LambdaMerge: Merging the Results of Query Reformulations

Daniel Sheldon
Oregon State University
sheldon@eecs.oregonstate.edu

Milad Shokouhi, Martin Szummer, and Nick Craswell
Microsoft Research
{milads, szummer, nickcr}@microsoft.com

Merging Query Reformulations

Query reformulation algorithms
• Improve retrieval by alleviating Q-D mismatch
  bill gates bio → bill gates biography
• Or can make things worse e.g. bio → biog
• Or even drift off-topic e.g. bill → melinda

Approach: Run multiple queries, merge results
• Post-retrieval it is easier to detect quality/drift
• Multiple queries give diverse relevance evidence

Contribution: New merging methods
1. CombRW: Weighted CombSUM (unsupervised)
2. LambdaMerge: Supervised merging
   • Trained to maximize target such as NDCG
   • Incorporating quality and drift features
   • Robust to bad reformulations

LambdaMerge

Issue k formulations to search engine:
• Original query \( q^{(1)} \) plus reformulations \( q^{(2)}, ..., q^{(k)} \)
• Get top-N lists \( D^{(1)}, ..., D^{(k)} \)

Generate features:
• Query-document features \( x^{(k)}_d \): relevance of document \( d \) specific to \( D^{(k)} \)
• Gating features \( z^{(k)} \): drift + overall quality of \( D^{(k)} \)

Scoring net assigns score \( f(x^{(k)}_d; \theta) \) to each formulation-document pair

Gating net assigns weights \( \alpha_1, ..., \alpha_k \) to formulations

Overall document score:
\[
s_d = \sum_k \alpha_k \cdot f(x^{(k)}_d; \theta)
\]

Parameters \( \theta \) (scoring) and \( \pi \) (gating) trained by backprop with LambdaRank gradients to optimize NDCG [1]

Experiments

Test collections: Bing data and GOV2

Reformulations: Click graph random walk [2]

Single-query methods:
• ORG: Original query
• RW1: Most likely alternative query from RW
• RAPP-L: Predict best query (lin. regression) [3]
• RAPP(O): Choose query via NDCG@5 (oracle)

Merging methods:
• CombSUM: Sum scores
• CombRW: CombSUM with random walk weight
• LambdaMerge: Using these features

Results:

<table>
<thead>
<tr>
<th></th>
<th>NDCG@5</th>
<th>NDCG@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORG</td>
<td>0.538</td>
<td>0.524</td>
</tr>
<tr>
<td>RW1</td>
<td>0.422</td>
<td>0.367</td>
</tr>
<tr>
<td>CombSUM</td>
<td>0.510</td>
<td>0.466</td>
</tr>
<tr>
<td>CombRW</td>
<td>0.542</td>
<td>0.516</td>
</tr>
<tr>
<td>RAPP-L</td>
<td>0.534</td>
<td>0.524</td>
</tr>
<tr>
<td>( \lambda )-Merge</td>
<td>0.555</td>
<td>0.530</td>
</tr>
<tr>
<td>RAPP(O)</td>
<td>0.556</td>
<td>0.530</td>
</tr>
</tbody>
</table>

| GOV2             |        |         |
| P@5              | 0.548  | 0.431   |
| CombSUM          | 0.584  | 0.438   |
| \( \lambda \)-Merge | 0.596 | 0.447 |
| RAPP(O)          | 0.592  | 0.457   |

Robustness Analysis

Unsupervised methods:

Supervised methods: