中文信息处理实验室-让机器理解中文
Chinese Information Processing Laboratory

ISCAS

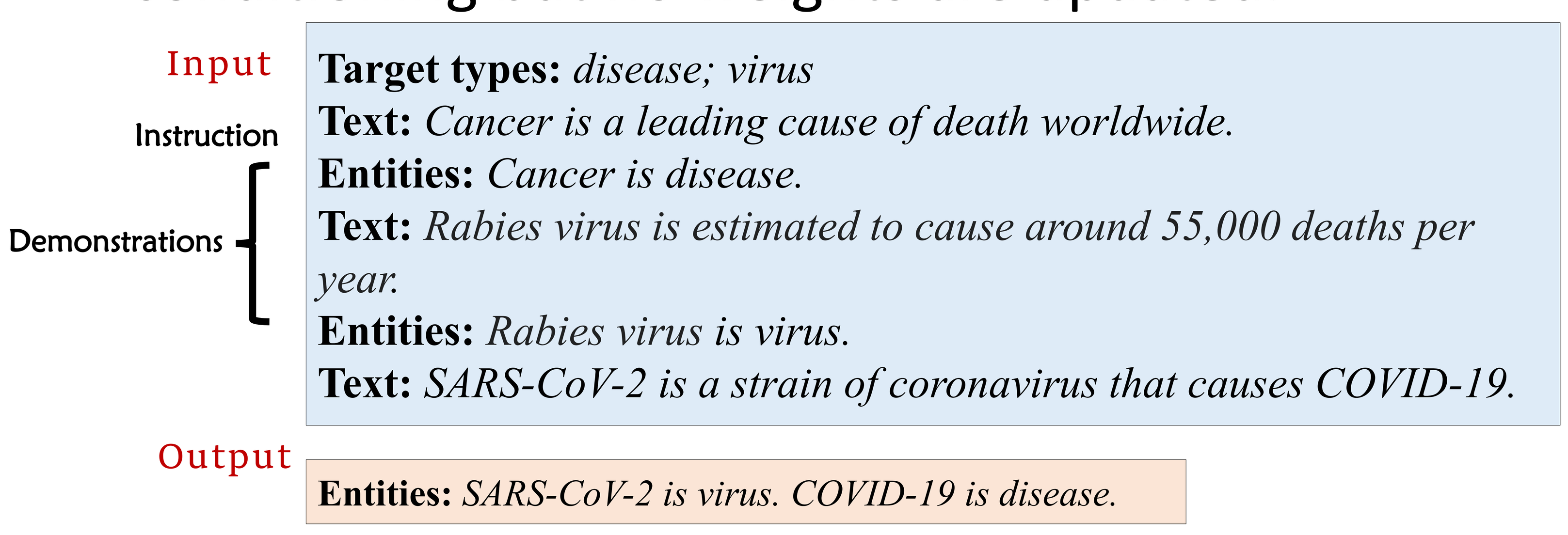
Learning In-context Learning for Named Entity Recognition

Jiawei Chen[1,4], Yaojie Lu[1,†], Hongyu Lin[1] , Jie Lou[3], Wei Jia[3], Dai Dai[3]
Hua Wu[3], Boxi Cao[1,4], Xianpei Han[1,2,†], Le Sun[1,2]

1Chinese Information Processing Laboratory 2State Key Laboratory of Computer Science
Institute of Software, Chinese Academy of Sciences
3Baidu Inc., Beijing, China
4University of Chinese Academy of Sciences

In-context Named Entity Recognition

- ■ **Named entity recogniztion**
  - Detect and classify named entities in text.
- ■ **In-context learning**
  - The model is given a few demonstrations (and instruction) of the task at inference time as conditioning but no weights are updated.



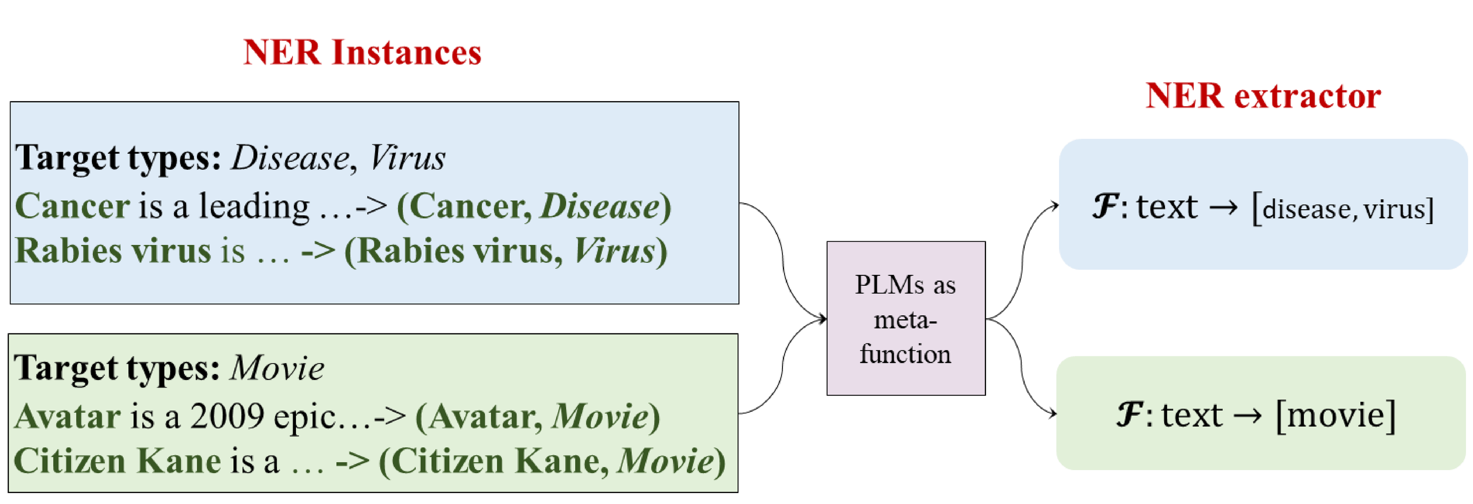
Meta-function View for in-context NER

- ■ We model pre-trained language models as a meta-function for NER:

$$\lambda_{instruction,demonstraions,text}.\ M$$

- ■ The new extractor can be implicitly constructed by instruction and demonstrations

$$(\lambda.M)(insturction, demonstraions) \rightarrow \{\mathcal{F}: text \rightarrow entities\}$$



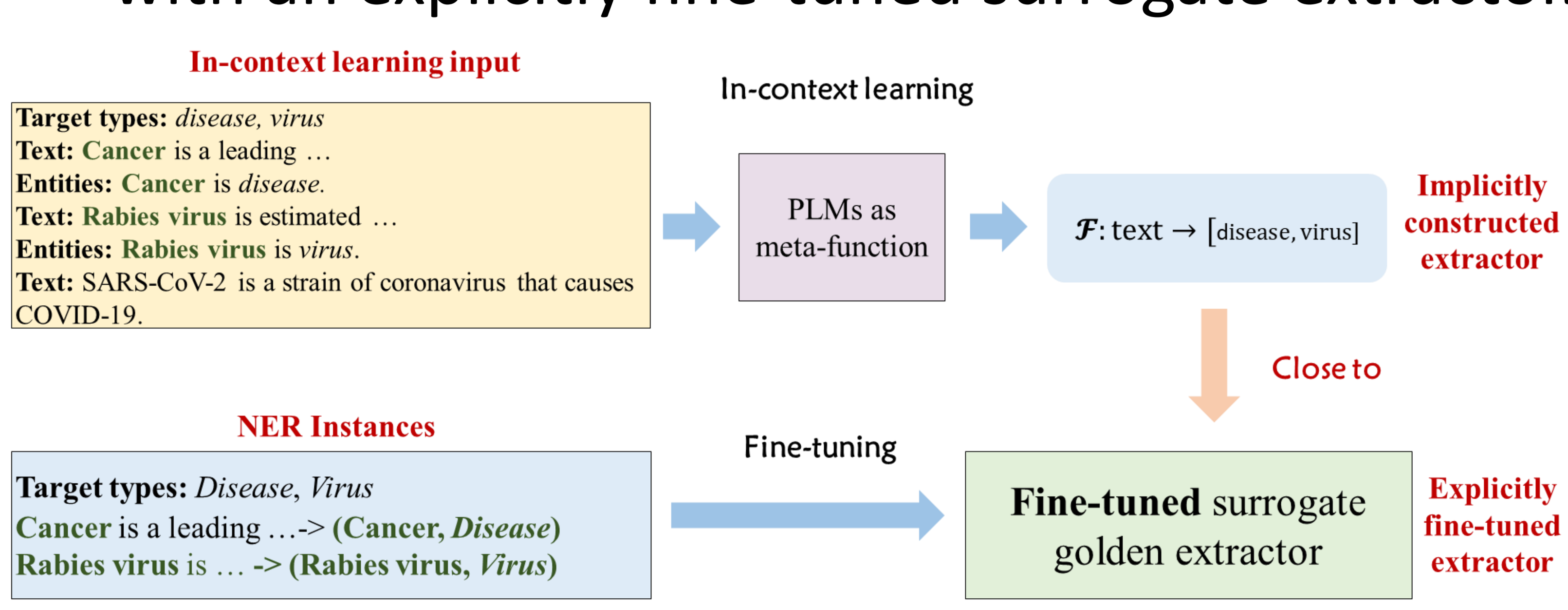
- ■ We model pre-trained language models as a meta-function for NER:

$$\lambda_{instruction,demonstraions,text}.\ M$$

- ■ The new extractor can be implicitly constructed by instruction and demonstrations

$$(\lambda.M)(insturction, demonstraions) \rightarrow \{\mathcal{F}: text \rightarrow entities\}$$

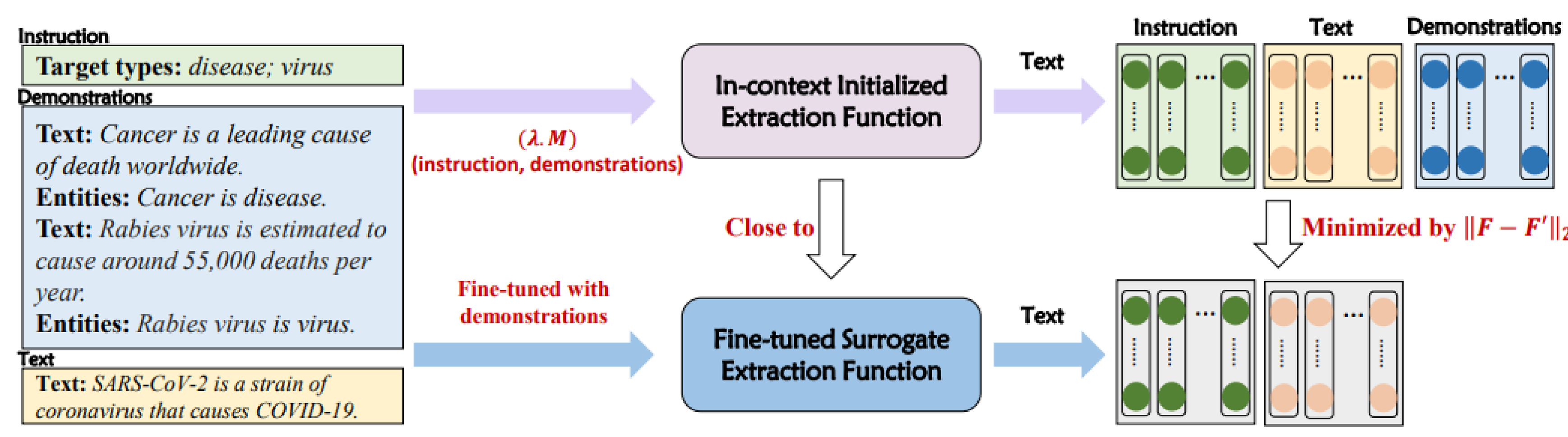
- ■ Meta-function pretraining: inject in-context NER ability into PLMs.
  - Comparing the implicitly constructed extractor with an explicitly fine-tuned surrogate extractor.



Meta-function pre-training

- ■ Meta-function pretraining: make the features of in-context model (PLMs) are as close as the features of surrogate golden extractor which is fine-tuned using instances in demonstrations.

$$\mathcal{L}_{meta-function} = Average(d(\boldsymbol{F}_{in-context}, \boldsymbol{F}'_{fine-tuned}))$$



- ■ Extraction function pre-training: pre-train to generate correct sequence in an auto-regressive way
- ■ Overall loss function:

$$\mathcal{L} = \alpha\mathcal{L}_{meta-function} + \mathcal{L}_{extraction}$$

Experimental Results

Models	#Param	CoNLL03		WNUT17		NCBI-disease		SEC-filings		AVE
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	
Pre-trained Language Models										
T5v1.1-large	770M	38.61	44.90	25.52	26.32	26.02	37.63	41.89	53.44	36.79
GPT2-xl	1.5B	33.69	39.55	22.63	24.86	25.54	33.25	42.83	**57.05**	34.93
T5-xl	3B	38.99	45.74	26.39	26.31	23.10	36.78	30.58	42.22	33.76
GPT-J-6B	6B	46.14	50.10	31.41	30.93	35.82	40.98	40.12	39.61	39.39
T5-xxl	11B	40.97	46.14	24.76	25.27	12.19	26.34	32.65	42.44	31.35
OPT-13B	13B	46.65	51.71	27.74	28.36	23.73	34.00	41.60	43.10	37.11
GPT-Neox-20B	20B	52.68	58.12	**36.29**	35.68	35.42	42.85	45.07	45.17	43.91
OPT-30B	30B	42.86	44.77	25.85	27.44	22.31	32.76	40.83	46.52	35.42
OPT-66B	66B	43.83	53.89	30.77	32.00	25.87	34.58	39.15	47.01	38.39
Pre-trained NER Models										
ProtoNet	345M	30.04	60.26	9.74	23.03	24.73	42.32	16.79	23.67	28.82
NNShot	345M	41.92	58.39	15.76	21.78	31.59	33.14	30.19	37.86	33.83
StructShot	345M	42.34	58.44	15.78	22.05	19.87	31.48	30.40	38.44	32.35
CONTAINER	345M	45.43	61.69	15.64	20.37	23.24	27.02	34.07	40.44	33.49
MetaNER-base	220M	53.94	62.59	25.55	30.41	35.00	37.24	46.88	51.39	42.88
MetaNER	770M	**57.40**	**63.45**	31.59	**36.52**	**40.01**	**44.92**	**52.07**	54.87	**47.60**

- ■ MetaNER can achieve good in-context NER performance.
- ■ In-context NER method can achieve robust performance, even under a large sourcetarget domain gap
- ■ Meta-function pre-training can effectively inject in-context learning ability into PLMs.

Conclusions

- ■ We model PLMs as a meta-function for in-context NER.
- ■ We propose the meta-function pre-training to inject in-context NER ability into PLMs.
- ■ Experimental results show that our method is effective for in-context NER.