DCU at the NTCIR-9 SpokenDoc Passage Retrieval Task

Maria Eskevich, Gareth J. F. Jones
Centre for Digital Video Processing,
Centre for Next Generation Localisation
School of Computing, Dublin City University, Ireland

Retrieval Methodology

Overview
- Investigate application of content-based segmentation for spoken passage retrieval
- Segmentation using standard TextTiling and C99 algorithms from text
- Standard Japanese text processing applied with language modelling information

Transcript Preprocessing
- Recognize individual morphemes of the sentences: ChaSen 2.4.0, based on Japanese morphological analyzer JUMAN 2.0 with ipadic grammar 2.7.0
- Form the text out of the base forms of the words
- Remove the stop words (SpeedBlog Japanese Stop-words) for one of the runs (NSW)

Segmentation
Transcripts are segmented using either:
- TextTiling (TT):
  - Cosine similarities between adjacent blocks of sentences
- C99:
  - Similarity between sentences cached using a cosine similarity measure to form a similarity matrix
  - Cosine scores replaced by the rank of the score in the local region
  - Segmentation points assigned using a clustering procedure

Retrieval System
SMART information retrieval system extended to use language modelling with a uniform document prior probability

Retrieval Results

<table>
<thead>
<tr>
<th>Transcript Type</th>
<th>Segmentation Type</th>
<th>uMAP</th>
<th>pwMAP</th>
<th>fMAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>BASELINE</td>
<td></td>
<td>0.0670</td>
<td>0.0520</td>
<td>0.0536</td>
</tr>
<tr>
<td>Manual</td>
<td>TT</td>
<td>0.0859</td>
<td>0.0429</td>
<td>0.0500</td>
</tr>
<tr>
<td>Manual</td>
<td>C99</td>
<td>0.0713</td>
<td>0.0209</td>
<td>0.0168</td>
</tr>
<tr>
<td>ASR</td>
<td>TT</td>
<td>0.0490</td>
<td>0.0329</td>
<td>0.0308</td>
</tr>
<tr>
<td>ASR</td>
<td>C99</td>
<td>0.0469</td>
<td>0.0166</td>
<td>0.0123</td>
</tr>
<tr>
<td>ASR_NSW</td>
<td>TT</td>
<td>0.0312</td>
<td>0.0141</td>
<td>0.0174</td>
</tr>
<tr>
<td>ASR_NSW</td>
<td>C99</td>
<td>0.0316</td>
<td>0.0138</td>
<td>0.0120</td>
</tr>
</tbody>
</table>

Results Analysis

Calculation of Average of Precision (sec)

<table>
<thead>
<tr>
<th>Relevant</th>
<th>Passage 1</th>
<th>Precision</th>
<th>Relevant</th>
<th>Passage 2</th>
<th>Precision</th>
<th>Relevant</th>
<th>Passage 3</th>
<th>Precision</th>
<th>Relevant</th>
<th>Passage 4</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

where: Precision = Length of the Relevant Part / Length of the Whole Passage

Average Length of Relevant Part and Segments

Conclusions
- Only runs on the manual transcript had higher scores than the baseline (uMAP metric only)
- TextTiling results are consistently higher than C99 for all the metrics for manual and ASR runs
- TextTiling has higher average of precision (in seconds) for all types of transcript, i.e. it locates topically coherent segments better
- High level of poor segmentation makes it harder to retrieve relevant content for C99 runs
- Removal of stop words before segmentation did not have any positive effect on the results