Context Matters: Towards Extracting a Citation’s Context Using Linguistic Features

Daniel Duma  
University of Edinburgh  
danielduma@gmail.com

Charles Sutton  
University of Edinburgh  
csutton@inf.ed.ac.uk

Ewan Klein  
University of Edinburgh  
ewan@inf.ed.ac.uk

Keywords
Citation context; context extraction; window of words; citation recommendation; information retrieval

1. INTRODUCTION

Keyword-based search engines are becoming increasingly sophisticated, and yet navigating the ever-increasing collection of academic knowledge remains an arduous task. Keeping abreast of relevant scientific literature is often a fragmented process that breaks the workflow of academic writing.

Wouldn’t it be helpful if your text editor automatically suggested papers that are contextually relevant? Our vision of future access to digital libraries is entirely integrated into the writing process and works to augment the writer’s knowledge and capabilities. We concern ourselves with the task of context-based citation recommendation: we desire to recommend contextually relevant citations. One example of this is getting relevant suggestions of related work at the early draft stage as the author is typing.

Citation contexts are a very important source of information for scientific discovery. The text that surrounds a citation to another paper inside an academic paper has been variously used to generate summaries of academic papers [8], to inform metrics of a paper’s impact [10], as “anchor text” in information retrieval scenarios [9], and within these, especially for context-based citation recommendation [4, 3, 2, 5].

Context extraction is a key sub-task in context-based citation recommendation, yet it has received painfully little attention in the literature to date. Previous approaches to context extraction fall into two big groups: symmetric window approaches and sentence selection approaches. Symmetric extraction uses for example a window of words, where the context is considered to be n tokens before the citation token and n tokens after it, or a window of sentences, where the citing sentence is included, plus n sentences before and/or after it.

The task of citation recommendation seems to have exclusively used symmetric windows so far. We propose that these methods are excessively simplistic and can be significantly improved upon. In this paper, we show that sentence selection methods are indeed superior to symmetric windows for the task of citation recommendation.

2. RELATED WORK

For the task of context-based citation recommendation, He et al. [4] used a symmetric window of words (50 before, 50 after) as did Liu et al. [7] (300 before, 300 after). He et al. [3] used passages (splitting the article into half-overlapping fixed-size windows of words). Huang et al. [5] used a window of sentences: citing sentence + 1 before + 1 after. Similarly, [9] used symmetric windows of words and sentences to build external document representations.

It is clear that always using a fixed window size and not dealing with coreference is guaranteed to introduce false positives and false negatives in extracted keywords, which leads to noise.

Instead of dealing with this noise exclusively by using weighting schemes based on topic modelling and word embeddings (e.g. [5]), we propose that those approaches will also benefit from a better selection of the context.

Sentence selection approaches have been applied primarily to summarization and sentiment analysis. Kaplan et al. [6] manually annotated a small corpus (50 citations) with relevant sentences to each citation and trained a coreference resolver on it in order to generate summaries of those papers. Similarly to this and also for summarization, Qazvinian et al. [8] manually annotated a corpus of 203 citations with relevant sentences to each citation within a 4-sentence window (2 up, 2 down) and trained a classifier which decided which sentences to include. More recently, Athar [1] built a larger annotated corpus and trained a classifier for sentiment analysis.

3. METHODOLOGY

3.1 Evaluation

We aim to recommend contextually relevant citations. To evaluate this, we exploit the human judgements that are already implicit in available resources, and so we avoid purpose-specific annotation. That is, we make it our task to recover the original citations in papers that have already been published and we judge our system’s accuracy at this task.

As others before, we frame this task as information retrieval, and we treat an existing citation’s context as the
query and the corpus of papers as our document collection. For all experiments, we use the ACL Anthology Corpus (AAC) enriched with AAN metadata. We select and separate a subset of documents in our collection as our test set. For each document in our test set (see 3.2 below), we:

1. select all references in the test document that can be resolved to documents inside our document collection (collection-internal references) and remove all other references we cannot match and the citations to them
2. substitute each citation token to a collection-internal reference with a citation placeholder
3. generate a query from the context of this placeholder
4. perform the query, aiming to rank the original cited reference as high in the results as possible

### 3.2 A corpus of annotated contexts

We employ the sentiment- and relevance-annotated corpus of Athar et al. [1] for our test set. In this corpus, 20 papers were selected from the ACL Anthology, and approximately 1700 citation contexts to these papers were manually annotated by a single annotator. Within a window of 2 sentences before the citing sentences and 2 after (2 up, 2 down), each sentence receives two annotations: a) whether it is relevant to the citation and b) its sentiment. The sentiment can be one of: p - positive, n - negative and o - objective.

### 4. EXPERIMENTS AND RESULTS

We have compared the following methods for extracting a citation’s context:

- **window**: a window of n tokens, the same number before and after the citation token
- **sentence**: a window of sentences
  - 1only: only the citing sentence.
  - n[up][down]: n sentences before (up) and m after the citing sentence (down). This window always includes the citing sentence.
  - paragraph: the full paragraph where the citation appears.
- **annotated_sentence**: sentences that were human-annotated as relevant to the citation.

The results are previewed in Table 1. They indicate that forming the context out of sentences that were manually annotated to be relevant to the citation leads to generating superior queries than using any other symmetric method. The minimal pair here is sentence_2up_3down and annotated_sentence_pno, showing that selecting which sentences to include within a 5-sentence window leads to higher scores. Interestingly, selecting sentences based on their annotated sentiment polarity produces worse results, leading us to conclude that sentiment classification, at least as present in this corpus, is not a useful feature.

### 5. REFERENCES


---

**Table 1: Experiment results.** Manually selecting sentences within a 5-sentence context is superior to symmetric methods, irrespective of sentiment annotation.

<table>
<thead>
<tr>
<th>Context extraction method</th>
<th>Avg. MRR score</th>
</tr>
</thead>
<tbody>
<tr>
<td>annotated_sentence_pno</td>
<td>0.1575</td>
</tr>
<tr>
<td>annotated_sentence_po</td>
<td>0.1533</td>
</tr>
<tr>
<td>annotated_sentence_n</td>
<td>0.1505</td>
</tr>
<tr>
<td>window500</td>
<td>0.147</td>
</tr>
<tr>
<td>sentence_0up_1down</td>
<td>0.1403</td>
</tr>
<tr>
<td>window50_50</td>
<td>0.1382</td>
</tr>
<tr>
<td>sentence_1up_1down</td>
<td>0.1378</td>
</tr>
<tr>
<td>sentence_2up_2down</td>
<td>0.136</td>
</tr>
<tr>
<td>window100_100</td>
<td>0.134</td>
</tr>
<tr>
<td>window30_30</td>
<td>0.134</td>
</tr>
<tr>
<td>sentence_paragraph</td>
<td>0.1313</td>
</tr>
<tr>
<td>sentence_1only</td>
<td>0.1309</td>
</tr>
<tr>
<td>sentence_1up</td>
<td>0.1287</td>
</tr>
<tr>
<td>annotated_sentence_p</td>
<td>0.0182</td>
</tr>
<tr>
<td>annotated_sentence_n</td>
<td>0.0134</td>
</tr>
</tbody>
</table>


