RUCIR at NTCIR-14 WWW-2 Task

Xue Yang, Shuqi Lu, Shijun Wang, Han Zhang and Zhicheng Dou
Renmin University of China, Beijing, P.R. China
{ruc_yangx, lusq, wangshijun, zhanghanjl, dou}@ruc.edu.cn

Data-Flow Overview

Step 1: Data Preprocessing
- Document Collection
  - Clueweb09, Clueweb12, SogouT-16, Sogou-QCL
- Relevance Judgement
  - TREC09 ~ TREC14, Sogou-QCL

Step 2: Feature Extraction
- Traditional Features
  - For different fields: (anchor), title, URL, body, whole
- Embedding Features
- Deep Neural Features
  - ARC-I
  - ARC-II
  - DRMM
  - aNMM
  - MV-LSTM
  - DUET

Step 3: Training & Evaluation
- Model
  - Ranklib: LambdaMART
- Metric Evaluation
  - NDCG@K, Q@K, nERR@K

Experiment Results (Chinese)

<table>
<thead>
<tr>
<th>Run</th>
<th>Query</th>
<th>Features</th>
<th>nDCG@10</th>
<th>Q@10</th>
<th>nERR@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RUCIR-1</td>
<td>Description</td>
<td>Traditional Embedding</td>
<td>0.4515</td>
<td>0.4228</td>
<td>0.5792</td>
</tr>
<tr>
<td>RUCIR-2</td>
<td>Content</td>
<td>Traditional Embedding</td>
<td>0.4866</td>
<td>0.4571</td>
<td>0.6044</td>
</tr>
<tr>
<td>RUCIR-3</td>
<td>Description</td>
<td>Traditional</td>
<td>0.4503</td>
<td>0.4223</td>
<td>0.5630</td>
</tr>
<tr>
<td>RUCIR-4</td>
<td>Description</td>
<td>Deep Neural</td>
<td>0.4458</td>
<td>0.4226</td>
<td>0.5619</td>
</tr>
<tr>
<td>RUCIR-5</td>
<td>Description</td>
<td>Deep Neural</td>
<td>0.2745</td>
<td>0.2404</td>
<td>0.3832</td>
</tr>
</tbody>
</table>

Experiment Results (English)

<table>
<thead>
<tr>
<th>Run</th>
<th>Query</th>
<th>Features</th>
<th>nDCG@10</th>
<th>Q@10</th>
<th>nERR@10</th>
</tr>
</thead>
<tbody>
<tr>
<td>RUCIR-1</td>
<td>Description</td>
<td>Traditional</td>
<td>0.3137</td>
<td>0.2973</td>
<td>0.4469</td>
</tr>
<tr>
<td>RUCIR-2</td>
<td>Content</td>
<td>Traditional</td>
<td>0.3489</td>
<td>0.3352</td>
<td>0.4917</td>
</tr>
<tr>
<td>RUCIR-3</td>
<td>Description</td>
<td>Traditional Embedding</td>
<td>0.3137</td>
<td>0.2973</td>
<td>0.4469</td>
</tr>
<tr>
<td>RUCIR-4</td>
<td>Description</td>
<td>Deep Neural</td>
<td>0.3293</td>
<td>0.3094</td>
<td>0.4602</td>
</tr>
<tr>
<td>RUCIR-5</td>
<td>Description</td>
<td>Deep Neural</td>
<td>0.2876</td>
<td>0.2659</td>
<td>0.4188</td>
</tr>
</tbody>
</table>

* Embedding features are cosine similarity between query representation and document representation, which are obtained by averaging the word vectors in the text. Word vectors are trained by word2vec.

Conclusion
- [CN] Traditional Features + Embedding Features
  - Other Runs; [EN] Traditional Features
- Query Content > Query Description (CO > DE)
- Deep Neural Features << Other Runs (5 << 1,2,3,4)