Section Classification in German Discharge Letters using Few-Shot Learning and Prompting

Phillip Richter-Pechanski[a,b,c,d,e], Philipp Wiesenbach[a,b,d], Dominic M. Schwab[b], Christina Kiriakou[b], Mingyang He[a,b,e], Nicolas A. Geis[b,d], Anette FRANK[e*], Christoph Dieterich[a,b,c,d*1]

[a]*Klaus Tschira Institute for Computational Cardiology, Heidelberg, DE*
[b]*Department of Internal Medicine III, University Hospital Heidelberg, DE*
[c]*German Center for Cardiovascular Research (DZHK) - Partner site Heidelberg/Mannheim, DE*
[d]*Informatics for Life, Heidelberg, DE*
[e]*Department of Computational Linguistics, Heidelberg University, DE*

To facilitate clinical routine, medical information extraction (IE) research aims to make use of unstructured information as found in clinical documents such as discharge letters. In order to do so, most supervised machine learning approaches rely on large amounts of manually annotated training data. However, recent developments in NLP show promising results in text classification tasks using pre-trained language models (PLM) and prompts [1], severely reducing the need of supervision.

In our work we evaluate pattern-exploiting training (PET, [2]), a semi-supervised few-shot learning method using prompts. Prompting exploits the ability of PLMs to infer knowledge from context without the cost-intense effort of model fine-tuning on large training datasets. In combination with lightweight supervision they achieve state-of-the-art results on various text classification tasks.

Hence, we explore this method in a German clinical IE text classification task, where a model classifies sentences from German discharge letters from the cardiovascular domain into eleven section categories, using section annotations of a publicly available German clinical corpus.

We evaluate PET using eight differently pre-trained PLMs and compare their results to baseline models used in current state-of-the-art approaches by running experiments on various few-shot training sizes. Our best performing PET model is based on a domain- and task-adapted BERT-large model [3], [4]. It achieves strong improvements over the baseline and publicly available PLMs for small shot sizes. We show that the performance gain diminishes with increase in sample sizes – an expected behavior that reinforces the incentives of our experiments. By studying interpretability methods, we were able to optimize the efficiency of our few-shot training sets, which improved the performance of our BERT-base model to a point where it achieves comparable classification results to the BERT-large model for selected section classes.

Overall, our results imply that PET can significantly improve the cost-efficiency of clinical text classification tasks on low-resource languages like German, by considerably reducing the need of annotated data *and* computational resources. We hope that our findings will pave the way to simplify the development of accurate section classification models to support, e.g. the automatic extraction of patient medication information or cardiovascular risk factors from discharge letters in clinical routine.

[1] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language Models are Unsupervised Multitask Learners,” OpenAI blog, vol. 1, no. 8, p. 9, 2019.

[1] *Corresponding authors: phillip.richter-pechanski@med.uni-heidelberg.de , christoph.dieterich@med.uni-heidelberg.de, frank@cl.uni-heidelberg.de . CD and AF share the last authorship.

[2] T. Schick and H. Schütze, “Exploiting Cloze Questions for Few Shot Text Classification and Natural Language Inference,” *EACL 2021 - 16th Conference of the European Chapter of the Association for Computational Linguistics, Proceedings of the Conference*, pp. 255–269, Jan. 2020, doi: 10.18653/v1/2021.eacl-main.20.

[3] S. Gururangan *et al.*, “Don’t Stop Pretraining: Adapt Language Models to Domains and Tasks,” in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Jul. 2020, pp. 8342–8360. doi: 10.18653/v1/2020.acl-main.740.

[4] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *Proceedings of the 2019 Conference of the North*, pp. 4171–4186, 2019, doi: 10.18653/V1/N19-1423.