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[1] and Petr Knoth[1]
The Knowledge Media Institute[1], The Open University, UK, Iris.ai[2], Norway

Abstract

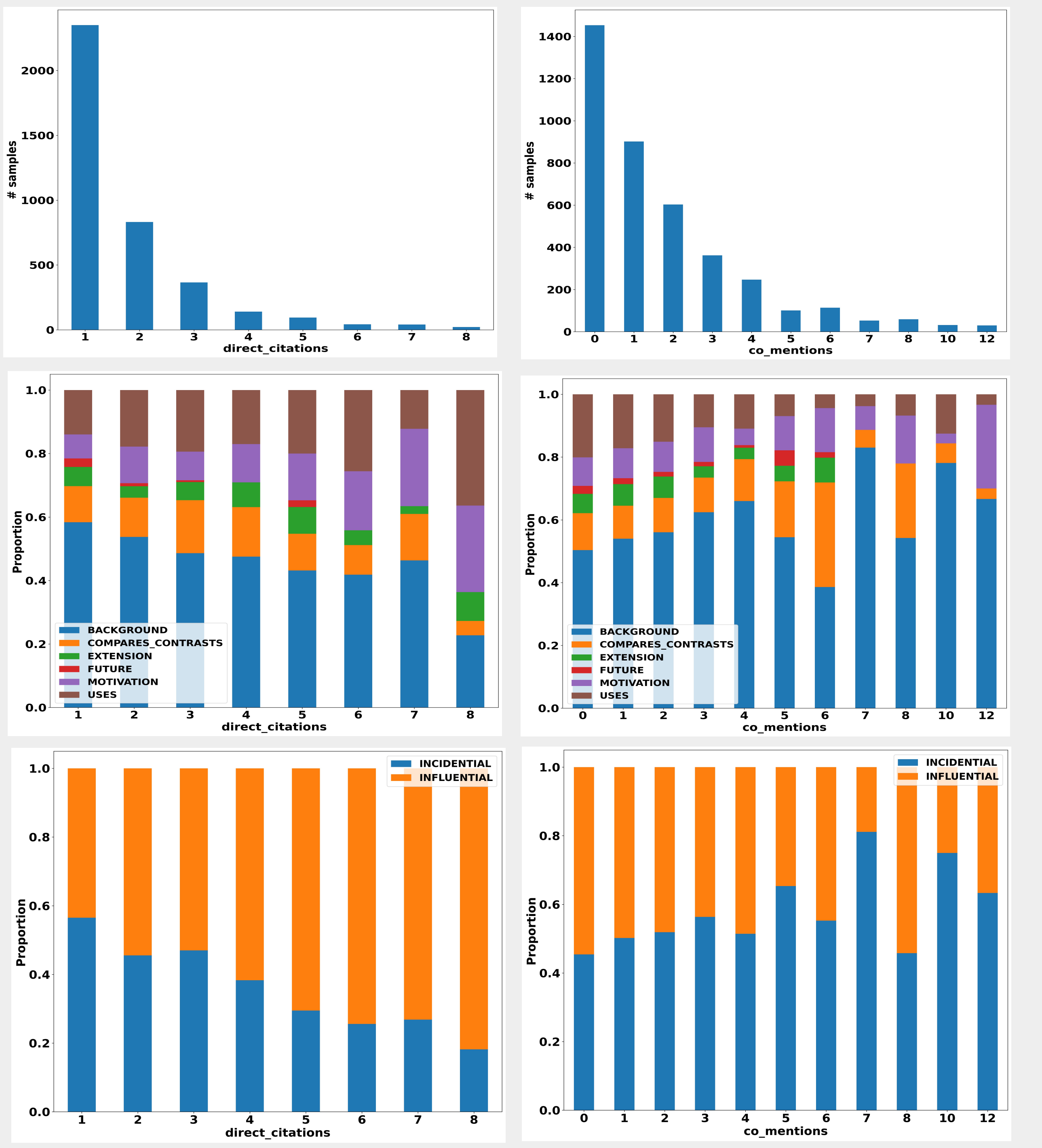
Classifying citations according to their purpose and importance is a challenging task that has gained considerable interest in recent years. This interest has been primarily driven by the need to create more transparent, efficient, merit-based reward systems in academia; a system that goes beyond simple 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**.

Motivation

- Existing datasets for citation classification are homogenous in nature and not feature enriched.
- ACT2 represents a multi-disciplinary, multi annotated corpus for citation classification. The dataset also includes 12 additional features extracted automatically, besides the existing 7 features.

Citation Functions	Examples
BACKGROUND	Most of the participatory models to design educational games are founded on educational theories and game design (see for example: Amory, 2007; #CITATION_TAG).
COMPARES_ CONTRASTS	The simplicity and validity of the numerator is perhaps misleading (#CITATION_TAG), especially in mental health.
EXTENSION	The items were derived from existing literature (#CITATION_TAG; Wagner and Schaltegger, 2004; Schoenherr, 2012; Zhang and Wang, 2014; Dubey et al. 2015).
FUTURE	We are thus exploring the option of using datasets such as CrossRef 12, Dimensions 13, OpenCitations [11], and Core [#CITATION_TAG].
MOTIVATION	To illustrate, consider the motivation given by #CITATION_TAG in developing their Bayesian account of word learning.
USES	For OTs I used the R package Oncotree [#CITATION_TAG] with its default settings.

Citation Importance	Examples
INCIDENTAL	The intervention was based on Body knowledging theory [20] [#CITATION_TAG].
INFLUENTIAL	In a related study, Mryglod et al. [#CITATION_TAG] used departmental h-index aggregation to predict REF rankings.

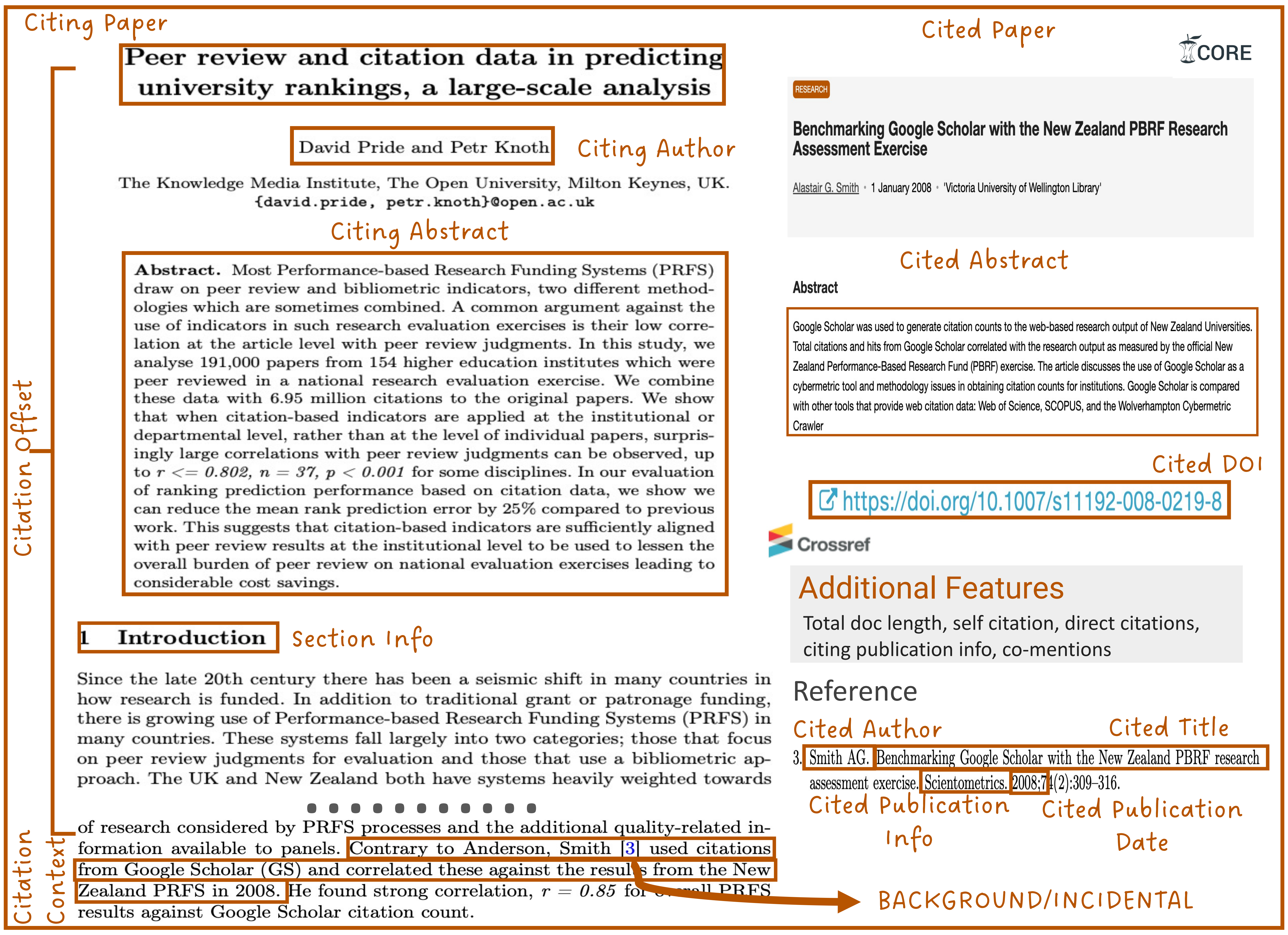


Distribution of (a) direct citations and (b) co-mentions with respect to citation function and importance classes

Conclusion

ACT2 – A new multi-annotated, multi-disciplinary, semi-structured feature-enriched open dataset.

Features



Peer review and citation data in predicting university rankings, a large-scale analysis

David Pride and Petr Knoth

The Knowledge Media Institute, The Open University, Milton Keynes, UK.
{david.pride, petr.knoth}@open.ac.uk

Abstract. Most Performance-based Research Funding Systems (PRFS) draw on peer review and bibliometric indicators, two different methodologies which are sometimes combined. A common argument against the use of indicators in such research evaluation exercises is their low correlation at the article level with peer review judgments. In this study, we analyse 191,000 papers from 154 higher education institutes which were peer reviewed in a national research evaluation exercise. We combine these data with 6.95 million citations to the original papers. We show that when citation-based indicators are applied at the institutional or departmental level, rather than at the level of individual papers, surprisingly large correlations with peer review judgments can be observed, up to $r <= 0.802$, $n = 37$, $p < 0.001$ for some disciplines. In our evaluation of ranking prediction performance based on citation data, we show we can reduce the mean rank prediction error by 25% compared to previous work. This suggests that citation-based indicators are sufficiently aligned with peer review results at the institutional level to be used to lessen the overall burden of peer review on national evaluation exercises leading to considerable cost savings.

1 Introduction

Since the late 20th century there has been a seismic shift in many countries in how research is funded. In addition to traditional grant or patronage funding, there is growing use of Performance-based Research Funding Systems (PRFS) in many countries. These systems fall largely into two categories; those that focus on peer review judgments for evaluation and those that use a bibliometric approach. The UK and New Zealand both have systems heavily weighted towards

● ● ● ● ● ● ● ● ● ● ●

of research considered by PRFS processes and the additional quality-related information available to panels. Contrary to Anderson, Smith [3] used citations from Google Scholar (GS) and correlated these against the results from the New Zealand PRFS in 2008. He found strong correlation, $r = 0.85$ for overall PRFS results against Google Scholar citation count.

RESEARCH

Benchmarking Google Scholar with the New Zealand PBRF Research Assessment Exercise

Alastair G. Smith • 1 January 2008 • 'Victoria University of Wellington Library'

Abstract

Google Scholar was used to generate citation counts to the web-based research output of New Zealand Universities. Total citations and hits from Google Scholar correlated with the research output as measured by the official New Zealand Performance-Based Research Fund (PBRF) exercise. The article discusses the use of Google Scholar as a cybermetric tool and methodology issues in obtaining citation counts for institutions. Google Scholar is compared with other tools that provide web citation data: Web of Science, SCOPUS, and the Wolverhampton Cybermetric Crawler

https://doi.org/10.1007/s11192-008-0219-8

Crossref

Additional Features

Total doc length, self citation, direct citations, citing publication info, co-mentions

Reference

3. Smith AG. Benchmarking Google Scholar with the New Zealand PBRF research assessment exercise. Scientometrics. 2008;74(2):309–316.

Comparison With Existing Datasets

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

Acknowledgement

This paper was supported by funding from NRC, Project ID: 309594, the AI Chemist under the coop- eration 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 Scientomet- rics PhD Studentship covering the contirbutions of Suchetha N. Kunnath and Petr Knoth.

References

[1] Simone Teufel, Advaith Siddharthan, and Dan Tidhar. 2006. Automatic classification of citation function. In Proceedings of the 2006 conference on empirical methods in natural language processing. Association for Computational Linguistics, 103–110.
[2] David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. 2018. Measuring the Evolution of a Scientific Field through Citation Frames. Transactions of the Association for Computational Linguistics 6 (2018), 391–406.
[3] Arman Cohan, Waleed Ammar, Madeleine van Zuylen, and Field Cady. 2019. Structural Scaffolds for Citation Intent Classification in Scientific Publications. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 3586–3596, Minneapolis, Minnesota. Association for Computational Linguistics.
[4] Pride, David, and Petr Knoth. "An authoritative approach to citation classification." Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020. 2020.
[5] Kunnath, Suchetha N., et al. "Overview of the 2021 SDP 3C citation context classification shared task." Association for Computational Linguistics, 2021.