

学術論文において研究データを参照する論文引用の識別

茂木光志¹ 伊藤滉一朗¹ 松原茂樹^{1,2}

¹ 名古屋大学大学院情報学研究科 ² 名古屋大学情報基盤センター

motegi.koshi.h9@es.mail.nagoya-u.ac.jp

{ito.koichiro.z5, matsubara.shigeki.z8}@f.mail.nagoya-u.ac.jp

概要

オープンサイエンスの一環である研究データの公開促進のために、被引用数などに基づき研究データを評価できる仕組みを整えることが考えられる。学術論文では、研究データの引用の代替として、その作成や公開を報じる論文が引用されることがある。本論文では、研究データの被引用数の算出に向けて、研究データを参照する論文引用を識別することの実現性について考察する。LLMを用いた識別実験を行い、その性能を評価した。

1 はじめに

近年、オープンサイエンスが世界的に推進されており [1, 2], 研究成果の公開が推奨されている。研究成果としては、論文のほか、データセットやツールなどの研究データが対象となる。このうち、論文についてはオープンアクセスの取り組みが進行しているのに対して、研究データのオープン化は十分に進んでいない。その要因の一つとして、論文では被引用数が代表的な評価指標として用いられ、その向上に研究者の関心が注がれているのに対し、研究データについては、そのような取り組みが少ないことが挙げられる。

この現状を打開するために、研究データについても、被引用数に基づき評価することが考えられる。しかし、学術論文における研究データの引用方法は、図 1 に示すように多岐にわたる。具体的には、

- **データ引用**：研究データに関する情報を参考文献リストに記載して、論文と同様に研究データを引用する方法
- **URL 引用**：研究データの公開先 URL を本文や脚注に記載する方法
- **論文引用**：研究データそのものを引用する代わりに、研究データの作成や公開を報じている論

(a) データ引用

In this study, we used Japanese Elder's Language Index Corpus (JELiCo) (Aramaki, 2016) as narrative data. This corpus includes speech data recorded from 30 elders.

References
Aramaki, Eiji. (2016). Japanese Elder's Language Index Corpus v2. <https://doi.org/10.6084/m9.figshare.2082706.v1>.

(b) URL 引用

...segmentation kit in Julius (Lee et al., 2001) to identify and Unidic response.

(Footnote)
¹<http://julius.osdn.jp/index.php?q=ouyoukit.html>
²<https://github.com/neologd/mecab-ipadic-neologd>
³<https://ja.osdn.net/projects/unidic/releases/58338>

(c) 論文引用

Here, we used MeCab (Kudo et al., 2004) for morphological analysis and the phone segmentation kit in Julius (Lee et al., 2001) to identify and Unidic response.

References
Kudo, T., Yamamoto, K., and Matsumoto, Y. (2004). Applying Conditional Random Fields to Japanese Morphological Analysis. In Proceedings of the 9th Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pages 230–237. Barcelona, Spain.

図 1: 研究データの引用方法 (文献 [5] から引用)

文を引用する方法

などがある。このうち、データ引用や URL 引用については、研究データの被引用数の集計に向けた取り組みが行われている [3, 4] 一方で、論文引用については、一般的には論文の被引用数の集計対象とされてきたこともあり、このような取り組みはほとんどない。

そこで本論文では、研究データの被引用数の算出に向けて、学術論文において間接的に研究データを参照する論文引用の識別可能性について考察する。具体的には、学術論文で行われる論文の引用を、その周辺テキストや書誌情報をもとに、研究データを参照する論文引用か否かの自動分類を試みる。大規模言語モデル (LLM) を用いた分類実験を実施し、その分類性能を評価した。

本論文の構成は以下の通りである。2 章では、研

究データの引用方法と被引用数の算出に関する取り組みの動向について述べる。3章では、本研究の問題設定を説明する。4章では、論文引用の分類可能性を実験により検証する。最後に、5章で本論文をまとめる。

2 研究データの引用と評価の動向

研究データの引用方法には、データ引用、URL引用、論文引用がある。本章では、研究データの引用方法の種別ごとに、被引用数の算出に関する既存の研究と取り組みを説明する。

2.1 データ引用

データ引用は、研究データの名称、作成者、識別子（DOI など）といった研究データの書誌情報を参考文献リストに記載して、論文と同様に研究データを引用する方法である（図 1(a) 参照）。この方法は、研究データの引用に関する共同宣言 [6] や Tromsø 勧告 [7] などで正式な引用方法として推奨されている。また、言語資源に関する国際会議 LREC¹⁾ や、研究データの作成と利用に関するオープンアクセスジャーナル Scientific Data²⁾ など、様々な学会や論文誌において、研究データを引用する方法として採用されている。

データ引用を対象に研究データの被引用数を算出する枠組みとして、Make Data Count³⁾、Data Citation Index [3]、EuropePMC API [4] などがあり、取得した被引用数に基づき、研究データとその作成者の評価が行われている。その具体的な取り組みとしては、OmicsDI Index [8] や Data-Index [9] がある。

2.2 URL 引用

URL 引用は、学術論文の本文や脚注に研究データの公開先 URL を記載することで、研究データを引用する方法である（図 1(b) 参照）。URL 引用による研究データの被引用数を算出するには、学術論文から研究データの公開先 URL を検出する必要がある。

学術論文中の URL について、正規表現や事前学習済み言語モデルを用いて抽出が行われている [10, 11, 12]。また、論文中で URL によって参照されたリソースの種類などの分類に取り組んだ研究が存在する [11, 13, 14]。

1) <https://lrec2026.info/authors-kit/>

2) <https://www.nature.com/sdata/submission-guidelines>

3) <https://makedatacount.org>

2.3 論文引用

論文引用による研究データの引用は、研究データそのものを引用する代わりに、研究データの作成や公開を報じている論文を引用する方法である（図 1(c) 参照）。一般に、論文引用は、その論文で提案された概念や手法などを参照するために行われるものであり、研究データを参照するものは必ずしも多くない。そのため、研究データの被引用数を正確に算出するには、学術論文中の論文引用から、研究データを参照する論文引用を識別する必要がある。

これまでに、研究データの作成や公開を報じている論文の引用回数を考慮した研究データの評価指標が提案されている [15]。ただし、研究データの作成や公開を報じている論文を PubMed のシソーラス MeSH (Medical Subject Headings)⁴⁾ の索引語を用いて特定しており、医学以外の分野での適用に課題がある。また、各論文引用が研究データを参照する論文引用であるか否かの識別は行われていない。

また、論文引用について、引用目的の自動分類に関する研究が行われている [16, 17, 18, 19, 20]。引用目的として、研究の背景を説明するための引用 (Background)、研究で利用した既存手法などを説明する際の引用 (Method) などを定義している。また、その定義に則ってコーパスを構築し、引用目的の自動分類を行っている。具体的には、SciBERT などの事前学習済み言語モデルを用いる手法 [21, 22] や、LLM を用いる手法 [23, 24] などが提案されている。しかし、これらの研究で定義している論文の引用目的の種別は、論文引用の参照先が研究データであるか否かに基づいたものではない。

3 問題設定

本論文では、学術論文における論文引用のうち、研究データを参照するものを識別することの実現可能性を検証する。具体的には、学術論文で行われる引用を、その周辺テキストや被引用論文の書誌情報を入力として、研究データを参照する引用 (Yes) か否か (No) を出力する二値分類問題として設定する。図 2 に、問題設定の概要を示す。

本研究で、学術活動で作成、生成、収集された研究成果物のうち、論文以外でかつ有形のものを研究データとする。データセットやソフトウェア、ツール、機械学習モデルなどが含まれる。一方、学術活

4) <https://www.nlm.nih.gov/mesh/meshhome.html>

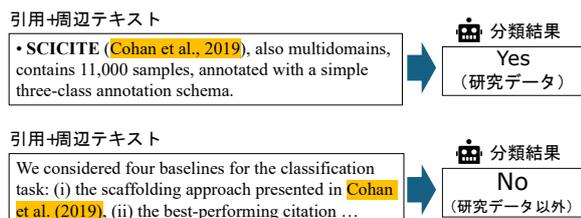


図 2: 問題設定の概要 (事例は文献 [23] から引用)

動において考案されたアルゴリズムや手法、機械学習のアーキテクチャ、タスク設定など、実体を伴わない無形の研究成果物は、研究データには含まないものとする。

本研究では、論文引用の周辺テキストにおいて、被引用論文の研究で作成された研究データについて言及しているものを、研究データを参照する論文引用と判定する。また、研究データの名称が明示的に記述されていない場合であっても、その引用対象が研究データを参照することが明らかであれば、研究データを参照する論文引用とみなす。

図 2 上の事例では、SCICITE データセットを参照するために論文引用が行われており、研究データを参照する引用であるといえる。図 2 下の事例では、手法 (the scaffolding approach) を参照するために論文引用が行われており、研究データを参照する引用とはいえない。

4 実験

研究データを参照する論文引用の識別可能性を検証するために、分類実験を行った。

4.1 実験データ

実験では、GSAP-NER データセット [12] を用いた。GSAP-NER データセットには、機械学習モデルあるいはデータセットを中心とする 10 種類のエンティティタイプについて、機械学習分野の学術論文 100 本を対象に、各エンティティの出現箇所がスパンとして付与されている。この中には、引用タグの出現位置に対するアノテーションが含まれる。

実験では、GSAP-NER データセットに含まれる論文 10 本を利用した⁵⁾。まず、GSAP-NER データセットのアノテーションをもとに、各論文に含まれる引用タグを含む段落を取得した。次に、各論文に含まれる引用タグについて、著者名などの情報をもとに、引用タグを各被引用論文の書誌情報と対応づけ

5) Otto ら [12] の実験において、交差検証を行うために 10 分割された論文集合の 1 つを利用した。

た。この後、著者の 1 人が、引用タグ、引用タグを含む段落、および書誌情報を参照し、論文引用が研究データを参照するか否かに分類した。その結果、論文 10 本に含まれる引用タグ 783 個のうち、191 個 (24.4%) が研究データを参照する引用であると判断された。

4.2 実験設定

実験では、指示チューニング済み LLM として Llama3.1 8B instruct [25], Ministral-3-{3B, 8B, 14B}-Instruct-2512 [26], GPT-4.1 シリーズ (GPT-4.1-nano, GPT-4.1-mini, GPT-4.1) [27], GPT-5 シリーズ (GPT-5-nano, GPT-5-mini, GPT-5) [28], GPT-5.2 [29] を使用した。GPT 以外のモデルでは vLLM⁶⁾ [30] を用いて実装を行った。構造化出力を利用し、各モデルの出力を Yes (研究データを参照する引用) と No (それ以外) の二値に限定した。また、温度パラメータ (temperature) を 0.0 に設定し、その他の設定ではデフォルト値を使用した⁷⁾。LLM への入力として、引用タグと引用タグを含む論文テキスト 1 段落の 2 つを用いる条件 (Cite + Con) と、それらに被引用論文の書誌情報を加えた条件 (Cite + Con + Bib) の 2 条件で実験を行った。また、比較対象として、Yes と No に等確率で分類を行う手法 (Random) を設けた。

評価指標には正解率 (Acc.), 適合率 (Prec.), 再現率 (Rec.), F 値 (F1) を利用した。

4.3 実験結果

各モデルの分類性能を表 1 に示す。Con + Cite において、GPT-5.2 の正解率と適合率が最も高く、その値はそれぞれ 0.891 と 0.886 であった。また、再現率と F 値は GPT-5-mini が最も高く、その値はそれぞれ 0.819 と 0.761 であった。

同一シリーズ内でのモデル性能に着目すると、パラメータ数が多いモデルほど性能が高い傾向にあった。中でも、GPT シリーズは全体的に高い値を記録しており、GPT-4.1-nano 以外のモデルは、0.7 程度の F 値を達成していた。以上の結果から、モデルによる性能差はあるものの、GPT シリーズの比較的パラメータ数が多いモデルを用いることで、研究データを参照する論文引用を高水準で識別できることを確認した。

6) <https://github.com/vllm-project/vllm>

7) Ministral については、vLLM で Mistral 社が開発したモデルに対して推奨されている設定を使用した。また、他モデルと演算精度を統一するため、演算精度を bfloat16 に統一した。

表 1: 研究データを参照する論文引用の分類性能

Model	Con + Cite				Con + Cite + Bib			
	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
Random	0.500	0.247	0.499	0.330	0.500	0.247	0.499	0.330
Llama3.1 8B	0.785	0.557	0.637	0.594	0.711	0.444	0.684	0.539
Ministral3 3B	0.799	0.821	0.238	0.369	0.780	0.920	0.119	0.211
Ministral3 8B	0.815	0.610	0.689	0.647	0.806	0.611	0.585	0.598
Ministral3 14B	0.839	0.694	0.622	0.656	0.826	0.665	0.596	0.628
GPT-4.1-nano	0.757	0.538	0.109	0.181	0.745	0.434	0.119	0.187
GPT-4.1-mini	0.854	0.661	0.767	0.710	0.844	0.673	0.715	0.693
GPT-4.1	0.880	0.777	0.720	0.747	0.888	0.797	0.731	0.762
GPT-5-nano	0.861	0.698	0.767	0.731	0.854	0.682	0.767	0.721
GPT-5-mini	0.874	0.712	0.819	0.761	0.874	0.724	0.788	0.754
GPT-5	0.885	0.841	0.658	0.738	0.881	0.813	0.674	0.737
GPT-5.2	0.891	0.886	0.642	0.745	0.884	0.892	0.601	0.718

Appendix Table 6 shows that our results on the validation set also transfer to the test set of WikiText-103. Currently, almost all models that present results on WikiText-103 use sliding window evaluation (defined in §B) to compute perplexities. We apply that method to our (and to the sinusoidal, rotary and T5 bias) models in Appendix Table 7. We find that our L = 3072 model surpasses the performance of Transformer-XL (Dai et al., 2019), the Sandwich (Press et al., 2020), and Shortformer (Press et al., 2021) models. Our results are similar to the ones obtained with staged training (Press et al., 2021) but fall short of results obtained by Routing Transformer (Roy et al., 2020) and kNN-LM (Khandelwal et al., 2020). The methods used in those models are orthogonal to ours, and we hypothesize that combining them with ours might lead to even larger performance increases.

Compressing & Distributing Optimizer States While 16-bit Adam has been used in several publications, the stability of 16-bit Adam was first explicitly studied for a text-to-image generation model DALL-E (Ramesh et al., 2021). They show that a stable embedding layer, tensor-wise scaling constants for both Adam states, and multiple loss scaling blocks are critical to achieving stability during training. Our work reduces the memory footprint of Adam further, from 16 to 8-bit. In addition, we achieve stability by developing new training procedures and non-linear quantization, both of which complement previous developments.

図 3: GPT-4.1 (Con+Cite+Bib) の分類の失敗例

また、書誌情報の有無による各モデルの性能を比較したところ、GPT-4.1-nano と GPT-4.1 では F 値が向上したものの、その他のモデルでは F 値が低下しており、書誌情報の利用の効果は認められなかった。

4.4 考察

最高の F 値を記録した Con + Cite + Bib における GPT-4.1 を対象に、分類におけるエラー分析を行う。GPT-4.1 は高い F 値を記録したものの、分類を誤る事例がいくつか存在した。代表的な失敗例を図 3 に示す。これらの事例は、それぞれ Transformer-XL と DALL-E という機械学習モデルを参照する引用であり、研究データを参照する引用 (Yes) と回答すべき事例であるものの、実際には No と出力していた。このように、GPT-4.1 は、機械学習モデルを参照す

る引用を正しく識別できない事例が相対的に多い傾向にあった。このような誤りを抑制するためには、引用タグの周辺に出現した機械学習モデルの名称の提示など、研究データに関する情報を追加することが有効であると考えられる。

5 おわりに

本論文では、研究データの被引用数の算出の自動化に向けて、学術論文から研究データを参照する論文引用の識別可能性を検証した。検証のために、論文内の引用を研究データを参照する引用か否かに分類する実験を行った。実験では、論文の引用の周辺テキストや被引用論文の書誌情報を入力として LLM に与え、その分類性能を評価した。実験を通して、LLM を用いることにより、研究データを参照する論文引用を高水準で識別できることを確認した。

論文においては、被引用論文が他研究に与える影響が一定でないことを踏まえ、引用の目的や重要度などを細分類することで、引用に重みづけを行う評価方法が考えられている。今後は、こうした取り組みを踏まえ、研究データを対象とした引用に対する重みづけ評価の実現に向けて、研究データの種別や使用用途の分類について検討したい。

謝辞

本研究は、JSPS 科研費 JP23K18506 の助成、文部科学省「AI 等の活用を推進する研究データエコシステム構築事業」の支援を受けたものです。

参考文献

- [1] UNESCO. Open science. <https://www.unesco.org/en/open-science> (Last Accessed: 2025-12-22).
- [2] Annex 1: G7 open science working group (OSWG). https://www8.cao.go.jp/cstp/kokusaiteki/g7_2023/annex1_os.pdf (Last Accessed: 2025-12-22).
- [3] Clarivate. Recommended practices to promote scholarly data citation and tracking: The role of the data citation index. White paper, 2023. Version 4.
- [4] Summer Rosonovski, Maria Levchenko, Rajat Bhatnagar, Umamageswari Chandrasekaran, Lynne Faulk, Islam Hassan, et al. Europe PMC in 2023. *Nucleic Acids Research*, Vol. 52, No. D1, pp. D1668–D1676, 2023.
- [5] Koichiro Ito, Masaki Murata, Tomohiro Ohno, and Shigeki Matsubara. Construction of responsive utterance corpus for attentive listening response production. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pp. 7244–7252, 2022.
- [6] FORCE11 Data Citation Synthesis Group. Joint declaration of data citation principles, 2014.
- [7] Helene N. Andreassen, Andrea L. Berez-Kroeker, Lauren Collister, Philipp Conzett, Christopher Cox, Koenraad De Smedt, Bradley McDonnell, and Research Data Alliance Linguistic Data Interest Group. Tromsø recommendations for citation of research data in linguistics, 2019.
- [8] Yasset Perez-Riverol, Andrey Zorin, Gaurhari Dass, Manh-Tu Vu, Pan Xu, Mihai Glont, Juan Antonio Vizcaino, Andrew F. Jarnuczak, Robert Petryszak, Peipei Ping, and Henning Hermjakob. Quantifying the impact of public omics data. *Nature Communications*, Vol. 10, No. 1, p. 3512, 2019.
- [9] Amelia S. C. Hood and William J. Sutherland. The data-index: An author-level metric that values impactful data and incentivizes data sharing. *Ecology and Evolution*, Vol. 11, No. 21, pp. 14344–14350, 2021.
- [10] Yasunori Yamamoto and Toshihisa Takagi. OREFiL: An online resource finder for life sciences. *BMC Bioinformatics*, Vol. 8, p. 287, 2007.
- [11] Masaya Tsunokake and Shigeki Matsubara. Classification of URL citations in scholarly papers for promoting utilization of research artifacts. In *Proceedings of the 1st Workshop on Information Extraction from Scientific Publications*, pp. 8–19, 2022.
- [12] Wolfgang Otto, Matthäus Zloch, Lu Gan, Saurav Karmakar, and Stefan Dietze. GSAP-NER: A novel task, corpus, and baseline for scholarly entity extraction focused on machine learning models and datasets. In *Findings of the Association for Computational Linguistics*, pp. 8166–8176, 2023.
- [13] He Zhao, Zhunchen Luo, Chong Feng, Anqing Zheng, and Xiaopeng Liu. A context-based framework for modeling the role and function of on-line resource citations in scientific literature. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 5206–5215, 2019.
- [14] Kazuhiro Wada, Masaya Tsunokake, and Shigeki Matsubara. Classification of URL citations on scholarly papers using intermediate task training. *IEICE Transactions on Information and Systems*, Vol. E108-D, No. 10, pp. 1183–1193, 2025.
- [15] Alison Callahan, Rainer Winnenburger, and Nigam H. Shah. U-index, a dataset and an impact metric for informatics tools and databases. *Scientific Data*, Vol. 5, No. 1, p. 180043, 2018.
- [16] Simone Teufel, Advait Siddharthan, and Dan Tidhar. Automatic classification of citation function. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pp. 103–110, 2006.
- [17] Amjad Abu-Jbara, Jefferson Ezra, and Dragomir Radev. Purpose and polarity of citation: Towards NLP-based bibliometrics. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 596–606, 2013.
- [18] David Jurgens, Srijan Kumar, Raine Hoover, Dan McFarland, and Dan Jurafsky. Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics*, Vol. 6, pp. 391–406, 2018.
- [19] Arman Cohan, Waleed Ammar, Madeleine van Zuylen, and Field Cady. 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*, pp. 3586–3596, 2019.
- [20] Suchetha Nambanoor Kunnath, Valentin Stauber, Ronin Wu, David Pride, Viktor Botev, and Petr Knoth. ACT2: A multi-disciplinary semi-structured dataset for importance and purpose classification of citations. In *Proceedings of the 13th Language Resources and Evaluation Conference*, pp. 3398–3406, 2022.
- [21] Zeren Shui, Petros Karypis, Daniel S. Karls, Mingjian Wen, Saurav Manchanda, Ellad B. Tadmor, and George Karypis. Fine-tuning language models on multiple datasets for citation intention classification. In *Findings of the Association for Computational Linguistics*, pp. 16718–16732, 2024.
- [22] Lorenzo Paolini, Sahar Vahdati, Angelo Di Iorio, Robert Wardenga, Ivan Heibi, and Silvio Peroni. CiteFusion: An ensemble framework for citation intent classification harnessing dual-model binary couples and shap analyses. *Scientometrics*, Vol. 130, No. 11, pp. 5911–5981, 2025.
- [23] Lasse M. Jantsch, Dong-Jae Koh, Seonghwan Yoon, Jisu Lee, Anne Lauscher, and Young-Kyoon Suh. FineCite: A novel approach for fine-grained citation context analysis. In *Findings of the Association for Computational Linguistics*, pp. 24525–24542, 2025.
- [24] Paris Koloveas, Serafeim Chatzopoulos, Thanasis Vergoulis, and Christos Tryfonopoulos. Can LLMs predict citation intent? An experimental analysis of in-context learning and fine-tuning on open LLMs. In *Linking Theory and Practice of Digital Libraries: 29th International Conference on Theory and Practice of Digital Libraries*, pp. 207–224, 2025.
- [25] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, et al. The Llama 3 herd of models, 2024.
- [26] Mistral. Ministral3. https://huggingface.co/docs/transformers/model_doc/ministral3 (Last Accessed: 2025-12-14).
- [27] OpenAI. Introducing GPT-4.1 in the api. <https://openai.com/index/gpt-4-1/> (Last Accessed: 2025-12-14).
- [28] OpenAI. Introducing GPT-5. <https://openai.com/index/introducing-gpt-5/> (Last Accessed: 2025-12-22).
- [29] Introducing GPT-5.2. <https://openai.com/index/introducing-gpt-5-2/> (Last Accessed: 2025-12-22).
- [30] Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*, pp. 611–626, 2023.

A プロンプト

図 4 にシステムプロンプトを, 図 5 にユーザプロンプトをそれぞれ示す。

You are a helpful assistant.

Your task is to determine whether, in the given context, the citation is made for referring to research data.

Here, “research data” includes:

- datasets (raw or processed data)
- software tools or code (usable tools or libraries)
- trained models (released weights, checkpoints, or model files)

Important:

The decision must be based on the citation intent in the provided Context, not on what the cited work provides in general. A citation counts as referencing research data (answer “Yes”) only if:

- the Context indicates that the cited work’s dataset, tool, or trained model is being used, accessed, downloaded, analyzed, applied, or employed in the study; or the cited work’s research data is directly referenced as a resource.

A citation should be labeled “No” if:

- the Context cites the work only for its method, algorithm, model architecture, theoretical idea, concept, or background explanation;
- the Context does not indicate the use or reference of any dataset, tool, or trained model provided by the cited work;
- the citation intent is ambiguous or unclear.

Respond with exactly one word:

Yes

No

図 4: システムプロンプト

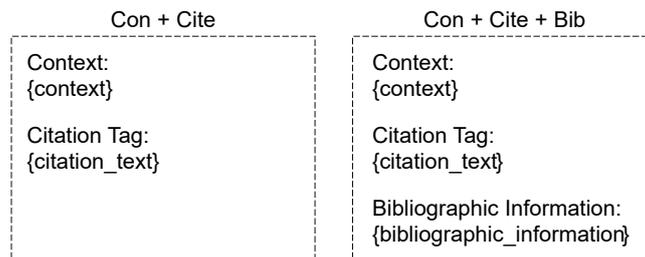


図 5: ユーザプロンプト