Responsible NLP Checklist

Paper title: *HMCL: Task-Optimal Text Representation Adaptation through Hierarchical Contrastive Learning*
Authors: *Zhenyi Wang, Yapeng Jia, Haiyan Ning, Peng Wang, Dan Wang, Yitao Cao*

How to read the checklist symbols:

☑ the authors responded 'yes'

☒ the authors responded 'no'

N/A the authors indicated that the question does not apply to their work

☐ the authors did not respond to the checkbox question

For background on the checklist and guidance provided to the authors, see the Responsible NLP Checklist page at ACL Rolling Review.

☑ **A. Questions mandatory for all submissions.**

☑ A1. Did you describe the limitations of your work?
This paper has a Limitations section.

☑ A2. Did you discuss any potential risks of your work?
We discuss the potential risks in Limitation part. However, this article introduces a general method that does not involve great risks related to ethics, society, or safety.

☑ **B. Did you use or create scientific artifacts? (e.g. code, datasets, models)**

☑ B1. Did you cite the creators of artifacts you used?
In Appendix A.1, we cite the creators of models, training methods and benchmarks.

N/A B2. Did you discuss the license or terms for use and/or distribution of any artifacts?
The datasets we used from other sources are well-known, publicly available datasets that are licensed for research purposes. Therefore, we did not extensively discuss the licensing information for these datasets.

☑ B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?
In Appendix A.1, we provided a detailed description of how the well-known publicly available datasets were used in this study. In Appendix A.4, we describe the original purpose of the data collected for our self-created dataset and its usage in this paper.

☑ B4. Did you discuss the steps taken to check whether the data that was collected/used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect/anonymize it?
In Appendix A.4, we discuss the security of the self-made dataset and the de-identification procedures applied.

☑ B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?
In Appendix A.1 and A.4, we provide the information of those datasets.

The Responsible NLP Checklist used at ACL Rolling Review is adopted from NAACL 2022, with the addition of ACL 2023 question on AI writing assistance and further refinements based on ARR practice.

☑ B6. Did you report relevant statistics like the number of examples, details of train/test/dev splits, etc. for the data that you used/created?
We introduce them in Appendix A.1 and A.4.

☑ **C. Did you run computational experiments?**

☑ C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?
We introduce the detailed information of used models in Appendix A.1.

☑ C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
We provide those information in Approach and Ablation studies, and also Appendix A.1.

☑ C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
We fairly compare all the methods and models for reporting the max score within the same training steps in a single run. Detailed information is introduced in Appendix A.1.

N/A C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation, such as NLTK, SpaCy, ROUGE, etc.), did you report the implementation, model, and parameter settings used?
(left blank)

☑ **D. Did you use human annotators (e.g., crowdworkers) or research with human subjects?**

☑ D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
The instruction, questions and choices to participants are listed in Appendix A.4.

☑ D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
In Appendix A.4, we state that all the annotators are formal employees of our institution. However, their specific employment compensation and benefits involve commercial matters and are not disclosed in the paper.

☑ D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating (e.g., did your instructions explain how the data would be used)?
In Appendix A.4, we state that the use of these data has been approved through a comprehensive security review by the company and has received the necessary usage permissions (the annotators are not external volunteers, so no separate disclosure is required).

N/A D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
The annotation for our data was conducted by internal employees performing job-related tasks under existing employment agreements. The task only involves objective, fact-based judgments derived from documents and does not constitute human subjects research, as it does not investigate annotator preferences, behaviors, or personal attributes. Therefore, formal ethics review board approval was not required. Privacy and data protection were ensured through company-internal governance mechanisms, including de-identification and access control.

☑ D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
In Appendix A.4, we explain that all the annotators have a preliminary understanding of the relevant

(merchant Q&A) knowledge, which is directly related to the information about the annotators for our dataset.

☑ **E. Did you use AI assistants (e.g., ChatGPT, Copilot) in your research, coding, or writing?**

☑ E1. If you used AI assistants, did you include information about their use?
In Appendix A.6.