Reviewer’s report

Title: Precursor-Induced Conditional Random Fields: Connecting Separate Entities by Induction for Improved Clinical Named Entity Recognition

Version: 0 Date: 03 Feb 2019

Reviewer: Tianyong Hao

Reviewer's report:

Focusing on NER from clinical text, this paper proposes a new type of CRF model called pi-CRF to improve its capability. The major methodology innovation lies on the additional labels of non-entities so as to build the sequence for entities and non-entities to express sentences. The non-entity labels are treated as an information transmission medium to deliver preceding entity information. The evaluation was based on three datasets, two in clinical and one in open domain. The results showed that the pi-CRF model improves conventional CRF model on clinical NER. The author also tried to reduce the computational complexity of pi-CRF in relation to the second-order CRF.

Overall, the paper is well written and organized. It has clear research meaning. However, there are some issues that should be addressed.

1) pi-CRF focuses on label sequences with the [entity, outside+, entity] pattern. The pattern is easily obtained based on labelling results using conventional CRF. The differences of utilizing the label sequences with pattern learning are not being clearly described.

2) In different types of clinical text, the expression styles are quite different. As a result, the "outside+" can be a punctuation or even a symbols. However, they should not affect the pattern or sequences. How this problem can be avoided or solved in pi-CRF?

3) In recent years, more and more deep learning models have been applied to clinical NER tasks and already achieved promising performance. I wonder that why the authors did not compare the pi-CRF with some existing NER methods, e.g., bi-LSTM-CRF.

4) The performance improvement compared with conventional one is limited. There are other variants of CRF models but lack of comparison.

5) Some important recent related work on medical NER is missing in references.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Unable to assess

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

Yes

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.

I am able to assess the statistics

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 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