Bad smells in reviewers' reports? Text-mining the MDPI Open Peer Review Corpus

Gilles Hubert, Guillaume Cabanac, Cyril Labbé

Bad smells in reviewers' reports? Text-mining the MDPI Open Peer Review Corpus

Gilles HUBERT,[1] Guillaume CABANAC[1, 2] and Cyril LABBÉ[3]

[1] Université Toulouse 3 – Paul Sabatier, IRIT UMR 5505 CNRS, Toulouse, France

[2] Institut universitaire de France (IUF), Paris, France

[3] Univ. Grenoble Alpes, CNRS, Grenoble INP, LIG, 38000 Grenoble, France

Objective. Malpractice affecting the reviewing process is detrimental to science. We introduce methods to reveal evidence of peer review manipulation, such as template usage, citation manipulations or botched and meaningless reviewer reports. We apply and evaluate these methods on a corpus of reports.

Method. We downloaded the "MDPI Open Peer Review Corpus 2" webscraped by Miłkowski et al. (2023, https://doi.org/10.18150/shkp7b). We focused on 'Round 1' reports, retaining those in plain text (excluding those uploaded as attached files). We computed statistics on report length, identified references suggested by reviewers with regular expressions, extracted frequent word sequences, and analysed the pairs of reports showing an inter-textual similarity higher than 90%.

Results. The MDPI corpus consists of reports for 135,653 accepted articles in MDPI journals from 2011 to 2022. We mined this 170 GB dataset and observed several shortcomings. Some reports appear to be truncated due to failed webscraping. The dataset we analysed contains 135,437 articles and their 339,387 associated reports. The average report has 270 words with a median length of 202. Microscopic reviews consist of one word only, such as 'accept', 'none', 'Nil' or 'N.A.' (n = 230). These seem to be reports (mis)presented by the publisher as 'Round 1' albeit resulting from a peer review ran for an undocumented earlier submission ('reject and resubmit' editorial act). Tiny reports of less than 20 words account for 3.5% of the dataset. We also searched for report templates being constantly reused. Report–report similarities show that 40 reports were almost identical. At least 10 articles share two identical reports. Large chunks of text were reused across unrelated 380 reports sharing at least 300 words in chunks of 10+ words. We also spotted potential coercive citations: some reports contain the same wording suggesting authors to add a set of DOIs/PMIDs.

Conclusion. We found very few evidence of questionable features in the MDPI reports. More research is needed to improve malpractice detection and assess its prevalence in (open) peer review reports. Once implemented into publishers' workflows, early misconduct detection should help to prevent botched reports, template reuse, and coercive citations.

Acknowledgements. GC and 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.