

Appendix for Paper: A Neural Span-Based Continual Named Entity Recognition Model

Permutations of Tasks

As described, we follow the existing 6 permutations of tasks on OntoNotes from recent works, and randomly sample 4 permutations on Few-NERD to factor out the influence of the learning order as shown in Tab. 1.

6 OntoNotes Permutations (\implies)							
1:	ORG	PER	GPE	DATE	CARD	NORP	
2:	DATE	NORP	PER	CARD	ORG	GPE	
3:	GPE	CARD	ORG	NORP	DATE	PER	
4:	NORP	ORG	DATE	PER	GPE	CARD	
5:	CARD	GPE	NORP	ORG	PER	DATE	
6:	PER	DATE	CARD	GPE	NORP	ORG	
4 Few-NERD Permutations (\implies)							
1:	LOC	PER	ORG	OTH	PROD	BUID	ART EVET
2:	ORG	PROD	ART	EVET	OTH	PER	LOC BUID
3:	PROD	EVET	OTH	PER	ART	LOC	BUID ORG
4:	BUID	OTH	PROD	PER	ORG	LOC	ART EVET

Table 1: Different permutations of tasks used on OntoNotes and Few-NERD.

Fine-grained Entity Types on Few-NERD

All the fine-grained entity types on Few-NERD will be used in our evaluation, as shown in Tab. 2. The corresponding coarse-grained entity types are bold.

location	bodiesofwater GPE island mountain other park road/railway/highway/transit
person	actor artist/author athlete director other politician scholar soldier
organization	company education government/governmentagency media/newspaper other politicalparty religion showorganization sportsleague sportsteam

other	astronomything award biologything chemicalthing currency disease educationaldegree god language law livingthing medical
product	airplane car food game other ship software train weapon
build	airport hospital hotel library other restaurant sportsfacility theater
art	broadcastprogram film music other painting writtenart
event	attack/battle/war/militaryconflict disaster election other protest sports

Table 2: The fine-grained entity types on Few-NERD

Performance w.r.t Different Permutations

To investigate the impact on the final performance of whether an entity type is learned earlier or later, we further plot the performances regarding all permutations on both datasets in Split-All setup of SpanKL. Similarly, we also use an extra black dashed curve to denote the Macro-F1 score over all entity types learned so far.

As shown in Fig. 1, we can observe that most entity types have a similar performance eventually although start to be learned at different steps. The final Macro-F1 scores in different orders also converge to a stable point. We can presume that the result at the final step of a certain permutation is already relatively representative for evaluation on CL.

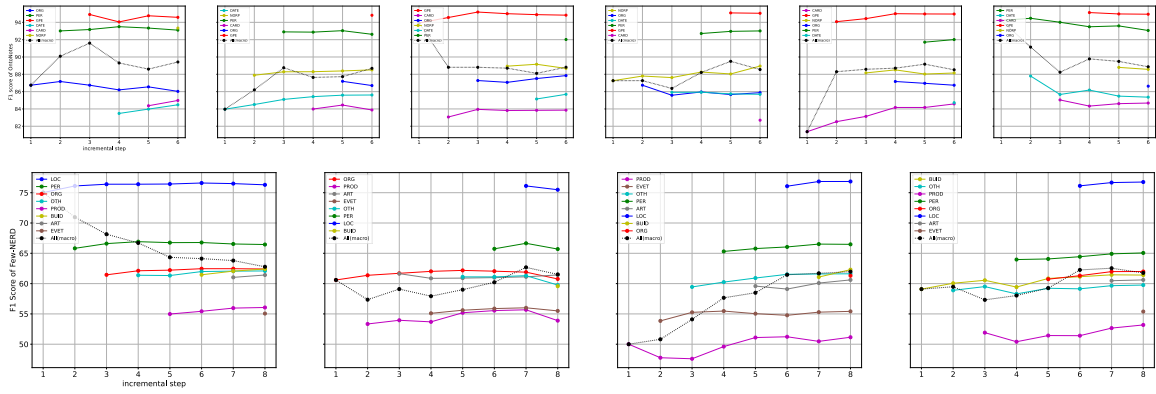


Figure 1: F1 curves of different entity types on OnteNotes and Few-NERD (rows) with different learning order (columns) indicated by downward in legends.