Reviewer’s report

Title: Constructing fine-grained entity recognition corpora based on clinical records of traditional Chinese medicine

Version: 0 Date: 07 Mar 2020

Reviewer: Lei Wang

Reviewer's report:

I think the new version has been improved. However, the following are my concerns:

1. The quality of written English is still a big issue. Here are some examples of grammar issues:

   - The misuse of articles
   In this manuscript, articles were misused many times. (e.g., Abstract: Line 39: "through sample annotation method;")

   - Redundant
   Page 8, Line 22 -23: "They first created build the draft guideline first,"

   - Missing words
   Page 2, Line 44 -45 : "TCM records have contain a large amount clinical information"
   Page 3, Line 49 -50 : "the characteristics of common texts the Chinese language"

   - verb form is not correct
   Page 8, Line 15 - 17 : "laboratory test entities is distinguished."

   - incorrect punctuation
   Page 9, Line 36: "the concepts in TCM clinical notes, e.g."

   - incorrect preposition
   Page 13, Line 17 -18 :"are of highly important"

I only list only a few mistakes here. Please polish and revise your manuscript carefully!

2. Will your data be publicly available?

3. The authors did not response the data privacy issue well. Protected Health Information (PHI) is a big concern in biomedical informatics. PHI information is not limited to patients' name, gender, age. Thus, please clarify your methods on de-identifying PHI information in your clinical notes. HHS website will help (https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html )

4. In Table 4, one entity would be classified as distinct class (e.g., tongue). If this corpus is applied in the future automatic annotation, I guess the machine will not recognize some entities properly.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Unable to assess

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

Not relevant to this manuscript

Quality of written English
Please indicate the quality of language in the manuscript:

Not suitable for publication unless extensively edited

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 organization 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