NLU 2021 Second assignment - Report

Lucie Nass - 219524

Evaluate spaCy NER on CoNLL 2003 data

There were some issues to deal with when comparing spaCy predictions and the CoNLL tags. The first is that in CoNLL there are only 4 tags:

- ORG for organisations
- PER for people
- LOC for locations
- MISC for all others

In spaCy, there are a variety of other tags, including MONEY and DATE for example. I decided to map the tags in the following way.

spaCy tag	spaCy signification	CoNLL tag
FAC	Buildings, airports, highways, bridges, etc.	LOC
GPE	Countries, cities, states	LOC
LOC	Non-GPE locations, mountain ranges, bodies of water	LOC
NORP	Nationalities or religious or political groups	ORG
ORG	Companies, agencies, institutions, etc.	ORG
PERSON	People, including	PER
Any other	Monye, quantities, product...	MISC

The second issue is that spaCy's tokenization is not the same as the one used in CoNLL. To be able to correctly compare the NE recognition performances, I mapped the spaCy tokens to the CoNLL, using the Alignment class from spaCy. By using the from_strings method and giving it the spaCy tokens and the CoNLL tokens for the same sentence, it is able to generate mappings in one way or another. In this assignment, I used x2y.lengths attribute, which gives me a list of integers giving the number of spaCy tokens to which corresponds the CoNLL token at that position in the CoNLL token lists. I then use this to fuse (if necessary) the spaCy tokens, keeping the tag of the first one.

	precision	recall	f1-score	support
B-LOC	0.76	0.68	0.72	1668
B-MISC	0.02	0.09	0.03	702
B-ORG	0.38	0.32	0.35	1661
B-PER	0.80	0.63	0.70	1617
I-LOC	0.54	0.56	0.55	257
I-MISC	0.05	0.36	0.08	216
I-ORG	0.42	0.52	0.46	835
I-PER	0.84	0.79	0.81	1156
O	0.94	0.86	0.90	38323
accuracy			0.80	46435
macro avg	0.53	0.53	0.51	46435
weighted avg	0.88	0.80	0.84	46435

On a token-level, as we can in the table, spaCy performs quite well on the people (0.8 and 0.84 precision, 0.63 and 0.79 recall, depending on the IOB position). However, its ability to correctly detect locations is lower, and goes low when we only look at the INSIDE tokens (a little above 50 % of precision and recall). Organizations show a low performance, but the worst category is the MISC tag. spaCy only has a 2 to 5 % precision in this category, but the recall is much higher, although still low for INSIDE miscellaneous tags (36 %). That may be explained by spaCy considering a wider range of types. For example, “Friday” is a named entity of type DATE in spaCy, but is not even tagged a named entity in CoNLL.

On a chunk-level, spaCy shows a good performance on locations (F1 = 0.7) and people (F1=0.68), while its performance on organisations is quite low (F1=0.31) and very low for miscellaneous (F1=0.02).

Grouping of entities

In this part, I decided to consider groups of entities independently from the order in which they are found. That means that [ORG, PERSON] is considered the same group as [PERSON, ORG]. This, to me, made sense because we are interested in which groups appear together, and they may appear together in different orders to mean the same type of thing.

When looking at the counts for each type of group, the 1-type groups are evidently the most common, but some groups with two types still have a high count. The most common is CARDINAL PERSON with 59 occurrences, which seems plausible. It may refer to the first king of some place for instance. The next three most common groups all also have PERSON, paired with NORP (44), GPE (39), and ORG (25). Not far behind are more groups with CARDINAL, paired with ORG (25), GPE (18), and NORP (16). Also among the most common is (GPE, ORG). These results make sense, as a cardinal and a nationality are very often used as modifiers to a person or organisation, while countries, people, and organizations very commonly have named entities modifiers.

Function that extends the entity span to cover the full noun-compounds

I didn’t succeed in writing this function, although I included in the notebook my attempt at writing it.