Reviewer’s report

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

Version: 0 Date: 04 Mar 2019

Reviewer: Reviewer 2

Reviewer's report:

PEER REVIEWER ASSESSMENTS:

OBJECTIVE - Full research articles: is there a clear objective that addresses a testable research question(s) (brief or other article types: is there a clear objective)?

Yes - there is a clear objective

DESIGN - Is the current approach (including controls and analysis protocols) appropriate for the objective?

No - there are minor issues

EXECUTION - Are the experiments and analyses performed with technical rigor to allow confidence in the results?

No - there are minor issues

Statistics - Is the use of statistics in the manuscript appropriate?

Yes - appropriate statistical analyses have been used in the study

INTERPRETATION - Is the current interpretation/discussion of the results reasonable and not overstated?

No - there are minor issues
OVERALL MANUSCRIPT POTENTIAL - Is the current version of this work technically sound? If not, can revisions be made to make the work technically sound?

Probably - with minor revisions

PEER REVIEWER COMMENTS:

GENERAL COMMENTS: The manuscript present a precursor induction conditional random fields (pi-CRF) method that enables the capture of specific high-order label transition factors to improve clinical named entity recognition performance. The proposed method explained with a few valid arguments and exemplary sentences. The authors gave the detail of conventional first, second, and high order CRF methods. Different data set for processing, training and testing were selected. However, more detail could be provided with some instance where the conventional first, second, and higher order CRF model don't work well and do not give desire result, more complex to determine dependency of entities. What type of consequences one can experience if long distanced entities are not captured or what type of information losses could be experienced. Is it possible or workable to use conventional first order CRF for shorter distanced entities or second/higher order CRF for long distanced entities? What type of relation could be defined between dependent entities and their distance? How relations could be determined for multiple entities, multiple outsides, and multiple entities? Why to use data of 2012? Why not any new and recent data for model training and testing purposes? Any possibility of loss of information during state reduction is not explained.

REQUESTED REVISIONS:

A few recommendations to improve the document: a parametric analysis of first, second, higher order CRF model should be included, consequences of not capturing distanced/dependent entities, and method(s) for state reduction, method(s) to check and compare complexities.

ADDITIONAL REQUESTS/SUGGESTIONS:

A few recommendations to improve the document: a parametric analysis of first, second, higher order CRF model should be included, consequences of not capturing distanced/dependent entities, and method(s) for state reduction, method(s) to check and compare complexities.

Note: This reviewer report can be downloaded - see attached pdf file.
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

No

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

No

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

No

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.

This reviewer has been recruited by a partner organization, Research Square. Reviewers with declared or apparent competing interests are not utilized for these reviews. This reviewer has agreed to publication of their comments online under a Creative Commons Attribution License attributed to Research Square and was paid a small honorarium for completing the review within a specified timeframe. Honoraria for reviews such as this are paid regardless of the reviewer recommendation.

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