



Assignment 2: sequence labelling

Text mining course

This is a **hand-in assignment for groups of two students**. Send in via Brightspace **before or on Tuesday November 7:**

- ☐ Submit your report as PDF and your python code as separate file. **Don't upload a zip file containing the PDF** (the Python code might be zipped if it consists of multiple files).
- ☐ Your report should not be longer than 3 pages.
- ☐ Do not copy text from sources (other groups, web pages, generative models such as chatGPT). Turnitin is enabled and a large overlap will be reported to the Board of Examiners.

Goals of this assignment

- ☐ You can pre-process existing annotated text data into the data structure that you need for classifier learning.
- ☐ You can perform hyperparameter optimization.
- ☐ You can perform a sequence labelling task with annotated data in Huggingface.
- ☐ You can evaluate sequence labelling with the suitable evaluation metrics.

Preliminaries

- ☐ You have followed the Huggingface tutorial on token classification <https://huggingface.co/learn/nlp-course/chapter7/2> and its preliminaries (**exercises week 6 and 7**).
- ☐ You have all the required Python packages installed and Python 3.10.

We are going to train an NER classifier for the task “Emerging and Rare entity recognition” from the Workshop on Noisy User-generated Text (W-NUT). The description of the task can be found at <https://noisy-text.github.io/2017/emerging-rare-entities.html> (I put the data itself on Brightspace)

Tasks

1. Download `W-NUT_data.zip` from the Brightspace assignment and unzip the directory. It contains 3 IOB files: `wnut17train.conll` (train), `emerging.dev.conll` (dev), `emerging.test.annotated` (test).
2. Convert the IOB data to the correct data structure for token classification in Huggingface (words and labels like the conll2023 data in the tutorial) and align the labels with the tokens. Note that since you are working with a custom dataset, the data conversion is a necessary step for using the Huggingface training function.



3. Set up the evaluation correctly for the W-NUT test set, following the tutorial.
4. Fine-tune a model with the default hyperparameter settings on the **train** set and evaluate the model on the **test** set. These are your baseline results.
5. Set up hyperparameter optimization with the AdamW optimizer. During optimization, use the **dev** set as validation. After the model has been optimized, evaluate the result on the **test** set.
6. Extend the evaluation function so that it shows the Precision, Recall and F-score for each of the entity types (person, location, etc.) on the **test** set. Include the metrics for the B-label of the entity type, the I-label, and the full entities. Look up the definitions of *macro- and micro-average scores* and compute the macro- and micro average F1 scores over all entities.

Write a report of at most 3 pages in which you:

- ☐ describe the task and the data (give a few statistics. What are the entity types?) and briefly describe two challenges of the task and the data.
- ☐ show your results:
 - a results table with both the baseline results and the results after hyperparameter optimization (do not report results on the dev set, only on the **test** set): a table with Precision, Recall, F-score for both settings.
 - a table with the results after hyperparameter optimization for the different entity types (Precision, Recall, F-score for B, I, and the full entities), and the macro- and micro F1 scores.
- ☐ write brief conclusions. Address the following questions:
 - What is the effect of hyperparameter optimization on the quality of the model?
 - What does the difference between scores for different entity types tell you?
 - Where does the difference between macro- and micro-averaged F1 scores come from?

Grading

Maximum 2 points for each of the following criteria:

- ☐ General: length correct (2-3 pages) and proper writing + formatting
- ☐ Description of the task and the data, with description of 2 challenges
- ☐ Baseline results with default hyperparameter settings and results with optimized hyperparameter settings
- ☐ Results after hyperparameter optimization for the different entity types
- ☐ Sensible conclusions, briefly addressing the questions listed above.