

## Correspondence

# On improving the implementation of automatic updating of systematic reviews

Anna Koroleva ,<sup>1,2,\*</sup> Camila Olarte Parra,<sup>2,3</sup> and Patrick Paroubek<sup>1</sup>

<sup>1</sup>LIMSI, CNRS, Université Paris-Saclay, Orsay, France, <sup>2</sup>Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands and <sup>3</sup>Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

\*Corresponding Author: Anna Koroleva, MSc, 44 Avenue Victor Hugo, 91440 Bures-sur-Yvette, France; aakorolyova@gmail.com; koroleva@limsi.fr

Received 9 July 2019; Editorial Decision 13 August 2019; Accepted 10 September 2019

To the Editor,

In their recent work, Martin et al.<sup>1</sup> proposed an open source online tool to help update systematic reviews. The authors use a combination of machine learning and crowd-sourcing approaches to propose and assess trials that might need to be included in the update. Bibliographic databases and the ClinicalTrials.gov registry are searched for new trials to complement the updates proposed by registered users.

We believe that this work provides a very useful tool for facilitating and automating some parts of the systematic review process, which are usually time-consuming. The proposed interface is clear, user-friendly, and easy to navigate; the authors make their data freely available for registered users, so that it can be re-used in future research.

We believe that this work is of high importance, and we would like to ask for some clarifications and to provide some suggestions on how to improve the system.

## MACHINE LEARNING

The authors employ a matrix factorization approach using a shared latent space to assess the relevance of trial registry entries for each systematic review. Matrix factorization is a well-established method used in recommender systems. However, it is not commonly used for automating the screening stage of systematic reviews; in the authors' previous work, it did not outperform the baseline approach (cosine similarity) in terms of work saved over sampling at 95% recall.

There are alternative approaches that can be more suitable for this task and could show better performance, including those using word embeddings (eg,<sup>2,3</sup>). In particular, Hashimoto et al.<sup>3</sup> developed an approach using paragraph vectors to represent documents, as

described by Le and Mikolov<sup>4</sup> who proposed to map every paragraph and word to a unique vector and to further concatenate a paragraph vector with several word vectors from this paragraph to predict the next word. Paragraph and word vectors are trained using stochastic gradient descent and backpropagation. Hashimoto et al. cluster the obtained paragraph vectors by a k-means clustering algorithm to detect latent topics in the data. The final representation of documents is calculated as a k-dimensional feature vector containing distances of the given document to the k cluster centroids. The authors showed that this method outperformed the Latent Dirichlet Allocation (LDA), used by Martin et al.<sup>1</sup>

## CROWD-SOURCING

Crowd-sourcing can be an efficient way of collecting human annotators' input for a particular task, but it has its drawbacks.

First of all, we would like to clarify how the quality of the contributions is controlled. The system allows only registered users to contribute, to avoid "noise" and random votes. To register, a user needs to answer some questions. However, the details on these questions and their impact on the quality control are not revealed, neither do we know whether any further requirements to qualify as a user exist. We suggest making the registration process more transparent.

One common way of crowd-sourcing quality control is to inject some gold standard data points at random intervals in the dataset and to check that they have been properly processed by the users (see eg, <https://www.ucomp.eu/data/sites/16/d2.2.pdf>). Of course the users do not know which of the data points are part of the gold standard dataset. This kind of quality control test could have been addressed by the authors.

Additionally, the authors do not mention if they have any system of profile management: eg, for reducing the weight of votes from a

user who always votes against the majority, or for taking into account the relative experience of each user. Simple majority vote is not always the best choice to resolve disagreements, especially for a task like systematic review, where one expert vote can be more valuable than a number of nonexpert votes. Also, it is not clear if there are any strategies to tackle possible conflict of interest from the voters.

Another point to raise is the difficulty of recruiting contributors, as the authors state in their Limitations section. This is indeed a problem often undermining the development of crowd-sourcing based work, but there are successful crowd-sourcing projects in the biomedical domain, such as Cochrane Crowd (<http://crowd.cochrane.org/index.html>). A well-elaborated communication and dissemination plan could help tackle this issue.

## DATA

Systematic reviews are included in the system if they match two criteria: have the words “systematic review” or “meta-analysis” in the title and include at least one link to a trial in the ClinicalTrials.gov registry. The authors deliberately chose this “conservative” approach, but we would like to point out that there are alternative approaches showing operational performance, eg, the rule-based algorithm proposed by Sarker, and Diego Mollá-Aliod<sup>5</sup> is reported to achieve the recall of 0.99 and precision of 1 for meta-analysis and systematic reviews. Adopting a similar approach would widen the selection of included reviews.

We would also like to raise the question of including systematic reviews that are not in open access. The researchers working on them might be interested in using the proposed tool, but it was not very clear to us from the article whether it is possible to include such a review in the system and how it is managed.

Another questionable choice is including the trial registration entries from ClinicalTrials.gov only. There is a number of other trial registries (see eg, <https://www.hhs.gov/ohrp/international/clinical-trial-registries/index.html>), as well as the WHO portal which provides access to a few primary registries. Including only one registry apparently limits the included trials.

The obvious reason for using ClinicalTrials.gov is the fact that metadata of articles in PubMed can contain a direct link to it, while links to other registries are not included in the metadata. However, trial registration numbers are often cited in the abstract and follow a fixed pattern including the registry name and the unique registration ID, which can be easily found with the help of regular expressions and used to automatically find the registry entry on the webpage of the corresponding registry.

## CONCLUSION

We commend the authors for their work on developing an open source online system to facilitate updating of systematic reviews. With this letter, we would like to encourage further work on this promising initiative to improve the results.

## FUNDING

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 676207.

## AUTHOR CONTRIBUTIONS

AK and COP contributed to the conception of this work and wrote the first version of the manuscript. PP supervised the project and contributed to the final version of the manuscript.

## ACKNOWLEDGMENTS

This work is a result of a Journal Club organised within the Methods in Research on Research (MiRoR) project. Special thanks to Christopher Norman, Melissa Sharp and David Blanco for their comments and suggestions. We would also like to thank the members of the Methods in Research on Research (MiRoR) Project (<http://mirror-ejd.eu/>) and Marie Skłodowska-Curie Actions for their support.

## CONFLICT OF INTEREST

The authors declare no conflicts of interest.

## REFERENCES

1. Martin P, Surian D, Bashir R, *et al.* Trial2rev: Combining machine learning and crowd-sourcing to create a shared space for updating systematic reviews. *JAMIA Open* 2019; 2 (1): 15–22.
2. Lerner I, Créquit P, Ravaud P, *et al.* Automatic screening using word embeddings achieved high sensitivity and workload reduction for updating living network meta-analyses. *J Clin Epidemiol* 2019; 108: 86–94.
3. Hashimoto K, Kontonatsios G, Miwa M, *et al.* Topic detection using paragraph vectors to support active learning in systematic reviews. *J Biomed Inform* 2016; 62: 59–65.
4. Le Q, Mikolov T. Distributed representations of sentences and documents. In: Proceedings of the 31st International Conference on Machine Learning (ICML-14); 2014: 1188–96; Beijing, China.
5. Sarker A, Mollá-Aliod D. A Rule-based approach for automatic identification of publication types of medical papers. Proceedings of the ADCS Annual Symposium; 2010; Melbourne, Australia.