Semantic Modelling of Citation Contexts for Context-aware Citation Recommendation

Tarek Saier and Michael Färber

Institute AIFB, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
{tarek.saier,michael.faerber}@kit.edu

Abstract. New research is being published at a rate, at which it is infeasible for many scholars to read and assess everything possibly relevant to their work. In pursuit of a remedy, efforts towards automated processing of publications, like semantic modelling of papers to facilitate their digital handling, and the development of information filtering systems, are an active area of research. In this paper, we investigate the benefits of semantically modelling citation contexts for the purpose of citation recommendation. For this, we develop semantic models of citation contexts based on entities and claim structures. To assess the effectiveness and conceptual soundness of our models, we perform a large offline evaluation on several data sets and furthermore conduct a user study. Our findings show that the models can outperform a non-semantic baseline model and do, indeed, capture the kind of information they're conceptualized for.

Keywords: Recommender Systems · Semantics · Digital Libraries

1 Introduction

Citations are a central building block of scholarly discourse. They are the means by which scholars relate their research to existing work—be it by backing up claims, criticising, naming examples, or engaging in any other form. Citing in a meaningful way requires an author to be aware of publications relevant to their work. Here, the ever increasing rate of new research being published poses a serious challenge. With the goal of supporting researchers in their choice of what to read and cite, approaches to paper recommendation and citation recommendation have been an active area of research for some time now [2].

In this paper, we focus on the task of context-aware citation recommendation (see e.g. [7,10,11,14]). That is, recommending publications for the use of citation within a specific, confined context (e.g. one sentence)—as opposed to global citation recommendation and paper recommendation, where publications are recommended with respect to whole documents or user profiles. Within context-aware citation recommendation, we specifically investigate the *explicit* semantic modelling of citation contexts. While *implicit* semantic information (such as what is captured by word embeddings) greatly benefits scenarios like keyword search, we

argue that the specificity of information needs in academia—e.g. finding publications that use a certain data set or address a specific problem—require a more rigidly modelled knowledge representations, such as those proposed in [21] or [8]. Regarding quality, such knowledge representations (e.g. machine readable annotations of scientific publications) would ideally be created manually by the researchers themselves (see [13]). However, neither have such ideals become the norm in academic writing so far, nor are large scale data sets with manually created annotations available. Thus, we create semantic models of citation contexts using NLP techniques to automatically derive such knowledge representations. Using our models we investigate if and when such novel representation formats are beneficial for context-aware citation recommendation.

Overall, we make the following contributions:

1. We propose novel ways of deriving semantically-structured representations from citation contexts, based on entities and claims, intended for context-aware citation recommendation. To the best of our knowledge, this is the first approach of its kind, as previous uses of semantically-structured representations for citation recommendation were only ever applied to whole papers (i.e. in a setting where richer information including authors, list of references, venue, etc. is available).
2. We perform a large-scale offline evaluation using four data sets, in which we test the effectiveness of our models.
3. We also perform a user study to further evaluate the performance of our models and assess their conceptual soundness.
4. We make the code for our models and details of our evaluation publicly available.[1]

The rest of the paper is structured as follows: In Sec. 2 we outline existing works on citation recommendation. We then describe in Sec. 3 the novel semantic approaches to citation recommendation. Sec. 4 is dedicated for the evaluation of our approaches. We conclude in Sec. 5.

2 Related Work

Citation recommendation can be classified into *global* citation recommendation and *context-aware* (sometimes also referred to as *"local")* citation recommendation [10]. Various approaches have been published in both areas, but there is, to the best of our knowledge, not one that both (a) is context-aware and (b) uses explicit semantic representations of citation contexts. We illustrate this in Table 1. In the following, we

Table 1: Overview of related work.

Semantic	*Citation recommendation type*	
	Context-aware	Global
yes	this paper	[19, 28, 29]
no	[5–7, 10–12, 14, 16]	(not considered)

[1] See `https://github.com/IllDepence/ecir2020`.

Table 2: Semantic approaches to global citation recommendation.

Paper	Recommendation approach	Semantic paper model
[19]	Content-based filtering and collaborative filtering	Topic ontology (used to classify papers)
[28]	Hybrid recommender system	Enrich metadata using LOD sources
[29]	Content-based filtering	Semantic distance measure based on relational features between papers

Table 3: Non-semantic approaches to context-aware citation recommendation.

Paper	Recommendation approach	Citation contex model
[10]	Content-based filtering	TF-IDF weighted VSM vectors
[12]	Translational model	"Source language" of translational model
[11]	Neural probabilistic model	Distributed word representation
[5]	Content-based filtering	TF-IDF weighted VSM vectors
[6]	Content-based filtering	TF-IDF weighted VSM vectors with weights for rethorical functions
[7]	Neural citation network	Word embeddings plus author embeddings
[16]	Content-based filtering	Word embeddings along three discourse facets **+1 given citation**
[14]	Graph Convolutional Network + BERT	Word embeddings

therfore outline the most related works on (semantic) global citation recommendation (upper right cell in Table 1), and (non-semantic) context-aware citation recommendation (lower left cell in Table 1).

Global Citation Recommendation. Global citation recommendation is characterized as a task for which the input of the recommendation engine is not a specific citation context but a whole paper. Various approaches have been published for global citation recommendation, some of which can also be used for paper recommendation, i.e., for recommending papers for the purpose of reading [1]. A few *semantic* approaches to global citation recommendation exist (see Table 2). They are based on a semantically-structured representation of papers' metadata (e.g., authors, title, abstract) [28, 29] and/or papers' contents [19]. Note that the approaches proposed in this paper are not using any of the papers' metadata or full text, as our goal is to provide fine-grained, semantically suitable recommendations for specific citation contexts.

Context-aware Citation Recommendation. Context-aware citation recommendation approaches recommend publications for a specific citation context and are thus also called "local" citation recommendation approaches. Existing context-aware citation recommendation approaches soly rely on lexical and syntactic features (n-grams, part-of-speech tags, word embeddings etc.) but do not

attempt to model citation contexts in an explicit semantic fashion. Table 3 gives an overview of context-aware citation recommendation approaches. We can mention SemCir [29] as the only approach we are aware of that *could* be regarded as a semantic approach to context-aware citation recommendation. The explicit semantic representations are, however, not generated from citation contexts (not context-aware), but from papers (global), that are textually (not necessarily semantically) similar to the citation contexts. We therefore categorize it as a semantic global approach.

3 Approach

To ensure wide coverage and applicability of our citation context models, we base our selection of structures to model on a typology of citation functions from the field of citation context analysis ([22], built upon [26]). Note that, because citation context analysis is primarily concerned with the *intent* of the author rather than the *content* of the citation context, we cannot use the functions as a basis for our models directly. Instead we inspect sample contexts of each function type and thereby identify *named entity* (NE) and *claim* as semantic structures of interest, as illustrated in Table 4. Note that the example contexts listed to have no structure ("-" in the *Structure* column) may contain named entities and claims as well (e.g. "DBLP" or "Lamers et al. base their definition of the author's name"), but these are (in the case of NEs) not representative of the cited work or (in the case of claims) just statements *about* a publication rather than statements being backed by the cited work.

The following sections will describe our entity-based and claim-based models for context-aware citation recommendation.

3.1 Entity-based Recommendation

The intuition behind an entity-based approach is that there exists a reference publication for a named entity mentioned in the citation context. For instance, this can be a data set ("CiteSeer^{x} [37]"), a tool ("Neural ParsCit [23]"), or a (scientific) concept ("Semantic Web [37]"). In a more loose sense this can also include publications being referred to as examples ("approaches to context-aware citation recommendation [5–7,10–12,14,16]"). Because names of methods, data sets, tools, etc. in academia often are neologisms and only the most widely used ones of them are reflected in resources like DBpedia, we use a set of noun phrases found in academic publications as surrogates for named entities (instead of performing entity linking). For this, we extract noun phrases from the arXiv publications provided by [24] and filter out items that appear only once. In doing so we end up with a set of 2,835,929 noun phrases[2] (NPs) that we use.

In the following, we define two NP-based representations of citation contexts, R_{NP} and R^{2+}_{NPmrk}. For this, $\mathcal{P}$ shall denote our set of NPs and c shall denote a citation context.

[2] See `https://github.com/IllDepence/ecir2020` for a full list.

Table 4: Semantic structures identified in citation contexts from a range of citation functions used in the field of citation context analysis (NE=named entity).

Function [22]	Structure	Examples (semantic structure *highlighted*)
Attribution	claim	"Berners-Lee et al. [37] argue that *structured collections of information and sets of inference rules are prerequisites for the semantic web to function.*"
	NE	"A variation of this task is '*context-based co-citation recommendation*' [16]."
	-	"In [5] Duma et al. test the effectiveness of using a variety of document internal and external text inputs with a TFIDF model."
Exemplification	NE	"We looked into approaches to *context-aware citation recommendation* such as [5–7, 10–12, 14, 16] for our investigation."
Further reference	-	"See [37] for a comprehensive overview."
Statement of use	NE	"We use *CiteSeer*^{x} [37] for our evaluation."
Application	NE	"Using this mechanism we perform '*context-based co-citation recommendation*' [16]."
Evaluation	-	"The use of DBLP in [37] restricts their data set to the field of computer science."
Establishing links between sources	claim	"A common motivation brought forward for research on citation recommendation is that *finding proper citations is a time consuming task* [7, 9, 10, 16]."
	-	"Lamers et al. [37] base their definition on the author's name whereas Thompson [26] focuses on the grammatical role of the citation marker."
Comparison of own work with sources	claim	"Like [37] we find that, albeit written in a structured language, *parsing LaTeX sources is a non trivial task.*"

$\boldsymbol{R_{\mathbf{NP}}}$ We define $R_{\mathrm{NP}}(c)$ as the set of maximally long NPs contained in c. Formally, $R_{\mathrm{NP}}(c) = \{t | t \text{ appears in } c \wedge t \in \mathcal{P} \wedge t^{+pre} \notin \mathcal{P} \wedge t^{+suc} \notin \mathcal{P}\}$ where t^{+pre} and t^{+suc} denote an extension of t using its preceding or succeeding word respectively. A context *"This has been done for language model training [37]"*, for example, would therefore have "language model training" in its representation, but not "language model".

$\boldsymbol{R^{2+}_{\mathbf{NPmrk}}}$ We define $R^{2+}_{\mathrm{NPmrk}}(c)$ as a subset of $R_{\mathrm{NP}}(c)$ containing, if present, the NP of minimum word length 2 directly preceding the citation marker which a recommendation is to be made for. Formally, $R^{2+}_{\mathrm{NPmrk}}(c) = \{t | t \in R_{\mathrm{NP}}(c) \wedge len(t) \geq 2 \wedge t \text{ directly precedes } m\}$ where m is the citation marker in c that a prediction is to be made for.

Listing 1.1: PredPatt example output.

```
?a shows ?b
    ?a: The paper
    ?b: SOMETHING := context-based methods can outperform
        global approaches
?a can outperform ?b
    ?a: context-based methods
    ?b: global approaches
```

Recommendation As is typical in context-aware citation recommendation [5, 6, 10] we aggregate citation contexts referencing a publication to describe it as a recommendation candidate. To that end, we define frequency vector representations for single citation contexts and documents as follows. A citation context vector is $V(R(c)) = (t_1, t_2, ..., t_{|\mathcal{P}|})$, where t_i denotes how often the ith term in $\mathcal{P}$ appears in $R(c)$. A document vector then is a sum of citation context vectors $\sum_{c \in \varrho(d)} V(R(c))$, where $\varrho(d)$ denotes the set of citation contexts referencing d. Similarities can then be calculated as the cosine of context and document vectors.

3.2 Claim-based Recommendation

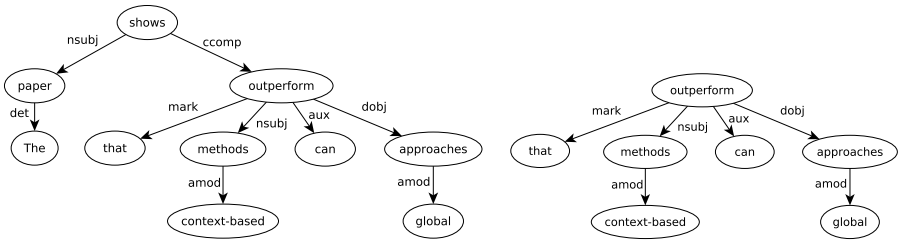
Our claim-based approach is motivated by the fact that citations are used to back up claims (see Table 4). These can, for example, be findings presented in a specific publication ("It has been shown, that ... [37].") or more general themes found across multiple works ("... is still an unsolved task [37-39].").

For the extraction of claims, we considered a total of four state of the art [30] information extraction tools (PredPatt [27], Open IE 5.0 [17], ClausIE [4] and Ollie [18]) and found PredPatt to give the best quality results[3]. For the simple sentence *"The paper shows that context-based methods can outperform global approaches."*, Listing 1.1 shows the user interface output of PredPatt and Figure 1 its internal representation using Universal Dependencies (UD) [20].

Because the predicates and especially arguments in the PredPatt user interface output can get very long—e.g. *"can outperform"* (including the auxiliary verb *"can"*) and *"context-based methods can outperform global approaches"* (unlikely to appear in another citation context with the exact same wording)—we build our claim-based representation R_{claim} from UD trees, as explained in the following section.

R_{claim} For each claim that PredPatt detects, it internally builds one UD tree. To construct our claim-based representation R_{claim}, we traverse each tree, identify the predicate and its arguments (subject and object) and save these in tuples.

[3] See `https://github.com/IllDepence/ecir2020` for details on the evaluation.



(a) Main clause. (b) Sub clause.

Fig. 1: UD trees as generated by PredPatt.

Algorithm 1 Construction of $R_{\text{claim}}(c)$

```
c ← strip_quotation_marks(c)                    ▷ remove quotation marks
c ← merge_citation_markers(c)                   ▷ e.g. "[x], [y]"⇒"[xy]"
pp_trees_c ← predpatt(c)                        ▷ get PredPatt output
output ← []
resolve_rels ← ['name', 'goeswith', 'mwe',      ▷ (b1) UD relations
               'compound', 'conj', 'amod',
               'advmod']
foreach t ∈ pp_trees_c do                       ▷ for all claims identified
    pred ← identify_predicate(t)                ▷ (a) resolve copula if present
    pred ← lemmatize(pred)
    foreach n ∈ traverse(t) do
        if pos_tag(n) == 'NOUN' then
            arg ← resolve_all(n, resolve_rels)  ▷ (b2) resolve compounds etc.
            output.append(pred + ':' + arg)     ▷ build pred:arg tuple
        end if
    end for
end for
return output
```

The exact procedure for this is given in Algorithm 1. If a sentence uses a copula (*be*, *am*, *is*, *are*, *was*), the actual predicate is a child node of the root with the relation type *"cop"*. This is resolved at marker **(a)**. For the identification of useful arguments (markers **(b_1)** and **(b_2)** in Algorithm 1), we look at all nouns within the UD tree and resolve compounds (*"compound"*, *"mwe"*, *"name"* relations), phrases split by formatting (*"goeswith"*), conjunctions (*"conj"*) as well as adjectival and adverbial modifiers (*"amod"*, *"advmod"*). To give an example for this, the noun *"methods"* in both trees in Figure 1 has the adjectival modifier *"context-based"*. In such a case our model would not choose *"methods"* as an argument to *"outperform"* but *"context-based methods"*. Listing 1.2 shows the complete representation generated for the example sentence.

Listing 1.2: $R_{\text{claim}}(c)$ for $c =$ *"The paper shows that context-based methods can outperform global approaches."*.

```
[
 'show:paper',
 'show:context based methods',
 'show:global approaches',
 'outperform:context based methods'
 'outperform:global approaches',
]
```

Recommendation For a set of predicate-argument tuples $\mathcal{T}$, we define frequency vector representations of citation contexts and documents as follows. A citation context vector is $V(R(c)) = (t_1, t_2, ..., t_{|\mathcal{T}|})$, where t_i denotes how often the ith tuple in $\mathcal{T}$ appears in $R(c)$. A document vector, again, is a sum of citation context vectors $\sum_{c \in \varrho(d)} V(R(c))$, where $\varrho(d)$ is the set of citation contexts referencing d. Similarities are then calculated as the cosine of TF-IDF weighted context and document vectors.

$\boldsymbol{R_{\text{claim+BoW}}}$ In addition to R_{claim}, we define a combined model $R_{\text{claim+BoW}}$ as a linear combination of similarity values given by R_{claim} and an bag-of-words model (BoW). Similarities in the combined model are calculated as $sim(A, B) = \sum_{m \in \mathcal{M}} \alpha_m sim_m(A, B)$ of the models $\mathcal{M} = \{R_{\text{claim}}, \text{BoW}\}$ with the coefficients $\alpha_{R_{\text{claim}}} = 1$ and $\alpha_{\text{BoW}} = 2$.

4 Evaluation

We evaluate our models in a large offline evaluation as well as a user study. In total, we compare four models—R_{NP}, R^{2+}_{NPmrk}, R_{claim} and $R_{\text{claim+BoW}}$—against a bag-of-words baseline (BoW). Our choice of a simple BoW model for a baseline is motivated as follows. Because our entity-based and claim-based models are, in their current form, string based, they can be seen as a semantically informed selection of words from the citation context. In this sense, they work akin to what is done in reference scope identification [15]. To evaluate the validity of the selections of words that our models lay focus on, we use the complete set of words contained in the context (BoW) to compare against. Comparing our models against deep learning based approaches (e.g. based on embeddings) would not provide a comparison in this selection behavior, and are therefore was not considered.

Table 5: Citation context sources and filter criteria.

Data source	Citing doc	Cited doc
arXiv [24]	computer science	≥5 citing docs
MAG [25]	computer science, English, abstract not NULL	≥50 citing docs
RefSeer [11]	title, venue, venuetype, abstract, and year in DB not NULL	title, venue, venuetype, abstract, and year in DB not NULL
ACL-ARC [3]	-	has a DBLP ID

Table 6: Key properties of data used for evaluation.

Data set	Train/test split	#Candidate docs	#Test set items	Mean CC/RC (SD)
arXiv	≤2016 / ≥2017	63,239	490,018	21.7 (51.2)
MAG	≤2017 / ≥2018	81,320	141,631	104.1 (198.6)
RefSeer	≤2011 / ≥2012	184,539	53,401	18.2 (47.0)
ACL-ARC	≤2005 / =2006	2,431	3,881	6.8 (9.5)

4.1 Offline Evaluation

Our offline evaluation is performed in a citation re-prediction setting. That is, we take existing citation contexts from scientific publications and split them into training and test subsets. The training contexts are used to learn the representations of the cited documents. The test contexts are stripped of their citations, used as input to our recommender systems and the resulting recommendations checked against the original citations.

Table 5 shows the four data sources we use as well as applied filter criteria. RefSeer and ACL-ARC are often used in related work (e.g. [5,7]), we therefore use *both* of them and *two additional* large data sets to ensure a thorough evaluation. Table 6 gives an overview of key properties of the training and test data for the evaluation. We split our data according to the citing paper's publication date and report *#Candidate docs*: the number of candidate documents to rank for a recommendation; *#Test set items*: the number of test set items (unit: citation contexts); *Mean CC/RC*: the mean number of citation contexts per recommendation candidate in the training set (i.e., a measure for how well the recommendation candidates are described, giving insight into how difficult the recommendation task for each of the data sets is).

Figure 2 shows the results of our evaluation. We measure NDCG, MAP, MRR and Recall at cut-offs from 1 to 10. Note that the evaluation using the arXiv data differs from the other cases in two aspects. First, it is the only case where we can apply R^{2+}_{NPmrk}, because citation marker positions are given. Second, because for citation contexts with several citations (cf. Table 4, "Exemplification")

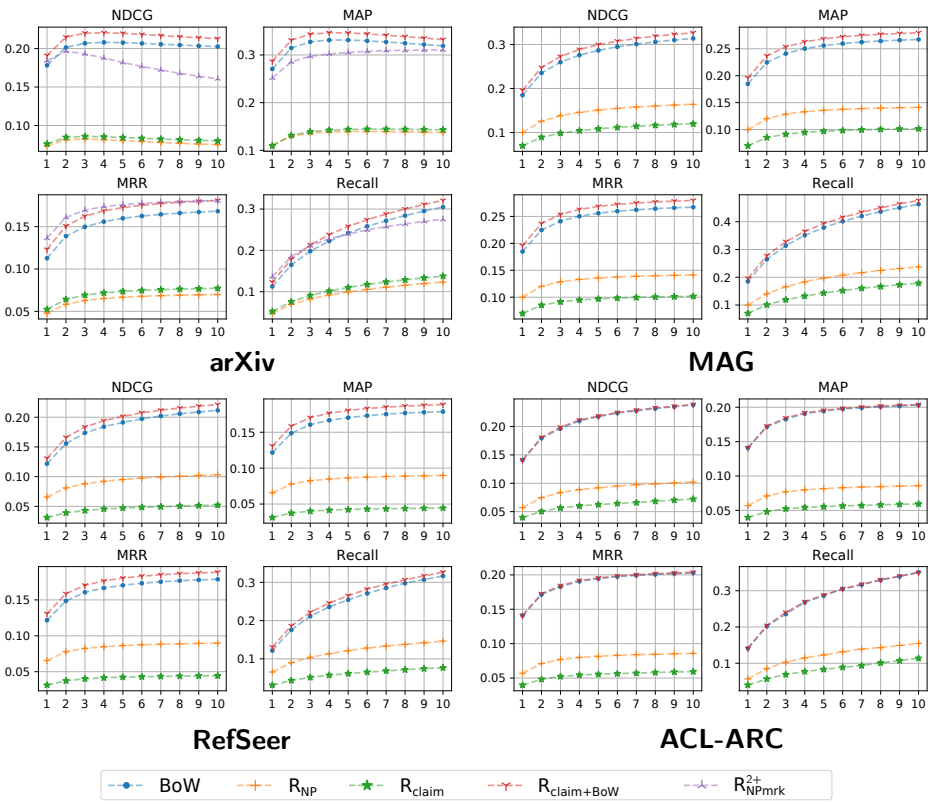


Fig. 2: Evaluation using arXiv, MAG, RefSeer and ACL-ARC. Showing NDCG, MAP, MRR and Recall scores at cut of values from 1 to 10.

the data set lists several cited documents (instead of just a single one), we are able to treat more than a single re-predicted citation as valid. We do this by counting re-predicted "co-citations" as relevant when calculating MAP scores and give them a relevance of 0.5 in the NDCG calculation. This also means that, looking at higher cut-offs, NDCG and MAP values can decrease because ideal recommendations require relevant re-predictions on all ranks above the cut-off.

As for the performance of our models shown in Figure 2, we see that for each of the data sets $R_{\text{claim+BoW}}$ outperforms the BoW baseline in each metric and for all cut-off values.[4] R_{claim} and R_{NP} do not compare in performance with the two aforementioned. This suggests that the claim structures we model with R_{claim} are not enough for well performing recommendations on their own, but do capture important information that non-semantic models (BoW) miss. R^{2+}_{NPmrk}, only present in the arXiv evaluation, gives particularly good results for lower

[4] To validate our findings, we further analyze the NDCG@5 results and note a statistically significant improvement for the arXiv, MAG and RefSeer data but no significant difference for the ACL-ARC data set.

cut-offs and performs especially well in the MRR metric. It performs the worst at high cut-offs measured by NDCG. Note that R^{2+}_{NPmrk} is only evaluated for test set items, where the model was applicable (i.e. where a noun phrase of minimum length 2 is directly preceding the citation marker; cf. Section 3.1). For our evaluation this was the case for 100,308 out of the 490,018 test set items (20.5%). The evaluation results for the citation marker-aware model R^{2+}_{NPmrk} indicate that it is comparatively well suited to recommend citations where there is one particularly fitting publication (e.g. a reference paper) and less suited for exemplifications (cf. Table 4).

4.2 User Study

To obtain more insights into the nature of our evaluation data, as well as a better understanding of our models, we perform a user study in which two human raters (the two authors) judge input-output pairs of our offline evaluation (i.e. citation contexts and the recommendations given for them). For this, we randomly choose 100 citation contexts from the arXiv evaluation, so that we can include R^{2+}_{NPmrk}. For each input context, we show raters the top 5 recommendations of the 3 best performing models of the offline evaluation, i.e., BoW, $R_{\text{claim+BoW}}$ and R^{2+}_{NPmrk} models (resulting in $100 \times 5 \times 3 = 1500$ items). Judgments are performed by looking at each citation context and the respective recommended paper. In addition, we let the raters judge the type of citation (Claim, NE, Exemplification, Other; cf. Table 4).

Table 7: User study evaluation scores at cut-off 5.

Model	Recall@5	MRR@5	MAP@5	NDCG@5
all contexts (138)				
Claim+BoW	**0.53**	0.44	0.41	0.46
BoW	0.51	**0.46**	**0.44**	**0.48**
NPmarker	0.35	0.35	0.33	0.34
only contexts of type "claim" (38)				
Claim+BoW	**0.63**	0.46	0.42	0.49
BoW	0.58	**0.48**	**0.46**	**0.51**
NPmarker	0.20	0.13	0.13	0.15
only contexts of type "NE" (45)				
Claim+BoW	0.46	0.44	0.41	0.44
BoW	0.47	0.45	0.41	0.35
NPmarker	**0.52**	**0.53**	**0.48**	**0.51**
only contexts of type "exemplification" (38)				
Claim+BoW	**0.56**	0.52	0.47	0.52
BoW	0.54	**0.53**	**0.49**	**0.54**
NPmarker	0.21	0.24	0.24	0.24
only contexts of type "other" (17)				
Claim+BoW	0.44	0.29	0.29	0.33
BoW	0.41	0.33	0.33	0.36
NPmarker	**0.50**	**0.50**	**0.44**	**0.47**

Table 7 shows the results based on the raters' relevance judgments. We present measurements for all contexts, as well as each of the citation classes on its own. We note that $R_{\text{claim+BoW}}$ and BoW are close, but in contrast to

the offline evaluation, $R_{\text{claim+BoW}}$ only outperforms BoW in the Recall metric. In the case of NE type citations, the R^{2+}_{NPmrk} model performs better than the other two models in all metrics. Furthermore, we can see that both $R_{\text{claim+BoW}}$ and R^{2+}_{NPmrk} achieve their best results for the type of citation they're designed for—Claim and NE respectively. This indicates that both models actually capture the kind of information they're conceptualized for. Compared to the offline evaluation, we measure higher numbers overall. While the user study is of considerably smaller scale and a direct comparison therefore not necessarily possible, the notably higher numbers indicate, that a re-prediction setting involves a non-negligible number of false negatives (actually relevant recommendations counted as not relevant).

4.3 Main Findings

The entity-based model R^{2+}_{NPmrk}, which captures noun phrases preceding the citation marker, performs best at low cut-offs and in the MRR metric. Low cut-offs and measuring the MRR can be interpreted as emulating citations for reference publications. This interpretation is also backed by the results of the user study, where R^{2+}_{NPmrk} outperformed all other models when recommending for citation contexts that referenced a named entity or concept. We therefore conclude that R^{2+}_{NPmrk} is well suited for recommending such types of citations. Our claim-based model R_{claim} does not compare in performance to a BoW baseline, but $R_{\text{claim+BoW}}$ outperforms aforementioned. We take this as an indication that the claim representation encodes important information which the non-semantic BoW model is not able to capture. In the user study $R_{\text{claim+BoW}}$ performs best for citation contexts, in which a claim is backed by the target citation. This suggests that the model indeed captures information related to claim structures.

5 Conclusion

In the field of context-aware citation recommendation, the explicit semantic modeling of citation contexts is not well explored yet. In order to investigate the merit of such approaches, we developed semantic models of citation contexts based on entities as well as claim structures. We then evaluated our models on several data sets in a citation re-prediction setting and furthermore conducted a user study. In doing so, we could demonstrate their applicability and conceptual soundness. The next step from hereon is to move from semantically informed text-based models to explicit knowledge representations. Our research also shows, that differentiating between different semantic representations of citation contexts due to varying ways of citing information is reasonable. Developing different citation recommendation approaches, depending on the semantic citation types, might therefore be a promising next step in our research.

References

1. Beel, J., Dinesh, S.: Real-World Recommender Systems for Academia: The Pain and Gain in Building, Operating, and Researching them. In: Proceedings of the Fifth Workshop on Bibliometric-enhanced Information Retrieval (BIR) co-located with the 39th European Conference on Information Retrieval (ECIR 2017). pp. 6–17 (2017)
2. Beel, J., Gipp, B., Langer, S., Breitinger, C.: Research-paper recommender systems: a literature survey. International Journal on Digital Libraries **17**(4), 305–338 (Nov 2016)
3. Bird, S., Dale, R., Dorr, B.J., Gibson, B.R., Joseph, M.T., Kan, M., Lee, D., Powley, B., Radev, D.R., Tan, Y.F.: The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics. In: Proceedings of the 6th International Conference on Language Resources and Evaluation. LREC'18 (2008)
4. Corro, L.D., Gemulla, R.: ClausIE: Clause-Based Open Information Extraction. In: Proceedings of the 22nd International World Wide Web Conference. pp. 355–366. WWW'13 (2013)
5. Duma, D., Klein, E.: Citation resolution: A method for evaluating context-based citation recommendation systems. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. pp. 358–363. ACL'14 (2014)
6. Duma, D., Klein, E., Liakata, M., Ravenscroft, J., Clare, A.: Rhetorical Classification of Anchor Text for Citation Recommendation. D-Lib Magazine **22** (2016)
7. Ebesu, T., Fang, Y.: Neural Citation Network for Context-Aware Citation Recommendation. In: Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 1093–1096 (2017)
8. Gábor, K., Buscaldi, D., Schumann, A., QasemiZadeh, B., Zargayouna, H., Charnois, T.: SemEval-2018 Task 7: Semantic Relation Extraction and Classification in Scientific Papers. In: Proceedings of The 12th International Workshop on Semantic Evaluation. pp. 679–688. SemEval@NAACL-HLT'18 (2018)
9. He, Q., Kifer, D., Pei, J., Mitra, P., Giles, C.L.: Citation recommendation without author supervision. In: Proceedings of the Forth International Conference on Web Search and Web Data Mining. pp. 755–764. WSDM'11 (2011)
10. He, Q., Pei, J., Kifer, D., Mitra, P., Giles, L.: Context-aware Citation Recommendation. In: Proceedings of the 19th International Conference on World Wide Web. pp. 421–430. WWW '10, ACM, New York, NY, USA (2010)
11. Huang, W., Wu, Z., Liang, C., Mitra, P., Giles, C.L.: A neural probabilistic model for context based citation recommendation. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. pp. 2404–2410. AAAI'15 (2015)
12. Huang, W., Wu, Z., Mitra, P., Giles, C.L.: Refseer: A citation recommendation system. In: Proceedings of the IEEE/ACM Joint Conference on Digital Libraries. pp. 371–374. JCDL'14 (2014)
13. Jaradeh, M.Y., Oelen, A., Farfar, K.E., Prinz, M., DSouza, J., Kismihók, G., Stocker, M., Auer, S.: Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge. In: Proceedings of the 10th International Conference on Knowledge Capture. p. 243246. K-CAP'19 (2019)
14. Jeong, C., Jang, S., Shin, H., Park, E., Choi, S.: A Context-Aware Citation Recommendation Model with BERT and Graph Convolutional Networks (2019), `http://arxiv.org/abs/1903.06464`

15. Jha, R., Jbara, A.A., Qazvinian, V., Radev, D.R.: NLP-driven citation analysis for scientometrics. Natural Language Engineering **23**(1), 93–130 (2017)
16. Kobayashi, Y., Shimbo, M., Matsumoto, Y.: Citation Recommendation Using Distributed Representation of Discourse Facets in Scientific Articles. In: Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries. pp. 243–251. JCDL'18 (2018)
17. Mausam: Open Information Extraction Systems and Downstream Applications. In: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. pp. 4074–4077. IJCAI'16 (2016)
18. Mausam, Schmitz, M., Bart, R., Soderland, S., Etzioni, O.: Open Language Learning for Information Extraction. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. pp. 523–534. EMNLP-CoNLL'12, Stroudsburg, USA (2012)
19. Middleton, S.E., Roure, D.D., Shadbolt, N.: Capturing knowledge of user preferences: ontologies in recommender systems. In: Proceedings of the First International Conference on Knowledge Capture. pp. 100–107. K-CAP'01 (2001)
20. Nivre, J., de Marneffe, M., Ginter, F., Goldberg, Y., Hajic, J., Manning, C.D., McDonald, R.T., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., Zeman, D.: Universal Dependencies v1: A Multilingual Treebank Collection. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation. pp. 1659–1666. LREC'16 (2016)
21. Peroni, S., Shotton, D.M.: FaBiO and CiTO: Ontologies for describing bibliographic resources and citations. Journal of Web Semantics **17**, 33–43 (2012)
22. Petrić, B.: Rhetorical functions of citations in high- and low-rated master's theses. Journal of English for Academic Purposes **6**(3), 238 – 253 (2007)
23. Prasad, A., Kaur, M., Kan, M.Y.: Neural ParsCit: A Deep Learning Based Reference String Parser. International Journal on Digital Libraries **19**, 323–337 (2018)
24. Saier, T., Färber, M.: Bibliometric-Enhanced arXiv: A Data Set for Paper-Based and Citation-Based Tasks. In: Proceedings of the 8th International Workshop on Bibliometric-enhanced Information Retrieval. pp. 14–26. BIR'19 (2019)
25. Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.P., Wang, K.: An Overview of Microsoft Academic Service (MAS) and Applications. In: Proceedings of the 24th International Conference on World Wide Web. pp. 243–246. WWW'15 (2015)
26. Thompson, P.: A pedagogically-motivated corpus-based examination of PhD theses: Macrostructure, citation practices and uses of modal verbs. Ph.D. thesis, University of Reading (2001)
27. White, A.S., Reisinger, D., Sakaguchi, K., Vieira, T., Zhang, S., Rudinger, R., Rawlins, K., Van Durme, B.: Universal Decompositional Semantics on Universal Dependencies. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 1713–1723. EMNLP'16 (2016)
28. Zarrinkalam, F., Kahani, M.: A multi-criteria hybrid citation recommendation system based on linked data. In: Proceedings of the 2nd International eConference on Computer and Knowledge Engineering. pp. 283–288. ICCKE'12 (2012)
29. Zarrinkalam, F., Kahani, M.: SemCiR: A citation recommendation system based on a novel semantic distance measure. Program: Electronic Library and Information Systems **47**, 92–112 (2013)
30. Zhang, S., Rudinger, R., Durme, B.V.: An Evaluation of PredPatt and Open IE via Stage 1 Semantic Role Labeling. In: Proceedings of the 12th International Conference on Computational Semantics. IWCS'17 (2017)