

少样本下的槽位提取探索

Exploration on Few-shot Slot Tagging

讲者：侯宇泰

导师：车万翔教授

单位：哈尔滨工业大学 · 哈工大社会计算与信息检索研究中心

主页：<https://atmahou.github.io/>



Few-shot Slot Tagging with Collapsed Dependency Transfer and Label-enhanced Task-adaptive Projection Network

Yutai Hou, Wanxiang Che, Yongkui Lai, Zhihan Zhou, Yijia Liu, Han Liu, Ting Liu



Few-shot Learning 介绍

□ 背景

- 深度学习方法很成功，但是需要大量的标注数据，消耗大量人工
- 人类非常擅长通过少量的样本识别一个新物体，
 - 比如小孩子只需要少量的图片就可以认识什么是斑马，什么是犀牛。



人类并非从头开始学，
从出生开始积累的大量经验辅助学习

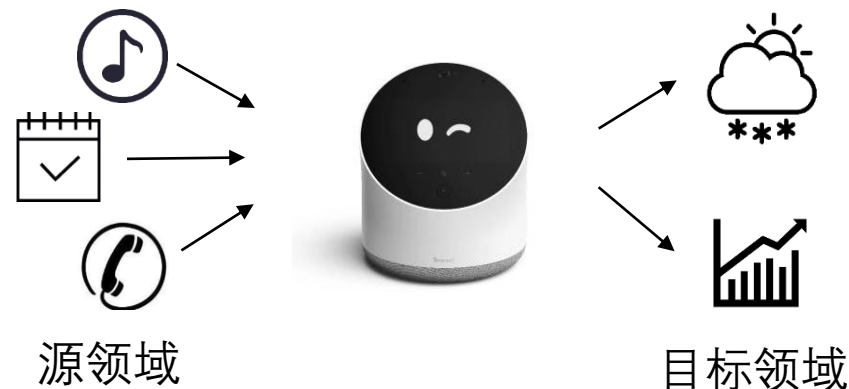
- 我们希望模型也可以利用领域外经验和少量样本进行学习
 - 让人工智能不那么人工，多一点智能



Few-shot Learning 介绍

□ 背景

- 真实应用往往需要频繁适应**新的领域**和**新需求**
 - 新的领域数据不足
 - 标签集的增、删、调整
 - 工业上，频繁重新训练模型的开销往往难以接受的



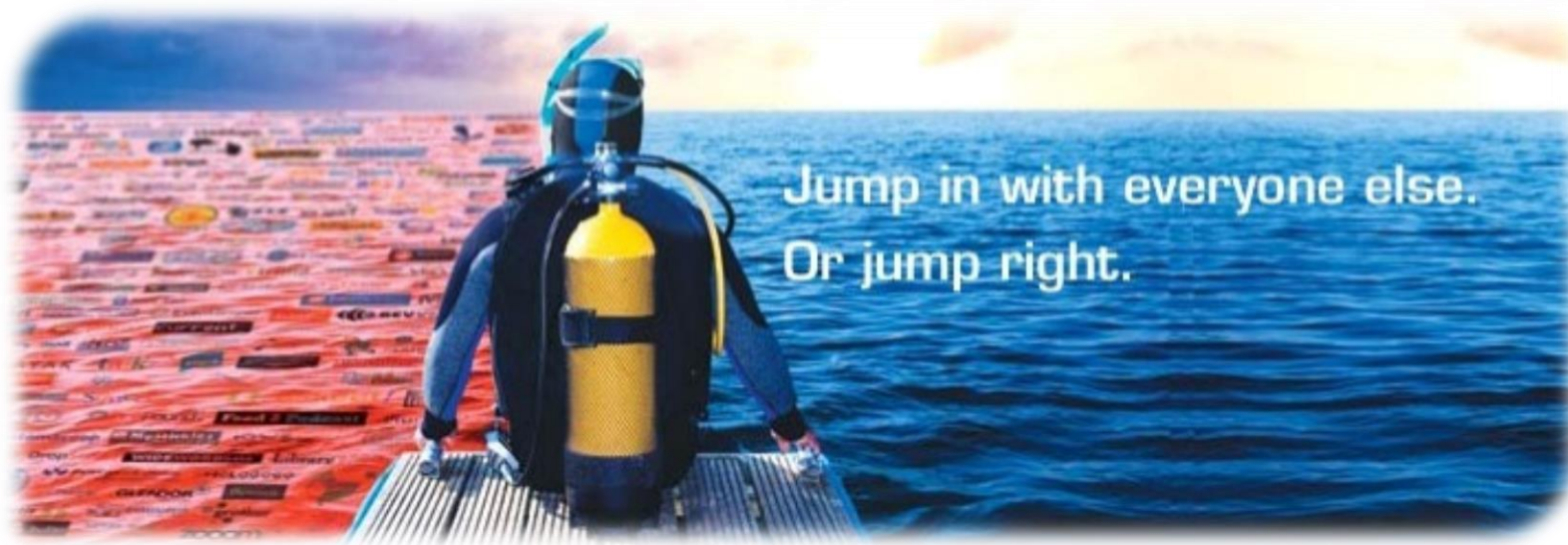
- 需求真实存在，众多公司投入研究精力





Few-shot Learning 介绍

- 少样本学习 (few-shot learning)
 - 少样本学习是专门解决前面提到问题的机器学习分支
 - 一般语境下, Few-shot Learning 指的是 Meta Learning 的一种应用场景
- 在自然语言尤其是对话领域的研究还方兴未艾
 - 针对序列标注的few-shot learning还有少有研究





Few-shot Learning 介绍

□ 任务设置: N-way K-shot

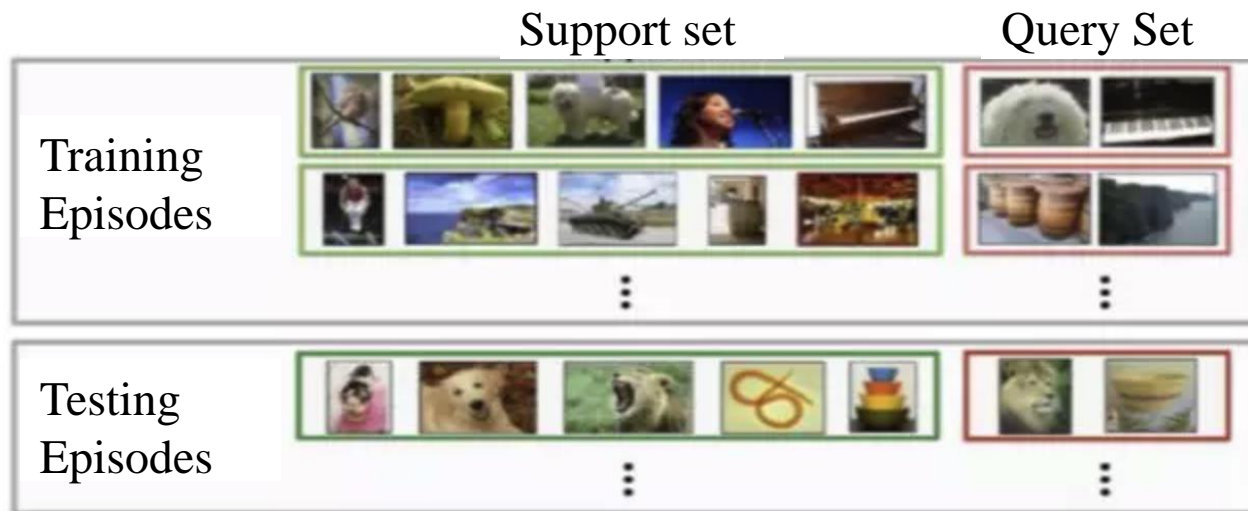
□ Episode —— 一个小样本学习任务

- 支撑集 Support set: **N** 个类别, 每类**K**个有标签样例
- 查询集 query set: 属于**N**类别, 随机**M**个待预测样本

$N \times K$ 个样本

M 个样本

□ Train & Test



训练时模拟数据不足的情况

多个小样本任务测试取平均

$$N=5 - K=1 - M=2$$



Few-shot Learning 介绍

□ 方法分类:

□ Model Based方法

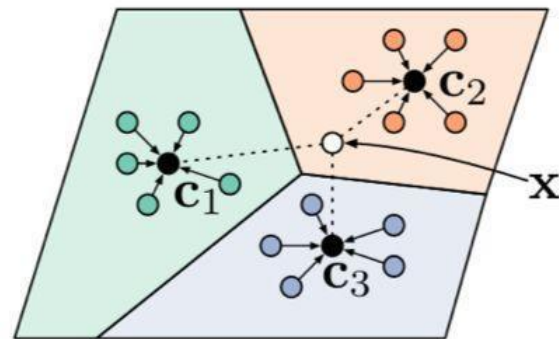
- 设计适用于小样本的模型架构

□ Similarity Based方法

- 利用样本间距离度量来预测新数据类别

□ Optimization Based方法

- 学习多任务通用初始化参数或者参数更新方法

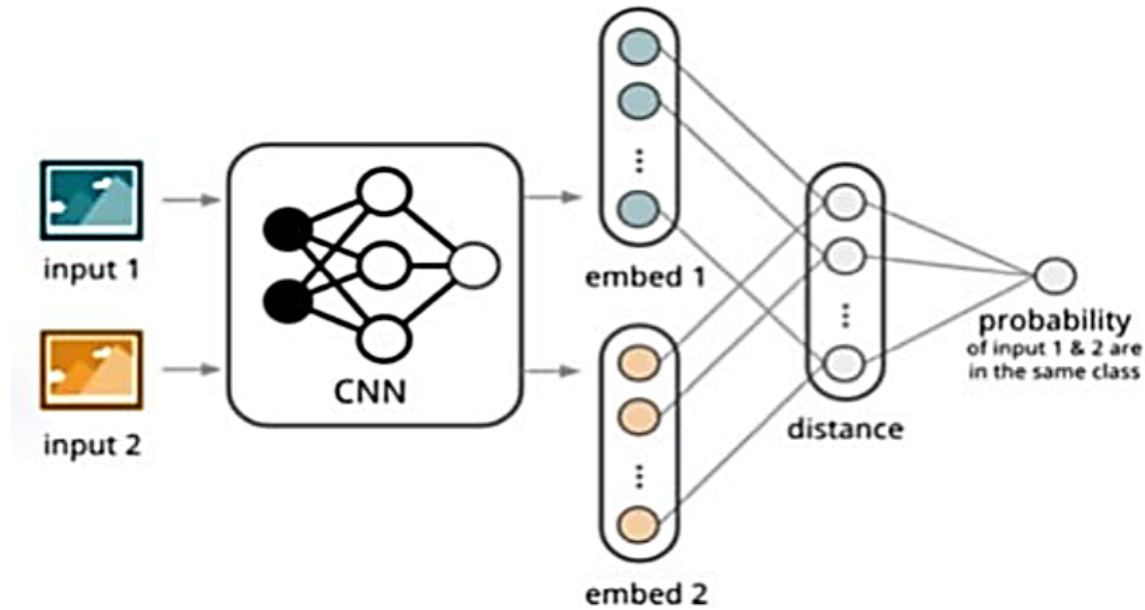


- 采用非参方法学习来避免拟合
- 在富数据领域学习相似度度量函数



Few-shot Learning 介绍

□ 孪生网络 (Siamese Network)



Chopra, S.; Hadsell, R.; LeCun, Y. "Learning a similarity metric discriminatively, with application to face verification". (CVPR2005)



Few-shot Learning 介绍

□ 孪生网络 (Siamese Network) —— 豆知识



Siamese 暹罗
泰国中部的一个地方

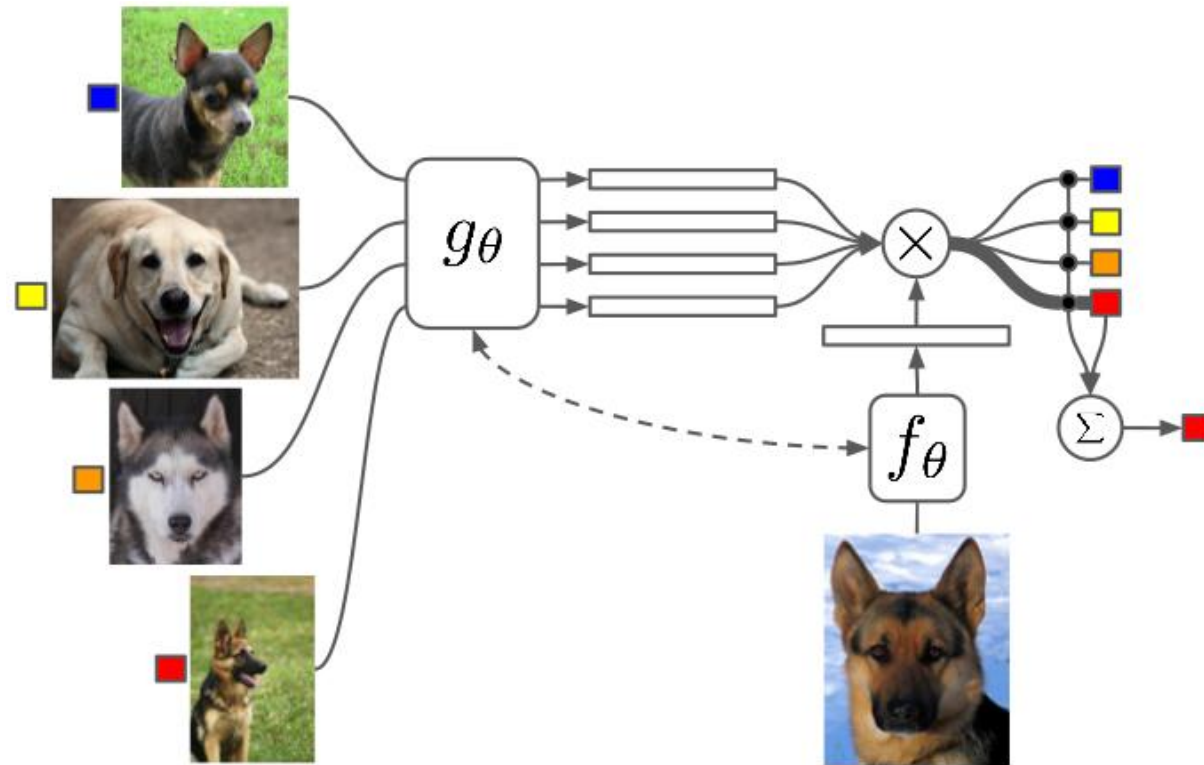


Siamese Twins
最著名的连体人之一，Siamese得名于此



Few-shot Learning 介绍

匹配网络 (Matching Network)

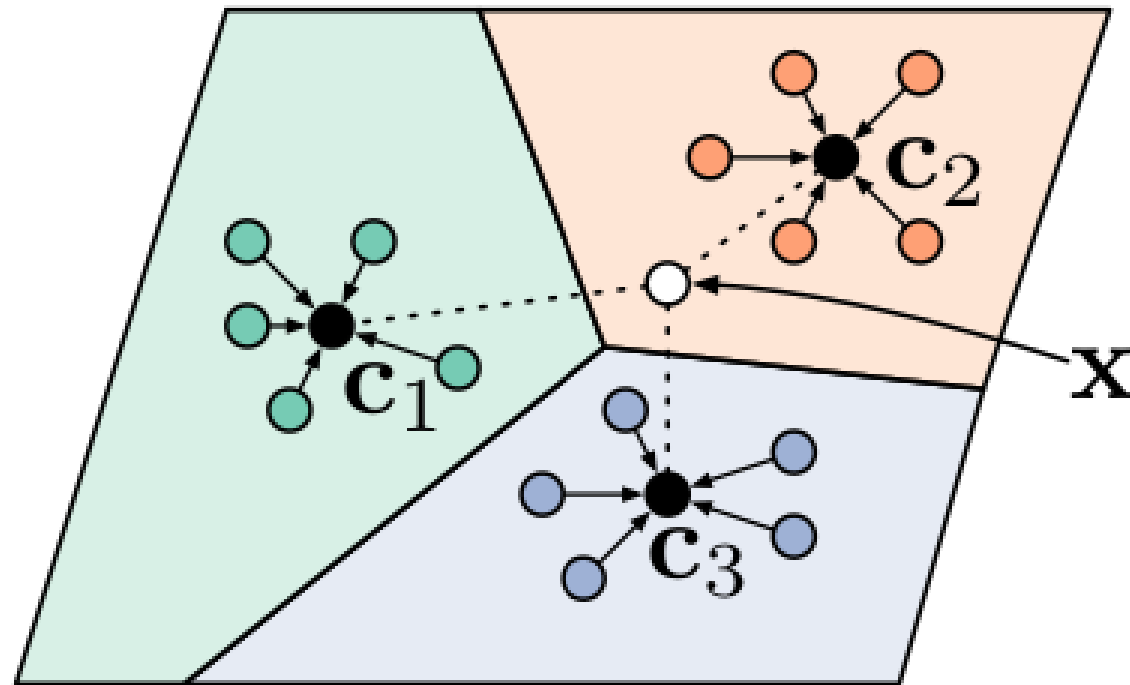


Vinyals, Oriol, Charles Blundell, Timothy Lillicrap, Koray Kavukcuoglu, and Daan Wierstra.
“Matching Networks for One Shot Learning.” (NIPS 2016)



Few-shot Learning 介绍

□ 原型网络 (Prototypical Network)



Snell, Jake, Kevin Swersky, and Richard S. Zemel. "Prototypical Networks for Few-shot Learning." (NIPS 2017)



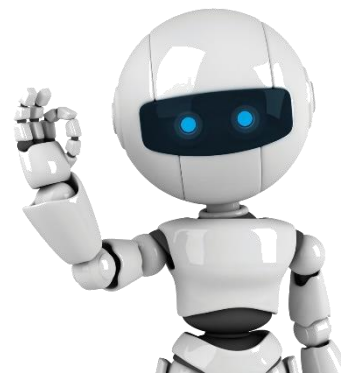
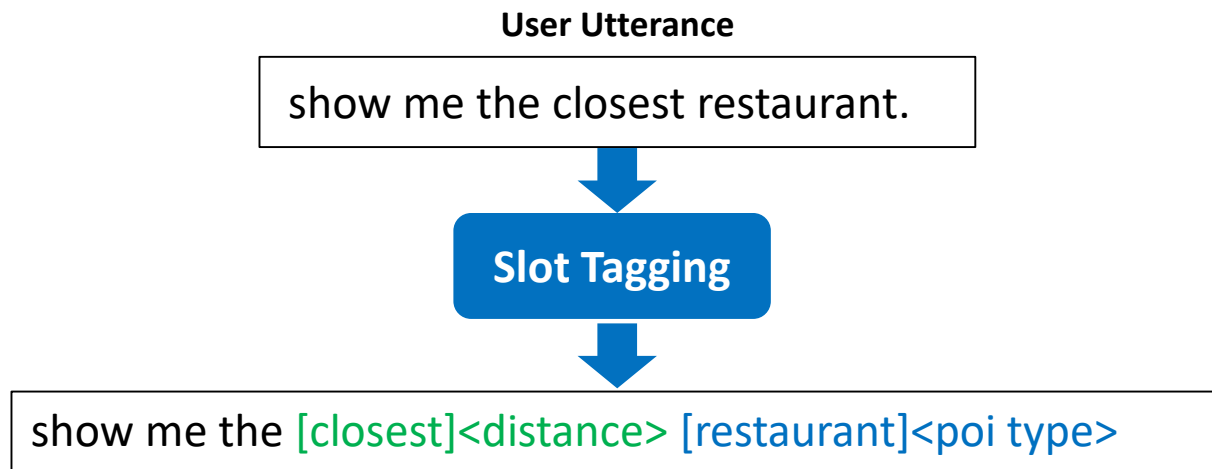
槽位提取任务

□ 任务型对话 Task-oriented Dialogue

- 帮助用户完成具体任务目标的对话系统，比如订酒店，订航班

□ 槽位提取 Slot Tagging

- 槽位提取是任务型对话系统的一个基础组成模块。
- 例子：





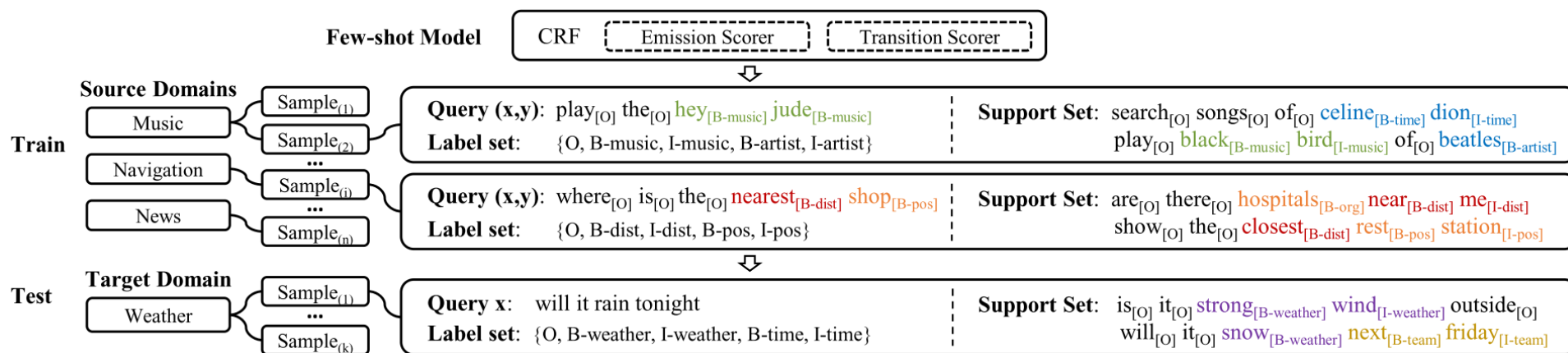
Few-shot 槽位提取任务

Train:

- 在data-rich domain上进行训练
- 训练时模拟数据稀缺的情况，和测试保持一致 (Vinyals et al., 2016)

Test:

- 在没见过的新domain上测试
- 每次给定一个few-shot支撑样例 (support set)，为没见过的query句子标记slot



Overviews of training and testing for Few-shot Slot Tagging



Few-shot learning 在槽位提取中的挑战

Transition: 对于label依赖关系的需求

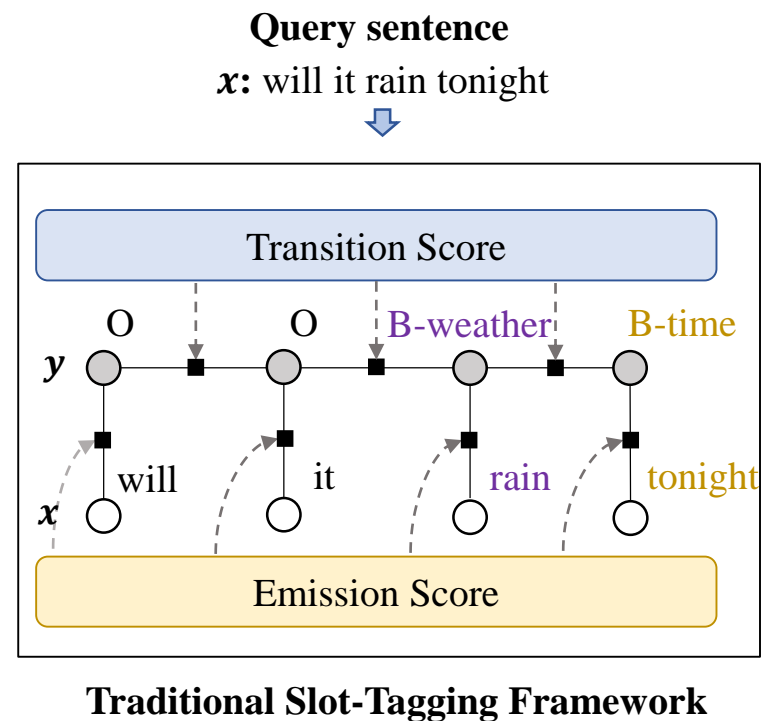
—— 序列标注任务中建模标签依赖关系能提升效果

- 不同domain标签集不同, 过往学到的transition没法直接用到新的domain上
- 从几个样本中无法学习合理的转移关系

Emissions: 用相似度来建模emission 有难度

—— 这里Emission为label表示和词的相似度

- Label表示在空间中表示很容易稠密不易区分
- 词在不同context和领域中的含义存在歧义





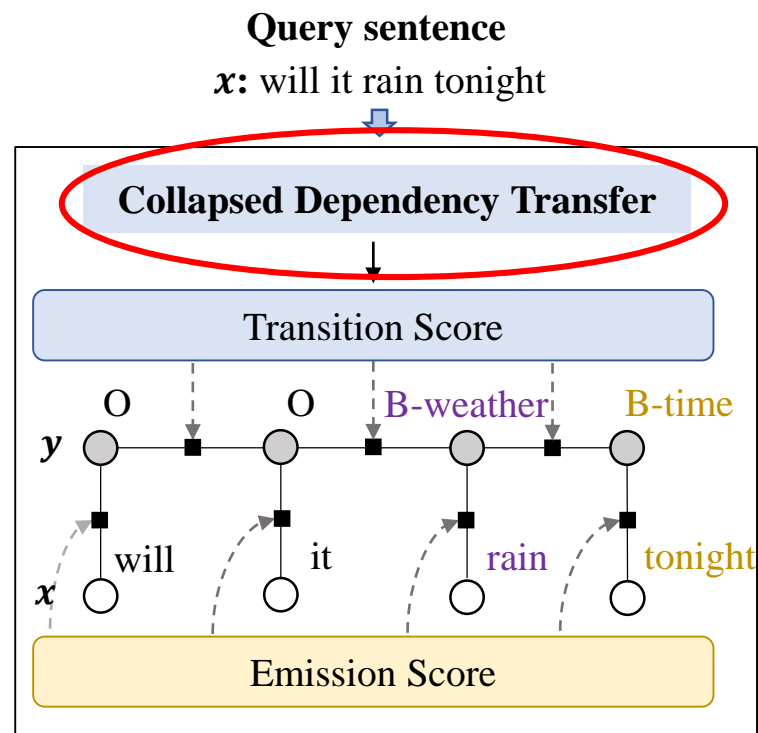
建模Transition的挑战与解决

□ Problem: 对于label 依赖关系的需求

- 不同domain标签集不同, 过往学到的transition没法直接用到新的domain上
- 从几个样本中无法学习合理的转移关系

□ Idea:

- 建模并迁移抽象的label依赖关系
- 在目标领域把抽象的label依赖关系展开为具体的关系



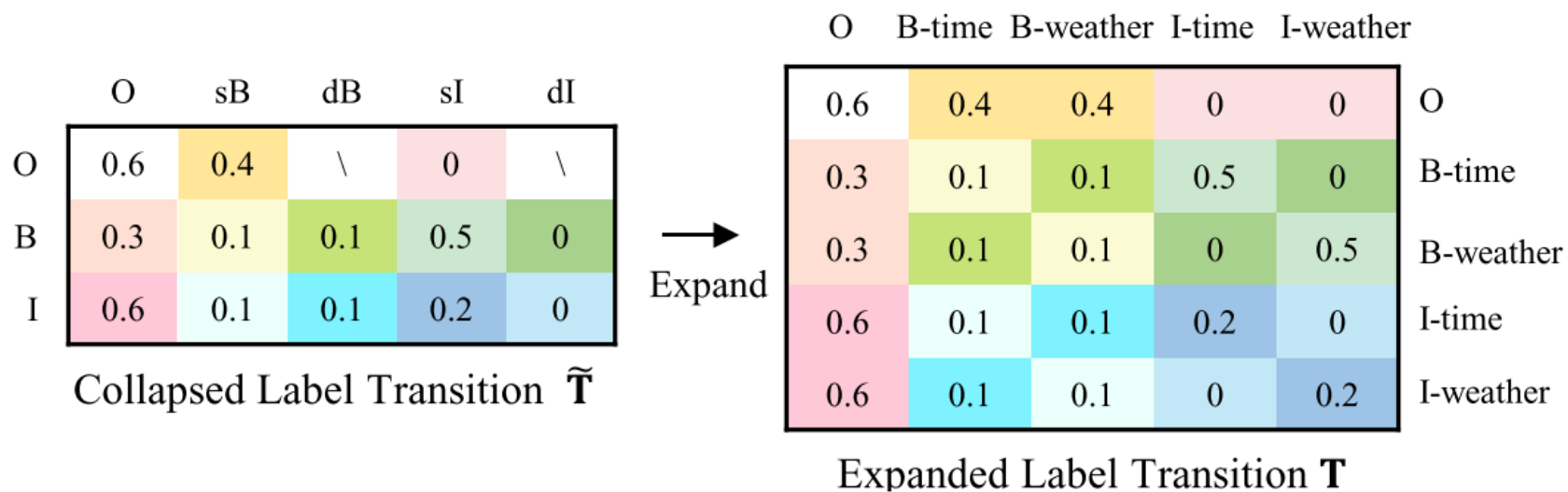
Proposed CRF Framework



建模Transition的挑战与解决

□ Solution: Collapsed Dependency Transfer

- 学习一个折叠坍塌的转移矩阵建模抽象标签转移概率
- 然后在新domain使用时硬根据具体label set展开 (填入颜色相容值)





解决Emission中的相似度计算挑战

□ Problem:

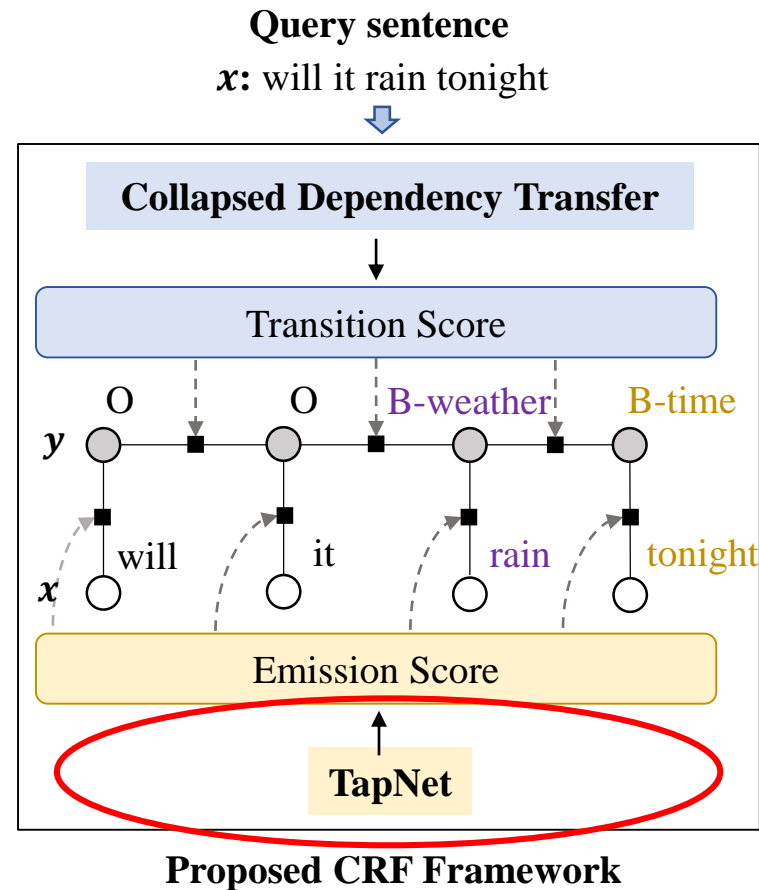
- Label表示在空间中表示很容易稠密不易区分

□ Idea:

- 拉开不同类别之间的距离
- 利用label 名中的语义信息

□ Solution: Task-adaptive Projection Net (TapNet)

- 构造一个为每个具体的Task构造一个映射空间
- 使得映射后不同类别能有效的分离
- 映射构造方法 Yoon et al. (ICML 2019)

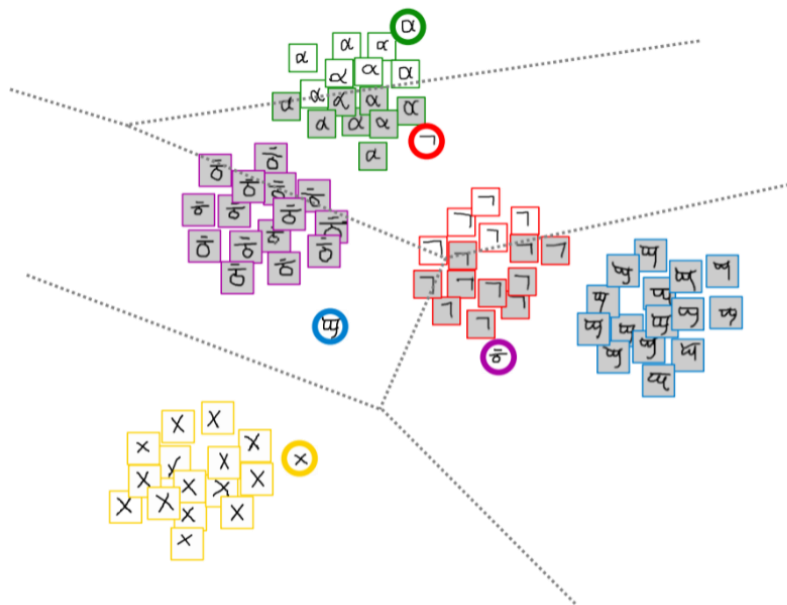




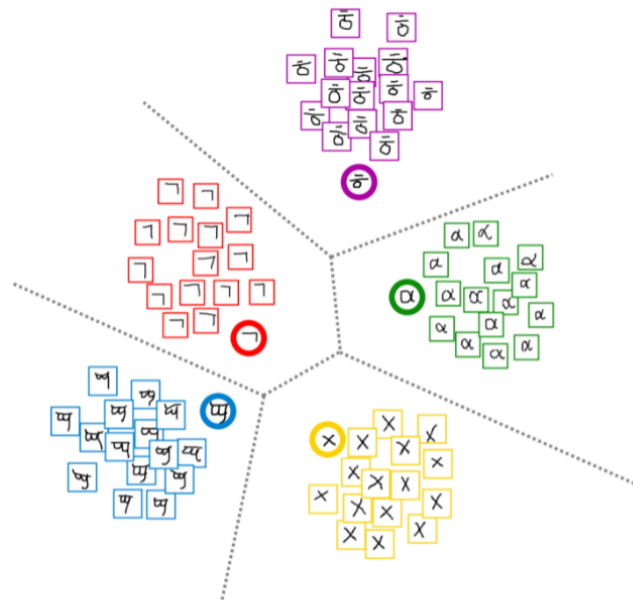
Task-adaptive Projection 构造

- 计算每个类别的prototype c_j : 相同类别词embedding取平均
- 准备一组彼此分离参考向量 $\Phi = [\varphi_1; \cdots; \varphi_N]$
- 计算得到一个映射空间 \mathbf{M} 使得:
 - 对应类别的 c_j 和 φ_j 在映射后能够对齐靠近
 - 不同类别的词在映射后能够有效分开

映射前



映射后





Task-adaptive Projection 构造

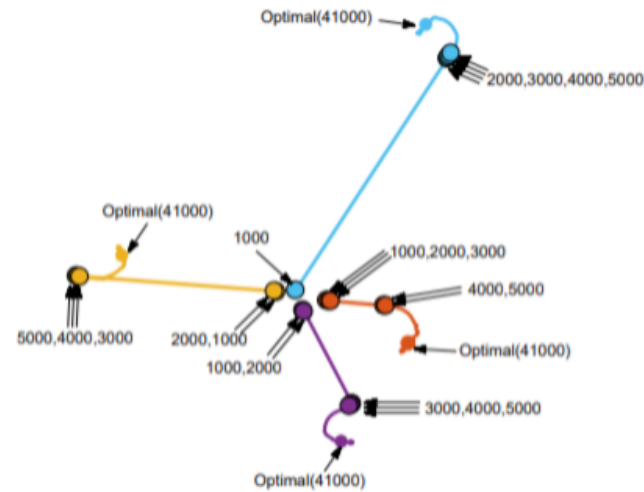
Details:

计算方法:

- 使用Linear Error Nulling的方法求解出一个M, 使得 Φ 和c的对齐偏差为0

参考向量 Φ 的得到:

- 随机初始化, 然后在source domain中训练
- 训练的过程会使这些 Φ 彼此逐渐分开



(a) t-SNE visualization for the trajectories of ϕ_k 's (b) t-SNE visualization of four selected references



解决Emission中的相似度计算挑战

Problem:

- Label表示在空间中表示很容易稠密不易区分

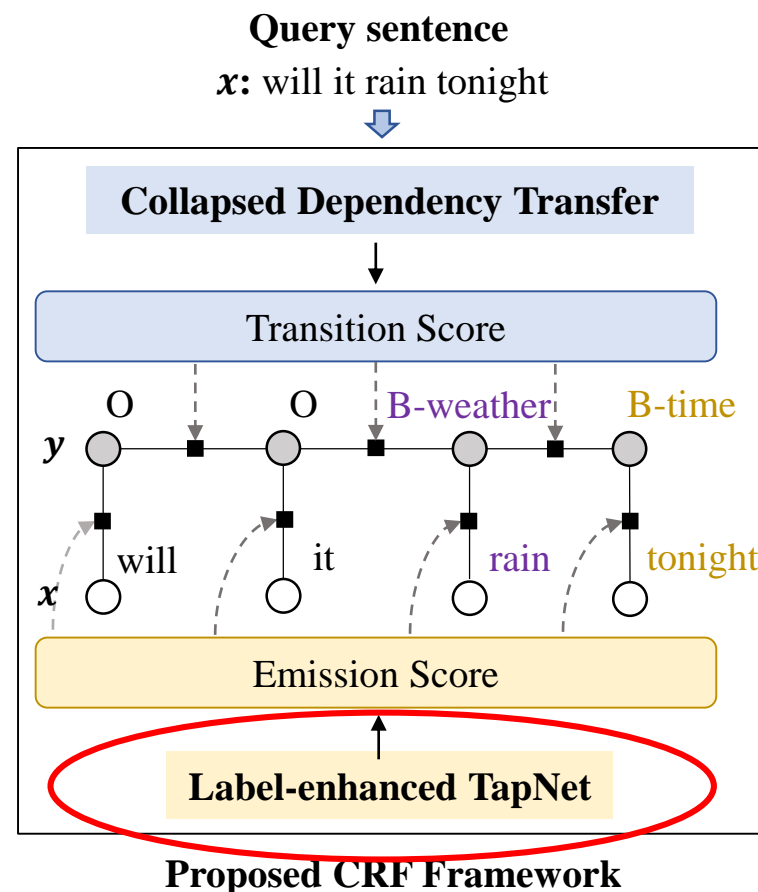
Idea:

- 拉开不同类别之间的距离
- 利用label 名中的语义信息

Solution: Label-enhanced TapNet

- 在构造映射空间时候, 要求和label名的表示对齐
- 类别表示以插值的方式加入 label 名的表示

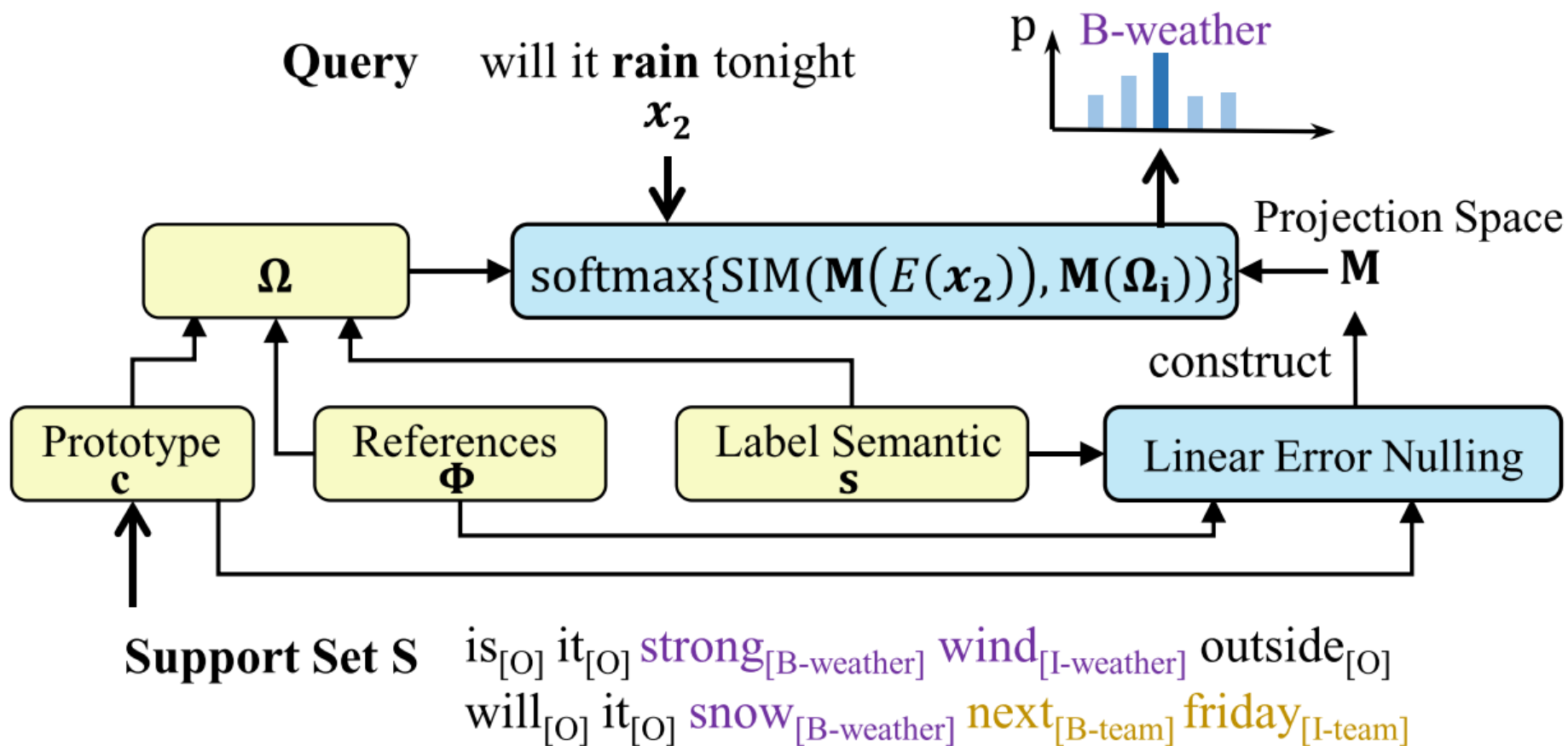
Label Semantics	
Name Semantics	Label
Ordinary	[O]
Begin + Weather	[B-weather]
Inner + Weather	[I-weather]
Begin + Time	[B-time]
Inner + Time	[I-time]





解决Emission中的相似度计算挑战

□ L-TapNet





解决Emission中的相似度计算挑战

□ Problem:

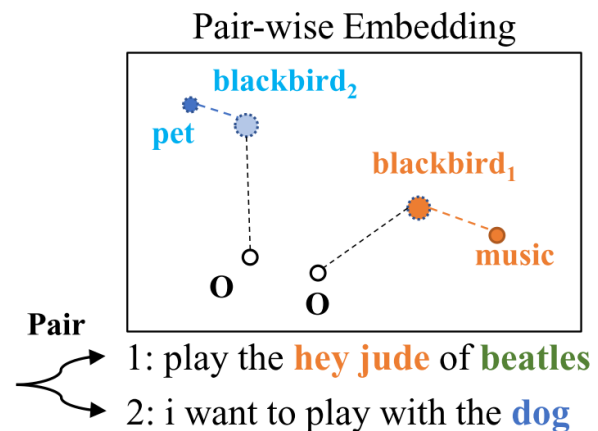
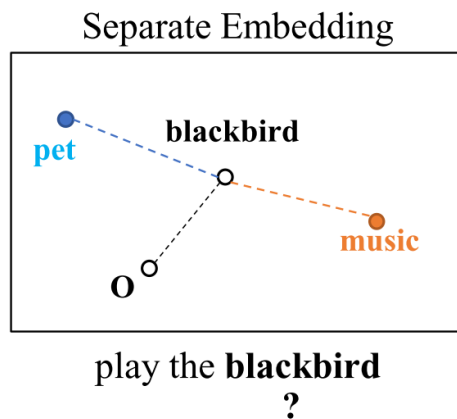
- 词在不同context和领域中的含义存在歧义

□ Idea:

- 利用上下文来消除部分歧义
- 利用support set 中的句子提供domain特定的上下文

□ Solution: Pair-wise Embedding

- 把query和support句子拼起来表示
- 通过bert捕获上下文信息





实验数据

- 我们在两个任务一共5个数据集上做了实验

Task	Dataset	Domain	# Sent	# Labels
Slot Tagging	Snips	Weather	2,100	10
		Music	2,100	10
		PlayList	2,042	6
		Book	2056	8
		SearchScreen	2,059	8
		Restaurant	2,073	15
		CreativeWork	2,054	3
NER	CoNLL	News	20679	5
	GUM	WiKi	3,493	12
	WNUT	Social	5,657	7
	OntoNotes	Mixed	159,615	19

Statistic of Raw Data



实验数据

- 我们从原始数据中构造得到few-shot数据
 - 构造1-shot和5-shot的数据
 - 取Cross-evaluation的平均结果
 - 每次: 1 test domain, 1 dev domain, rest train domain

Domain	Slot Tagging							Named Entity Recognition			
	We	Mu	Pl	Bo	Se	Re	Cr	News	Wiki	Social	Mixed
Ave. $ S $ (1-shot)	6.15	7.66	2.96	4.34	4.29	9.41	1.30	3.38	6.50	5.48	14.38
Samples (1-shot)	2,000	2,000	2,000	2,000	2,000	2,000	2,000	4,000	4,000	4,000	4,000
Ave. $ S $ (5-shot)	28.91	34.43	13.84	19.83	19.27	41.58	5.28	15.58	27.81	28.66	62.28
Samples (5-shot)	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Statistic of Few-shot Data



Slot Tagging 实验结果

Model	1-shot Slot Tagging							Ave.	5-shot Slot Tagging							Ave.
	We	Mu	Pl	Bo	Se	Re	Cr		We	Mu	Pl	Bo	Se	Re	Cr	
Bi-LSTM	10.54	16.93	17.75	54.01	17.48	23.10	9.75	21.37	25.44	39.69	45.36	73.58	55.03	40.30	40.49	45.70
SimBERT	36.10	37.08	35.11	68.09	41.61	42.82	23.91	40.67	53.46	54.13	42.81	75.54	57.10	55.30	32.38	52.96
TransferBERT	45.66	27.80	39.27	9.28	13.59	35.12	9.97	25.81	56.01	43.85	50.65	14.19	23.89	36.99	14.29	34.27
MN	20.85	10.91	40.22	58.12	24.48	32.98	70.17	36.82	38.80	37.98	51.97	70.61	37.24	34.29	72.34	49.03
WPZ	4.34	7.12	13.51	40.10	11.85	8.02	9.36	13.47	9.35	14.04	16.71	47.23	19.56	11.45	13.41	18.82
WPZ+GloVe	17.95	22.08	19.90	42.67	22.28	22.74	16.86	23.50	27.15	34.09	22.15	50.40	28.58	34.59	23.76	31.53
WPZ+BERT	47.54	39.77	50.01	69.69	60.05	54.32	67.14	55.51	69.06	57.97	44.44	71.97	74.62	51.01	69.22	62.61
TapNet+CDT	64.07	53.21	60.21	81.99	64.15	68.34	62.74	64.96	67.83	68.72	73.74	86.94	72.12	69.19	66.54	72.15
L-WPZ+CDT	73.62	49.80	61.03	81.06	72.61	64.27	66.93	67.05	78.23	62.36	59.74	76.19	83.66	69.69	71.51	71.62
L-TapNet+CDT	70.44	60.06	66.62	78.69	71.98	68.88	69.95	69.53	69.58	64.09	74.93	85.37	83.76	69.89	73.80	74.49



NER 实验结果

Model	1-shot Named Entity Recognition					5-shot Named Entity Recognition				
	News	Wiki	Social	Mixed	Ave.	News	Wiki	Social	Mixed	Ave.
Bi-LSTM	2.60	3.07	0.56	1.91	2.04	6.59	8.34	0.87	12.20	7.00
SimBERT	19.22	6.91	5.18	13.99	11.35	32.01	10.63	8.20	21.14	18.00
TransferBERT	1.62	0.58	1.38	3.08	1.67	4.93	0.91	3.71	15.64	6.30
MN	19.44	4.93	13.39	14.04	12.95	20.33	5.73	7.84	8.04	10.49
WPZ	3.63	2.00	0.92	0.67	1.80	4.15	3.13	0.89	0.90	2.27
WPZ+GloVe	9.36	3.24	2.30	2.55	4.36	16.84	5.26	5.42	3.51	7.76
WPZ+BERT	31.85	3.69	9.16	6.15	12.71	49.85	9.60	18.66	13.17	22.82
L-TapNet+CDT	41.64	9.81	18.89	18.40	22.19	43.20	10.98	23.13	19.15	24.12



Analysis

□ 各个模块各有多少贡献？

□ 我们进行Ablation Test, 依次去掉提出模块并观察性能损失

Model	1-shot	5-shot
Ours	69.53	74.49
- dependency transfer	-8.32	-6.41
- pair-wise embedding	-9.24	-5.33
- label semantic	-7.27	-5.17
- prototype reference	-0.85	-2.81

□ 是否真的学到了Label的转移关系 or 只是学到了一些简单规则？

Model	1-shot	5-shot
L-TapNet	61.21	68.08
L-TapNet+Rule	62.77	69.26
L-TapNet+CDT	69.53	74.49



准确率案例分析

- 我们分析了几种转移类型，以及我们方法在这些case上的效果

Bi-gram Type		Proportion	L-TapNet	+CDT
Border	O-O	28.5%	82.7%	83.7%
	O-B	24.5%	78.3%	81.5%
	B-O	8.2%	72.4%	74.8%
	I-O	5.8%	76.7%	81.7%
	I-B/B-B	7.8%	65.0%	72.5%
Inner	B-I	13.3%	78.5%	83.6%
	I-I	12.1%	77.8%	82.7%

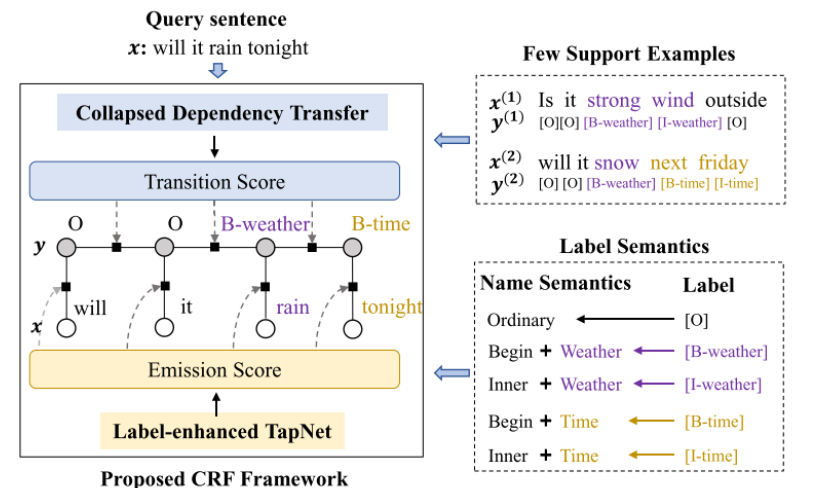
- 结果表明，我们的抽象关系转移：
 - 能隐式地帮助模型判断slot边缘
 - 能学到slot内部的成一致性



感谢您的聆听!

总结:

- 我们提出了一种基于CRF的Few-shot Slot Tagging框架.
- 我们提出了L-TapNet来利用label名增强并拉远不同类别的表示
- 我们提出Collapsed Dependency Transfer来解决few-shot情景下的label转移关系建模问题
- 我们使用Pair-wise Embedding来提供更好的词表示



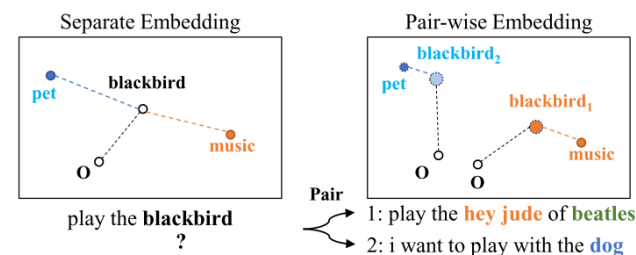
	O	sB	dB	sI	dI
O	0.6	0.4	\	0	\
B	0.3	0.1	0.1	0.5	0
I	0.6	0.1	0.1	0.2	0

Collapsed Label Transition \tilde{T}

Expand

	O	B-time	B-weather	I-time	I-weather
O	0.6	0.4	0.4	0	0
B-time	0.3	0.1	0.1	0.5	0
B-weather	0.3	0.1	0.1	0	0.5
I-time	0.6	0.1	0.1	0.2	0
I-weather	0.6	0.1	0.1	0	0.2

Expanded Label Transition T



Thanks