Ranking Model Selection and Fusion for Effective Microblog Search

Zhongyu Wei1,2, Wei Gao1, Tarek El-Ganainy1, Wald Magdy1, Kam-Fai Wong3

1Qatar Computing Research Institute, Doha, Qatar 2The Chinese University of Hong Kong, Hong Kong, China

1. Introduction

- Our research studies how to improve re-ranking microblog search results by leveraging ranked lists from multiple rankers that are selected automatically
- Our proposed solution:
  - **Query sensitive ranker selection**: Select 1 best performed rankers from candidate ranking models in a query-sensitive manner
  - **Result fusion of selected rankers**: Aggregate the ranked lists of selected rankers via different fusion techniques

2. Framework overview

3. Base ranker: Web-search-based query expansion

4. Ranker selection

- Performance estimation for candidate ranker $R_i$ given test query $q$
  
  \[ \frac{1}{L} \sum_{j=1}^{L} ps(q_j^{(t)}, R) \]

- L-Nearset neighbor queries of $q$
  
  \[ \text{sim}(q_i, q_j^{(t)}) = KLDiv(D(R, q), D(R_j, q_j^{(t)})) \]

- Distribution of divergence between ranked lists of $R$, and other rankers:
  
  \[ D(R, q) = \{D(R_1 \parallel R, q), D(R_2 \parallel R, q), ... \} \]

- Normalized divergence score between ranked lists of two rankers:
  
  \[ D(R_1 \parallel R_2, q) = \frac{1}{Z} \sum_j s_j(t) - s_j(t) \]

5. Rank aggregation

- CombMNZ (Fox and Shaw, 1994):
  
  \[ \text{CombMNZ}(t) = \{r \in R: \text{rank}(t) \leq c \} \times \sum \text{score}(t) \]

- Weighted Borda-fuse (Aslam and Montague, 2001)

- Weighted Condorcet-fuse (Montague and Aslam, 2002)

- Reciprocal rank fusion (RRF) (Cormack et al., 2009):
  
  \[ \text{RRF}(t) = \sum_{k=r}^{K} \frac{1}{k+\text{rank}(t)} \]

6. Candidate rankers

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Validation %</th>
<th>Optimized metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>RankBoost (Freund, 2003)</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>MART (Friedman, 2001)</td>
<td>20%</td>
<td>--</td>
</tr>
<tr>
<td>RandomForest (Breiman, 2001)</td>
<td>20%</td>
<td>--</td>
</tr>
<tr>
<td>RankNet (Burge, et al., 2005)</td>
<td>20%</td>
<td>--</td>
</tr>
<tr>
<td>Coordinate Ascent (Masteller and Croft, 2007)</td>
<td>20%</td>
<td>P@30 MAP</td>
</tr>
<tr>
<td>LambdaMART (Wu, et al., 2010)</td>
<td>20%</td>
<td>P@30 MAP</td>
</tr>
</tbody>
</table>

Table 1. Eight ranking models trained for ensemble

7. Datasets

<table>
<thead>
<tr>
<th>Collection</th>
<th># of tweets</th>
<th># of terms</th>
<th>Average length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweets2011</td>
<td>16,141,809</td>
<td>155,562,640</td>
<td>9.64</td>
</tr>
<tr>
<td>Tweets2013</td>
<td>243,711,538</td>
<td>2,928,413,436</td>
<td>12.04</td>
</tr>
</tbody>
</table>

Table 2. The statistics of tweets collections

<table>
<thead>
<tr>
<th>Query Set</th>
<th># of queries</th>
<th># of annotated</th>
<th># of relevant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q52011</td>
<td>50</td>
<td>40,855</td>
<td>2,864</td>
</tr>
<tr>
<td>Q52012</td>
<td>60</td>
<td>73,073</td>
<td>6,286</td>
</tr>
<tr>
<td>Q52013</td>
<td>60</td>
<td>71,279</td>
<td>9,011</td>
</tr>
</tbody>
</table>

Table 3. The statistics of relevance judgment

8. Experiments and Results

- **LM/LM$_{subprf}$**: baseline rankers
- **BestSingle**: Use best single ranker among all the candidate rankers
- **PMO**: Model selection method proposed by Peng et al. (2009) (top-1)
- **BestSel**: Our model selection method (top-1)
- **x-all**: x-fuse of the results of all candidate rankers (without selection)
- **x.sel**: x-fuse of the results of selected candidate rankers (top-L)

9. Conclusions

- Explore model selection and fusion methods for re-ranking microblog search results based on multiple learned ranking models
- An extension of query-sensitive model selection
- With a moderate base ranker, our re-ranking significantly outperforms the best single ranker and the existing model selection method
- Simply aggregating all rankers is not better than the best single ranker