Requirements:

python 3.7

pip3 install -r requirements.txt

python3 -m spacy download en_core_web_sm

Data:

The train, validation (development), test, and ARC_test CoNLL format data are in the directory /ScienceExamCER_data/conll_data/.

Usage:

Option 1: Train the model by yourself (This will take much time)

$ python run_ner.py --data_dir=ScienceExamCER_data/conll_data/ --bert_model=bert-base-cased --task_name=ner --output_dir=output_base --max_seq_length=64 --num_train_epochs 140 --do_train --do_eval --warmup_proportion=0.1

The output model and evaluation results on the development set (eval_results.txt) will be in the directory /output_base/.

Option 2: Download pre-trained model

download the modle from: http://www.cognitiveai.org/dist/ScienceExamCER-Pretrained-Model.zip

unzip the archive and put the files in the directory /output_base/ (If it does not exist, create this directory in the same folder)

Do Named-Entity-Recognition in ScienceExamCER test dataset:

python run_ner.py --data_dir=ScienceExamCER_data/conll_data/ --bert_model=bert-base-cased --task_name=ner --output_dir=output_base --max_seq_length=64 --do_eval --do_eval_test --warmup_proportion=0.1

The test set evaluation results (test_results.txt) will be output in the directory /output_base/ . The output will include precision, recall, and F1 performance per semantic class, as well as overall performance

	precision	recall	f1-score	support
CelestialMovement	0.7931	1.0000	0.8846	46
UnderwaterEcosystem	0.7500	0.8571	0.8000	7
RelativeNumber	0.8324	0.8780	0.8546	164
ActionsForAnimals	0.8032	0.8830	0.8412	171
StateOfBeing	0.9374	0.9718	0.9543	1064
Negations	0.9516	0.9593	0.9555	123
Minerals	0.9167	0.7458	0.8224	59
RelativeTime	0.9048	0.9280	0.9162	389
...	...	...	...	...
TimeMeasuringTools	1.0000	1.0000	1.0000	1
Vehicle	1.0000	1.0000	1.0000	1
BusinessNames	0.0000	0.0000	0.0000	1
avg / total	0.8479	0.8625	0.8519	25374

Use the model to tag ARC test set:

python run_ner.py --data_dir=ScienceExamCER_data/conll_data/ --bert_model=bert-base-cased --task_name=ner --output_dir=output_base --max_seq_length=64 --do_eval --do_eval_ARCtest --warmup_proportion=0.1

Examples like this:

Token	Semantic Class
Jennifer	B-PERSON
measured	B-Examine B-Measurements
air	B-Pressure
temperature	B-Temperature
and	O
air	B-Pressure
pressure	I-Pressure
outside	B-RelativeLocations
her	O
house	B-ManmadeLocations
at	B-RelativeTime
noon	B-TimesOfDayDayNight
for	O
five	B-CardinalNumber
days	B-TimeUnit
.	O

Inspired by JerryZeyu and Thank you.