Why Does This Entity Matter? Support Passage Retrieval for Entity Retrieval.

Shubham Chatterjee (sc1242@cs.unh.edu) and Laura Dietz (dietz@cs.unh.edu)
University Of New Hampshire, Durham NH, USA

Abstract
Our goal is to complement an entity ranking with human-readable explanations of how those retrieved entities are connected to the information need. While related to the problem of support passage retrieval, in this paper, we explore two underserved indicators of relevance: contextual entities and entity salience. The effectiveness of the indicators are studied within a supervised learning-to-rank framework on a dataset from TREC Complex Answer Retrieval. We find that salience is a useful indicator, but it is often not applicable. In contrast, although performance improvements are obtained by using contextual entities, using contextual words still outperforms contextual entities.

Entity Context Document

Entity Salience
- We use SMIR to annotate passages with salient entities and salience score and class (whether salient or not).
- We consider Score(p|e) = Salience(e|p).
- Score Passages in two ways:
  - Score(p|e) = Score(e|p): Score(e|p)
  - Score(p|e) = Score(e|p): Score(e)
- Example: For query Brexit and entity Theresa May, consider the two passages below:
  Passage 1: It Labour and the Conservatives fail to reach an agreement, MPs will face a series of votes on Brexit options, which could include another referendum. Theresa May has said her government stands ready to abide by the decision of the House if Labour does the same.
  Passage 2: British Prime Minister Theresa May offered a new Brexit plan on Tuesday, in a last-ditch effort to get her still-suspicious Brexit deal approved.

Baselines
Baseline 1: Frequency of relevant entity links. We rank passages for a query-entity pair by the number of relevant entities in the passage.
Baseline 2: Compound entity-query score. We retrieved passages using a compound query, where the query is a combination of the original query and the target entity.

Features

Results

<table>
<thead>
<tr>
<th>Feature</th>
<th>Baseline 1</th>
<th>Baseline 2</th>
<th>ECN</th>
<th>QEE</th>
<th>SSPE</th>
<th>L2R-All-Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP</td>
<td>0.09±0.002</td>
<td>0.07±0.003</td>
<td>0.31±0.004</td>
<td>0.35±0.004</td>
<td>0.30±0.003</td>
<td>0.02±0.003</td>
</tr>
<tr>
<td>P@1</td>
<td>0.06±0.003</td>
<td>0.06±0.003</td>
<td>0.30±0.004</td>
<td>0.34±0.004</td>
<td>0.25±0.004</td>
<td>0.03±0.004</td>
</tr>
</tbody>
</table>

Table 1: Performance with standard error of features individually and combined with L2R, including subsets/ablations.

Conclusion and Future Work
We propose a joint query-entity-passage ranking method and present some initial results. In particular, we show that co-occurring entities are an important indicator of which passages might support an entity for a query. In contrast, retrieving passages using a compound query of the original query and the entity is not sufficient for the problem. We also experiment with entity salience, and find that it is a highly informative feature when it is applicable, however 95% of target entities do not have any passage with a salient mention in the candidate set. We thus identify a need for developing high-recall salience techniques that are applicable to a larger number of entities, as well as new indexing and retrieval methods that integrate entity salience in an early phase.

Research Questions and Discussions

RQ1 What is the effect of contextual entities?
- We observe in Table 1 that ranking passages with contextual entities (ECN) achieves a MAP of 0.31 and query expansion with entities (QEE) a MAP of 0.30. This is a significant improvement over the two baselines which have a MAP of 0.09 and 0.07 respectively. This demonstrates the benefit of contextual entities.
- Although the combination of all variations of QEE performs well on its own, the ablation study shows that one would obtain even better results by leaving QEE features out to improve to a MAP 0.35.
- QEW features perform well on their own as well as with others.

RQ2 What is the effect of entity salience?
- Salience is a useful indicator—however it is only applicable for entities which have a salient passage in the candidate pool.

RQ3 Is it better to expand queries with entities or words?
- Contextual entities performs well.
- Retrieval with contextual words is better than retrieval with contextual entities.
- When L2R is trained to optimize MAP on all contextual word and entity features (Subset-5) the trained parameter assigns the maximum weight to contextual word features.