

# **COLING 2024**

# A Novel Three-stage Framework for Few-shot Named Entity Recognition



#### Introduction



- Recent research(SpanProto, DecomposedNER ...) decomposes the task into two phases: span extraction and entity classification. This paradigm simplifies the difficulty of the task, so it tends to give good results.
- We summarize our main contributions as follows:
  - 1. We propose a novel three-stage framework for FS-NER, including: Teacher span recognizer, Student span recognizer and Entity classifier.
  - 2. We are the pioneers in applying the idea of soft label learning to the few-shot field, which is helpful for obtaining a stronger span recognizer.
  - 3. We are the first to integrate prompt learning with prototypical network as our Entity classifier.



# **Examples**



Types	(1). Location (2). Person		
Support	<ul> <li>(1).The nearest tube station is [covent garden]<sub>Location</sub>.</li> <li>(2).[Shami]<sub>Person</sub> was born into a shiite family in 1945.</li> </ul>		
Query	Culbertson came back to fort union in 1840.	12 13	
Output	Location: fort union Person: Culbertson	e e	



#### Framework





## Span recognizer

Construct key and query vectors:

$$q_i = W_q h_i + b_q, \qquad k_i = W_k h_i + b_k$$

Calculate span score:

$$s(i,j) = q_i^T k_j$$

Loss function:

$$\mathcal{L} = \log\left(1 + \sum_{(i,j)\in Pos} e^{-s(i,j)}\right) + \log\left(1 + \sum_{(i,j)\in Neg} e^{s(i,j)}\right)$$









	Intra					Inter				
Models	1∼2-shot		5~10-shot		Aug	1∼2-shot		5~10-shot		Aug
	5 way	10 way	5 way	10 way	Avg.	5 way	10 way	5 way	10 way	Avg.
$ProtoBERT^{\dagger}$	20.76±0.84	15.05±0.44	42.54±0.94	35.40±0.13	28.44	38.83±1.49	32.45±0.79	58.79±0.44	52.92±0.37	45.75
$\mathrm{NNShot}^\dagger$	25.78±0.91	18.27±0.41	36.18±0.79	27.38±0.53	26.90	47.24±1.00	38.87±0.21	55.64±0.63	49.57±2.73	47.83
$\mathrm{StructShot}^\dagger$	30.21±0.90	21.03±1.13	38.00±1.29	26.42±0.60	28.92	51.88±0.69	43.34±0.10	57.32±0.63	49.57±3.08	50.53
CONTaiNER(Das et al., 2021)	40.43	33.84	53.70	47.49	43.87	55.95	48.35	61.83	57.12	55.81
SpanProto <sup>*</sup>	39.76±1.72	31.62±0.73	51.05±0.96	46.05±0.31	42.12	55.72±1.21	50.22±1.03	62.65±0.11	57.64±0.45	56.56
$\mathrm{ESD}^{\dagger}$	36.08±1.60	30.00±0.70	52.14±1.50	42.15±2.60	40.09	59.29±1.25	52.16±0.79	69.06±0.80	64.00±0.43	61.13
$DecomposedMetaNER^{\dagger}$	49.48±0.85	42.84±0.46	62.92±0.57	57.31±0.25	53.14	64.75±0.35	58.65±0.43	71.49±0.47	68.11±0.05	65.75
Ours	56.35±0.64	50.51±0.36	65.22±0.52	58.35±0.19	57.61	68.20±0.79	64.72±0.23	72.86±0.46	68.62±0.27	68.60

Table 1: F1 scores on FewNERD. The best results are in **bold**.<sup>†</sup> denotes the result reported in Ma et al.  $(2022)^4$ .\* represents the results we reproduce with the same dataset version.

-	1-shot				5-shot					
	News	Wiki	Social	Mixed	Avg.	News	Wiki	Social	Mixed	Avg.
$Transfer BERT^{\dagger}$	4.75±1.42	0.57±0.32	2.71±0.72	3.46±0.54	2.87	15.36±2.81	3.62±0.57	11.08±0.57	35.49±7.60	16.39
$Sim BERT^{\dagger}$	19.22	6.91	5.18	13.99	11.33	32.01	10.63	8.20	21.14	18.00
Matching Network <sup>†</sup>	19.50±0.35	4.73±0.16	17.23±2.75	15.06±1.61	14.13	19.85±0.74	5.58±0.23	6.61±1.75	8.08±0.47	10.03
$ProtoBERT^{\dagger}$	32.49±2.01	3.89±0.24	10.68±1.40	6.67±0.46	13.43	50.06±1.57	9.54±0.44	17.26±2.65	$13.59 \pm 1.61$	22.61
L-TapNet+CDT(Hou et al., 2020)	44.30±3.15	12.04±0.65	20.80±1.06	15.17±1.25	23.08	45.35±2.67	11.65±2.34	23.30±2.80	20.95±2.81	25.31
$\rm Decomposed Meta NER^{\dagger}$	46.09±0.44	17.54±0.98	25.14±0.24	34.13±0.92	30.73	58.18±0.87	31.36±0.91	31.02±1.28	45.55±0.90	41.53
Ours	57.42±0.28	30.89±0.75	27.91±0.44	37.72±0.83	38.49	62.44±0.56	38.57±0.64	31.23±1.02	46.64±0.49	44.62

Table 2: F1 scores on Cross-DataSet.The best results are in **bold**.† denotes the result reported in Ma et al. (2022).



	Intra	Inter
Ours	57.48	68.60
1)w/o Soft Boundary Learning	56.12	66.63
2)w/o Prompt	53.28	65.07
3)w/o Soft Boundary Learning w/o Prompt	51.85	61.90

Table 3: The average F1 scores of ablation study on Few-NERD.







Figure 5: Impact of  $\lambda$  on the F1 score of student span recognizer on 5-way 5~10-shot Few-NERD.





Figure 6: The t-SNE visualization of entity representations on Few-NERD Intra, 5-way  $5\sim$ 10-shot validation set. We randomly choose 5 classes which include a total of 9722 samples. The left section shows the conventional ProtoNet's 2D visualization, while the right depicts Prompt-ProtoNet's 2D visualization.

#### Conclusion



- We are the pioneers in applying the idea of soft label learning to the few-shot field, which is helpful for obtaining a stronger span recognizer.
- We are the first to integrate prompt learning with prototypical network as our Entity classifier.
- Extensive experiments demonstrate that our method outperforms previous methods with a large margin on two widely used benchmarks.



