Context matters: Towards extracting a citation’s context using linguistic features

Aim: recommend potential citations at a particular location in a draft paper

Task: select the context for which to recommend citations

Evaluation: attempt to recover original citations in existing published papers from the whole document collection

Previous work: traditionally all contexts are extracted using symmetric windows over words or sentences

Approach: compare symmetrical methods for extracting a citation’s context: window-of-words and window-of-sentences with a human oracle selecting relevant sentences

Corpus: ACL Anthology Corpus (AAC)

1. Motivation

• All previous work on recommendation uses symmetric methods to extract the context of a citation
• Are symmetric methods optimal?

2. Annotated citation contexts

Athar and Teufel (2012) – Context-Enhanced Citation Sentiment Detection
• Corpus: ACL Anthology
• Annotated contexts: ~1800 (citations to 20 selected papers)
• Per-sentence annotations:
  • relevant (3115 sentences)
  • sentiment:
    • (p)ositive (261)
    • (n)egative (365)
    • (o)objective (2489)
• Most sentences containing a citation are labelled objective. (1929)

3. Evaluation

1. Index document collection
AAC: ~28k documents, excluding annotated documents

2. Generate queries
From each of the annotated citation contexts, remove stopwords and generate one query using:
• Window of words (30, 50, 100, 500)
• Window of sentences (1 only, 1 up, 1 down, 1up + 1down, 2up + 2down, paragraph)
• Oracle / human annotations (all relevant, combinations of positive, negative and objective)

3. Evaluate queries
Run queries, attempt to retrieve original citation from document collection, measure Mean Reciprocal Rank (MRR)

4. Context extraction methods

Extraction methods
• Window of tokens (30 up, 30 down)
• Window of sentences (2 up, 2 down)
• Oracle - annotated sentences (p + n + o)

5. Results

<table>
<thead>
<tr>
<th>Annotated sentence n</th>
<th>Evaluation: Mean Reciprocal Rank</th>
<th>Annotated sentence p</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences 1up + citing</td>
<td>0.0134</td>
<td>0.1287</td>
<td>0.134</td>
</tr>
<tr>
<td>Sentences only (citation)</td>
<td>0.0182</td>
<td>0.1309</td>
<td>0.136</td>
</tr>
<tr>
<td>Sentences paragraph</td>
<td>0.1313</td>
<td>0.134</td>
<td>0.138</td>
</tr>
<tr>
<td>Tokens 30up-30down</td>
<td>0.134</td>
<td>0.1378</td>
<td>0.1403</td>
</tr>
<tr>
<td>Tokens 100up-100down</td>
<td>0.134</td>
<td>0.1382</td>
<td>0.147</td>
</tr>
<tr>
<td>Sentence 2up-2down</td>
<td>0.136</td>
<td>0.1403</td>
<td>0.1505</td>
</tr>
<tr>
<td>Sentence 1up-1down</td>
<td>0.1378</td>
<td>0.147</td>
<td>0.1575</td>
</tr>
<tr>
<td>Tokens 500up-500down</td>
<td>0.1382</td>
<td>0.1505</td>
<td>0.1575</td>
</tr>
<tr>
<td>Sentence citing + 1down</td>
<td>0.1403</td>
<td>0.1575</td>
<td>0.1575</td>
</tr>
<tr>
<td>Tokens 500up-500down</td>
<td>0.147</td>
<td>0.1575</td>
<td>0.1575</td>
</tr>
<tr>
<td>Annotated sentence n</td>
<td>0.1505</td>
<td>0.1575</td>
<td>0.1575</td>
</tr>
<tr>
<td>Annotated sentence p+n+o</td>
<td>0.1575</td>
<td>0.1575</td>
<td>0.1575</td>
</tr>
</tbody>
</table>

Findings:
• Human oracle outperforms all symmetrical methods. Symmetrical windows of either tokens or sentences are therefore not optimal.
• The annotated sentiment of sentences was not useful for context extraction. The more sentences we include that were annotated as relevant, the higher the score.
• More query terms is not always better. Carefully selecting relevant text spans for context extraction improves results.

6. Discussion

Future work: keyword extraction using linguistic features. Train a machine learning classifier to generate queries from sub-sentence-length spans.