



HAL
open science

Toward identifying mis-citations in scientific papers

Qinyue Liu, Amira Barhoumi, Cyril Labbé

► **To cite this version:**

Qinyue Liu, Amira Barhoumi, Cyril Labbé. Toward identifying mis-citations in scientific papers. 2023.
hal-04462254

HAL Id: hal-04462254

<https://hal.science/hal-04462254v1>

Preprint submitted on 19 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Toward identifying mis-citations in scientific papers

Qinyue Liu¹, Amira Barhoumi¹, and Cyril Labbé¹

¹Univ. Grenoble Alpes, LIG, CNRS, Grenoble INP, 38000 Grenoble, France

Objective. Inaccurate citations in scientific publications are detrimental to science and are difficult to detect. We present various methods to automatically evaluate citation reliability by measuring the correlation between citation context and the cited paper. We report the first results of automatically detecting mis-citations in scientific literature.

Method. We hypothesized that a reliable citation context would be “semantically” close to a textual sequence in the cited work. We restrict our study to cited paper’s abstract considered either as one piece or as a set of sentences. Using language models, semantic similarity is measured by a cosine-based similarity or a paraphrase classifier-based method. We collect a dataset of mis-citation contexts to evaluate the efficiency of these methods.

Results. The collection of mis-citations is divided into 2 categories: reliable citations and mis-citations. Among the mis-citations, we identified different levels of inaccuracy and thus divided them into two subcategories: in-domain mis-citations and out-domain mis-citations. The latter when the cited paper is from a completely different topic, and nothing appears relevant to justify the citation. The former captures more subtle cases like overgeneralization, mis-understanding, errors in numbers... We annotated 64 citation contexts manually: 31 are annotated as reliable (thus in-domain) and 33 as mis-citation. Among these mis-citations, 21 are in-domain and 12 are out-domain. The cosine similarity method considering the abstract as one piece correctly predicted 39 out of 64 citations (61%), based on segmented abstracts, it correctly predicted 43 citations (67.1%). The paraphrase classifier method considering the abstract as one piece correctly predicted 39 citations. The best result comes from the paraphrase classifier method using segmented abstracts achieving a rate of 75% citations correctly predicted. For this configuration, further analysis on subcategories of mis-citations reveals an improved prediction of in-domain citations: 37 were correctly predicted as in-domain compared to 28 when using the abstract as one piece.

Conclusion. We investigate automatic detection of citation accuracy. We built a dataset of citation context with different kinds of mis-citation. We tested two methods based on text vector representations and suggest that large language models are promising to address this task.

Acknowledgements. CL acknowledge the [NanoBubbles](#) project that has received Synergy grant funding from the European Research Council (ERC), within the European Union’s Horizon 2020 program, grant agreement no. [951393](#).