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

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 approaches use 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.

JCDL '16 June 19-23, 2016, Newark, NJ, USA

ACM ISBN 978-1-4503-4229-2/16/06.

DOI: http://dx.doi.org/10.1145/2910896.2925431

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 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_[m]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_2down* 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.

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

Context extraction method	Avg. MRR score
annotated_sentence_pno	**0.1575**
annotated_sentence_po	0.1533
annotated_sentence_no	0.1505
window500_500	0.147
sentence_0up_1down	0.1403
window50_50	0.1382
sentence_1up_1down	0.1378
sentence_2up_2down	0.136
window100_100	0.134
window30_30	0.134
sentence_paragraph	0.1313
sentence_1only	0.1309
sentence_1up	0.1287
annotated_sentence_p	0.0182
annotated_sentence_n	0.0134

5. REFERENCES

[1] A. Athar and S. Teufel. Context-enhanced citation sentiment detection. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 597–601. Association for Computational Linguistics, 2012.

[2] D. Duma and E. Klein. Citation resolution: A method for evaluating context-based citation recommendation systems. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 2014.

[3] J. He, J.-Y. Nie, Y. Lu, and W. X. Zhao. Position-aligned translation model for citation recommendation. In *String Processing and Information Retrieval*, pages 251–263. Springer, 2012.

[4] Q. He, J. Pei, D. Kifer, P. Mitra, and L. Giles. Context-aware citation recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 421–430. ACM, 2010.

[5] W. Huang, Z. Wu, C. Liang, P. Mitra, and C. L. Giles. A neural probabilistic model for context based citation recommendation. In *In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.

[6] D. Kaplan, R. Iida, and T. Tokunaga. Automatic extraction of citation contexts for research paper summarization: A coreference-chain based approach. In *Proceedings of the 2009 Workshop on Text and Citation Analysis for Scholarly Digital Libraries*, pages 88–95. Association for Computational Linguistics, 2009.

[7] X. Liu, Y. Yu, C. Guo, Y. Sun, and L. Gao. Full-text based context-rich heterogeneous network mining approach for citation recommendation. In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 361–370. IEEE Press, 2014.

[8] V. Qazvinian and D. R. Radev. Identifying non-explicit citing sentences for citation-based summarization. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 555–564. Association for Computational Linguistics, 2010.

[9] A. Ritchie. Citation context analysis for information retrieval. Technical report, University of Cambridge Computer Laboratory, 2009.

[10] S. Teufel, A. Siddharthan, and D. Tidhar. Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 103–110. Association for Computational Linguistics, 2006.