Redundancy Removal to Selectively Diversify Information Retrieval Results

Xiaolin Wang, Hai Zhao, Baoliang Lu
BCMI Lab, Shanghai Jiao Tong University
arthur.xl.wang@gmail.com  {zhaohai; blu}@cs.sjtu.edu.cn

The Brain-like Computing and Machine Intelligence Lab (BCMI) of Shanghai Jiao Tong University takes part in the NTCIR-9 Intent Chinese subtask. A redundancy removal (RedRem) algorithm is proposed to diversify the top-N retrieval results.

score(d, q) = \frac{BM25F(d, q)}{\max_{d' \in D} BM25F(d', q)} + \lambda \frac{PR(d)}{\max_{d' \in D} PR(d')}

where
- d is a document
- q is a query
- BM25F is BM25F similarity score
- PR is SogouT-Rank score

Require: retrieved documents \( S = \{(d_i, s_i)|i = 1, \ldots, n\} \)
where \( d_i \) is a document and \( s_i \) is its normalized confidence score s.t. \( s_1 = 1 \) and \( s_i \geq s_{i+1} \).

Ensure: re-ranked documents \( U = \{(p_{k_j}, u_j)|j = 1, \ldots, n\} \)
where \( d_{k_j} \) is the j-th page and \( u_j \) is its updated score.

\( \lambda \) for all \( j, j = 2, \ldots, n \) do
  for all \( t, t = i, \ldots, n \) do
    \( u_t = u_t - f_{\text{RED}}(d_t, U) \)
  end for
end for

\( k_j \leftarrow \text{argmax}_{\{u_j\}} \)
add \( (d_{k_j}, u_j) \) into \( U \)
end for
return \( U \)

where
\( f_{\text{RED}}(d, U) = \alpha \frac{|d \cap U|}{|d|} + \beta \frac{|\{w|w \in d, w \notin U\}|}{|d|} \)

<table>
<thead>
<tr>
<th>Runs</th>
<th>Page retrieval</th>
<th>Page rank(( \lambda ))</th>
<th>RedRem? (( \alpha, \beta ))</th>
<th>I-rec@10</th>
<th>D-nDCG@10</th>
<th>D#-nDCG@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>SJTUBCMI-D-C-1</td>
<td>BM25F</td>
<td>Yes(0.4)</td>
<td>Yes(0.1,-0.9)</td>
<td>0.6038</td>
<td>0.2654</td>
<td>0.4346</td>
</tr>
<tr>
<td>SJTUBCMI-D-C-2</td>
<td>BM25F</td>
<td>Yes(0.4)</td>
<td>No</td>
<td>0.6008</td>
<td>0.3317</td>
<td>0.4663</td>
</tr>
<tr>
<td>SJTUBCMI-D-C-3</td>
<td>BM25F</td>
<td>No</td>
<td>No</td>
<td>0.5856</td>
<td>0.3288</td>
<td>0.4572</td>
</tr>
<tr>
<td>SJTUBCMI-D-C-4</td>
<td>BM25</td>
<td>Yes(0.4)</td>
<td>Yes(0.1, -0.9)</td>
<td>0.6108</td>
<td>0.2756</td>
<td>0.4432</td>
</tr>
<tr>
<td>SJTUBCMI-D-C-5</td>
<td>BM25</td>
<td>No</td>
<td>Yes(0.0,-0.9)</td>
<td>0.6228</td>
<td>0.2816</td>
<td>0.4522</td>
</tr>
</tbody>
</table>