

MultiNERD with RoBERTa report

Anastasios Lamproudis
e-mail: analampr1@gmail.com

December, 2023

In this project the MultiNERD [1] dataset is used to finetune a RoBERTa-base [2] model in a Named Entity Recognition (NER) task. The model is finetuned in two different setups, one using the complete set of annotated classes in the dataset and one using only a subset of the set of classes. The results from these experiments are summarized in table 1.

Models	P-mic	P-mac	R-mic	R-mac	F1-mic	F1-mac	Leakage
Model A	0.93	0.78	0.95	0.81	0.94	0.79	0.00387
Model B	0.95	0.87	0.97	0.93	0.96	0.89	0.00182

Table 1: The performance of the models in terms of **Precision** micro and macro (P-mic/P-mac), **Recall** micro and macro (R-mic/R-mac), and **F1** micro and macro (F1-mic/F1-mac) scores.

The results show that reducing the label set in which the model is finetuned to a set of the three most populous classes and two of the least populous classes allows for a better performance (**Model B**), compared to when using the complete set of classes (**Model A**). This also holds true for the two classes that are not as popular, Animal (ANIM) and Diseases (DIS) that in both of which **Model B** outperforms **Model A** as shown in table 2. However it is noticeable that for diseases, the data leakage is higher for **Model B** compared to **Model A**.

Models	ANIM	DIS	Leakage ANIM	Leakage DIS
Model A	0.73	0.74	0.16	0.14
Model B	0.75	0.76	0.12	0.17

Table 2: The performance of the models in terms of **F1** micro score in the Animal (ANIM) and Disease (DIS) classes.

The limitations of this approach is that the main improvement in performance is achieved through limiting the class set and not through any other technique such as enriching the data for the classes in which the performance is limited. This leads to two models that are very well performing with the most populous classes but are limited when it comes to the other classes.

References

- [1] Simone Tedeschi and Roberto Navigli. MultiNERD: A multilingual, multi-genre and fine-grained dataset for named entity recognition (and disambiguation). In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 801–812, Seattle, United States, July 2022. Association for Computational Linguistics.
- [2] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.