Reviewer’s report

Title: A classification framework for exploiting sparse multi-variate temporal features with application to adverse drug event detection in medical records

Version: 0 Date: 02 Jul 2018

Reviewer: Guoqian Jiang

Reviewer's report:

In this manuscript, authors develop a framework that aims to detect adverse drug events (ADEs) using EHR data. Unlike other similar studies, this study focused on the problem of feature sparsity. This work seems promising but it should be better to provide sufficient detail to explain the difference between this study and Zhao's study (Ref 30 of the paper) and to clarify the specific contribution of this work.

Also, there are some confusion and errors in this paper that need to be addressed and revised carefully.

Major comments:

1. In this paper, authors developed a three-phase framework to transform multi-variate time series features into a single-valued feature. It seems that this framework is an extensive work of Zhao's study. Although the authors have mentioned the difference between Zhao's study in Phase B of the framework, it would be better to explain the difference in Phase A and Phase C.

2. The choice of evaluation baseline is not clear in the study. In Methods Section, Strategy 1, the last paragraph, the authors stated that they chose plain approach as the baseline for evaluation. However, in Benchmarked methods paragraph, it seems like sl method was used as a baseline. It would be better to illustrate which one and why choose it as the baseline of evaluation.

3. Methods Section, Strategy 1, the authors said that this sl approach is "a modified and improved version of random dynamic subsequence used in Zhao et al." So, what modifications have been made?

4. Methods Section, Exploiting sparsity, as the most important part of the framework, it is not clear why authors did write this part in Phase B?
5. The number of table in the paper is a little bit confusing. Table 5 was written before Table 4 in the text as well as Table 4 appeared before Table 3.

6. Results Section, in Figure 4(a), only 5 ADEs was selected and shown for their AUC. It would better to explain why the authors chose them?

Minor comments:
1. First author's affiliation was not shown in the footnote of first page.

2. Method Section, the paragraph after equation 6, line 2, there is a typo error: "Ye at al." should be "Ye et al."

3. Method Section, the paragraph after algorithm 1, line 5, there may be a word after "In Section". The same problem also appeared in the last paragraph of Phase B and the last paragraph of Strategy 3.

4. The explanation of Figure 2, line 2, the "threshold doxp(abba)" may be "threshold δoxp(abba)"?

5. Table 5, all of the ICD code in the first column should have a ",," before the last number.

6. Figure 1, the last equation of Phase C may be "\(\hat{O} = \tau^* (O)\)"?

**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.

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.

I recommend additional statistical review

**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