Importance weighting on the passage representations.
• PARADE: Building a Transformer block on top of the passage representations and using a [CLS] token to aggregate the representations.

More details about PARADE can be found in the original work [1].

Results

**Effectiveness.** We trained PARADE on the NTCIR WWW-1 and WWW-2 queries. The results support the findings in [1] that aggregating passage representations is more effective than aggregating passage scores and that the full PARADE model is more effective than the simpler variants.

<table>
<thead>
<tr>
<th>Run Name</th>
<th>PARADE variant</th>
<th>nDCG@10</th>
<th>Q@10</th>
<th>nERR@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>mpii-E-CO-NEW-3</td>
<td>PARADE_MAX</td>
<td>0.6337</td>
<td>0.6556</td>
<td>0.7395</td>
</tr>
<tr>
<td>mpii-E-CO-NEW-2</td>
<td>PARADE_ATTN</td>
<td>0.6743</td>
<td>0.6905</td>
<td>0.7787</td>
</tr>
<tr>
<td>mpii-E-CO-NEW-1</td>
<td>PARADE</td>
<td>0.6897</td>
<td>0.7016</td>
<td>0.8090</td>
</tr>
</tbody>
</table>

Results from WWW-3.

**Error Analysis.** Models fail severely in the following queries:

- You want to visit the website "www.freeweblayouts.net".
- You want to find the official website of Akron Beacon Journal
- You want to know how Zeus is described in the Greek Mythology.

Observation: For the queries seeking websites, regarded as known-item search, it might not be necessary to employ a full-document ranking model.