

微調 NER 模型：<https://github.com/huggingface/transformers/tree/main/examples>  
需要先下載 transformers

## Important note

### Important

To make sure you can successfully run the latest versions of the example scripts, you have to **install the library from source** and install some example-specific requirements. To do this, execute the following steps in a new virtual environment:

```
git clone https://github.com/huggingface/transformers
cd transformers
pip install .
```



transformers >

名稱	修改日期	類型	大小
.circleci	2023/8/2 下午 04:02	檔案資料夾	
.git	2023/8/2 下午 04:02	檔案資料夾	
.github	2023/8/2 下午 04:02	檔案資料夾	
build	2023/8/2 下午 04:04	檔案資料夾	
docker	2023/8/2 下午 04:02	檔案資料夾	
docs	2023/8/2 下午 04:02	檔案資料夾	
examples	2023/8/2 下午 04:02	檔案資料夾	
model_cards	2023/8/2 下午 04:02	檔案資料夾	
notebooks	2023/8/2 下午 04:02	檔案資料夾	
scripts	2023/8/2 下午 04:02	檔案資料夾	
src	2023/8/2 下午 04:04	檔案資料夾	
templates	2023/8/2 下午 04:02	檔案資料夾	
tests	2023/8/2 下午 04:02	檔案資料夾	
utils	2023/8/2 下午 04:02	檔案資料夾	
.coveragerc	2023/8/2 下午 04:02	COVERAGERC 檔...	1 KB
.gitattributes	2023/8/2 下午 04:02	文字文件	1 KB
.gitignore	2023/8/2 下午 04:02	文字文件	2 KB
awesome-transformers.md	2023/8/2 下午 04:02	MD 檔案	38 KB
CITATION.cff	2023/8/2 下午 04:02	CFF 檔案	3 KB
CODE_OF_CONDUCT.md	2023/8/2 下午 04:02	MD 檔案	6 KB
conftest.py	2023/8/2 下午 04:02	Python File	4 KB
CONTRIBUTING.md	2023/8/2 下午 04:02	MD 檔案	19 KB
hubconf.py	2023/8/2 下午 04:02	Python File	9 KB
ISSUES.md	2023/8/2 下午 04:02	MD 檔案	19 KB
LICENSE	2023/8/2 下午 04:02	檔案	12 KB
Makefile	2023/8/2 下午 04:02	檔案	4 KB
pyproject.toml	2023/8/2 下午 04:02	TOML 檔案	1 KB
README.md	2023/8/2 下午 04:02	MD 檔案	94 KB
README_es.md	2023/8/2 下午 04:02	MD 檔案	94 KB
README_hd.md	2023/8/2 下午 04:02	MD 檔案	144 KB
README_ja.md	2023/8/2 下午 04:02	MD 檔案	99 KB
README_ko.md	2023/8/2 下午 04:02	MD 檔案	94 KB
README_zh-hans.md	2023/8/2 下午 04:02	MD 檔案	90 KB
README_zh-hant.md	2023/8/2 下午 04:02	MD 檔案	90 KB
setup.cfg	2023/8/2 下午 04:02	CFG 檔案	1 KB
setup.py	2023/8/2 下午 04:02	Python File	16 KB

Desktop -> transformers -> examples -> pytorch -> token-classification

把需要微調的訓練集與驗證集（測試集）放入

transformers > examples > pytorch > token-classification >

名稱	修改日期	類型	大小
wandb	2023/9/3 下午 03:57	檔案資料夾	
README.md	2023/8/2 下午 04:02	MD 檔案	5 KB
requirements.txt	2023/8/2 下午 04:02	文字文件	1 KB
run.sh	2023/8/2 下午 04:02	Shell Script	1 KB
run_ner.py	2023/8/2 下午 04:44	Python File	27 KB
run_ner_no_trainer.py	2023/8/2 下午 04:02	Python File	34 KB
run_no_trainer.sh	2023/8/2 下午 04:02	Shell Script	1 KB
train_1.json	2023/7/3 下午 03:40	JSON 檔案	3,683 KB
val_1.json	2023/7/3 下午 03:40	JSON 檔案	723 KB

利用命令提示字元（cmd）進入此資料夾，輸入以下指令微調模型（建議：以 GPU 訓練會更好）

```
python run_ner.py --model_name_or_path ckiplab/bert-base-chinese-ner --tokenizer_name bert-base-chinese --train_file train_1.json --validation_file val_1.json --output_dir /NER_TEST/MODEL --per_device_train_batch_size 4 --per_device_eval_batch_size 4 --num_train_epochs 10 --metric_for_best_model f1 --do_train --do_eval --overwrite_output_dir --report_to wandb
```

我們以此微調出六種案件類型的 NER 模型

Case	train_val (Drug)	train_val (Forged_Documents)	train_val (Gamble)
Epoch	10	10	10
Accuracy	0.98451	0.99175	0.9889
F1	0.76971	0.82421	0.84115
Eval / Loss	0.07542	0.04712	0.05886
Train / Loss	0.0542	0.0438	0.0419
Train / Train_loss	0.03563	0.02873	0.02857
Train	329	325	326
Eval	82	81	82
Case	train_val (Negligent_Injury)	train_val (Public_Danger)	train_val (Theft)
Epoch	10	10	10
Accuracy	0.97428	0.98626	0.97379
F1	0.78008	0.79748	0.48406
Eval / Loss	0.14542	0.07394	0.13843
Train / Loss	0.0861	0.0045	0.0644
Train / Train_loss	0.05137	0.02436	0.04369
Train	374	1065	327
Eval	93	266	82