Paper review by tywang

What is the major weakness of this paper?

As far as I am concerned, the formula of the contrastive learning objective might have small problem. The correct formula might be:

$$l_{cl}^{x} = -\frac{1}{\binom{|P|}{2}} \sum_{\substack{y_i, y_j \in P \\ y_j \neq y_i}} \log \frac{\exp(\mathrm{sim}(\boldsymbol{h}_i, \boldsymbol{h}_j)/\tau)}{\sum_{\substack{y_k \in P \cup N \\ y_k \neq y_i}} \exp(\mathrm{sim}(\boldsymbol{h}_i, \boldsymbol{h}_k)/\tau)} \tag{1}$$

To be specific, y_k should be selected from N, instead of being selected from $P \cup N$. This is because the purpose of the CL loss is to pull positive samples together and set apart those different ones. Because $l_{cl}^{x} \propto sim(h_i, h_k)$, minimizing l_{cl}^{x} means $sim(h_i, h_k)$ should be minimized, which is a process of setting samples from P and N apart. However, if y_k is selected from P, $sim(h_i, h_k)$ would be expected to be maximized and to be a value close to 1. Thus, it would be contradictory to be placed in the denominator.

In addition, Contrastive loss is also used in many fields such as textual classification. Zhangs[1] used CL in textual classification and the usage of CL is similar to yours, but they did not select samples from the union of positive and negative set like you.

Therefore, that's why I think this formula might be problematic. Of course, maybe you are right and I have misunderstood the paper. If yes, please let me know.

How would you improve on the weakness? What would you propose as a follow-up project?

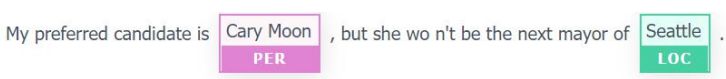
As mentioned above, selecting y_k particularly from N set might help improve on the weakness.

Things that I want to propose as a follow-up project.

1, TF-IDF word replacement may also could be used to construct positive sample.

2, Named entities in in summary and document could be marked.

Take this sentence as an example: *My preferred candidate is Cary Moon, but she won't be the next mayor of Seattle*. Through pre-trained model, we can obtain its named entities and their types, which is



Then the sentence with marked entity could be: *My preferred candidate is <PER> Cary*

Moon</PER>, but she won't be the next mayor of <LOC> Seattle </LOC>.

By doing this may reduce misconstructing phrases(belong to **intrinsic errors**) to a certain extend, because this might inform the model what is a complete phrase (We don't want any word in the phrase *Cary Moon* to be missing or replaced)

In this way, the summary's representation h_* could be the average of the tags <PER> and <LOC>.

This method[2] is usually used in relation extraction whose entity and its type are important.

References

[1] Zhang, D., Nan, F., Wei, X., Li, S., Zhu, H., McKeown, K., ... & Xiang, B. (2021). Supporting clustering with contrastive learning. *arXiv preprint arXiv:2103.12953*.

[2] Zhong, Z., & Chen, D. (2020). A frustratingly easy approach for entity and relation extraction. *arXiv preprint arXiv:2010.12812*.