

# Review of: "Enhancing Student Writing Skills: Leveraging Transfer Learning and Fine-tuned Language Models for Automated Essay Structure Recognition"

Matteo Ciniselli<sup>1</sup>

<sup>1</sup> Università della Svizzera Italiana

Potential competing interests: No potential competing interests to declare.

## Paper summary

The paper investigates the usage of Deep Learning models (BigBird and Longformer) to evaluate the essay structure. The goal of the paper is to correctly classify the argumentative and rhetorical elements present in the essay.

## Evaluation summary

- + the related work section is complete and exhaustive
- + the preprocessing steps are clear
- the way in which the analysis has been performed can be improved
- the results section is really short and can be improved

## Soundness

The paper is overall sound and shows the effectiveness of Deep Learning models for NER task. However, a baseline is missing, and I cannot fully judge the goodness of the results. I recommend you to compare your results with Alkabool et al work in order to make your improvement clearer. If the dataset used and task are the same, you should stress it in the discussion.

I also suggest you to be more rigorous in the analysis:

- how did you choose the hyperparameters? Did you perform hyperparameter tuning?
- With a small difference in the F1 score you can not claim that Longformer outperforms Bigbird (at least a significance test is required)
- You can discuss Fig3 in the results section and not just mention it in the Conclusions. This analysis can be interesting.
- The table you are referring to in the results section is not reported in the paper
- Figure 1 and 2 does not add any meaningful information to the paper

## Relevance

The paper is relevant, showing that we can achieve a high F1 score for the NER task. However, adding a tagging example from your model in the paper may be very useful from the reader perspective. At least, I can easily judge the performance of the model.

### **Novelty**

The paper is not novel, and it seems to be a sort of extension of Alkabool et al. I suggest you to better argument this in order to stress your contribution for the NER task.

### **Presentation**

The paper presentation and structure can be improved.

Some words are capitalized without any reason, sometimes you used we (“we performed the following steps”) and I (“test my models”, “I believe that”). I suggest you to always use the plural form everywhere.

In the abstract the acronym is wrong (AWE rather than AWS). In the related work section, I suggest you to always cite the authors (“Vaswani et al [12] introduced the transformer architecture” rather than “the transformer architecture introduced in [12]”) and to improve the ending (avoid “and Hence our approach”).

You can avoid to report all the BERT configuration in the related section (simply say that this is the BERT-base config).

I also believe that the fact that you used Pandas dataframes in the analysis is not relevant and you can avoid to mention it

Since you applied tagging before stopwords removal and stemming, you can change the order of the statements in the data cleaning section (you can add a bullet point starting from “Next, we converted” and name it “Tagging”).

Reporting the F1 formula in that way can be misleading (you can simply say that it is the armonic mean of precision and recall)

### **Replicability**

The work is not replicable since there is no replication package provided