Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 3398–3406
Marseille, 20-25 June 2022

ACT2: A multi-disciplinary semi-structured dataset for importance and purpose classification of citations

Suchetha N. Kunnath[1], Valentin Stauber[2], Ronin Wu[2], David Pride[1], Viktor Botev[2], Petr Knoth[1]
[1]Knowledge Media Institute, The Open University, UK, [2]Iris.ai, Norway
snk56@open.ac.uk, valentin@iris.ai, ronin@iris.ai, david.pride@open.ac.uk, victor@iris.ai, petr.knoth@open.ac.uk

Abstract

Classifying citations according to their purpose and importance is a challenging task that has gained considerable interest in recent years. This interest has been primarily driven by the need to create more transparent, efficient, merit-based reward systems in academia; a system that goes beyond simple bibliometric measures and considers the semantics of citations. Such systems that quantify and classify the influence of citations can act as edges that link knowledge nodes to a graph and enable efficient knowledge discovery. While a number of researchers have experimented with a variety of models, these experiments are typically limited to single-domain applications and the resulting models are hardly comparable. Recently, two Citation Context Classification (3C) shared tasks (at WOSP2020 and SDP2021) created the first benchmark enabling direct comparison of citation classification approaches, revealing the crucial impact of supplementary data on the performance of models. Reflecting from the findings of these shared tasks, we are releasing a new multi-disciplinary dataset, ACT2, an extended SDP 3C shared task dataset. This modified corpus has annotations for both citation function and importance classes newly enriched with supplementary contextual and non-contextual feature sets the selection of which follows from the lists of features used by the more successful teams in these shared tasks. Additionally, we include contextual features for cited papers (e.g. Abstract of the cited paper), which most existing datasets lack, but which have a lot of potential to improve results. We describe the methodology used for feature extraction and the challenges involved in the process. The feature enriched ACT2 dataset is available at `https://github.com/oacore/ACT2`.

Keywords: Citation Classification, Citation Importance, Citation Purpose, Research Evaluation

1. Introduction

Characterising citations according to their type and importance has been extensively studied by researchers since 2006 (Teufel et al., 2006). While the former gives insights about author citation behaviours, the latter emphasises how influential a referenced work is to the citing article. These semantic aspects of citations have the potential to address some of the shortcomings of the currently widely used bibliometric measures which consider all citations equally (Garfield, 1972; Shotton, 2010). Moreover, being able to quantify and classify the influence of a cited work allows linking research works systematically which can lead to more efficient knowledge discovery (Ding et al., 2014).

Classifying citations based on purpose or importance is, however, challenging in many ways (Kunnath et al., 2021). As with many other classification problems, the primary difficulty is in procuring annotated datasets. The reluctance of authors in expressing their real motivation for citing a work makes annotation difficult, even for subject experts (Athar and Teufel, 2012). Likewise, choosing a suitable best classification schema that is better capable of capturing specific citation attributes without being too generic or causing any overlap between classes is critical for this task (Radoulov, 2008). Another dataset specific challenge involved is in determining what attributes from citing and cited papers should be incorporated in the corpus, besides the citation context, which is typically the sentence surrounding citation in citing article.

To address the above mentioned challenges, (Pride and Knoth, 2020; Pride et al., 2019) introduced a multi-disciplinary, author annotated dataset, known as Academic Citation Typing (ACT), with annotations for both citation function and citation importance. Following the classification taxonomy from (Jurgens et al., 2018), this dataset groups citations into one of the six functions, which expresses the role of the cited paper in the citing publication. Additionally, for predicting the academic influence, the dataset was annotated using binary classification schemes proposed by (Valenzuela et al., 2015; Zhu et al., 2015). A portion of the ACT dataset was used in two consecutive 3C shared tasks which were collocated with the 8th International Workshop on Mining Scientific Publications (WOSP), 2020[1] (Kunnath et al., 2020) and the 2nd workshop on Scholarly Document Processing (SDP) 2021[2] (N. Kunnath et al., 2021).

In this work, we provide an extended SDP 3C shared task dataset (ACT-SDP) with new auxiliary features, selected as they are seen as highly promising based on the experience from the two 3C shared task itera-

[1]`https://wosp.core.ac.uk/jcdl2020/shared_task.html#3ctask`

[2]`https://sdproc.org/2021/sharedtasks.html#3c`

tions. Using automated methods, we extract additional contextual and non-contextual information for both citing and cited papers to create the new dataset, known as ACT2. We believe this newly updated corpus, will serve as a unique, common resource for reinforcing future research in citation classification.

2. Related Work

Previously developed datasets for citation classification are domain specific, with data predominantly from computational linguistics and biomedical sciences (Teufel et al., 2006; Valenzuela et al., 2015; Jurgens et al., 2018; Cohan et al., 2019). Moreover, earlier classification schemes have a varying granularity; some taxonomies are too fine-grained for machine learning models (Shotton, 2010), while others are quite broad (Cohan et al., 2019; Nicholson et al., 2021). Concerning the methods used for citation classification, earlier systems tend to rely on supervised models with hand engineered contextual and non-contextual features. Recent deep language models like SciBERT (Beltagy et al., 2019), pre-trained from a large un-annotated corpus, is among some of the most promising approaches for citation classification.

In an attempt to facilitate formal comparison of existing methods, (Kunnath et al., 2020) organised the first citation context classification (3C) shared task[3][4]. The participants were evaluated on citation purpose and citation influence classification tasks on a common platform. A second version of the shared task organised as part of SDP witnessed significant participation of more than 20 teams, thus serving as a benchmark for evaluating different systems[5][6] (N. Kunnath et al., 2021). The overall results from both editions of the shared task provides an indication of the more successful approaches.

3. Methodology

To determine additional features required for creating the new ACT2 dataset, we relied on the following two approaches:

- **Results from 3C shared task** – The response of participating teams from both versions of the shared task gave various insights about the features that are critical for citation purpose and influence classification tasks (Maheshwari et al., 2021; Baig et al., 2021; Varanasi et al., 2021; de Andrade and Gonçalves, 2020; Mishra and Mishra, 2020). Table1 illustrates the features used by teams, which finished first and second. Submitted results show the use of citation context for extracting semantic information for both tasks. However, cited and citing titles were also found valuable for performance improvement. Besides, most of the participants commented on the need to include additional information about the cited paper as part of the shared task dataset.
- **Insights from meta-analysis** – A comprehensive meta-analysis (Kunnath et al., 2021) on citation classification reveals the use of a diverse set of features by the existing methods. Table 2 shows the most frequently used attributes for both purpose and influence classification systems. There is a significant overlap of features used for both tasks. However notable difference is in the use of more contextual information from citing paper for citation purpose classification, whereas methods for influence task relies on features from cited work (e.g. abstract) as well. Similar feature preferences was also observed from the submitted results of the 3C shared task.

3.1. ACT Dataset

ACT is the largest multi-disciplinary dataset for citation classification with citations annotated by 883 authors for both purpose and influence. $11,233$ citing sentences are labelled according to the following purpose classes: BACKGROUND, USES, COMPARES_CONTRASTS, MOTIVATION, EXTENSION and FUTURE. The class COMPARES_CONTRASTS was further extended to reflect similarities, differences and disagreements between citations (Pride and Knoth, 2020). To represent influence, citations are annotated using the two classes: INCIDENTAL and INFLUENTIAL.

The original dataset has the following fields:

- **unique_id** – Unique identifier
- **core_id** – COREID of citing caper
- **citing_title** – Citing paper title
- **citing_author** – Citing paper first Author
- **cited_title** – Cited paper title
- **cited_author** – Cited paper first author
- **citation_context** – Citing sentence with current citation masked using *#AUTHOR_TAG*
- **citation_class_label** – Annotated citation functions
- **citation_influence_label** – Annotated importance labels

Here, core_id represents the unique identifier assigned to papers from the world's largest open access dataset provided by the CORE[7] (Knoth and Zdrahal, 2012) aggregator. All the citing papers in this dataset are obtained from CORE.

[3]https://www.kaggle.com/c/3c-shared-task-purpose
[4]https://www.kaggle.com/c/3c-shared-task-influence
[5]https://www.kaggle.com/c/3c-shared-task-purpose-v2
[6]https://www.kaggle.com/c/3c-shared-task-influence-v2
[7]https://core.ac.uk/

Year	Subtask	Team*	Field Used	Features Used
2021	Purpose	Team 1	citation context	Sentence embeddings from SciBERT, RoBERTa
		Team 2	citation context, fulltext	TF-IDF, GLoVe + ELMo, citation frequency, offset
	Influence	Team 1	citation context, citing title	Sentence embeddings from SciBERT, RoBERTa
		Team 2	citation context, cited & citing title	TF-IDF, Word Mover's Distance, self-citation, Polarity, keyword overlap length of cited title and context
2020	Purpose	Team 1	citation context	TF-IDF, GLoVe, LDA for topic extraction
		Team 2	citation context, citing & cited title	TF-IDF
	Influence	Team 1	citation context, citing & cited title	fasttext embeddings, TF-IDF
		Team 2	citation context, cited	TF-IDF, self-citation

* Top 2 teams

Table 1: Features used by winning shared task participants

Task	**Features Used**						
	Contextual				**Non-Contextual**		
	Syntactic	**Semantic**			**Positional-based**	**Frequency-based**	**Other**
		Textual-based	**Similarity-based**	**Polarity-based**			
Purpose	Dependency Relations, POS Tags	Cue phrases, n-grams	Topic similarity with cited paper	Polarity lexicons	Citation location within Article, Paragraph, Section	Citation count	Self-citation
Influence	Co-mentions, Explicit Vs Implicit citations	Cue words	Abstract similarity with cited paper	Polarity lexicons	Citation location within Article, Paragraph, Section	Citation count	Self-citation, Author Overlap

Table 2: Most frequently used features by existing works

From the $11,233$ instances, a subset constituting $4,000$ instances ($3,000$ for training set and $1,000$ for test set) with citation class distributions matching original dataset (Pride and Knoth, 2020) was extracted and used for the first 3C shared task. The same dataset was extended with fulltext for citing papers to create ACT-SDP. This dataset was used for the second edition of 3C shared task.

3.2. ACT2

Based on the findings from the shared task and the field meta-analysis, we inspect the possibility of enriching the 3C shared task dataset with the following contextual and non-contextual features:

- **citation_offset**
- **total_doc_length**
- **section_info**
- **citing_abstract**
- **cited_abstract**
- **self_citation**
- **direct_citations**
- **citing_publication_info**
- **cited_publication_info**
- **cited_publication_date**
- **co_mentions**
- **cited_doi**

Table 3 shows the newly introduced 12 features and their definitions. All the structural and publications related features are included based on the insights obtained from meta-analysis, whereas features like abstracts are added as a result of the feedback we received from the shared task participants.
The following subsections describe the methodology we used for updating the shared task dataset and also explains the tools and approaches used for extracting the above mentioned features.

Feature name	Definition	Source
citation_offset (F1)	Number of characters in the citing paper before the appearance of the citation	meta-analysis/ shared task
total_doc_length (F2)	Total number of characters in the citing paper	meta-analysis
section_info (F3)	A mapping in the format of {section_title: character_offset}, where character_offset indicates the number of characters before the appearance of the first character in section_title	meta-analysis
citing_abstract (F4)	Abstract of the citing paper	meta-analysis
cited_abstract (F5)	Abstract of the cited paper	meta-analysis
self_citation (F6)	The citing paper cites at least one citing author's own work (1: True; 0: False)	meta-analysis/ shared task
direct_citations (F7)	The number of times the cited paper is cited within the citing paper	meta-analysis
citing_publication_info (F8)	The journal/conference title of the citing paper	shared task
cited_publication_info (F9)	The journal/conference title of the cited paper	shared task
cited_publication_date (F10)	The publication date of the cited paper in YYYY(-MM) format	meta-analysis
co_mentions (F11)	The number of other citations co-occurring in the citation context with the cited paper	meta-analysis
cited_doi (F12)	DOI of the cited paper	meta-analysis

Table 3: New features in ACT2, their definitions and reason for including in the dataset.

3.3. Document Parsing

Following (Lo et al., 2020), we used a combination of both GROBID (Lopez, 2009) and ScienceParse[8] for parsing PDFs associated with citing papers. To this end, we extracted all the citing paper PDFs from CORE, which were further parsed using GROBID to obtain XML/TEI encoded documents, and ScienceParse to obtain json encoded documents.

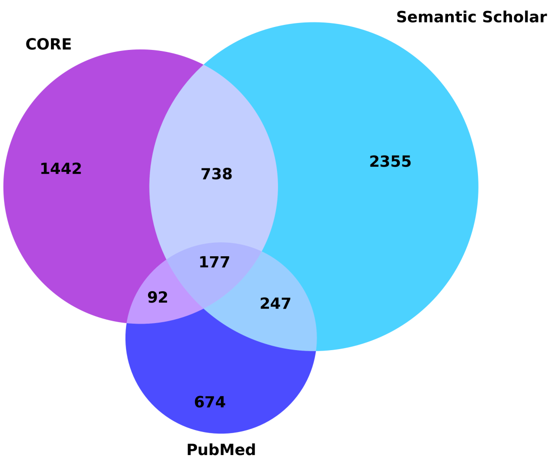


Figure 1: Distribution of cited abstracts extracted from CORE, Semantic Scholar and PubMed

[8]https://github.com/allenai/science-parse

3.4. Features extracted using GROBID

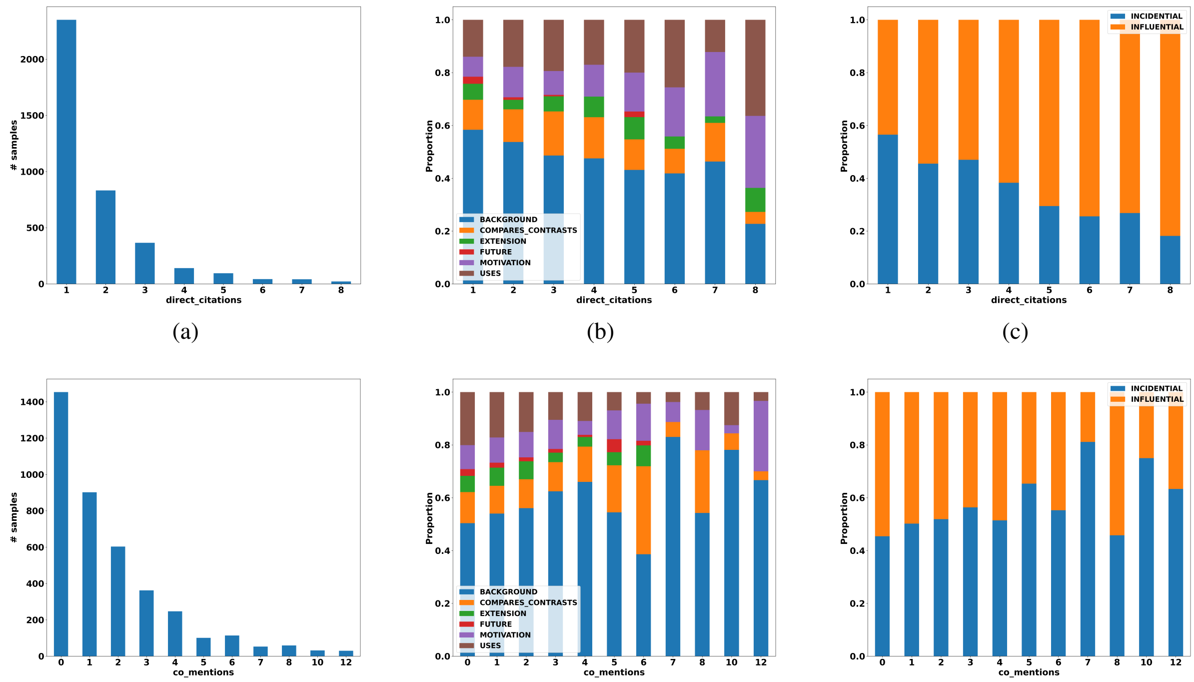
From the parsed GROBID XML output, we extracted the following meta-data:

1. citing paper abstract
2. citing paper title
3. citing authors
4. citing publication information
5. cited paper title
6. cited authors
7. cited publication date
8. cited publication information
9. cited DOI

If any of 5, 6 or 7 was missing from the GROBID extracted data, we used the CrossRef API[9] to extract the missing data if 9 was available.

For updating the ACT dataset we needed to match the citation information extracted from GROBID with the entries in the original dataset. To this end we used a sequence matcher to compare cited title, cited author and citation context of each dataset entry with possible candidate citations extracted from GROBID. We calculated an aggregate overall matching score as a weighted sum of the 3 sequence matching score contributions with ratios cited title/cited author/citation context = 2/2/1. We then chose the candidate with the highest overall matching score for updating the dataset entry.

[9]https://github.com/CrossRef/rest-api-doc



(a) (b) (c)

(d) (e) (f)

Figure 2: Distribution of direct citations and co-mentions with respect to citation function and importance classes: (a) Number of samples per direct citations, (b) Proportion of citation function per direct citations, (c) Proportion of citation influence per direct citations, (d) Number of samples per co-mentions, (e) Proportion of citation function per co-mentions, (f) Proportion of citation influence per co-mentions

3.5. Refining citation context

We also improved the citation context present in the dataset. Using the sentence segmentation feature from GROBID and the sentence splitter (sentencizer) from SpaCy[10] we extracted the citing sentence, i.e. the sentence in which the citation occurs and updated the citation context with this citing sentence. Additionally, we mask the entire citation using *#CITATION_TAG*.

3.6. Section title and offset extraction using ScienceParse

We used ScienceParse[11] to extract the full text segmented into sections from the paper PDFs. As intermediary data we extracted the paper full texts by concatenating all extracted sections, prepended with their extracted titles. We then recorded the absolute character offset of each section (more specifically, its title) in these full texts. We added the extracted section information as a json string of the mapping {section_title: character_offset} under the column section_info.
Additionally we added the absolute character offset of each citation in these full texts under citation_offset, as well as the document length of the citing paper as total characters of the extracted full texts under total_doc_length.

[10] https://github.com/explosion/spaCy
[11] https://github.com/allenai/science-parse

3.7. Cited and Citing abstract

The ACT dataset contains citing papers that are open access and hence present in CORE. However, this was not necessarily the case for all the cited articles, wherein fulltext was not publicly available in most cases. Consequently, we used multiple resources including Semantic Scholar[12] and PubMed Central (PMC) [13] API services for extracting information about cited papers in addition to the CORE API [14]. Initially, all the API services were queried using cited titles from the shared task dataset. For Semantic Scholar and PubMed, we used the response to get respective IDs, which were then queried for extracting cited_abstract. For extracting biomedical cited abstracts from PMC, we utilised the python package, Biopython[15] (Cock et al., 2009). Finally, for citing_abstract extraction, we mainly used parsed XML files from GROBID. However, some of the missing abstracts were obtained using ScienceParse.

[12] https://www.semanticscholar.org/product/api
[13] https://www.ncbi.nlm.nih.gov/home/develop/api/
[14] https://core.ac.uk/services/api
[15] https://biopython.org/

3.8. Cited DOIs

If cited DOIs were not available from GROBID, we extracted them using the CrossRef API[16]. Here, the cited paper title was used to query the API, and the retrieved result was used, if the sequence matching scores between the queried and retrieved titles were above 0.8 only.

3.9. Self-citations

For extracting self_citations, we analysed the intersection between citing and cited authors. We matched both forenames and surnames extracted from GROBID. In case of missing citing author details, we additionally queried the CORE API for retrieving author names. In the dataset, we represent this as a binary feature with values 1 and 0 indicating the presence or absence of a self-citation.

3.10. Direct Citations and Co-Mentions

Frequency based features like direct_citations and co_mentions are obtained as aggregates from the raw data extracted with GROBID mentioned in Sec. 3.4.

4. Results

We were able to extract all values for the features extracted with ScienceParse, i.e. citation_offset, total_doc_length and section. However, for the rest of the features, we encountered several issues resulting in missing values. The missing values for the features extracted using GROBID stem from the fact that for some dataset entries we couldn't make a clear match to any corresponding GROBID extracted citation. In these cases the cited_author from the original dataset is not contained within the newly extracted authors.[17]

We have identified the following issues contributing to above mentioned mismatches:

- Paragraphs that are completely missing from the XML extracted from GROBID
- Citations in the text that are not detected as such by GROBID
- Citations that are marked as such by GROBID, but are not linked against any entry in the bibliography
- Mistakes in the original dataset features used for matching, which are too severe for a matching attempt
- Mistakes in the citations or bibliography entries made by the paper author

In total we encountered 68 such mismatches, resulting in at least 68 missing values for features which rely on GROBID data alone.

Table 4 shows the number of missing values for each newly extracted feature. The highest number of missing values were reported for cited abstracts. Figure 1 illustrates the number of abstracts that were obtained from all three APIs. The maximum number of abstracts was extracted using Semantic Scholar. Due to the frequent non-availability of fulltexts in the repositories, extracting cited abstracts was challenging. Out of $4,000$ total instances present in the dataset, we managed to extract $3,040$ abstracts.

To represent the correlation between direct citations and co-mentions with citation function and the importance categories, we plotted graphs as shown in Figure 2. In order to get better insights from the graphs, we removed all null values and also removed direct_citation and co_mentions values with less than 20 instances. As the graphs indicate, the probability of a citation belonging to the INFLUENTIAL class is higher as the number of direct citations increases. Similarly, for function classification, as direct citation count increases, it is more likely that the citation does not belong to the BACKGROUND class. For co-mentions however, it becomes more likely for the citation to belong to the BACKGROUND or INCIDENTAL class as the comention value increases.

Features	# missing
cited_abstract	960
citing_abstract	8
self_citation	422
section_info	0
total_doc_length	0
citation_offset	0
cited_doi	354
citing_publication_info	315
cited_publication_info	72
cited_publication_date	112
co_mentions	68
direct_citations	68

Table 4: Extracted features and their statistics

5. Discussion

While there are multiple datasets for citation classification, the available datasets are restricted to specific domains (Computer Science and Bio Medicine). However, academic citation styles differ across disciplines and these differences are not fully represented in the

[16] https://github.com/CrossRef/rest-api-doc

[17] We count cases, where the cited_author from the original dataset is partially contained within one of the newly extracted authors, also as match (e.g. "'hara" vs. "O'hara"), as long as the original author name is longer than 2 characters. Cases with shorter original author names stem from a faulty extraction in the original datset and have been replaced by their best match from GROBID if the overall matching score is higher than 0.8.

Dataset	Annotations		Multi-Disciplinary?	Size	Enhanced Feature Set
	Citation Function	Citation Importance			
CFC Corpus (Teufel et al., 2006)	✓	✗	✗	548	Structural and contextual citing info
ACL-ARC (Jurgens et al., 2018)	✓	✗	✗	$1,969$	Structural and contextual citing info
SciCite (Cohan et al., 2019)	✓	✗	✗	$11,020$	Structural and contextual citing info
ACT ((Pride and Knoth, 2020))	✓	✓	✓	$11,233$	Contextual citing info
ACT-SDP ((N. Kunnath et al., 2021))	✓	✓	✓	$4,000$	Contextual citing info
ACT2 (This work)	✓	✓	✓	$4,000$	Structural and contextual citing info, contextual and frequency based cited info

Table 5: ACT2: Comparison with existing datasets

existing homogeneous datasets. This makes the models trained on these datasets not necessarily directly deployable on research literature more broadly. Focusing on heterogeneity and domain diversity is thus one of the main advantages of our approach over other existing in the literature.

Additionally, the other datasets are not necessarily enriched with features that are used by a variety of state-of-the-art approaches. This means that they first need to be enriched before a range of methods can be applied. Which on the other hand makes performance comparisons rather complicated, a key problem our work addresses.

Table 5 provides a comparison of ACT2 with other annotated citation classification datasets currently available. The multi-disciplinary nature[18] and enriched feature set make ACT2 an all-inclusive unique corpus for citation purpose and importance for researchers working in this domain.

As all citing documents in ACT2 are open access, it was possible for us to produce and allow free redistribution of this dataset including all the enrichments. Looking at commercial offerings in this area, such as scite[19] which classifies citations into the three classes, SUPPORTING, MENTIONING and CONTRASTING (previously somewhat controversially called DISPUTING), the annotated datasets underpinning these models are not openly available for scrutiny. However, we strongly believe that keeping citation classification datasets openly available is a necessary precondition for being able to use them in a variety of use cases, but especially in highly sensitive use cases such as in research assessment.

ACT2 introduces 12 new features which have been automatically extracted and have the potential to improve classification prediction. For instance, structural attributes like section details allow to narrow down the motivation for citing. The three frequency-based features are directly related to the influence or type of a citation. For example, self-citation helps trace the research trajectory of one or a group of researchers. On the other hand, the number of direct citations in a citing paper for a cited work is a quantitative indication of the direct impact of the cited work (Wang et al., 2020). The position of the citation within the citing paper helps identify and quantify how often and where different citation classes tend to appear in citing documents, providing possibilities to use location information in classification models.

However, these features can also be extracted and used more broadly in other research and applications areas. For instance, co-mentions can be effectively used in recommender systems, identifying papers with a similar research focus. Citing and cited publication information can be used to detect research collaboration evolution, and to match researchers with similar interests. Contextual similarity between citing and cited abstracts allows more in-depth automated analyses of citation classifications and/or citing motivations.

Beyond research assessment, the models that can be trained and derived from ACT2 can also be applicable in a variety of use cases including but not limited to research information retrieval, recommender systems and scholarly knowledge graphs.

Although using the best performing scientific publication parsers currently available, extracting information from multi-disciplinary research papers remains challenging. The number of missing values shown in Table 4 is an indication of this fact and constitutes a known limitation of this work. A future improvement in the performance of scientific parsers will also help improve the quality of extracted features and minimise missing values.

6. Conclusion

Motivated by the experience from running two shared tasks for citation context classification and a meta-analysis of state-of-the-art approaches, we openly re-

[18]The instances present in the full ACT dataset are from one of the 19 top-level domains from Microsoft Academic Graph (MAG), with Psychology (21.99%), Medicine (13.48%), Biology (10.91%) and Computer Science (10.27%) dominating the corpus.

[19]https://scite.ai/

lease an extended ACT2 dataset for citation classification. We use automated methods to enrich the ACT-SDP dataset with 12 additional features valuable in citation classification. These structural, discursive and frequency based features are extracted from citing and cited articles. We believe this new multi-annotated, multi-disciplinary, semi-structured open dataset will serve as a standard corpus for training, benchmarking and experimenting with citation classification models.

7. Acknowledgements

This paper was supported by funding from NRC, Project ID: 309594, the AI Chemist under the cooperation of Iris.ai AS with The Open University, UK, covering the contributions of Valentin Stauber, Ronin Wu, David Pride, Viktor Botev and partly Petr Knoth.
Additionally, we received funding contribution from Jisc under Grant Reference: 4133, OU Scientometrics PhD Studentship covering the contirbutions of Suchetha N. Kunnath and Petr Knoth.

8. Bibliographical References

Athar, A. and Teufel, S. (2012). Detection of implicit citations for sentiment detection. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse*, pages 18–26, Jeju Island, Korea, July. Association for Computational Linguistics.

Baig, Y. M., Oesterling, A. X., Xin, R., Yu, H., Ghosal, A., Semenova, L., and Rudin, C. (2021). Multitask learning for citation purpose classification. In *Proceedings of the Second Workshop on Scholarly Document Processing*, pages 134–139, Online, June. Association for Computational Linguistics.

Beltagy, I., Lo, K., and Cohan, A. (2019). SciBERT: A pretrained language model for scientific text. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3615–3620, Hong Kong, China, November. Association for Computational Linguistics.

Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., et al. (2009). Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11):1422–1423.

Cohan, A., Ammar, W., Van Zuylen, M., and Cady, F. (2019). Structural scaffolds for citation intent classification in scientific publications. *arXiv preprint arXiv:1904.01608*.

de Andrade, C. M. V. and Gonçalves, M. A. (2020). Combining representations for effective citation classification. In *Proceedings of the 8th International Workshop on Mining Scientific Publications*, pages 54–58, Wuhan, China, 05 August. Association for Computational Linguistics.

Ding, Y., Zhang, G., Chambers, T., Song, M., Wang, X., and Zhai, C. (2014). Content-based citation analysis: The next generation of citation analysis. *Journal of the Association for Information Science and Technology*, 65(9):1820–1833. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23256.

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(4060):471–479.

Jurgens, D., Kumar, S., Hoover, R., McFarland, D., and Jurafsky, D. (2018). Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics*, 6:391–406.

Knoth, P. and Zdrahal, Z. (2012). Core: three access levels to underpin open access. *D-Lib Magazine*, 18(11/12).

Kunnath, S. N., Pride, D., Gyawali, B., and Knoth, P. (2020). Overview of the 2020 WOSP 3C citation context classification task. In *Proceedings of the 8th International Workshop on Mining Scientific Publications*, pages 75–83, Wuhan, China, 05 August. Association for Computational Linguistics.

Kunnath, S. N., Herrmannova, D., Pride, D., and Knoth, P. (2021). A meta-analysis of semantic classification of citations. *Quantitative Science Studies*, pages 1–46, 12.

Lo, K., Wang, L. L., Neumann, M., Kinney, R., and Weld, D. (2020). S2ORC: The semantic scholar open research corpus. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4969–4983, Online, July. Association for Computational Linguistics.

Lopez, P. (2009). Grobid: Combining automatic bibliographic data recognition and term extraction for scholarship publications. In *International conference on theory and practice of digital libraries*, pages 473–474. Springer.

Maheshwari, H., Singh, B., and Varma, V. (2021). SciBERT sentence representation for citation context classification. In *Proceedings of the Second Workshop on Scholarly Document Processing*, pages 130–133, Online, June. Association for Computational Linguistics.

Mishra, S. and Mishra, S. (2020). Scubed at 3C task a - a simple baseline for citation context purpose classification. In *Proceedings of the 8th International Workshop on Mining Scientific Publications*, pages 59–64, Wuhan, China, 05 August. Association for Computational Linguistics.

N. Kunnath, S., Pride, D., Herrmannova, D., and Knoth, P. (2021). Overview of the 2021 SDP 3C citation context classification shared task. In *Proceedings of the Second Workshop on Scholarly Document Processing*, pages 150–158, Online, June. Association for Computational Linguistics.

Nicholson, J. M., Mordaunt, M., Lopez, P., Uppala, A., Rosati, D., Rodrigues, N. P., Grabitz, P., and Rife, S. C. (2021). scite: A smart citation index that displays the context of citations and classifies their intent using deep learning. *Quantitative Science Studies*, 2(3):882–898, 11.

Pride, D. and Knoth, P. (2020). An authoritative approach to citation classification. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, pages 337–340.

Pride, D., Knoth, P., and Harag, J. (2019). Act: an annotation platform for citation typing at scale. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 329–330. IEEE.

Radoulov, R. (2008). Exploring automatic citation classification. Master's thesis, University of Waterloo.

Shotton, D. (2010). Cito, the citation typing ontology. In *Journal of biomedical semantics*, volume 1, pages 1–18. Springer.

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

Valenzuela, M., Ha, V., and Etzioni, O. (2015). Identifying meaningful citations. In *Workshops at the twenty-ninth AAAI conference on artificial intelligence*.

Varanasi, K. K., Ghosal, T., Tiwary, P., and Singh, M. (2021). IITP-CUNI@3C: Supervised approaches for citation classification (task a) and citation significance detection (task B). In *Proceedings of the Second Workshop on Scholarly Document Processing*, pages 140–145, Online, June. Association for Computational Linguistics.

Wang, W., Wang, Z., Yu, T., Pak, C., and Yu, G. (2020). Research on citation mention times and contributions using a neural network. *Scientometrics*, 125(3):2383–2400, December.

Zhu, X., Turney, P., Lemire, D., and Vellino, A. (2015). Measuring academic influence: Not all citations are equal. *Journal of the Association for Information Science and Technology*, 66(2):408–427.