Reviewer's report

Title: Machine learning algorithms for systematic reviews: reducing workload in a preclinical review of animal studies and reducing human screening error

Version: 1 Date: 25 Oct 2018

Reviewer: Gabriela Nunez-Mir

Reviewer's report:

In this study, the authors test the performance of two ML approaches in selecting classifications for inclusion in a systematic literature review. This study is timely, interesting and necessary, as it attempts to resolve two important issues in literature reviews—the massive amount of time and human resources needed to sift through the ever-growing corpora of literature and the unavoidable human bias and error. Although the study itself appears well-planned and nicely executed, the manuscript has a few shortcomings despite having gone through a round of revisions. Addressing these will greatly elevate the manuscript by making it much more attractive to a broader audience of researchers.

First, there are instances in the manuscript where the writing feels a little alienating to the reader. As mentioned by previous reviewers, abbreviations for institutions and tools are used throughout the manuscript. These have been spelled out in the new version, but as someone unfamiliar with these institutions, I fail to see the purpose of mentioning them so often, especially since this paper has the potential to be useful to a very broad audience. In a similar vein, a tool is mentioned towards the end of the manuscript—the Systematic Review Facility tool (SyRF). It is unclear if this tool was used to perform the feature generation and classifying in the study. If this is the case, and if the authors want to encourage the use of this tool, SyRF should be introduced and described much earlier in the manuscript.

Second, the manuscript addresses two separate issues—using ML approaches to select citations and using ML to identify and correct human error in selecting citations. I believe both issues are worthwhile research topics; however, it may be too daunting of a task to address both in the same paper. I believe this is the reason why previous reviewers expressed confusion regarding the study's methodology. In the Methods section, the authors separate both goals into two "Steps." At the beginning of Step 1, the authors dedicate a couple of manuscript pages to describe different feature generation and classifier methods. These sections seem out of place in the current format of the manuscript, but they are not without merit. In my opinion, this manuscript would benefit from being separated into two manuscripts. The first manuscript would explore different feature generation methods and classifiers, and then test how different combinations of these perform (in the current manuscript, the authors do not clearly explain how they arrived at the feature generation/classifier combinations they used as ML approaches). The second manuscript would explore the issue of using ML approaches to detect and correct human error (this part is underdeveloped in the current manuscript). If the authors choose to keep the
manuscript as one, I would suggest paring down the "Feature Generation" and "Classifiers" sections and using the text in these sections to describe the two approaches in the "Approaches" section.

Specific comments:

L24-26: Run-on sentence. Please divide in two.

L39: Switch "assigned" and "the."

L57: Add comma after "intensive."

L99: The goal as stated in this sentence does not accurately describe the purpose of the study. I don't believe the authors identified the amount of training data needed to use ML algorithms effectively.

L107: With "systematic error" do you mean human biases/error?

L113-114: "as part of a preclinical systematic review framework at the classification stage" this phrase is confusing. You may simply say "as part of the citation classification stage in a preclinical systematic review."

L119: You mean screening of a large body of literature FOR systematic review. The review itself wasn't screened.

L142: Period after EMBASE.

L169: This sentence is vague and unclear.

L312: Five consecutive records seem very low.

L365: How is human error determined? This is an important question the authors need to answer. Once incongruences are found in human and machine decisions, how did the authors confirm that the human had made an error? Was an additional reviewer brought in to inspect the record?

L394: What is "This"?

L473: Period after "important."

Level of interest
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests
Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organisation that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal

Were you mentored through this peer review?

No