Author's response to reviews

Title: Recognition of medication information from discharge summaries using ensembles of classifiers

Authors:

Son Doan (doan@nii.ac.jp)
Nigel Collier (collier@nii.ac.jp)
Hua Xu (Hua.Xu@vanderbilt.edu)
Hoang Duy Pham (duyph@ptit.edu.vn)
Minh Phuong Tu (phuongtm@ptit.edu.vn)

Version: 12 Date: 3 April 2012

Author's response to reviews: see over
Dear Editor,

Thank you and the reviewers very much for your constructive comments. Attached please find our revised manuscript. All changes are highlighted in the text and we list below point-by-point answers to all criticisms raised by reviewers.

Yours sincerely,

Son Doan, PhD
National Institute of Informatics
2-1-2 Hitotsubashi
Chiyoda-ku, Tokyo 101-8430, Japan
Phone: +81 (0)3 4212 2536
Fax: +81 (0)3 3556 1916
Email: doan@nii.ac.jp
Reviewer #2

Major comments

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Major Compulsory Revisions</td>
</tr>
<tr>
<td></td>
<td>1. The authors note that simple “majority voting provides an effective way to combine classifiers and has been shown to deliver good performance in a number of applications including NER from biomedical texts,” acknowledging that their conclusions appear to be well represented elsewhere in the literature.</td>
</tr>
<tr>
<td></td>
<td>And from comment 2: While these results do show that systems which combine classifiers can outperform individual classifiers, this is well trod ground.</td>
</tr>
<tr>
<td></td>
<td>Answer: We agree with the reviewer that combining multiple classifiers has been shown good performance in a number of NER tasks in biomedical text. However, to the best of our knowledge, such a conclusion has not been validated in the clinical entity recognition systems, which is the motivation of this study. To avoid confusion and make our conclusion more precise, we have rewritten the sentence mentioned in the review on page 12 as follows: “Simple and easy to implement, majority voting provides an effective way to combine classifiers and has been shown to deliver good performance in a number of applications, for example, in recognizing protein/gene names from research articles [28]. However, the effectiveness of this voting scheme in recognizing clinical entities has not been investigated, which we are addressing in this study.”</td>
</tr>
<tr>
<td></td>
<td>We also added a sentence on page 5 to clarify our contribution in this study: “To the best of our knowledge, this is the first study on investigating ensemble classifiers in recognizing medication relevant entities in clinical text.”</td>
</tr>
<tr>
<td>2</td>
<td>2. With the inclusion of statistical significance in the current draft, the paper clarifies the results of the reported experiments. Unfortunately, this information does not provide strong support for the claim.</td>
</tr>
<tr>
<td></td>
<td>The results for the first setting show that at p=0.05, there is no difference in performance between the simple majority voting, the local CRF-based</td>
</tr>
</tbody>
</table>
voting, and the local SVM-based voting systems for all of the categories. Any variation noted must be attributed to random chance. Therefore, the authors should not claim a hierarchy of these systems based upon the results.

Similarly, the results for the second setting show that at p=0.05, there is no difference in the performance between the simple majority voting, the local CRF-based voting, and the local SVM-based voting systems for the six individual categories. It is only under the combined “All” category that any difference can be seen between the simple majority voting and the other combined systems. The discussion section only describes the results from the “All” category, ignoring the fact that the combined systems showed no difference in performance on the six individual categories. The authors should fully describe their results and provide insight for the observed results.

Answer:

Thanks for your constructive comments. We were worried about the inconsistence between the “ALL” and individual categories regarding the significance test results. We checked the statistical test program and found an error in the source code that was used to prepare the data for statistical test. The error came from incorrectly reading fields from input files. We corrected and double-checked it carefully to make sure it is correct this time. We are really sorry for this mistake.

The results now show significant differences in some individual categories, in addition to the “All” category. We are grateful for this valuable comment. In the revised manuscript, we updated Tables 6 and 8, as well as some text in the “Results and discussion” (see pp. 16,17,18).

Our main conclusion is that ensemble classifiers could achieve better performance than a single classifier in recognizing medication related entities. We did not specify which ensemble classifier performs the best. We have added a paragraph to the end of “Results and discussion” to clarify this point.