Named Entity Recognition on German Medical Data

Johanna Bohn, Prof. Dr. Bettina Harriehausen-Mühlbauer, Prof. Dr. Jutta Groos
Darmstadt University of Applied Sciences – Department of Natural Science and Mathematics & Computer Science

Motivation

Medical data can be used to perform analyses and research in the medical domain. However, large percentages of the data remain unstructured and thus unused after they are created. Medical reports are one example of this: they exist to transmit humanly comprehensible information from one doctor to another in natural language. But in order to allow automated systems to take advantage of the information enclosed in natural language, the information must first be extracted.

The task of NER is a computational linguistics subtask of information extraction and deals with classifying every word in a document as belonging to one of a set of predefined categories. The categories we extract are diagnosis, treatment, and medication. Our goal is to advance research in the field of German NLP in the medical domain by finding the best approach to achieve good performance on a benchmark dataset.

Experiments

We do this by first obtaining and analysing the dataset called Berlin-Tübingen-Oncology corpus (BRONCO) [Kittner2021]. We train, validate, and evaluate six transformer models on it: the monolingual models GBERT, MedBERT, and GELECTRA, as well as the multilingual models mBERT, XLM-RobERTa, and XLM-RoBERTa GER. We train them using both a feature extraction and a fine-tuning approach, and perform hyperparameter optimization (HPO).
Additionally, we examine the feasibility of performing domain-adaptive pre-training on one model.

Results

When comparing the feature extraction approach and the fine-tuning approach, the second consistently outperformed the first for all models except GBERT. For this model, we experienced instabilities that discourage further use.
Performing HPO increases performances on all accounts – apart, again, from GBERT because of its instability. A higher increase was achieved for the feature extraction approach compared to the fine-tuning approach. We believe that this higher increase, however, is only interesting in a research setting: we consider the performances of the models trained under the unoptimised feature extraction approach as too low to be helpful in any realistic setting.

While F1-scores of approximately 64.0, as in the optimised feature extraction approach, might be useful in some contexts, it is unlikely that a situation exists in which there are no adequate computing resources to leverage the fine-tuning approach, but sufficient resources to perform HPO. Even though the increase for the fine-tuning approach is smaller, it is more relevant, because the smaller increase of up to 4.0 points directly increases the top performances.

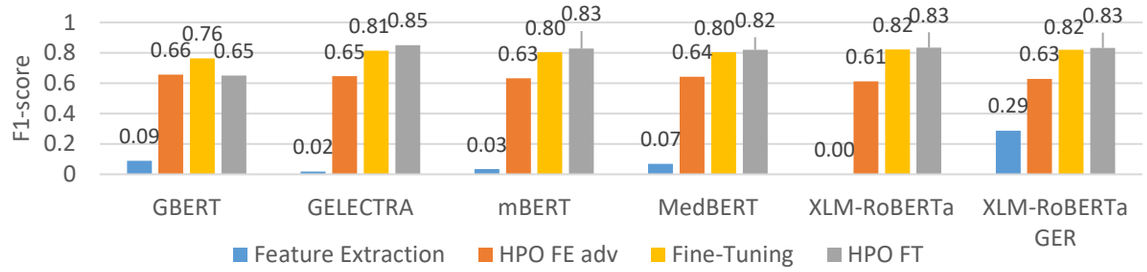

Fig. 1: F1-scores for feature extraction, the advanced feature extraction HPO, fine-tuning, and optimised fine-tuning.

Out of all six models, we identify GELECTRA as the best monolingual, and XLM-RoBERTa GER as the best multilingual model.
Examining the feasibility of domain-adaptive pre-training produced only ambiguous results.

Conclusion

To our knowledge, our fine-tuned GELECTRA model achieves the highest F1-score so far on the held back BRONCO50 testing dataset with an overall F1-score of 82.2. With that, we manage to outperform the results originally published for the dataset by Kittner et al. by 2.3-7.7 depending on the predefined category [Kittner2021].

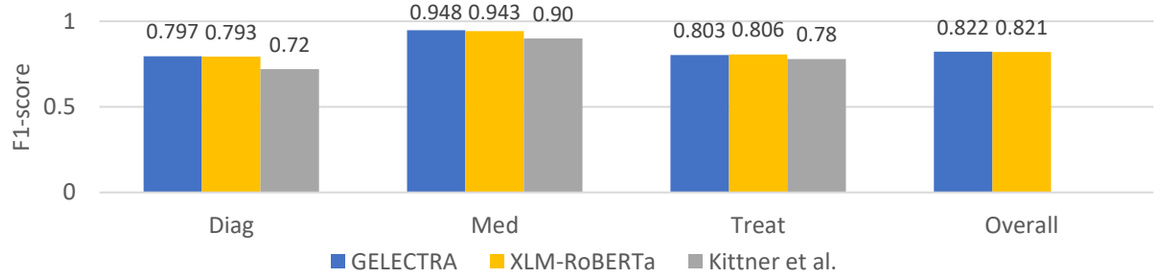

Fig. 2: Performances achieved by GELECTRA and XLM-RoBERTa on the held back BRONCO50 dataset compared to those achieved by Kittner et al., the publishers of the BRONCO dataset, by use of a CRF [Kittner2021].

The results show that adapting state-of-the-art models for German medical NER is an open research question with promising results.

Referenzen

[Kittner2021] - Kittner, M., Lamping, M., Rieke, D. T., Götze, J., Bajwa, B., Jelas, I., Rüter, G., Hautow, H., Sänger, M., Habibi, M., Zettwitz, M., Bortoli, T. de, Ostermann, L., Ševa, J., Starlinger, J., Kohlbacher, O., Malek, N. P., Keilholz, U., & Leser, U. (2021). Annotation and initial evaluation of a large annotated German oncological corpus. *JAMIA Open*, *4*(2), 1–9. https://doi.org/10.1093/jamiaopen/ooab025