

Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心



小样本命名实体识别

硕士研究生 刘小丫

2023年08月27日



- 总结反思
 - 内容讲的比较快，主线不太清晰
 - 内容深度不足

- 相关内容
 - 林朝坤 《小样本命名实体识别》——**2021.05.30**
 - 吴杭颐 《匮乏资源命名实体识别》——**2021.11.02**



- 预期收获
 - 1.了解小样本命名实体识别的基本概念
 - 2.了解小样本命名实体识别的常用方法
 - 3.理解小样本命名实体识别的算法原理
 - 4.了解小样本命名实体识别的前沿发展和应用



- 背景简介
- 基础概念
 - 命名实体识别
 - 小样本命名实体识别
 - 常见解决方法
- 算法原理
 - **CONTaiNER**
 - **COPNER**
- 前沿发展
- 参考文献



- 命名实体识别 (Named entity recognition, NER)
 - 提取文本中具有特定意义的实体并对这些实体进行分类, 包括人名、地名、机构名、日期、时间、货币、百分比等
 - 不仅需要找出实体的位置, 还需要对实体进行分类

[小明] 在 [北京理工大学] 的 [信息科学实验楼] 听了一场学术报告

PER

ORG

LOC

General domain(news)

- Person name
- Organization name
- Location name
- Numeric expression

Biomedical domain

- Gene
- Protein
- Disease
- Chemical
- Cell

不同领域的实体

Paragraph

*London*_[Art-Music] is the fifth album by the *British*_[Loc-GPE] rock band *Jesus Jones*_[Org-ShowOrg] in 2001 through *Koch Records*_[Org-Company]. Following the commercial failure of 1997's "*Already*_[Art-Music]" which led to the band and *EMI*_[Org-Company] parting ways, the band took a hiatus before regathering for the recording of "*London*_[Art-Music]" for Koch/Mi5 Recordings, with a more alternative rock approach as opposed to the techno sounds on their previous albums. The album had low-key promotion, initially only being released in the *United States*_[Loc-GPE]. Two EP's were released from the album, "*Nowhere Slow*_[Art-Music]" and "*In the Face Of All This*_[Art-Music]".

NER任务: 找+分类



- 命名实体识别难点

- 歧义多
- 边界界定困难
 - 武汉市长江大桥
- 标注数据缺失
- 特定领域命名实体识别局限性

General domain(news)

- Person name
- Organization name
- Location name
- Numeric expression

Biomedical domain

- Gene
- Protein
- Disease
- Chemical
- Cell

不同领域的实体

- 目前NER只是在**有限的领域和实体类型**中取得了较好的成绩，但这些技术无法更好地迁移到其他特定领域中，比如医学，生物，化学领域
 - 原因一方面是，不同领域的数据往往具有**领域独特特征**
 - 另一方面，由于**领域资源匮乏**造成标注数据缺失，收集大量标注数据**昂贵、困难**

➡ 如何使用**少量标注数据**进行命名实体识别？

➡ **小样本**命名实体识别任务

- 小样本命名实体识别
 - 仅使用几个标记示例泛化到**没见过的实体类**
- 常用方法
 - 模型微调：源数据集- \rightarrow 目标数据集，过拟合
 - 度量元学习
 - 利用**以往的知识经验**来指导新任务的学习
 - **N-way K-shot**:
 - N代表类别数量，
 - K代表每一类别中样本数量
 - **Support Set**: 支持集，已知样本
 - **Query Set**: 查询集，待预测样本

Training task 1

Support set



Query set



Training task 2 . . .

Support set



Query set

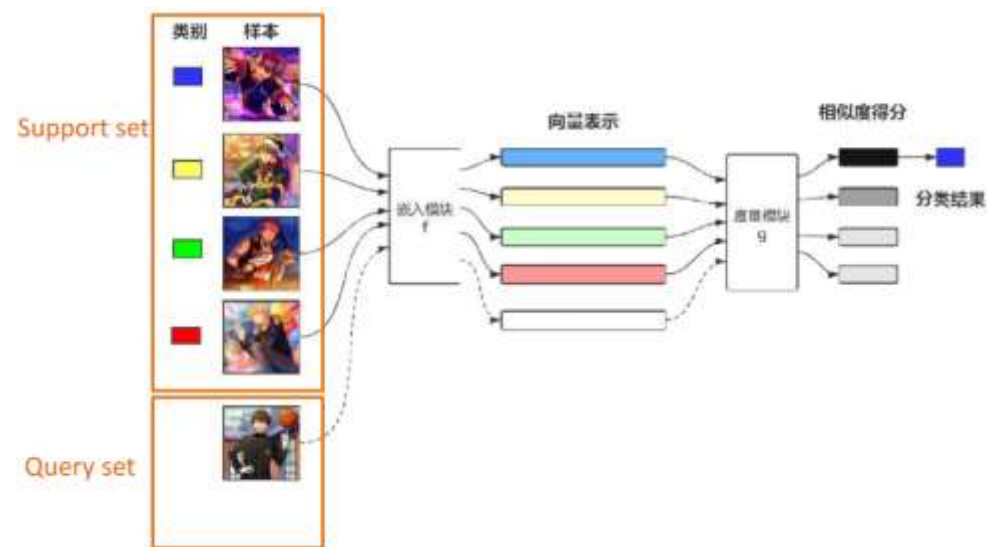


Test task 1 . . .

Support set



Query set



CONTaiNER



【CONTaiNER】



TIPO PCDL

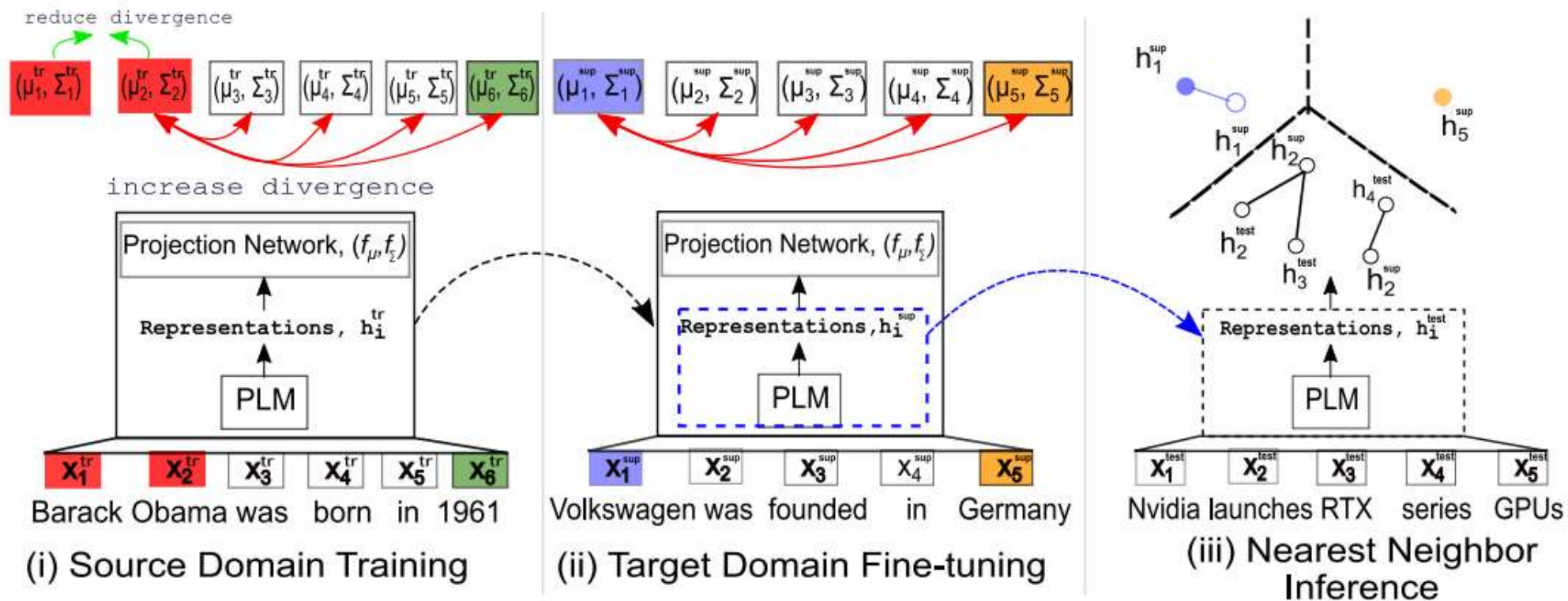
T	目标	仅使用少量标注数据进行命名实体识别
I	输入	少量标注了实体位置和类别的句子和待预测的无标注句子 (Few-NERD数据集, 句子188.2K条, 实体188K个)
P	处理	<ol style="list-style-type: none">1. 使用预训练语言模型编码得到序列中token的中间表示2. 生成token的高斯嵌入3. 使用对比损失目标函数在源域训练4. 使用支持集样本在目标域微调模型5. 使用实例级最近邻推理
O	输出	句子中的实体位置和其类别标签

P	问题	在源域上训练容易过拟合, 在目标域上泛化性差
C	条件	存在少量标注语料
D	难点	如何优化token间分布距离而不是只优化特定于类别的属性
L	水平	CCF-A ACL 2022



- 算法原理图

CONTaiNER



Source Tags: ■ PER ■ DATE Target Tags: ■ ORG ■ LOCATION



框架与原理

• 高斯嵌入

– 预训练编码器编码 (BERT)

- 对于n个token的序列, 得到中间表示为

$$[\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_n] = \text{PLM}([x_1, x_2, \dots, x_n])$$

– 高斯分布嵌入

- 映射得到均值和方差, 得到高斯分布嵌入

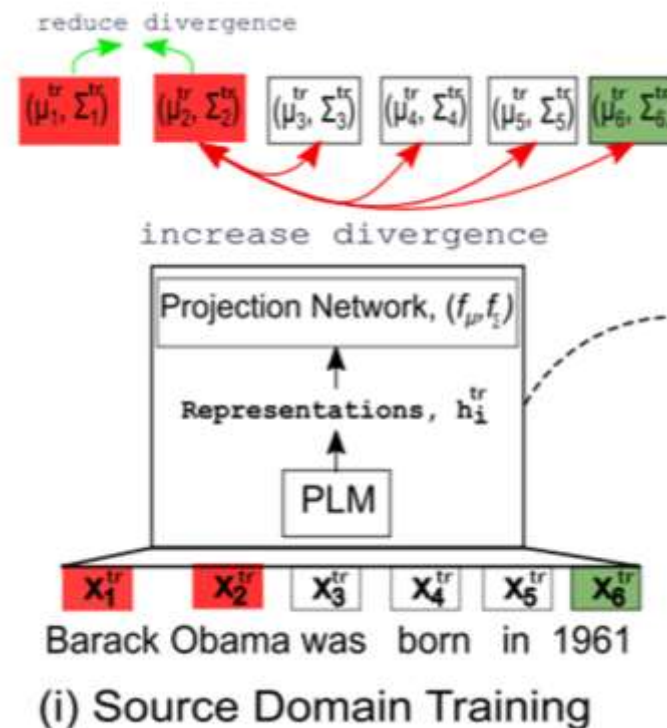
$$\boldsymbol{\mu}_i = f_{\mu}(\mathbf{h}_i), \boldsymbol{\Sigma}_i = \text{ELU}(f_{\Sigma}(\mathbf{h}_i)) + (1 + \epsilon)$$

- 点嵌入简单优化样本距离, 高斯嵌入通过方差估计显式地建模实体类分布, 不仅促进了广义特征表示, 而且有助于小样本目标域的自适应
- 先前工作表明:

- 映射到密度空间捕获了表示的不确定性

- 高斯嵌入之间的KL散度允许考虑不对称距离, 更好地代表了包含、相似或者蕴含关系, 保留了单词间的层次结构

Source Tags: ■ PER ■ DATE





框架与原理

- 源域训练

- 计算token对之间的KL散度

$$D_{KL}[N_q || N_p] = [N(\mu_q, \Sigma_q) || N(\mu_p, \Sigma_p)]$$

$$d(p, q) = \frac{1}{2} (D_{KL}[N_q || N_p] + D_{KL}[N_p || N_q])$$

- 对比学习

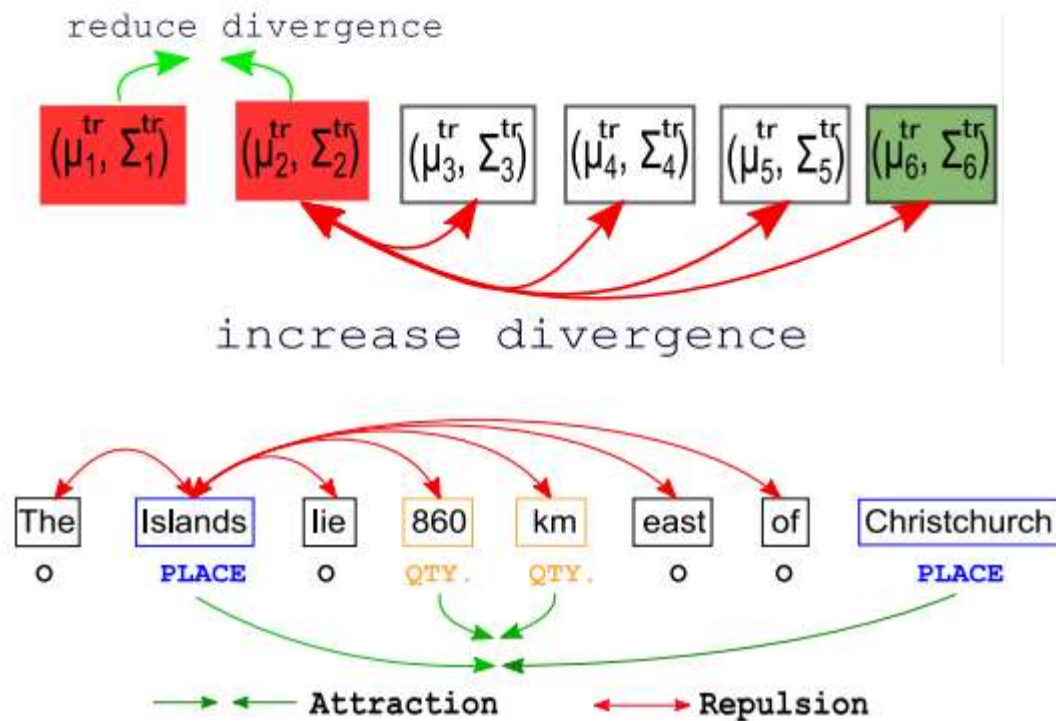
$$l(p) = -\log \frac{\sum_{(x_q, y_q) \in X_p} \frac{\exp(-d(p, q))}{|X_p|}}{\sum_{(x_q, y_q) \in X, p \neq q} \exp(-d(p, q))}$$

$$X_p = \{(x_q, y_q) \in X | y_p = y_q, p \neq q\}$$

- 目的：减少相似实体token嵌入距离，增加不相似实体token嵌入距离

- 意义：1) 更好地捕获**标签依赖**关系

- 2) 有效避免先前方法的**O-tokens陷阱**：O类可能是测试集中有效目标实体



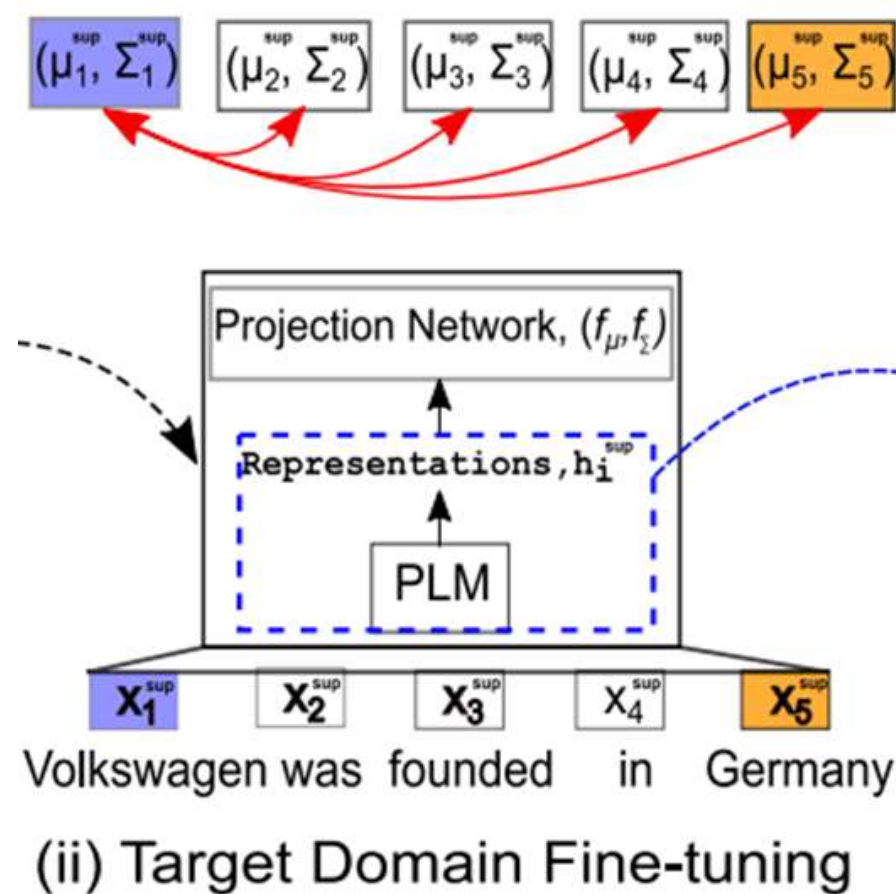


框架与原理

- 目标域微调
 - 5shot场景有效, 1shot场景效果不佳
 - 原因: 模型没有关于目标类的先验知识, 示例难以推断目标类分布的方差
 - 解决: 1shot场景使用嵌入分布均值的欧式距离的平方

$$d'(p, q) = \|\mu_p - \mu_q\|_2^2$$

- 早停策略:
 - 损失阈值, 小于阈值停止微调
 - 防止用小的支持集微调有过拟合的风险



Target Tags: ■ ORG ■ LOCATION



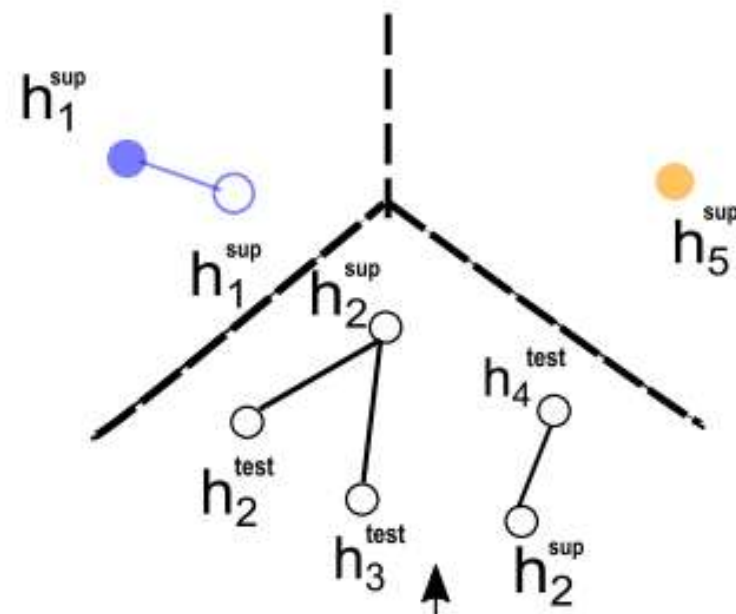
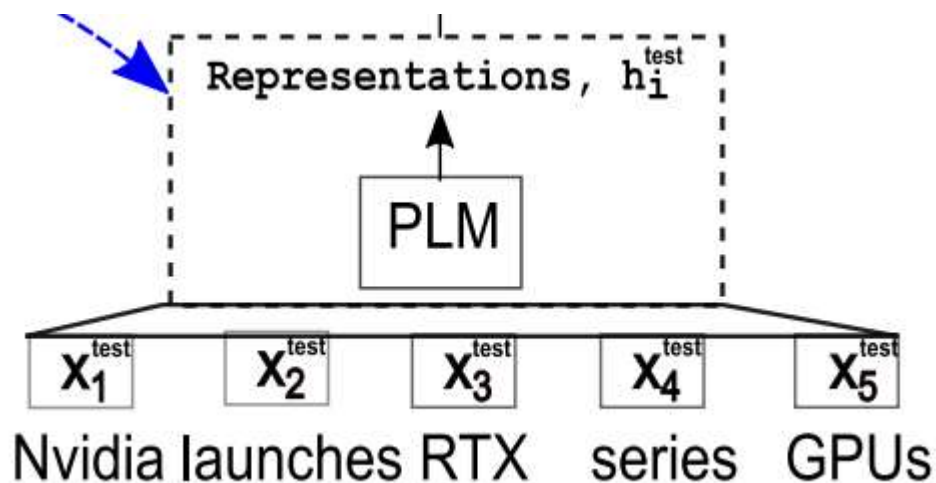
- 实例级最近邻推理

- 使用训练好的PLM得到表示
- 最近邻推理:

Target Tags: ■ ORG ■ LOCATION

- 测试集token的标签为最近的支持集token的标签

$$y_i^{test} = \underset{y_k^{sup} \text{ where } (x_k^{sup}, y_k^{sup}) \in X_{sup}}{\operatorname{argmin}} \|h_i^{test} - h_k^{sup}\|_2^2$$





实验设计

- 数据集

- **Ontonotes 5.0**数据集：通用领域
- **I2B2**数据集：医药领域
- **CoNLL'03**数据集：新闻领域
- **WNUT'17**数据集：社交媒体领域
- **GUM**数据集：采访、新闻、工具性文本、旅游指南
- **Few-NERD**数据集：
 - 大规模专用数据集，66细粒度，8粗粒度

- 实验设计

- 标签集扩展任务
- 域转移任务
- **Few-NERD**设置

Dataset	Domain	# Class	# Sent
OntoNotes	General	18	76K
I2B2'14	Medical	23	140K
CoNLL'03	News	4	20K
WNUT'17	Social	6	5K
GUM	Mixed	11	3.5K
FEW-NERD	Wikipedia	66	188K

Table 1: Summary Statistics of Datasets





- 标签集扩展任务

- 新的实体类型可能出现在相同的现有文本域中

- 使用OntoNotes数据集实验标签集扩展能力，18个实体类分为ABC三组不相交的实体类型集合
- 在每个组上测试，另外2个组作为训练集，测试组的实体被替换为O标签

Model	1-shot				5-shot			
	Group A	Group B	Group C	Avg.	Group A	Group B	Group C	Avg.
Proto	19.3 ± 3.9	22.7 ± 8.9	18.9 ± 7.9	20.3	30.5 ± 3.5	38.7 ± 5.6	41.1 ± 3.3	36.7
NNShot	28.5 ± 9.2	27.3 ± 12.3	21.4 ± 9.7	25.7	44.0 ± 2.1	51.6 ± 5.9	47.6 ± 2.8	47.7
StructShot	30.5 ± 12.3	28.8 ± 11.2	20.8 ± 9.9	26.7	47.5 ± 4.0	53.0 ± 7.9	48.7 ± 2.7	49.8
CONTaiNER	32.2 ± 5.3	30.9 ± 11.6	32.9 ± 12.7	32.0	51.2 ± 5.9	55.9 ± 6.2	61.5 ± 2.7	56.2
+ Viterbi	32.4 ± 5.1	30.9 ± 11.6	33.0 ± 12.8	32.1	51.2 ± 6.0	56.0 ± 6.2	61.5 ± 2.7	56.2

- 结论：各个情况优于对比算法，表现良好



迁移任务

- 域转移任务
 - 源域上训练的模型->新的文本域
 - **OntoNotes**作为源文本域，评估**I2B2**，**CoNLL**，**WNUT**，**GUM**上的性能

Model	1-shot					5-shot				
	I2B2	CoNLL	WNUT	GUM	Avg.	I2B2	CoNLL	WNUT	GUM	Avg.
Proto	13.4 ± 3.0	49.9 ± 8.6	17.4 ± 4.9	17.8 ± 3.5	24.6	17.9 ± 1.8	61.3 ± 9.1	22.8 ± 4.5	19.5 ± 3.4	30.4
NNShot	15.3 ± 1.6	61.2 ± 10.4	22.7 ± 7.4	10.5 ± 2.9	27.4	22.0 ± 1.5	74.1 ± 2.3	27.3 ± 5.4	15.9 ± 1.8	34.8
StructShot	21.4 ± 3.8	62.4 ± 10.5	24.2 ± 8.0	7.8 ± 2.1	29.0	30.3 ± 2.1	74.8 ± 2.4	30.4 ± 6.5	13.3 ± 1.3	37.2
CONTaiNER	16.4 ± 1.7	57.8 ± 10.7	24.2 ± 2.9	17.9 ± 1.8	29.1	24.1 ± 1.9	72.8 ± 2.0	27.7 ± 2.2	24.4 ± 2.2	37.3
+ Viterbi	21.5 ± 1.7	61.2 ± 10.7	27.5 ± 1.9	18.5 ± 4.9	32.2	36.7 ± 2.1	75.8 ± 2.7	32.5 ± 3.8	25.2 ± 2.7	42.6

– 结论:

- 大多数情况优于对比算法
- 尤其是有挑战性的**GUM**数据集，其他对比算法表现很糟糕，**CONTaiNER**好得多
- **CoNLL**较好的原因是：所有其他领域与**OntoNotes**几乎没有交集，**CoNLL**中的目标实体包含在**OntoNotes**中



• Few-NERD设置

– INTRA: 根据粗粒度类型划分

- 训练集: 属于People, Art, Product, MISC

粗粒度类型的细粒度实体类型

- 验证集: Event, Building

- 测试集: ORG, LOC

- 粗粒度类型3个集之间没有重叠, 更有挑战性

– INTER:

- 粗粒度类型共享, 细粒度互不相交

– 结论:

- 在有挑战性的INTRA场景上, 表现良好
- 各个任务, 尤其是5shot情况,

– 通过对比高斯嵌入优化的分布建模, CONTaiNER能够更好利用多个少样本

Model	5-way		10-way		Avg.
	1~2 shot	5~10 shot	1~2 shot	5~10 shot	
StructShot	35.92	38.83	25.38	26.39	31.63
ProtoBERT	23.45	41.93	19.76	34.61	29.94
NNShot	31.01	35.74	21.88	27.67	29.08
CONTaiNER	40.43	53.70	33.84	47.49	43.87
+ Viterbi	40.40	53.71	33.82	47.51	43.86

Table 4: F1 scores in FEW-NERD (INTRA).

Model	5-way		10-way		Avg.
	1~2 shot	5~10 shot	1~2 shot	5~10 shot	
StructShot	57.33	57.16	49.46	49.39	53.34
ProtoBERT	44.44	58.80	39.09	53.97	49.08
NNShot	54.29	50.56	46.98	50.00	50.46
CONTaiNER	55.95	61.83	48.35	57.12	55.81
+ Viterbi	56.1	61.90	48.36	57.13	55.87

Table 5: F1 scores in FEW-NERD (INTER).



实验二：微调的作用和目标

– 是否需要微调：在标签集扩展任务上案例研究

- 结论：微调提高了性能，**5shot**更明显
- 原因：微调可以最大限度利用**目标域中的少量样本**

– 微调的目标：**KL-散度** or 欧氏距离？

- 现象：前面实验发现使用**KL-散度**，**1shot**效果不好
- 分析原因：模型**没有关于目标类的先验知识**，**1shot**示例可能无法提供关于目标类分布的足够信息

• 解决：

- **1shot**用均值嵌入的欧式距离作为优化的目标
- **5shot**用高斯嵌入的**KL-散度**，因为能更好利用多样本

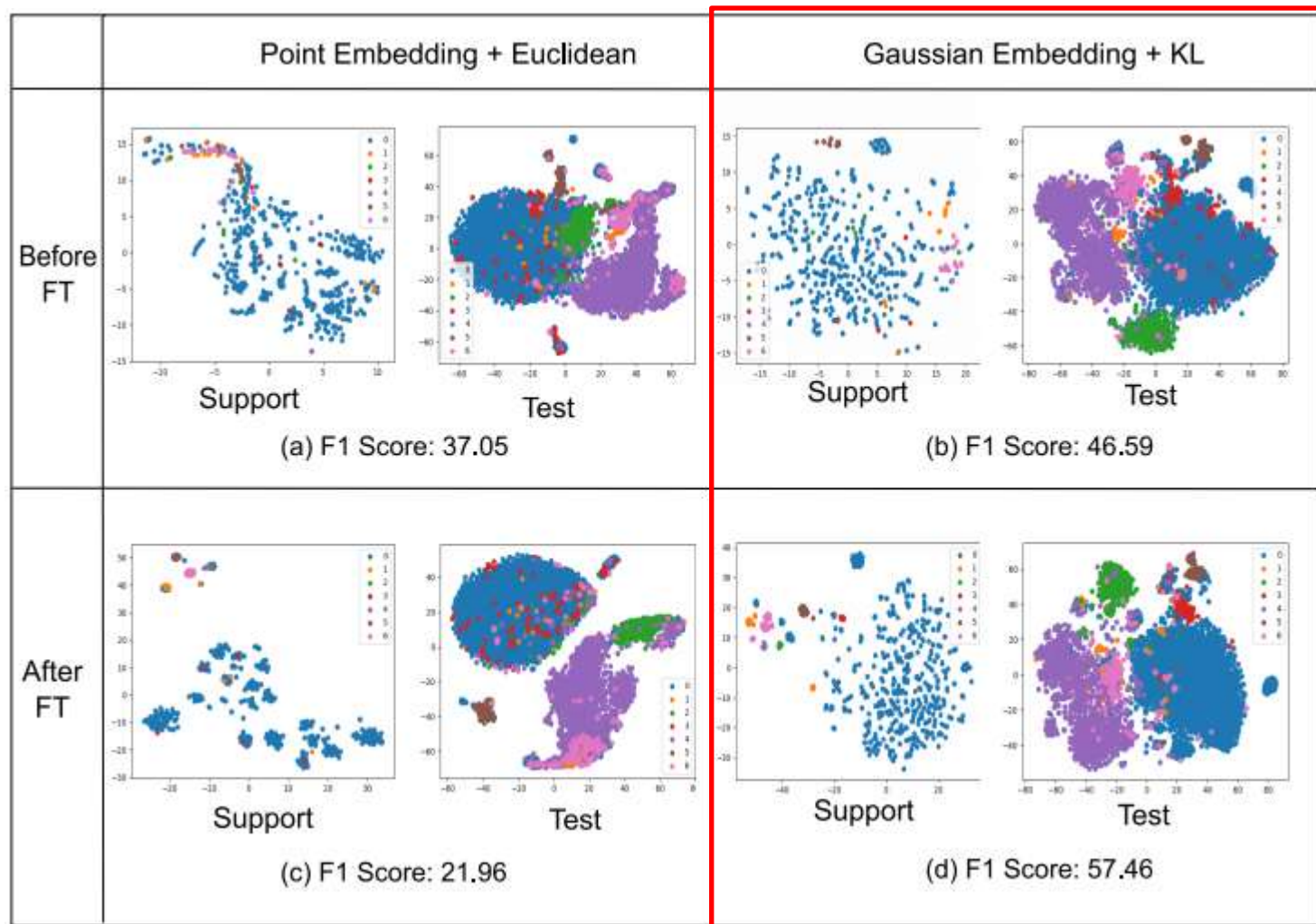
	W/O Finetuning	W/ Finetuning
1-shot	31.76	32.90
5-shot	56.99	61.48

	KL-Gaussian	Euclidean-mean
1-shot	18.78	27.48
5-shot	32.50	31.12



实验结果

- 实验三：
 - 欧氏距离+点嵌入
 - 不同类支持示例会在一起，意味着较差的泛化
 - 微调后，欧氏距离会过度拟合小的支持数据，损害性能
 - 高斯嵌入
 - 在微调前后有更好性能
 - 微调前，对没见过的类已经有较好的泛化能力
 - 微调后，集群分离更清晰





优劣分析

- 优势
 - 将对比学习技术引入小样本命名实体识别领域
 - 优化了小样本NER的token间分布距离，有效缓解了源自训练域的过拟合问题
- 劣势
 - 整体性能仍然很低，没有准备好部署在高风险领域（如医学，I2B2数据集）
 - 尤其是1shot的效果提高较小

COPNER



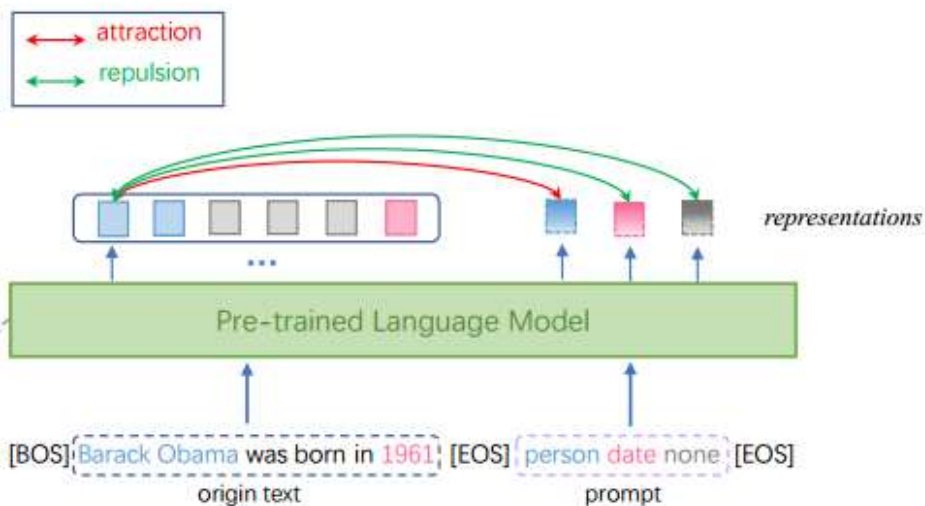
【COPNER】



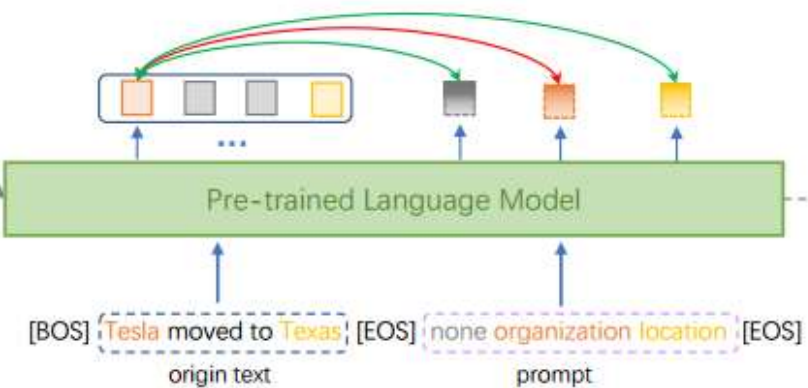
TIPO BCDL

T	目标	仅使用少量标注数据进行命名实体识别
I	输入	少量标注了实体位置和类别的句子和待预测的无标注句子 (Few-NERD数据集, 句子188.2K条, 实体188K)
P	处理	<ol style="list-style-type: none">1. 构造提示符, 得到扩展的输入序列2. 使用预训练语言模型编码得到序列表示3. 计算对比损失4. 目标标签空间微调5. 最近的类特定词推理
O	输出	句子中的实体位置和其类别标签

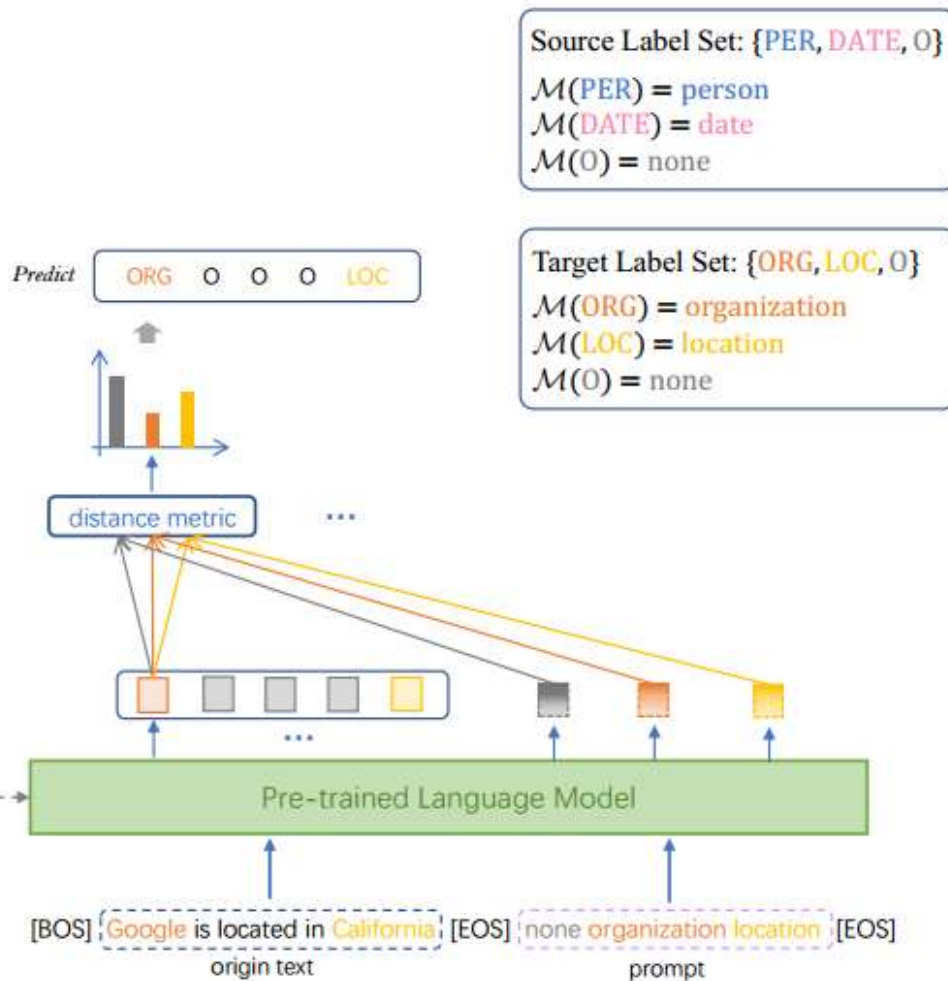
P	问题	度量参照物可能远远不能表示相应的实体类
C	条件	存在少量标注语料
D	难点	如何找到更合适和稳定的度量参照物
L	水平	CCF-B COLING 2022



(a) Training in source label space



(b) Adapting to target label space with support set



(c) Inference in target label space



提示引导

- 构造提示符

- 定义类特定词的映射 M

- 目的：为标签集中每个标签得到一个类特定的词 v_i
- 使用给定实体的类名作为CWs (class-specific words)

$$M(LOC) = \text{location}$$

- 优点

- 包含相关实体类的一般语义信息
- 避免来自有限标注数据的偏差

- 构造提示符，得到扩展的输入序列

- t 为原始输入句子长度， n 为实体类个数，还添加了 v_{n+1} 表示非实体类

$$X' = \{x_1, x_2, \dots, x_t, v_1, \dots, v_n, v_{n+1}\}$$

Few-NERD	
#Class	#Class-specific word
location-bodiesofwater	water
location-island	island
person-athlete	athlete
person-director	director
organization-show	show
organization-company	company
building-airport	airport
building-hospital	hospital
art-painting	painting
art-film	film
...	...



源标签空间训练

- 源标签空间训练
 - 生成上下文表示

$$\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_t, \mathbf{h}'_1, \dots, \mathbf{h}'_n]$$

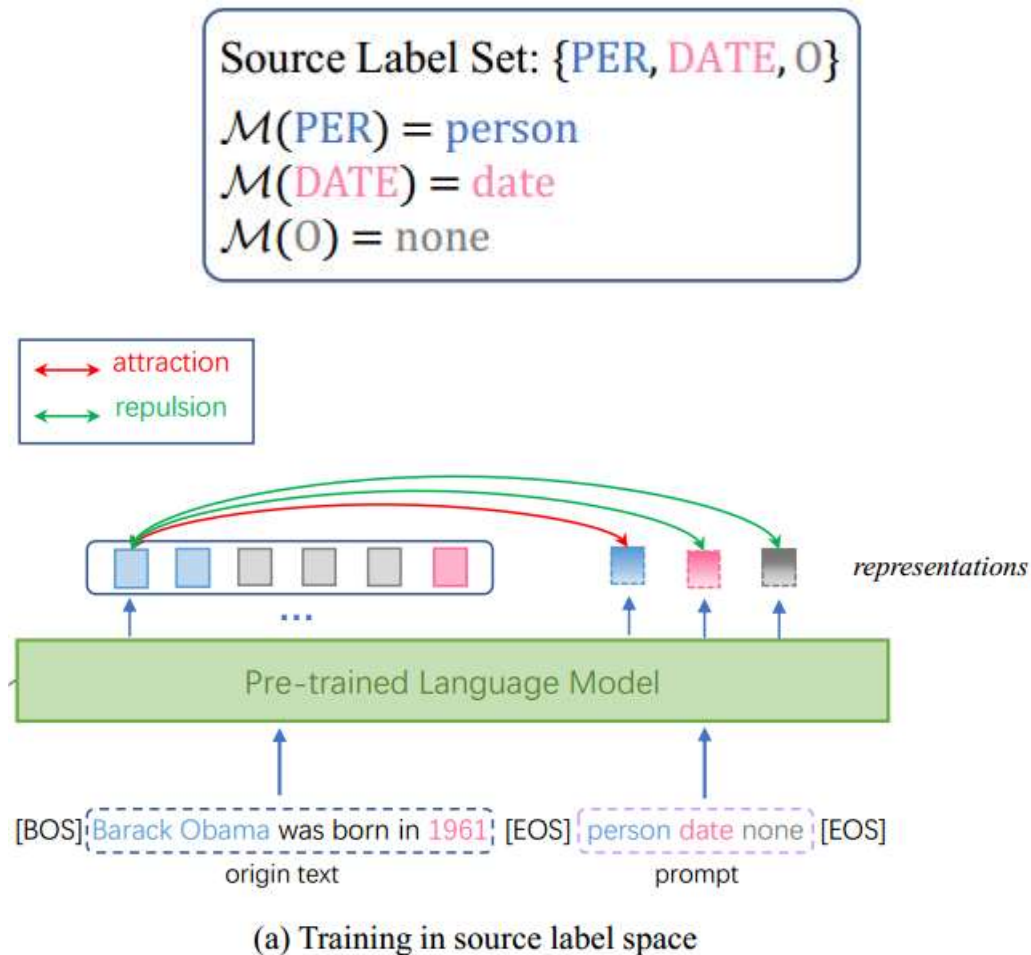
$$= \text{PLM}([x_1, x_2, \dots, x_t, v_1, \dots, v_n])$$
 - 提示符中CWs的表示可以作为训练原始句子中token表示的引导

– 计算对比损失：拉近相关CWs，拉远不相关CWs

- 构建正负对：

正对 (x_p, v_p) ，负对是 x_p 和其他类CWs

$$l(x_p) = -\log \frac{\exp(-d(\mathbf{h}_p, \mathbf{h}'_p)/\tau)}{\sum_{q=1}^{n+1} \exp(-d(\mathbf{h}_p, \mathbf{h}'_q)/\tau)}$$





微调&推理

- 目标标签空间微调
 - 使用支持集微调, 适应目标标签空间
 - 早停策略: 损失阈值, 小于阈值停止微调
 - 防止模型适应不足或过拟合

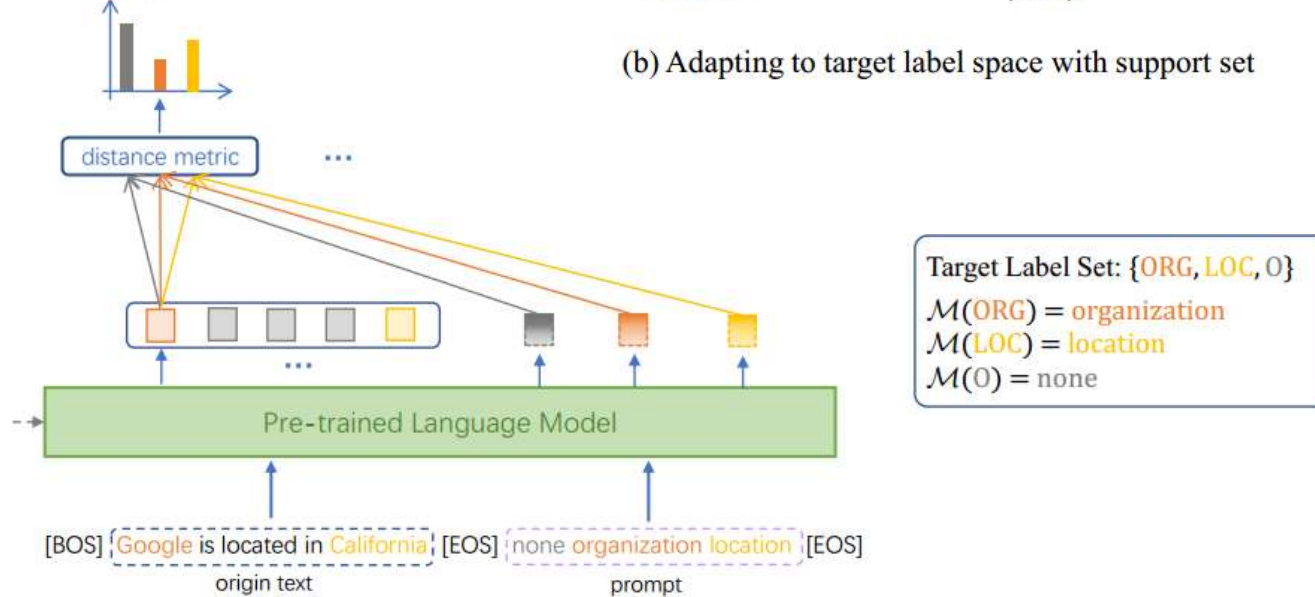
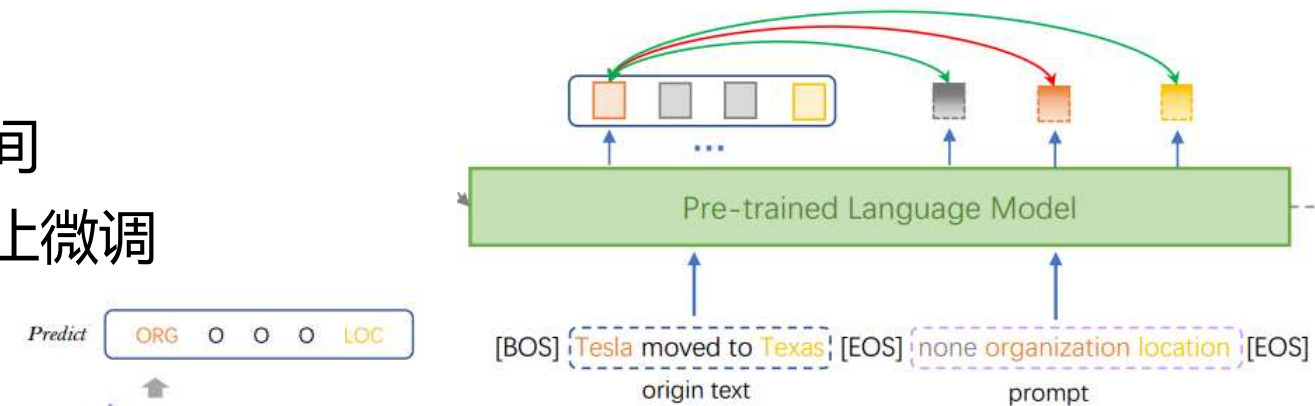
- 推理

- 得到测试样本序列表示

$$H^{test} = [\mathbf{h}_1, \dots, \mathbf{h}_t, \mathbf{h}'_1, \dots, \mathbf{h}'_n, \mathbf{h}'_{n+1}]$$

- 对于每个token, 分配给距离它最近的CW的标签 c_j

$$y_i^{test} = \underset{c_j}{\operatorname{argmin}} \|h_i - h'_j\|_2^2$$



Target Label Set: {ORG, LOC, O}
 $\mathcal{M}(\text{ORG}) = \text{organization}$
 $\mathcal{M}(\text{LOC}) = \text{location}$
 $\mathcal{M}(O) = \text{none}$



数据标注

- 数据集

- **OntoNotes5.0** 数据集：通用领域
- **I2B2** 数据集：医药领域
- **CoNLL'03** 数据集：新闻领域
- **WNUT'17** 数据集：社交媒体领域
- **MIT-Movie** 数据集：评论领域
- **Few-NERD** 数据集：大规模专用数据集，**66**细粒度，**8**粗粒度

Datasets	Domain	#Class	#Sent	#Entity
OntoNotes	General	18	76.7k	104.2k
WNUT'17	Social	6	5.7k	3.9k
I2B2'14	Medical	23	140.8k	29.2k
CoNLL'03	News	4	20.7k	35.1k
MIT-Movie	Review	12	12.2k	26.6k
Few-NERD	General	66	188.2k	491.7k

- 3种评估设置

- 跨标签空间
- 域转移
- 同标签空间





实验结果

- 实验一：对比实验-跨标签空间
 - Few-NERD数据集的2个不同任务
 - **INTER**: 粗粒度类别在训练、测试集共享，细粒度不共享
 - **INTRA**: 不同集合汇总细粒度实体类属于不同的粗粒度类别

Model	5-way		10-way		Avg.
	1~2 shot	5~10 shot	1~2 shot	5~10 shot	
ProtoBERT	44.44	58.80	39.09	53.97	49.08
NNShot	54.29	50.56	46.98	50.00	50.46
StructShot	57.33	57.16	49.46	49.39	53.34
CONTaiNER	55.95	61.83	48.35	57.12	55.81
+Struct	56.10	61.90	48.36	57.13	55.87
COPNER	65.39	67.59	59.69	62.32	63.75
+Struct	65.98	67.70	59.56	62.37	63.90

Model	5-way		10-way		Avg.
	1~2 shot	5~10 shot	1~2 shot	5~10 shot	
ProtoBERT	23.45	41.93	19.76	34.61	29.94
NNShot	31.01	35.74	21.88	27.67	29.08
StructShot	35.92	38.83	25.38	26.39	31.63
CONTaiNER	40.43	53.70	33.84	47.49	43.87
+Struct	40.40	53.71	33.82	47.51	43.86
COPNER	53.52	58.74	44.13	51.55	51.99
+Struct	54.26	58.84	44.26	51.18	52.14

- 结论：在2个任务上f1值均明显优于对比算法；尤其是1shot情况优势更明显
 - 原因是1shot可能不能给出关于目标类分布的足够信息，COPNER中的CWs补充了与类相关的语义信息



实验内容

- 实验一：对比实验-域转移
 - 在通用领域OntoNotes5.0数据集上训练
 - 在I2B2, CoNLL'03, WNUT'17这3个数据集上微调后测试

Model	1 shot				5 shot			
	CoNLL	WNUT	I2B2	Avg.	CoNLL	WNUT	I2B2	Avg.
ProtoBERT	49.9±8.6	17.4±4.9	13.4±3.0	26.9	61.3±9.1	22.8±4.5	17.9±1.8	34.0
NNShot	61.2±10.4	22.7±7.4	15.3±1.6	33.1	74.1±2.3	27.3±5.4	22.0±1.5	41.1
StructShot	62.4±10.5	24.2±8.0	21.4±3.8	36.0	74.8±2.4	30.4±6.5	30.3±2.1	45.2
CONTaiNER	57.8±10.7	24.2±2.9	16.4±1.7	32.8	72.8±2.0	27.7±2.2	24.1±1.9	41.5
+Struct	61.2±10.7	27.5±1.9	21.5±1.7	36.7	75.8±2.7	32.5±3.8	36.7±2.1	48.3
COPNER	67.0±3.8	33.8±2.5	34.6±1.8	45.1	74.9±2.9	34.8±3.1	41.1±1.6	50.2
+Struct	66.5±2.1	34.9±1.8	35.8±1.3	45.7	74.6±3.1	34.2±2.6	43.7±1.5	50.8

- 结论：在大多数情况下f1值优于对比算法；尤其是1shot情况优势更明显



实验设置

- 实验一：对比实验-同标签空间
 - 通过限制可用的训练样本来检验COPNER的泛化能力

Model	CONLL					MIT-Movie				
	5 shot	10 shot	20 shot	50 shot	Avg.	5 shot	10 shot	20 shot	50 shot	Avg.
BERT-tagger	41.9±12.1	59.9±10.7	68.7±5.1	73.2±3.1	60.9	39.6±6.4	50.6±7.3	59.3±3.7	71.3±3.0	55.2
NNShot	42.3±8.9	59.2±11.7	66.9±6.1	72.6±3.4	60.3	39.0±5.5	50.5±6.1	58.9±3.5	71.2±2.9	54.9
StructShot	45.8±10.3	62.4±11.0	69.5±6.5	74.7±3.1	63.1	41.6±9.0	53.2±5.5	61.4±3.0	72.0±6.4	57.1
TemplateNER	43.0±6.2	57.9±5.7	66.4±6.1	72.7±2.1	60.0	46.0±3.9	49.3±3.4	59.1±0.4	65.1±0.2	54.9
EntLM	49.5±8.3	64.8±3.9	69.5±4.5	73.7±2.1	64.4	46.6±9.5	57.3±3.7	62.4±4.1	71.9±1.7	59.6
+Struct	51.3±7.7	66.9±3.0	71.2±3.9	74.8±1.9	66.1	49.2±8.9	59.2±4.0	63.9±3.7	73.0±1.8	61.3
COPNER	54.9±4.1	65.3±2.4	70.7±1.8	75.0±1.5	66.5	50.9±4.4	59.7±0.4	66.7±1.8	73.8±0.6	62.8
+Struct	54.2±7.9	66.2±2.9	71.8±1.8	77.0±1.4	67.3	50.1±3.6	61.9±1.4	68.9±2.4	74.6±0.3	63.9

- 结论：在大多数情况下f1值优于对比算法；并且f1值的标准差低于其他算法，表明COPNER比这些基线更稳定



实验设计

- 实验二：零样本能力实验

- 在一个资源丰富的数据集上训练后，COPNER学习了CWs和输入句子的标记之间隐藏的上下文关联
- 猜想：即使没有任何支持数据，这些习得的上下文关联也有助于COPNER对未见过的实体类别进行分类
- 在跨标签空间任务和域转移任务上实验

Model	1 shot		
	CoNLL	WNUT	I2B2
ProtoBERT	49.9±8.6	17.4±4.9	13.4±3.0

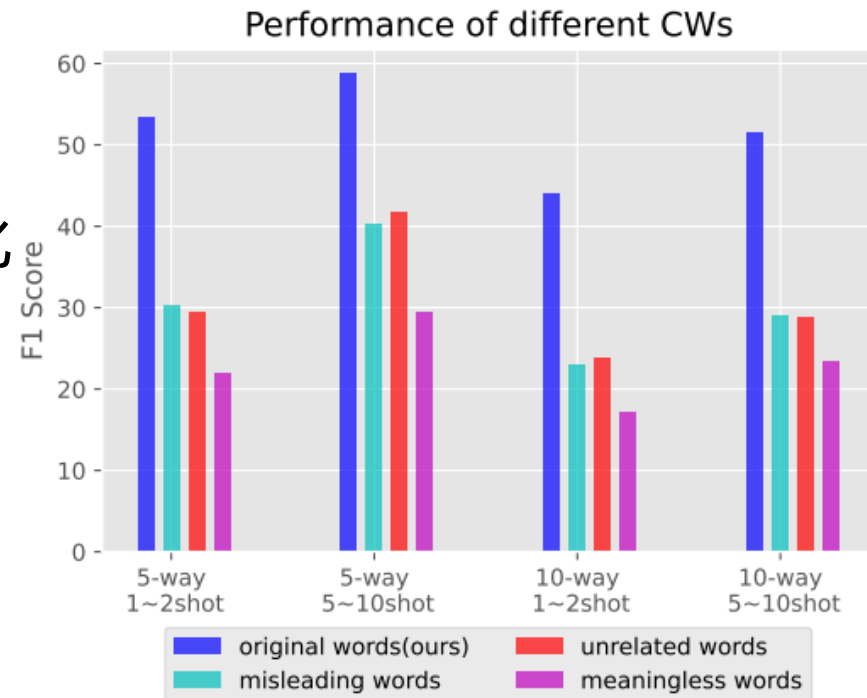
Model	Few-NERD				Domain Transfer		
	INTER		INTRA		CONLL	WNUT	I2B2
	5 way	10 way	5 way	10 way	-	-	-
COPNER	31.95	19.52	14.72	8.73	46.26	17.58	17.29
+Struct	33.97	20.92	16.06	9.64	49.39	17.41	17.47

- 结论：具有良好性能，在域转移任务中，0shot设置下甚至可以与1shot下的原型网络相媲美



实验三

- 实验三：类特定词 (CWs) 的影响
 - 目的：为了证明类名语义信息的影响
 - Few-NERD (INTRA) 数据集, f1值与不同CWs的变化
 - 误导词 (misleading)：在标签间随机交换CWs:
 $M(PER) = location, M(ORG) = person$
 - 不相关词 (meaningless)：在词汇表中随机选择
 - 无意义的词 (unrelated)：[unused0], [unused1]



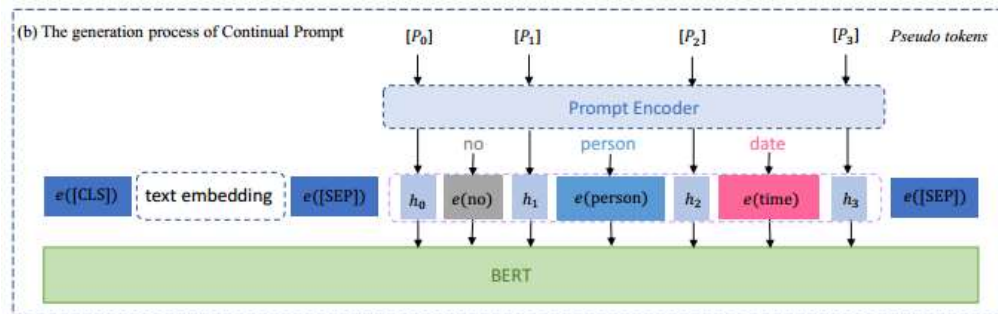
- 结论:

- 与原始CWs相比, 3个变体性能明显下降: 原始词>>误导词和不相关词>无意义词
- 误导词和不相关词中的错误语义信息可能会误导实体表示学习, 造成巨大性能损失
- 证明了匹配实体类的语义作为度量锚点更有效
- 无意义词进一步下降表明语义信息在小样本度量学习中是至关重要的



实验设置

- 实验四：探究不同形式提示的效果
 - 条件：不同形式的提示对基于提示的方法有不同的影响
 - 3种提示符
 - 队列提示符：直接以随机顺序组合CWs
 - 分区提示符：使用额外的token分割每个CWs
 - 连续提示符：使用连续表示分割CWs



[CLS] Barack Obama was born in 1961 [SEP] no person date [SEP]
Queue Prompt

[CLS] Barack Obama was born in 1961 [SEP] [S] no [S] person [S] date [S] [SEP]
Partition Prompt

[CLS] Barack Obama was born in 1961 [SEP] Prompt [SEP]
Continual Prompt



实验四

• 实验四：探究引入提示的影响

– 3种提示符

- **FG**: 不加任何提示的基线
- **QP**: 队列提示符
- **PP**: 分区提示符
- **CP**: 连续提示符

Model	5-way		10-way		Avg.
	1~2 shot	5~10 shot	1~2 shot	5~10 shot	
FG	43.55	51.85	37.49	48.79	45.42
QP	52.15	57.34	42.79	50.99	50.82
PP	53.52	58.74	44.13	51.55	51.99
CP	53.38	58.81	44.40	51.63	52.06

• 结论:

- 3种方法都比FG好的多，这表明当模型执行实体表示计算时，提示在提供与类别相关的信息方面是有效的，提示符的引入可以有效提高模型能力
- PP和CP都有很好的性能，但是后者引入了额外训练参数，所以主要实验中选择PP方法
- QP性能较差表明，将CWs放在一起会对模型性能产生负面影响

实验设计

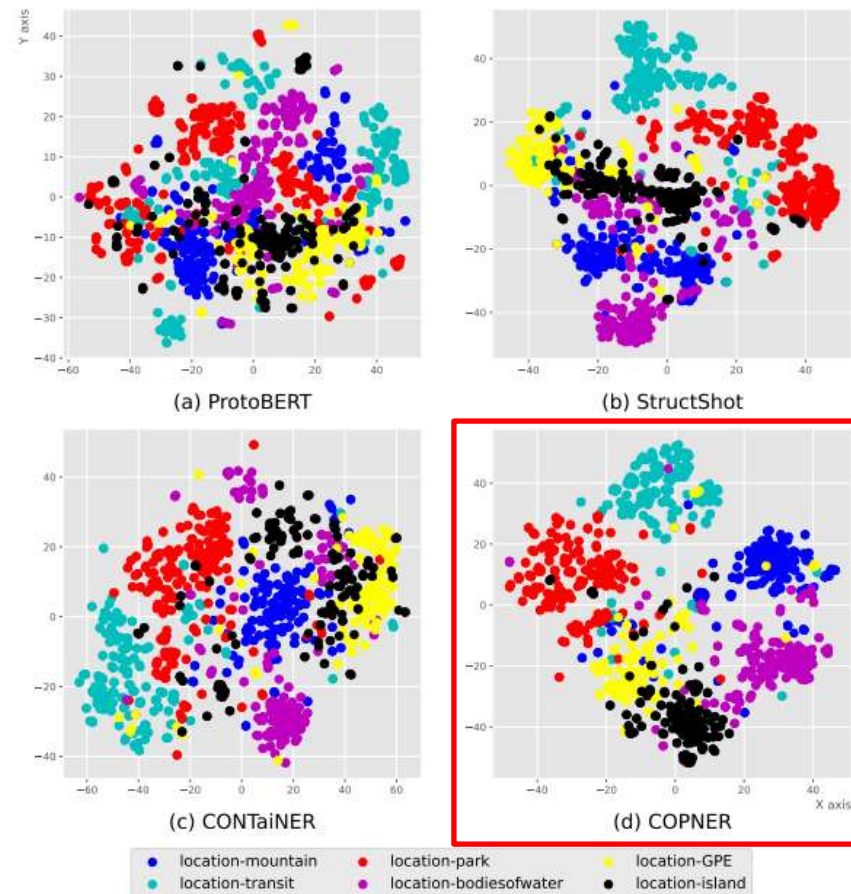
- 实验五：有效性分析

- 是否能增强实体表示？

- **Few-NERD (INTRA)** 测试集中，从位置类别中的6个细粒度类中随机抽取示例，可视化不同方法获得的嵌入

- 结论：

- **ProtoBERT**表示能力最弱，无法区分不同类别实体表示
- **StructShot**改进了，但倾向于将同一类的实体表示分发到多个集群
- **CONTaiNER**进一步增强了实体表示，但在一些类中很弱
- **COPNER**表现最好，**不同类别的实体分布差异更大，相似的实体分布更聚集**





实验五：有效性分析

– 探究在推理过程中度量参照物对不同模型的影响

• 类别判别能力

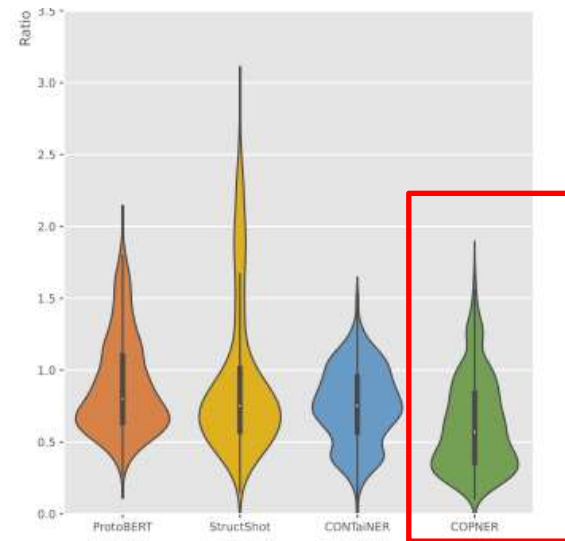
– 在基于距离的推理中，当测试样本离正度量参照物近，离负度量参照物远时，测试样本容易被正确区分

– 结论：COPNER对正负对区分能力更强，正负距离比更低，说明最接近的类特定词推理更合适

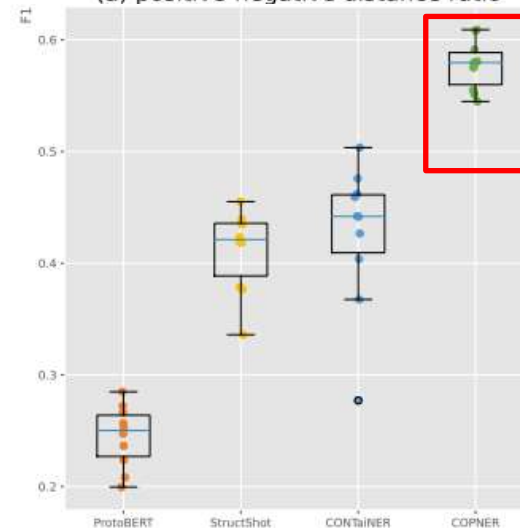
• 度量稳定性

– 在基于距离的度量方法中，一个好的度量推理结果应该对不同的支持集不敏感，选不同的支持集进行测试，看f1值

– 结论：COPNER中最接近的类特定词推理最稳定，而最近邻推理对支持集差异敏感，导致预测结果差异较大



(a) positive-negative distance ratio



(b) F1 scores with 10 different support sets



优劣分析

- 优势
 - 利用对比学习和提示引导的新颖方法获得增强的实体表示和稳定的度量参照物
 - 处理零样本NER任务的能力
- 劣势
 - 添加了一个额外的特定于类的词来表示非实体，处理过于简单粗糙，O类嘈杂
 - 在自适应数据上微调存在过拟合的问题



应用总结



- 应用总结

- 对话系统

- 针对用户回答，通过命名实体识别找到需要的信息
 - 如，地点实体，时间实体，根据用户回答去查数据库

- 机器翻译

- 需要准确识别文本中的命名实体再对其进行正确翻译

- 知识挖掘、知识应用

- 知识挖掘：用户询问客服商品详情页没有的属性，从客服回复中抽取出来，补充商品图谱数据
 - 知识应用：利用知识库问答去回答用户询问商品属性值的相关问题

出发地和目的地是哪里呢？

从北京出发，到上海

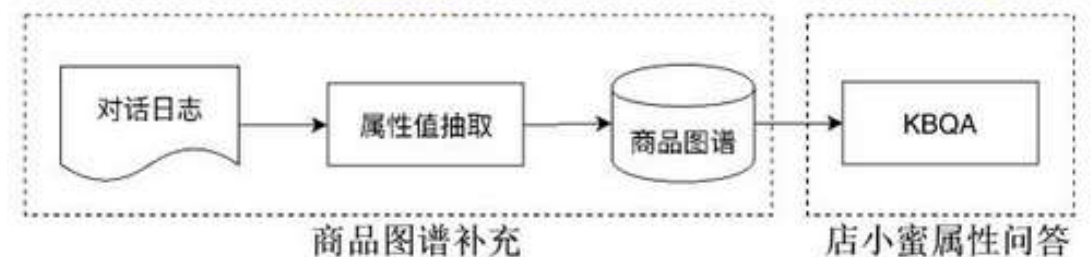
日期呢？打算什么时候出发？

打算11月16日出发，最好晚上

找到了一个晚上9点东航的航班，9:10分出发，10:20到虹桥机场，可以吗？

知识挖掘

知识应用





- [1] Das S S S, Katiyar A, Passonneau R J, et al. CONTaiNER: Few-Shot Named Entity Recognition via Contrastive Learning[C]. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. 2022: 6338-6353.
- [2] Huang Y, He K, Wang Y, et al. Copner: Contrastive learning with prompt guiding for few-shot named entity recognition[C]. Proceedings of the 29th International conference on computational linguistics. 2022: 2515-2527.

谢谢!

大成若缺，其用不弊。大盈若冲，其用不穷。大直若屈。大巧若拙。大辩若讷。静胜躁，寒胜热。清静为天下正。

