

Tailoring Scientific Argument Mining for Scientific Literature Correction

Yagmur Ozturk, Cyril Labbé, François Portet

▶ **To cite this version:**

Yagmur Ozturk, Cyril Labbé, François Portet. Tailoring Scientific Argument Mining for Scientific Literature Correction. Journée Natural Language Argumentation – GDR TAL – GDR RADIA, Nov 2023, Valbonne, France. hal-04346693

HAL Id: hal-04346693
https://hal.science/hal-04346693v1

Submitted on 15 Dec 2023

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.



HAL Authorization

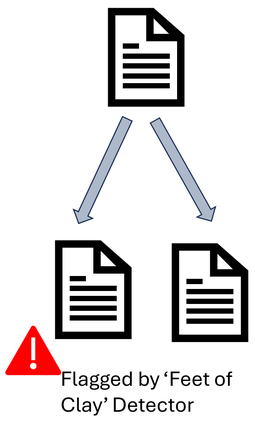
Tailoring Scientific Argument Mining for Scientific Literature Correction

Yagmur Ozturk
Cyril Labbé, François Portet
Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000, Grenoble, France

What is scientific literature correction?

PRACTICES
- Retraction of articles
- Post publication peer review





PLATFORMS
- Problematic paper screener [1]

PUBPEER
The online journal club

AIM: Providing a tool that helps with the correction process

Can **SAM** be used to assess if retracted citations **weaken the validity of claims and results** in a paper?

Scientific Argument Mining (SAM)

Extraction of the argument structure of scientific articles

- **Argumentative Units**: Labels to identify and classify parts of argument structure **(BACKGROUND/AIM…)**
- **Argumentative Relations**: Labels to classify relations between units **(SUPPORT/ELABORATION…)**
- **Citation Functions**: Purpose of a citation, not always included in SAM frameworks

Example: Argument structure with SAM

Unit Type:
[1]BACKGROUND
[2]BACKGROUND

RELATED WORK
Brown et. al. [1] mentions the usage of … in their work. Smith et al. [2] demonstrated that endosomal escape is possible under the conditions…first ones to prove that…

Relation Type:
[2]ELABORATION

[2] USE/BASIS

METHODOLOGY
We used the same conditions as Smith et.al. in our experiments…and explore extensions on their baseline…

[2] BY-MEANS

[2] OWN RESULT

RESULTS
We prove that building on [2], we are able to explore the application of endosomal escape towards…

When Smith and Brown are retracted

Unit Type:
[1]BACKGROUND
[2]BACKGROUND

RELATED WORK
Brown et. al. [1] mentions the usage of … in their work. Smith et al. [2] demonstrated that endosomal escape is possible under the conditions…first ones to prove that…

1
2

Relation Type:
[2]ELABORATION

[2] USE/BASIS

METHODOLOGY
We used the same conditions as Smith et.al. in our experiments…and explore extensions on their baseline…

2

[2] BY-MEANS

[2] OWN RESULT

RESULTS
We prove that building on [2], we are able to explore the application of endosomal escape towards…

2

Related Works on SAM

Author	Domain	Corpus Size	Unit Spans	Relation Labels
Teufel [2]	Comp. Ling.	80 articles	Sentence	✗
Teufel [3]	Chem	50 articles	Sentence	✗
Lauscher [4]	Comp. Imaging	40 articles	Token-level	In corpus but not automatized
Binder [5]	Comp.Imaging	39 articles	Mixed	✓

Why tailoring of these works is necessary for our task
- Scarce gold data
- Fine grained annotations [2,3] : complex human annotation tasks
- Coarse-grained annotations [4,5] **:** Not enough to reflect the argumentative complexity
- Token level: span detection of argument units is weak
- Sentence level: Usually one sentence IS NOT one argument unit

Applying SAM to a new domain

- **New dataset** annotated for argumentation using guidelines of [2,3] which overlaps with [4]
- Applying SAM to **nanobiology**
- Applying SAM to scientific literature correction practices
- **Evaluating** automatic SAM system of [5] on new data
- Leveraging from citation function systems

This work is funded by The NanoBubbles [6] project that has received Synergy grant funding from the European Research Council (ERC), within the European Union's Horizon 2020 programme, grant agreement number 951393.

References

[1] G. Cabanac, C. Labbé, and A. Magazinov, 'The 'Problematic Paper Screener' automatically selects suspect publications for post-publication (re)assessment', in *7th World Conference on Research Integrity (WCRI 2022)*, 2022
[2] S. Teufel, J. Carletta, and M. Moens, 'An annotation scheme for discourse-level argumentation in research articles', in *Ninth Conference of the European Chapter of the Association for Computational Linguistics*, 1999
[3] S. Teufel, A. Siddharthan, and C. Batchelor, 'Towards Domain-Independent Argumentative Zoning: Evidence from Chemistry and Computational Linguistics', in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 2009
[4] A. Lauscher, G. Glavaš, and K. Eckert, 'ArguminSci: A Tool for Analyzing Argumentation and Rhetorical Aspects in Scientific Writing', in *Proceedings of the 5th Workshop on Argument Mining*, 2018
[5] A. Binder, L. Hennig, and B. Verma, 'Full-Text Argumentation Mining on Scientific Publications', in *Proceedings of the first Workshop on Information Extraction from Scientific Publications*, 2022
[6] F. Bordignon *et al.*, 'Nano bubbles: how, when and why does science fail to correct itself?' 2023.

European Research Council
Established by the European Commission

NANO BUBBLES