

基于深度学习的联合实体关系抽取

孙长志

华东师范大学

计算机科学与技术学院



指导老师：孙仕亮 吴苑斌

2019 年 11 月 22 日

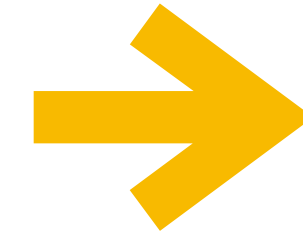
如何获得更好的性能



数据



模型



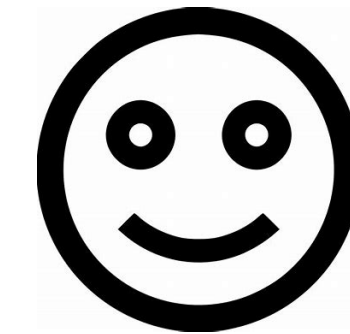
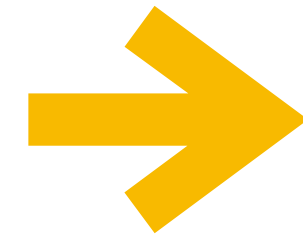
性能



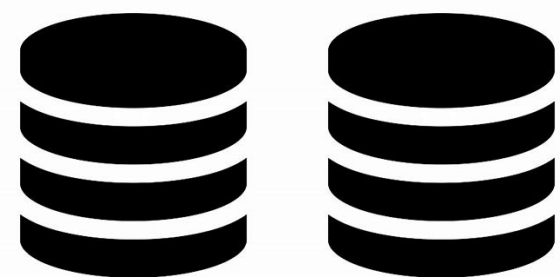
数据



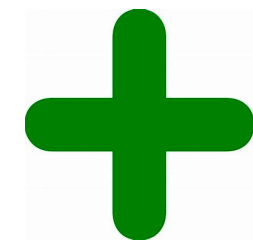
更好的模型



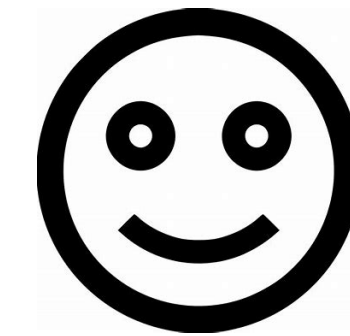
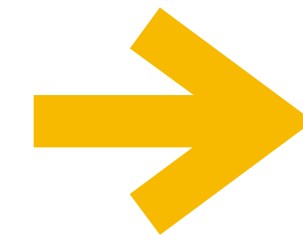
更好的性能



更多的数据



模型



更好的性能

提纲

- 实体关系抽取研究背景
- 如何收集和使用更多的数据
- 如何训练一个更好的联合模型
- 总结

实体关系抽取

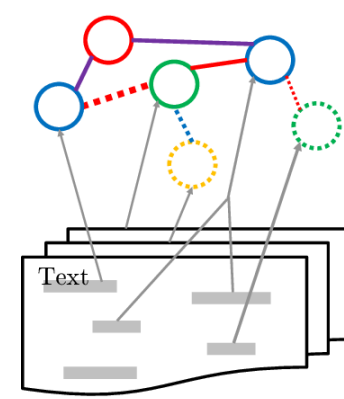
- 实体识别 & 关系抽取

自由文本

Steve Jobs was the co-founder
and CEO of Apple

结构化三元组

(Apple, Steve Jobs, founder)



知识库填充

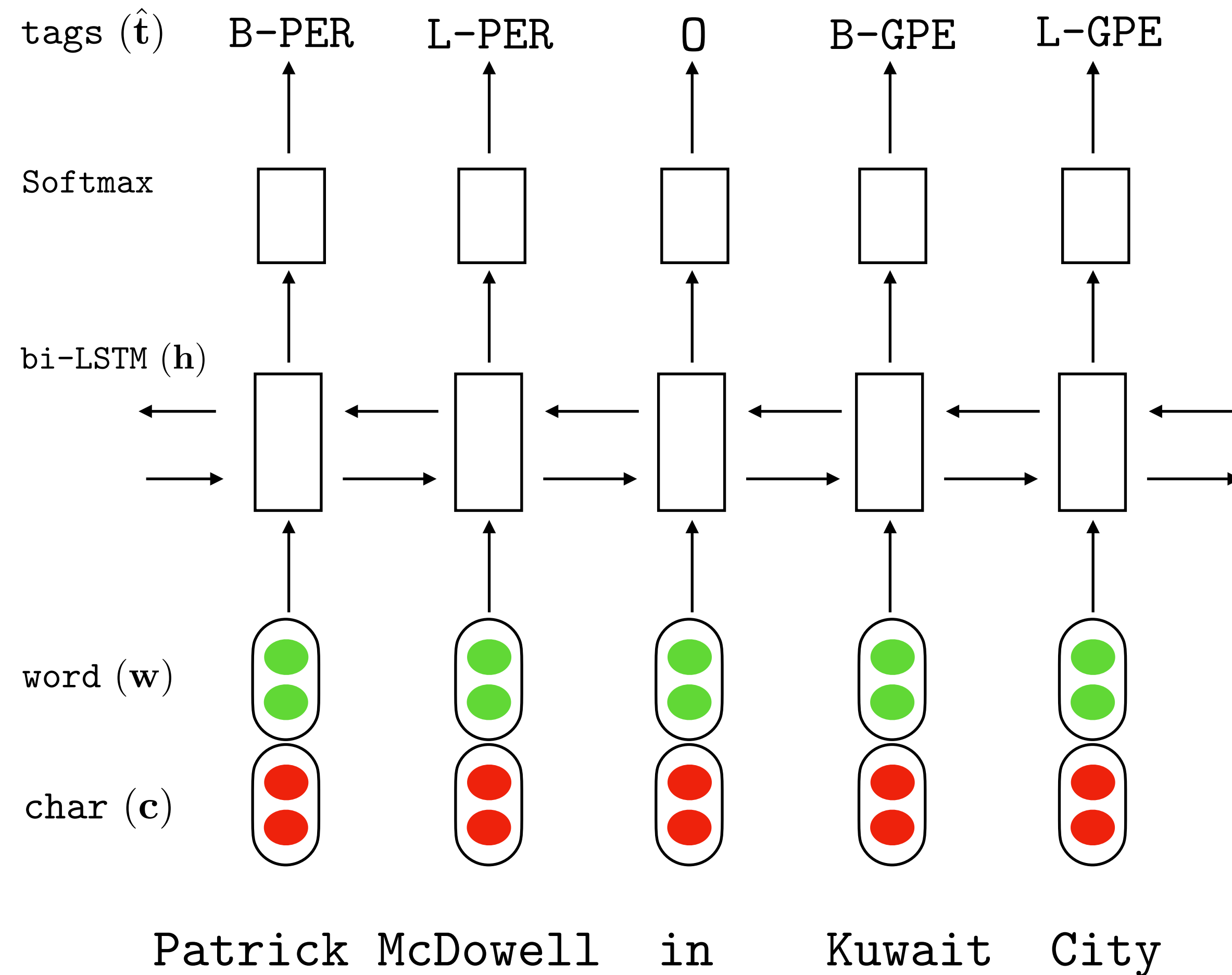


自动问答



信息检索

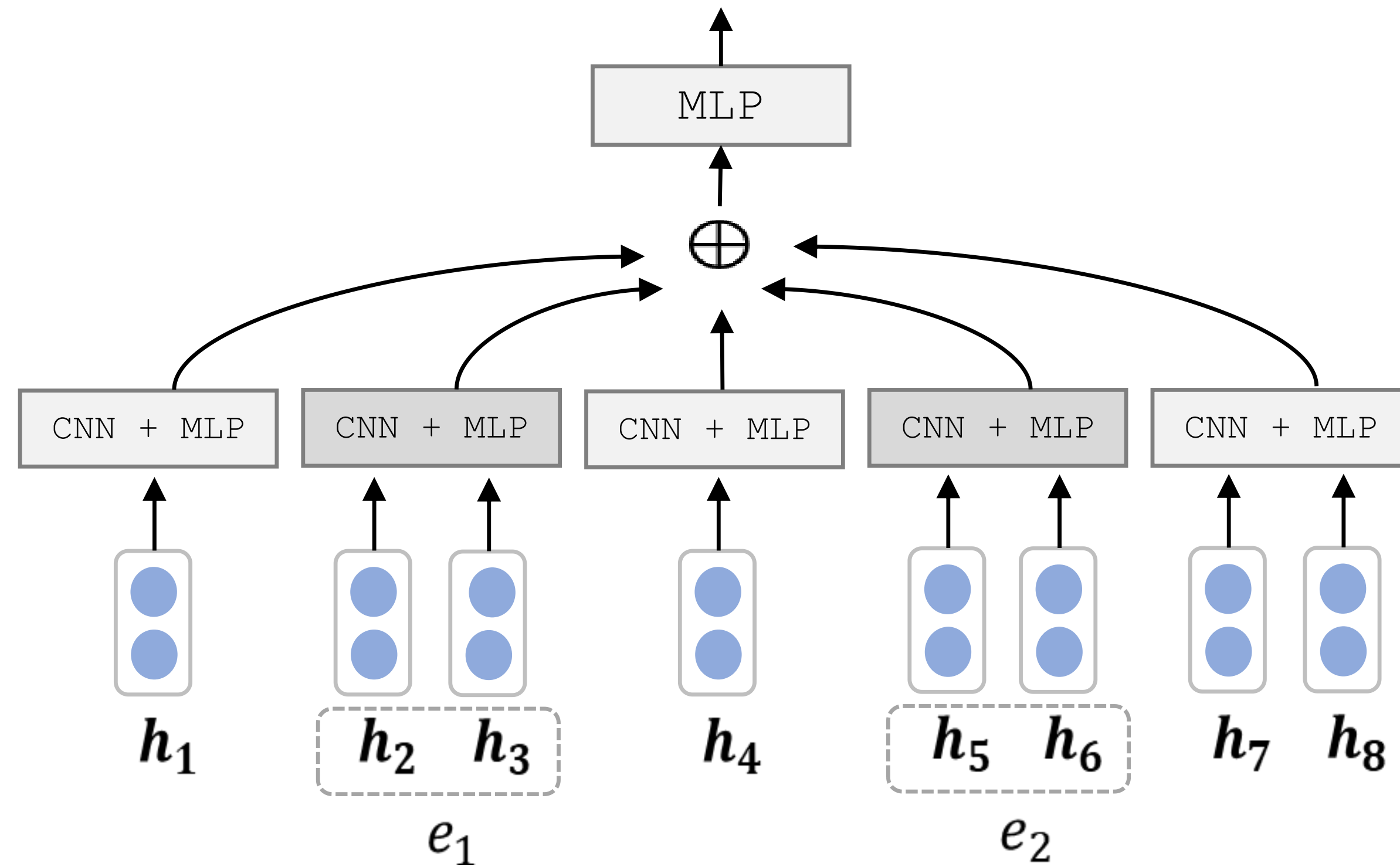
实体识别



- 序列标注框架

- 输入是词序列，输出是 BILOU/BIO 标签序列

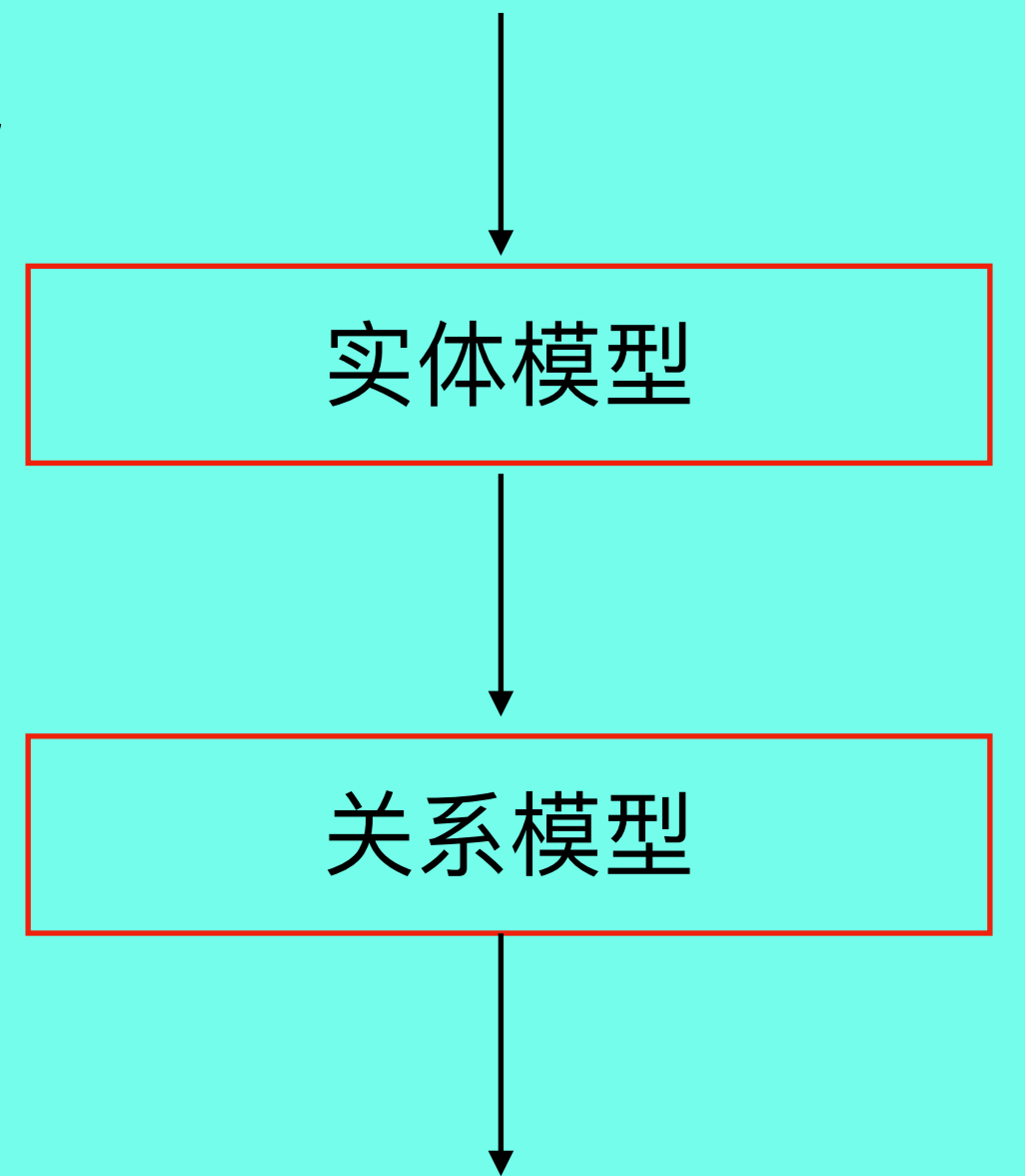
关系抽取



- 多类分类任务
- 特征
 - 两个实体，两个实体的上下文

主流方法

流水线

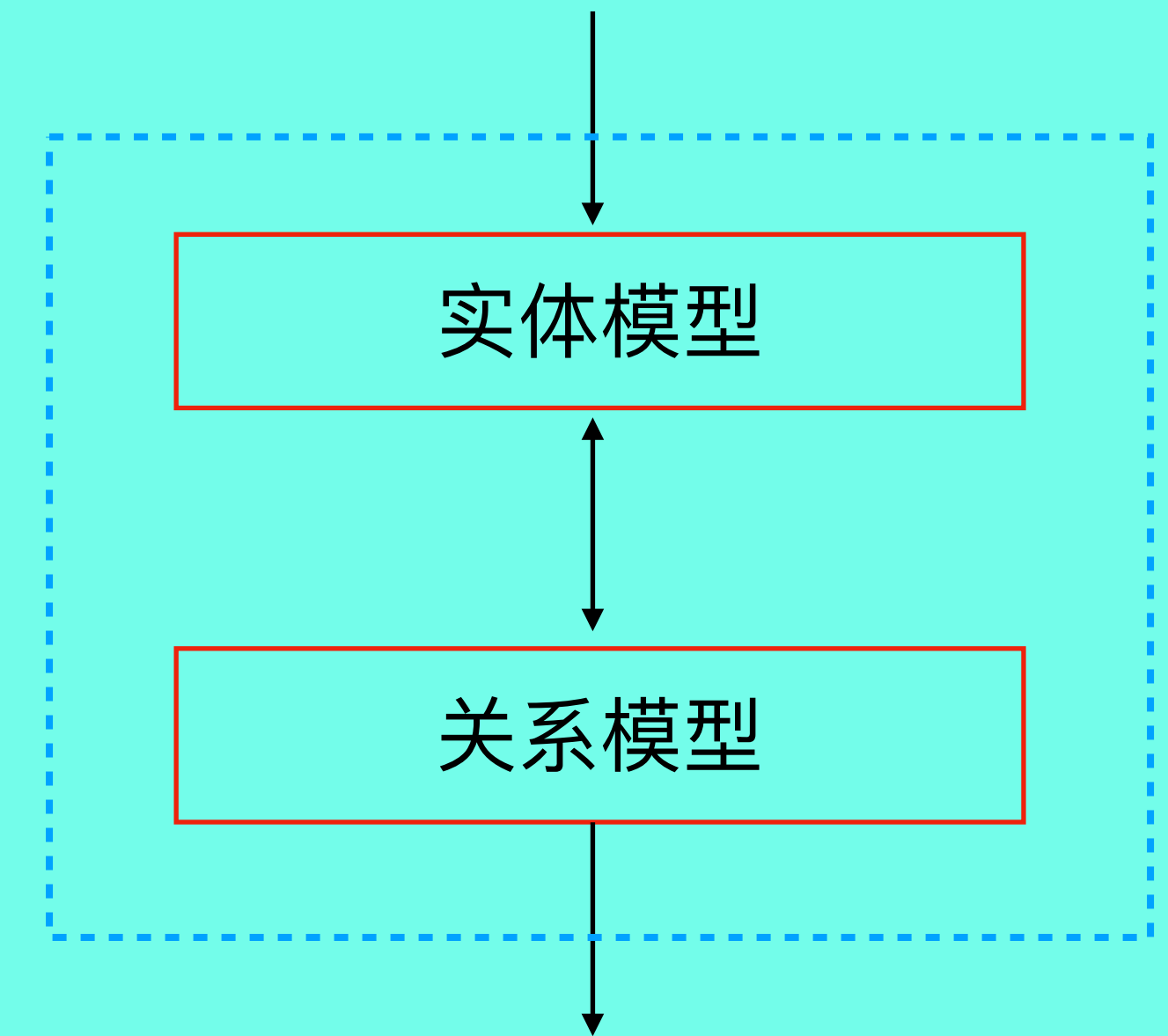


简单有效
易于结合各自的数据集

错误传播

独立

联合



利用两个子任务之间的潜在信息
缓解错误传播

缺乏标注数据

- 我们关注于 联合模型

联合模型

共享参数

[Miwa and Bansal, 2016, Katiyar and Cardie, 2017]

每个子模型可以抽取丰富的特征 (没有限制)

子模型之间没有显式的交互 (隐式共享参数)

联合解码

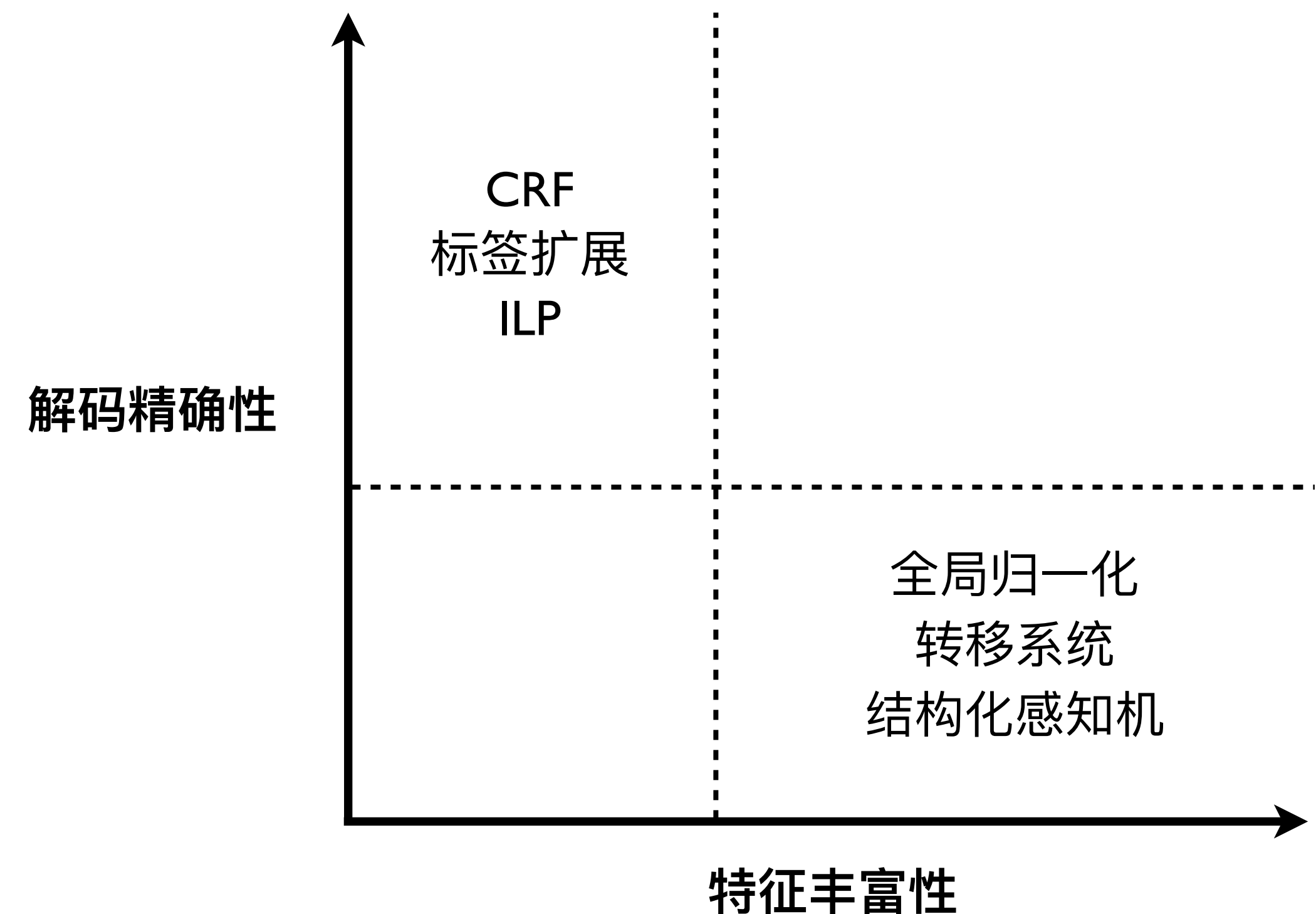
更进一步加强两个子模型
之间的交互

可以直接对子模型的输出做约束

权衡 : 特征丰富性 & 解码精确性

联合解码

- 限制特征 → 精确解码
 - CRF [Katiyar and Cardie, 2016]
 - 标签扩展 [Zheng et al., 2017]
 - ILP [Yang and Cardie, 2013]
- 不限制特征 → 近似解码
 - 全局归一化 [Zhang et al., 2017]
 - 转移系统 [Wang et al., 2018]
 - 结构化感知机 [Li and Ji, 2014]



缺乏标注数据

- 人工标注费时费力
- 代价大

是否存在一种自动生成训练集的方法？

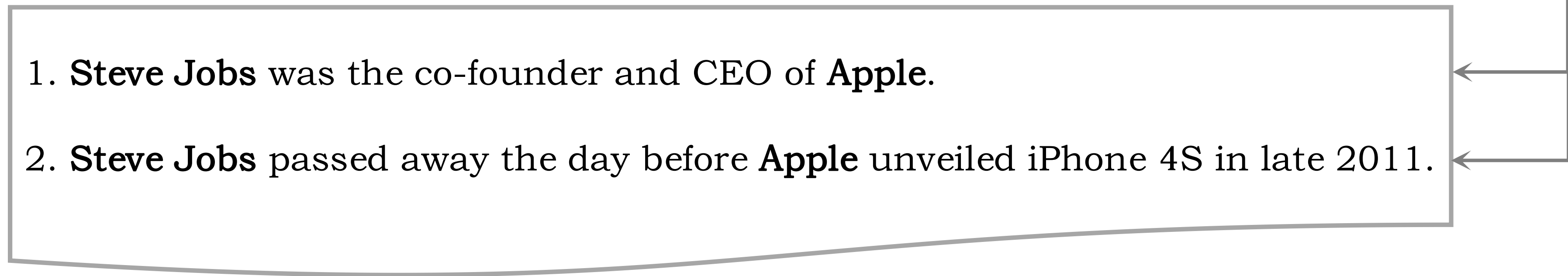
YES

远程监督

Freebase

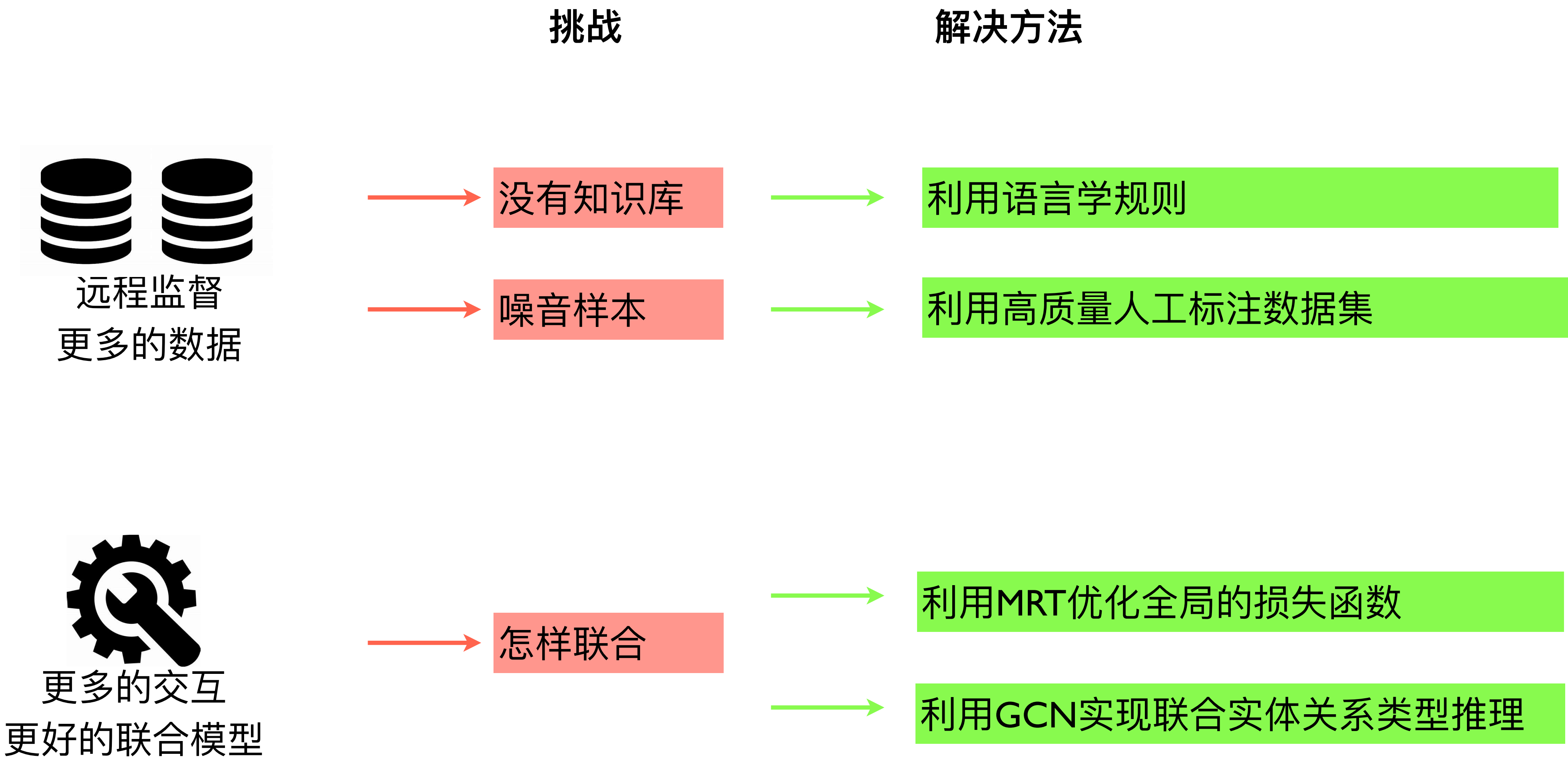


Free texts



- 如果知识库中的实体对存在某种关系，那么包含这个实体对的每个句子均表达了此关系
- 噪音样本；有些应用场景不存在现成的知识库

挑战及我们的解决方法



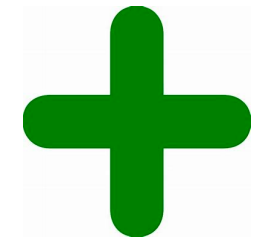
解决方法

- 利用语言学规则
- 利用高质量人工标注数据集
- 利用MRT优化全局的损失函数
- 利用GCN实现联合类型推理

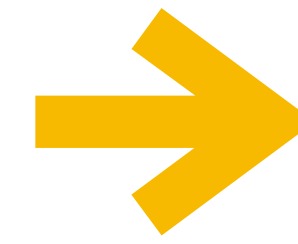
动机



自由文本



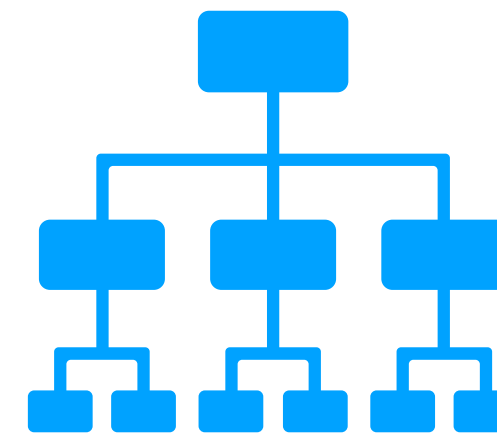
知识库



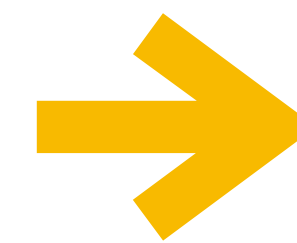
训练数据



自由文本



语言学结构



训练数据



倾向性关系抽取

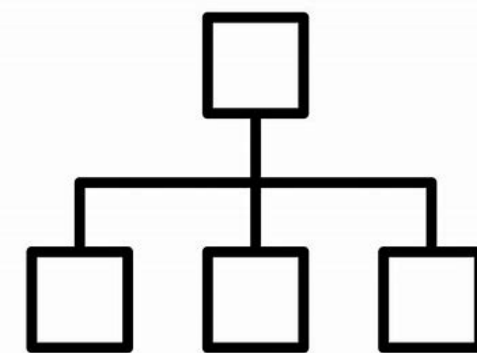
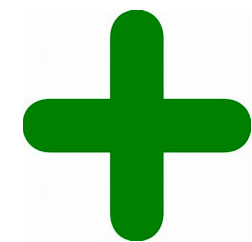
The unit is [well designed] and [perfect reception]

- 评价词: well designed, perfect reception
- 评价对象: unit
- 倾向性关系: (well designed, unit), (perfect reception, unit)
- 没有现成的知识库可以用来进行远程监督

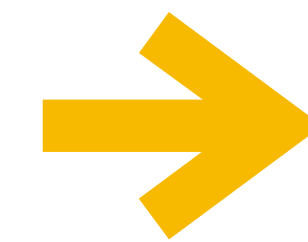
基于语言学规则的远程监督



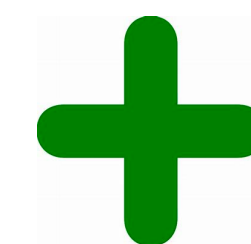
自由文本



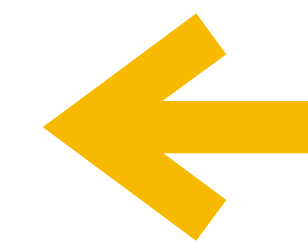
语言学结构



训练数据

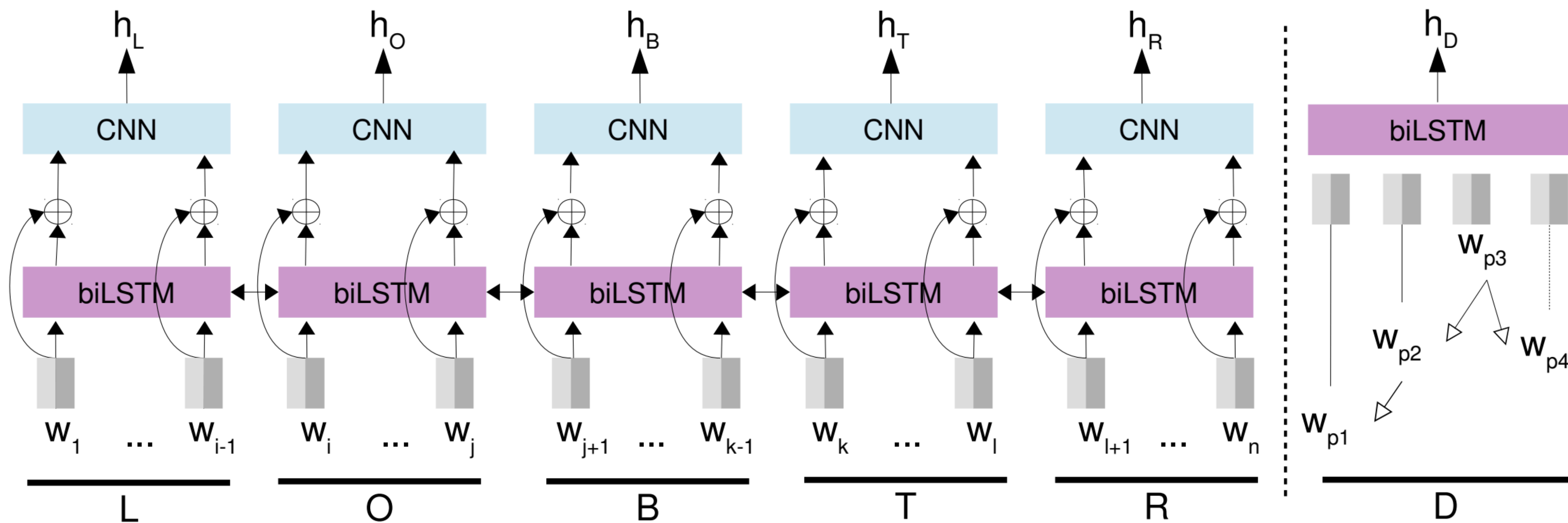


性能



模型

模型分类器



语言学规则

Name	Pattern	Output	Example
P1	$w_1 \xrightarrow[\text{dep}]{\text{amod}} w_2$ $w_1.\text{pos} \in \text{NOUN}, w_2.\text{pos} \in \text{ADJ}$	$T = w_1.\text{np}$ $O = w_2$	It is a [cool] <u>case</u> . $\text{case} \xrightarrow{\text{amod}} \text{cool}$
P2	$w_1 \xrightarrow[\text{xcomp}]{\text{acomp}} w_2$ $w_1.\text{pos} \in \text{VERB}, w_2.\text{pos} \in \text{ADJ}$	$T = w_1.\text{vp}$ $O = w_2$	The case <u>looks</u> [great]. $\text{looks} \xrightarrow{\text{xcomp}} \text{great}$
P3	$w_1 \xrightarrow{\text{advmod}} w_2$ $w_1.\text{pos} \in \text{VERB}, w_2.\text{pos} \in \text{ADV}$	$T = w_1$ $O = w_2$	The cover <u>matches</u> [perfectly]. $\text{matches} \xrightarrow{\text{advmod}} \text{perfectly}$
P4	$w_1 \xrightarrow{\text{nsbj}} w_2$ $w_1.\text{pos} \in \text{NOUN},$ $w_2.\text{pos} \in \text{NOUN or VERB or ADJ}$ has a copula verb between w_1 and w_2	$T = w_1.\text{np}$ $(O = w_2.\text{np}$ $O = w_2.\text{vp}$ $O = w_2.\text{adjp})$	<u>This case</u> is [an excellent choice]. $\text{case} \xleftarrow{\text{nsbj}} \text{choice}$
C1	$w_1 \xleftarrow{\text{conj}} (O', T')$ $O'.\text{pos.} \in \text{ADJ or ADV}$ $w_1.\text{pos} \in \text{ADJ or ADV}$	$T = T'$ $O = w_1$	The case <u>looks</u> [great] and very [cute]. $\text{cute} \xleftarrow{\text{conj}} (\text{great}, \text{looks})$
C2	$(O', T') \xrightarrow{\text{nsbj}} w_1$ $w_1.\text{pos} \in \text{NOUN}, T'.\text{pos} \in \text{VERB}$ $O'.\text{pos} \in \text{ADV}, T' \text{ and } O' \text{ are adjacent}$	$T = w_1$ $O = T' O'$	<u>The case</u> [fits perfectly]. $(\text{perfectly}, \text{fits}) \xrightarrow{\text{nsbj}} \text{the case}$

- 高精确率，低召回率
- 处理复杂的倾向性关系

实验

System	Phone			Movie			Food			Pet		
	P	R	F	P	R	F	P	R	F	P	R	F
Adjacent	38.6	65.7	48.6	30.0	58.8	39.7	31.4	46.5	37.5	28.4	62.2	39.0
Bootstrapping	44.0	62.9	51.8	26.9	49.3	34.8	43.6	54.0	48.2	33.6	57.7	42.5
Pattern	69.4*	64.4	61.1	62.2*	42.4	50.4	76.0*	41.9	54.1	59.9*	51.3	55.3*
LR	60.1	64.7	62.4	55.6	57.0	55.3	65.5	49.2*	56.2	47.6	59.3*	52.8
NN	63.4	67.9*	65.6*	56.8	58.2*	57.5*	70.5	47.7	56.9*	51.4	58.0	54.5
NN+Pattern	64.4	70.5	67.3	58.2	59.9	59.1	68.4	50.8	58.3	54.9	58.2	56.5

- 语言学规则可以帮助远程监督
- 分类器可以提高性能($NN > LR$)
- 合并NN 和 Pattern 的结果可以进一步提升性能

实验

Systems	P	R	F
Klinger et al. (2014)	-	-	65.0
Jebbara et al. (2016)	87.0	75.0	81.0
Pattern	51.4	20.7	29.5
LR (end-to-end)	49.5	27.8	35.6
NN (end-to-end)	43.3	40.1	41.6
LR (gold)	89.1	47.9	62.3
NN (gold)	81.4	62.8	70.9

- 可以与有监督方法的性能相比较

解决方法

- 利用语言学规则
- 利用高质量人工标注数据集
- 利用MRT优化全局的损失函数
- 利用GCN实现联合类型推理

使用高质量人工标注数据



- 使用少量高质量人工标注数据帮助远程监督实体关系抽取

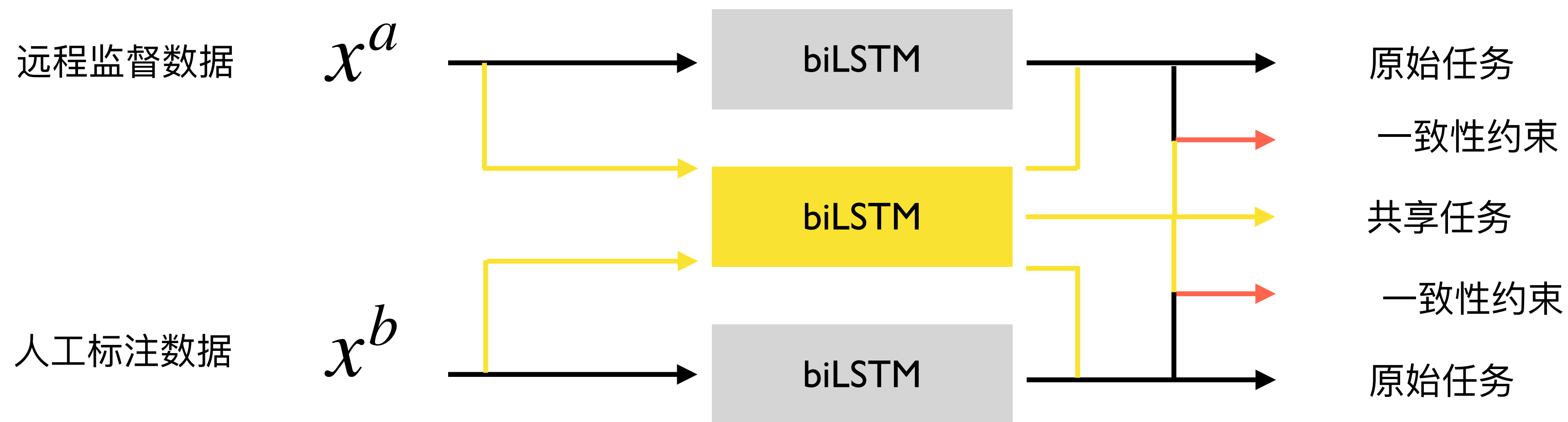
挑战

ACE05	Entity Types	NYT
PER (person) LOC (location) ORG (organizaion) GPE (geographical entities)	FAC (facility) VEH (vehicle) WEA (weapon)	PER (person) LOC (location) ORG (organizaion)

ACE05	Relation Types	NYT
ART (agent-artifact) GPE-AFF (gpe-affiliation) ORG-AFF (organization-affiliation) PART-WHOLE (part-whole) PER-SOC (person-social) PHYS (phisical)	/business/company/founders /business/company/industry /location/country/capital /location/location/contains /people/person/nationality /people/person/place_of_birth	

- 异质数据集
 - 实体类型和关系类型有差异
 - 标注准则也不太一样

解决方法



- 多任务学习框架 (共享-私有)
- 两个共享任务: 实体边界识别, 二元关系分类
 - 假设两个共享任务对不同的数据集不太敏感

- 共享任务与原始任务之间的一致性约束
$$P_{\text{seq}}^c(B|s) = \sum_{*} P_{\text{seq}}^a(B- * |s)$$
- 转移矩阵

转移损失函数

- $M_{\text{seq}}^a, M_{\text{seq}}^b$ 是转移矩阵

$$\mathcal{L}_{\text{trans}}^{\nabla, \text{seq}} = \|P_{\text{seq}}^c - M_{\text{seq}}^{\nabla} P_{\text{seq}}^{\nabla}\|_2, \quad \nabla \in \{a, b\}.$$

- 相似地

$$\mathcal{L}_{\text{trans}}^{\nabla, \text{rel}} = \|P_{\text{rel}}^c - M_{\text{rel}}^{\nabla} P_{\text{rel}}^{\nabla}\|_2, \quad \nabla \in \{a, b\}.$$

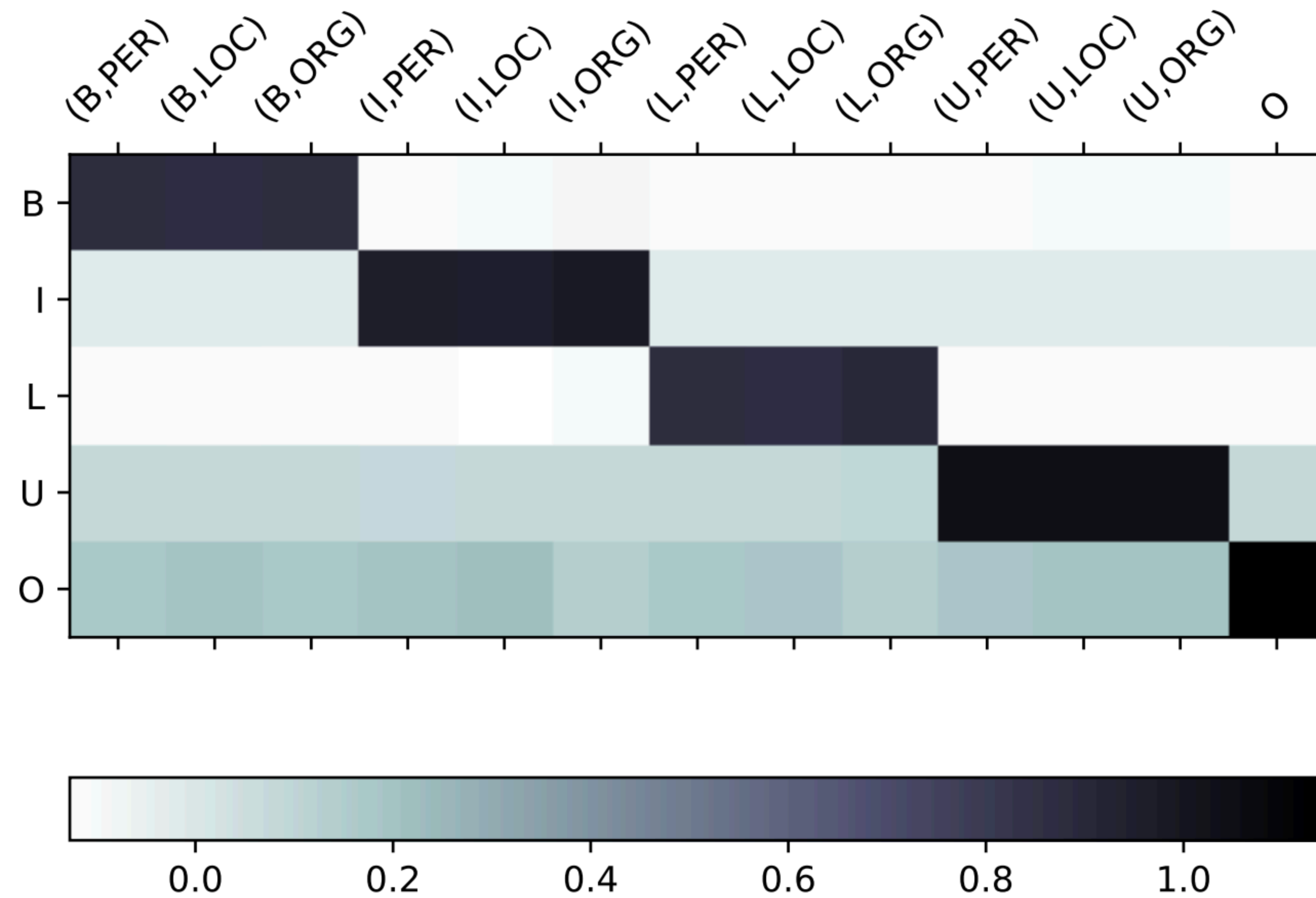
实验

Model	Relation		
	P	R	F
Gormley (2015)	55.3	15.4	24.0
Mintz (2009)	25.8	39.3	31.1
Tang (2015)	33.5	32.9	33.2
Hoffman (2011)	33.8	32.7	33.3
L&J (2014)	57.4	25.6	35.4
Ren (2017)	42.3	51.1	46.3
Zheng (2017)	61.5	41.4	49.5
Wang (2018)	64.3	42.1	50.9
Sun (2018)	67.4	42.0	51.7
Our Model	70.4	45.6	55.4
Sun (2018)(exactly match)	65.2	40.6	50.0
Our Model(exactly match)	68.3	44.2	53.7

Model	Entity			Relation		
	P	R	F	P	R	F
only D^a	82.6	91.2	86.7	61.8	43.3	50.9
$D^a \cup D^b$	84.5	91.9	88.1	60.7	42.3	49.8
only h^c	83.5	93.1	88.0	65.6	42.0	51.2
$h^c + \mathcal{L}^c$	86.2	92.5	89.2	66.5	45.4	53.9
+ $\mathcal{L}_{trans}$	82.8	89.6	86.1	64.8	46.2	54.0
+ $\mathcal{L}_{trans}^{seq}$	86.6	92.9	89.6	70.4	45.6	55.4
+ $\mathcal{L}_{trans}^{rel}$	87.2	93.9	90.4	72.9	40.9	52.4

- 利用人工标注数据集帮助远程监督实体关系抽取
- 共享任务比共享表示更好
- 一致性约束是有效的

实验

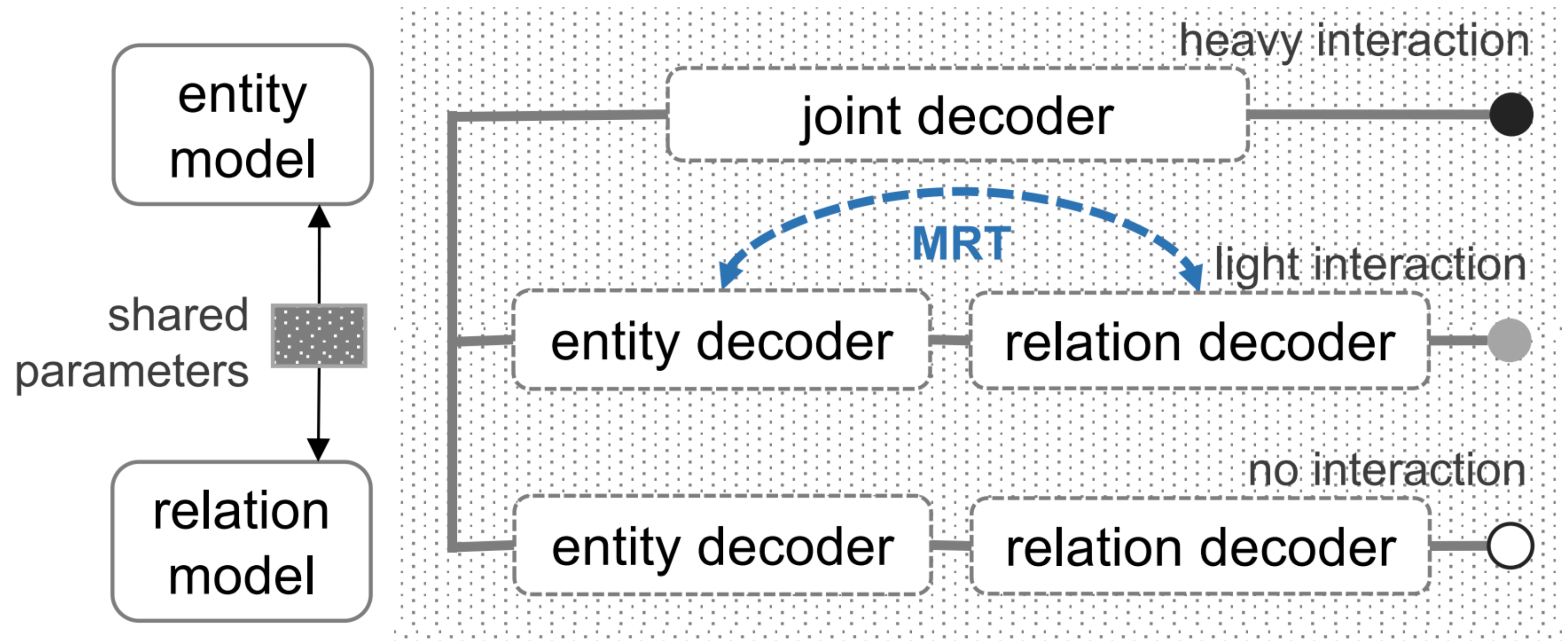


- 近似满足一致性约束

解决方法

- 利用语言学规则
- 利用高质量人工标注数据集
- 利用**MRT**优化全局的损失函数
- 利用GCN实现联合类型推理

联合模型



- 加强实体模型和关系模型的交互，同时不影响特征的丰富性
- MRT 是一个不错的解决方案

Minimum Risk Training

$$\arg \min_P \mathbf{E}_{\hat{\mathbf{y}} \sim P(\hat{\mathbf{y}}|\mathbf{x})} \Delta(\hat{\mathbf{y}}, \mathbf{y})$$

- $\Delta(\hat{\mathbf{y}}, \mathbf{y})$ 衡量模型输出 $\hat{\mathbf{y}}$ 和真实标注 $\mathbf{y}$ 之间的差异
- 可以直接优化全局句子级别的损失函数，比如F1分数

常用策略

优化 $\mathcal{L}_{\text{ent}} + \mathcal{L}_{\text{rel}}$ (局部)

- 基于局部的损失函数
 - 实体模型：词级别
 - 关系模型：实体对级别
- 实体模型和关系模型均意识不到对方的损失函数

联合概率
$$P(\hat{y}|s; \Theta) = P(\hat{\mathcal{E}}|s; \Theta_E)P(\hat{\mathcal{R}}|s, \hat{\mathcal{E}}; \Theta_R)$$
$$= \prod_i P_{\text{ent}}(\hat{t}_i|s; \Theta_E) \prod_{\substack{e_1, e_2 \in \hat{\mathcal{E}} \\ e_1 \neq e_2}} P_{\text{rel}}(\hat{l} \mid s, e_1, e_2; \Theta_R),$$

联合MRT目标函数

$$\mathbf{E}_{\hat{\mathbf{y}} \sim P(\hat{\mathbf{y}}|s; \Theta)} \Delta(\hat{\mathbf{y}}, \mathbf{y}) = \sum_{\hat{\mathbf{y}} \in \mathcal{Y}(s)} P(\hat{\mathbf{y}}|s; \Theta) \Delta(\hat{\mathbf{y}}, \mathbf{y})$$

- $\Delta_E(\hat{\mathbf{y}}, \mathbf{y}) = 1 - F_{\text{ent}}(\hat{\mathcal{E}}, \mathcal{E})$
- $\Delta_R(\hat{\mathbf{y}}, \mathbf{y}) = 1 - F_{\text{rel}}(\hat{\mathcal{R}}, \mathcal{R})$
- $\Delta_{E+R}(\hat{\mathbf{y}}, \mathbf{y}) = 1 - \frac{1}{2} [F_{\text{ent}}(\hat{\mathcal{E}}, \mathcal{E}) + F_{\text{rel}}(\hat{\mathcal{R}}, \mathcal{R})]$

微调损失函数

$$\min \sum_{\hat{\mathbf{y}} \in \mathcal{Y}(s)} P(\hat{\mathbf{y}}|s; \Theta) (\Delta(\hat{\mathbf{y}}, \mathbf{y}) + \Gamma(\hat{\mathbf{y}})) + \xi$$

$$\text{s.t. } \Gamma(\mathbf{y}^*) - \Gamma(\mathbf{y}) \geq 1 - \xi, \quad \xi \geq 0,$$

$$\mathbf{y}^* = \arg \min_{\hat{\mathbf{y}} \in \mathcal{Y}(s)} \Gamma(\hat{\mathbf{y}})$$

在这里，我们使用 $\Gamma(\hat{\mathbf{y}}) = 1 - P(\hat{\mathbf{y}}|s; \Theta)$ ，目标函数为

$$\sum_{\hat{\mathbf{y}} \in \mathcal{Y}(s)} P(\hat{\mathbf{y}}|s; \Theta) (\Delta(\hat{\mathbf{y}}, \mathbf{y}) - P(\hat{\mathbf{y}}|s; \Theta)) + [1 - P(\mathbf{y}|s; \Theta) + P(\mathbf{y}^*|s; \Theta)]_+.$$

其中 $[u]_+ = \max(u, 0)$ 是 Hinge loss

微调损失函数

使用子集 $\mathcal{Y}'(s)$ 代替 $\mathcal{Y}(s)$ (呈指数增长)

$$\mathcal{L}_{\text{mrt}}(\Theta) = \sum_{\hat{\mathbf{y}} \in \mathcal{Y}'(s)} Q(\hat{\mathbf{y}}|s; \Theta, \mu, \alpha) \Delta(\hat{\mathbf{y}}, \mathbf{y})$$

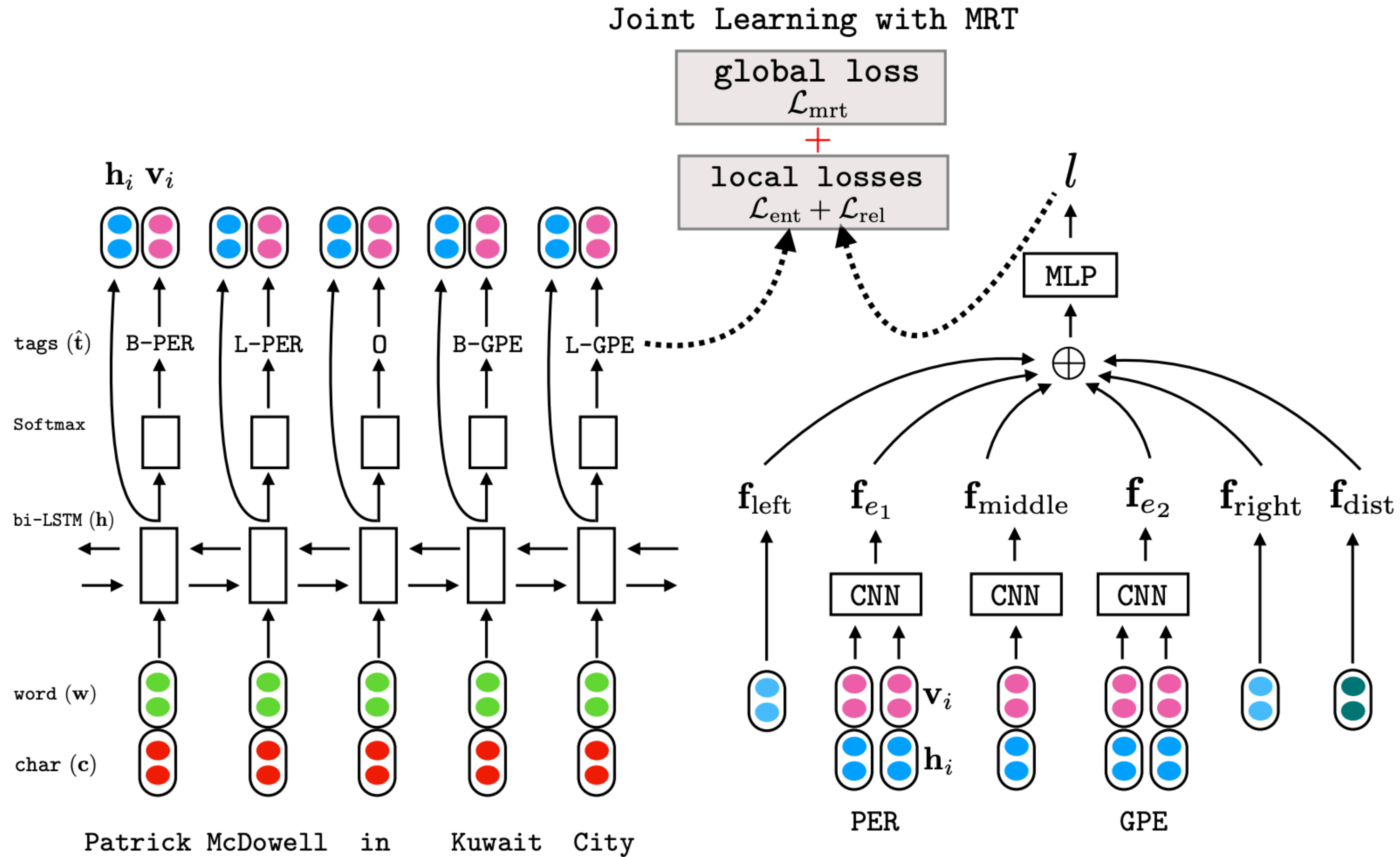
其中 $Q(\hat{\mathbf{y}}|s; \Theta, \mu, \alpha)$ 是 $P(\hat{\mathbf{y}}|s; \Theta)$ 在子集 $\mathcal{Y}'(s)$ 的重新归一化

$$Q(\hat{\mathbf{y}}|s; \Theta, \mu, \alpha) = \frac{1}{Z} [P(\hat{\mathcal{E}}|s, \Theta_E)^\mu P(\hat{\mathcal{R}}|s, \hat{\mathcal{E}}, \Theta_R)^{1-\mu}]^\alpha$$

$$Z = \sum_{(\mathcal{E}', \mathcal{R}') \in \mathcal{Y}'(s)} [P(\mathcal{E}'|s, \Theta_E)^\mu P(\mathcal{R}'|s, \mathcal{E}', \Theta_R)^{1-\mu}]^\alpha$$

- 超参数 α 控制 Q 分布的尖锐程度
- 超参数 μ 衡量实体分布和关系分布在 Q 中的重要程度

基于MRT的联合模型



实验

Model	Entity			Relation		
	P	R	F	P	R	F
L&J (2014)	85.2	76.9	80.8	65.4	39.8	49.5
M&B (2016)	82.9	83.9	83.4	57.2	54.0	55.6
Zhang (2017)	-	-	83.5	-	-	57.5
K&C (2017)	84.0	81.3	82.6	55.5	51.8	53.6
NN	84.0	82.9	83.4	59.5	56.3	57.8
MRT	83.9	83.2	83.6	64.9	55.1	59.6

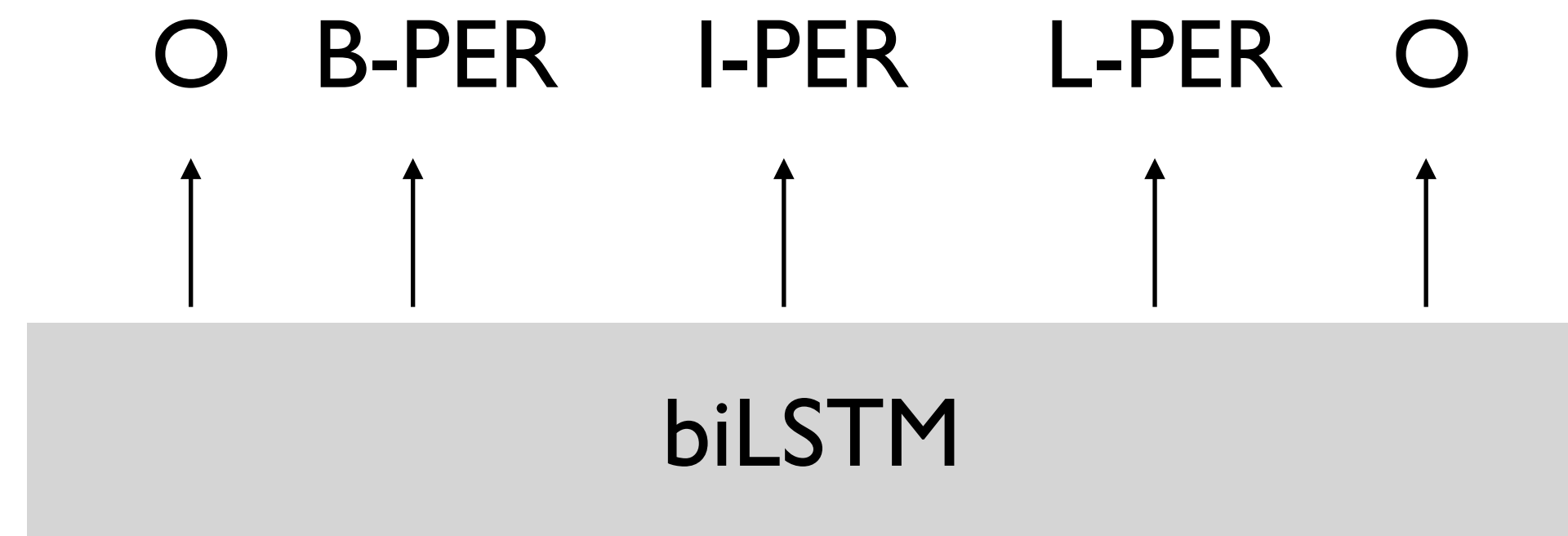
Model	Relation		
	P	R	F
(Zheng et al., 2017)	61.5	41.4	49.5
NN	61.8	43.3	50.9
MRT	67.4	42.0	51.7
(Ren et al., 2017)	42.3	51.1	46.3
NN (exact match)	59.4	41.7	49.0
MRT (exact match)	65.2	40.6	50.0

- MRT 可以取得最好的性能
- MRT是一种轻量级方法并且适用于任何的联合模型

解决方法

- 利用语言学规则
- 利用高质量人工标注数据集
- 利用MRT优化全局的损失函数
- 利用**GCN**实现联合类型推理

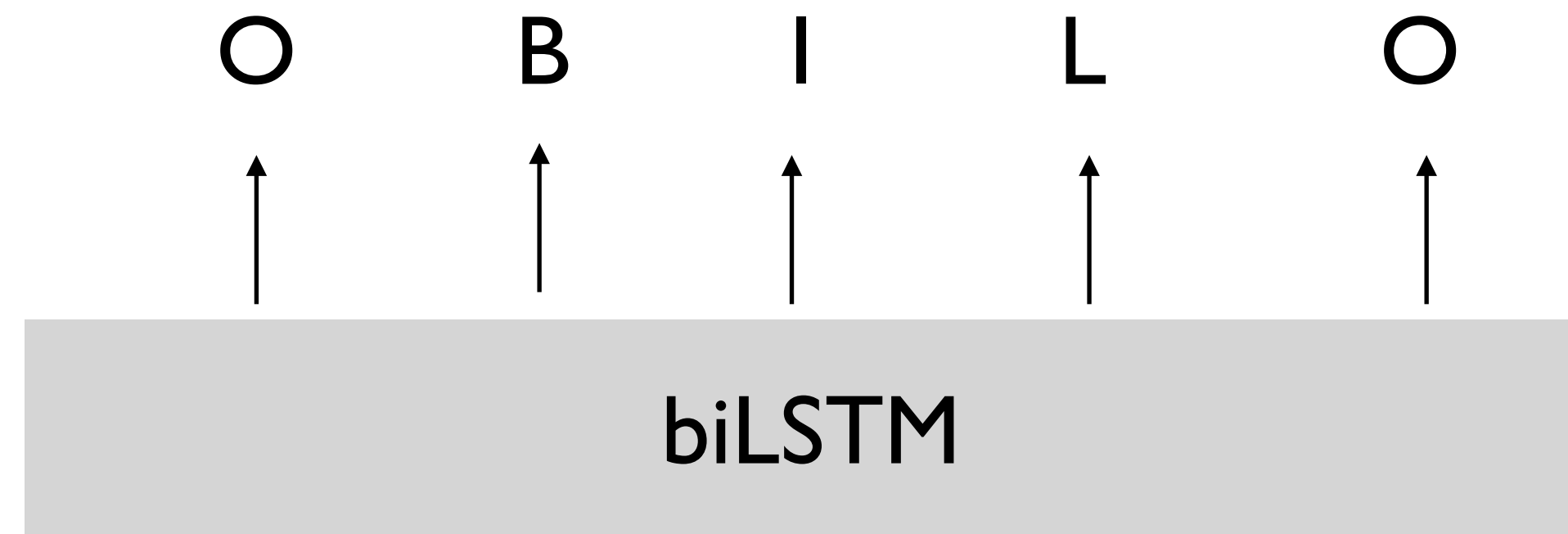
实体识别



没有关系的特征

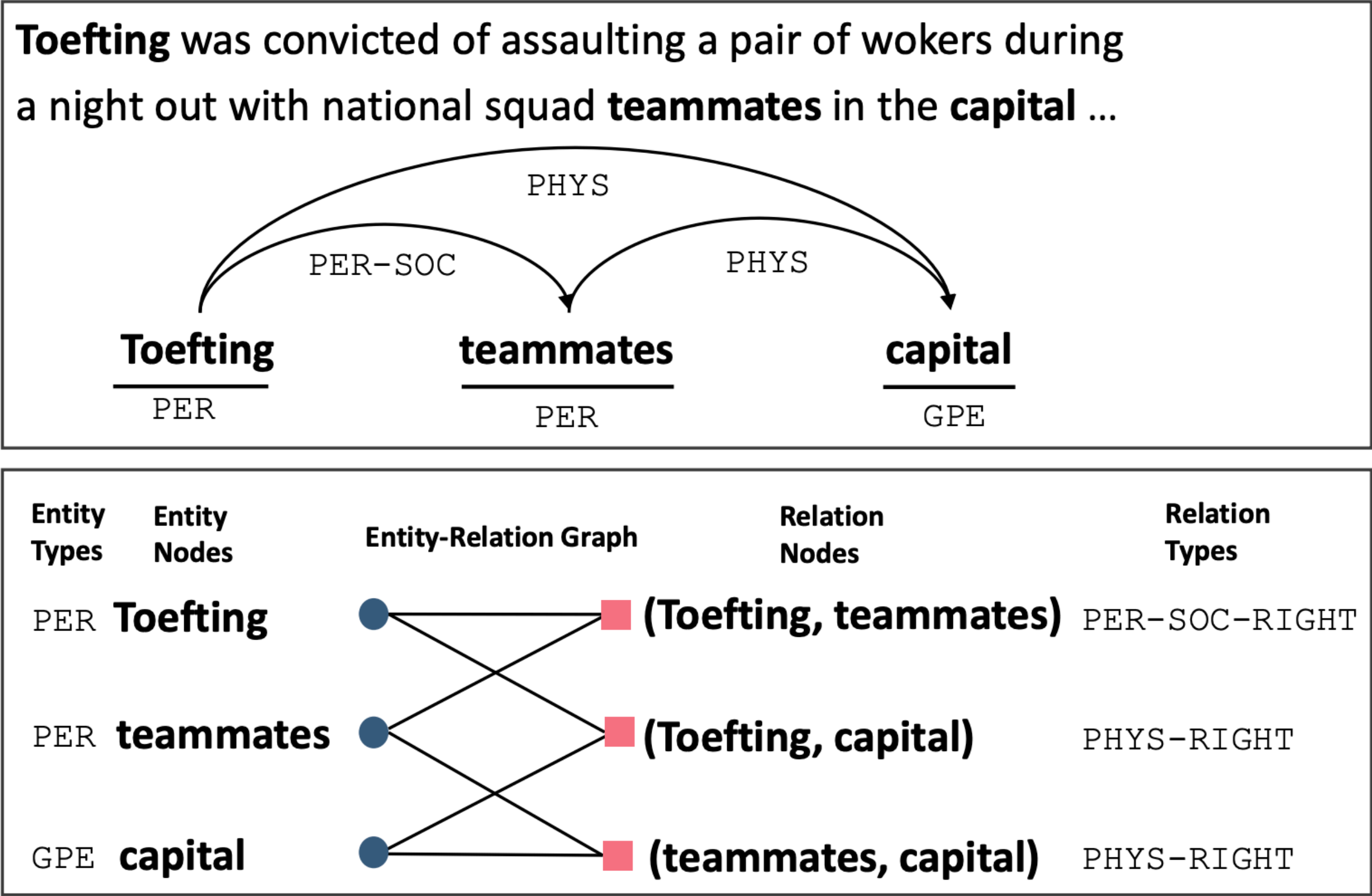
- 联合处理边界识别和类型推理
- 预测实体类型时很难利用到关系的信息

基于GCN的联合实体关系抽取



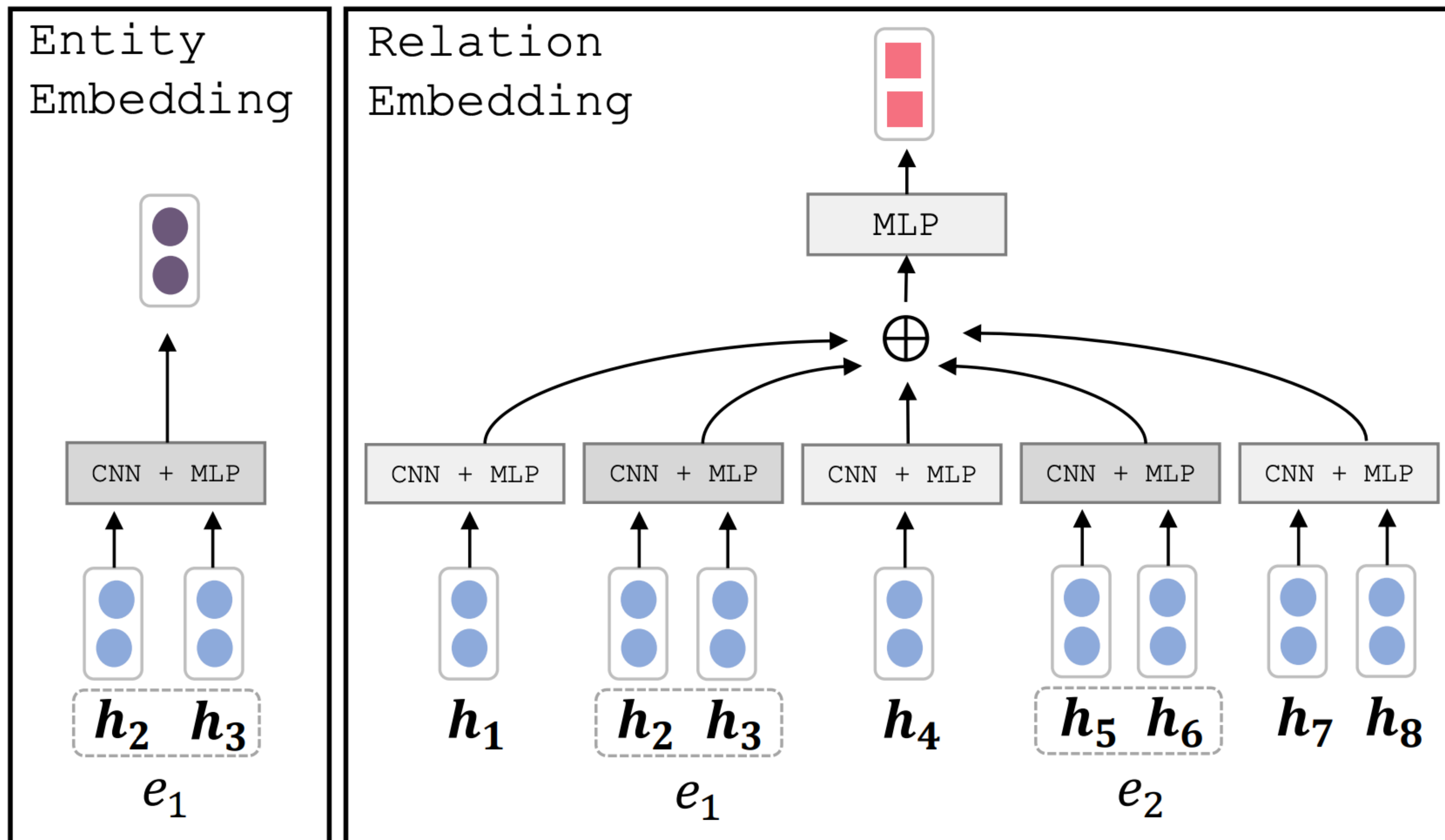
- 将实体边界识别任务从实体识别任务中分离
- 重新定义联合实体关系抽取
 - 实体边界识别
 - 实体关系类型推理
- 利用图卷积网络融合实体和关系的特征

实体-关系二分图



- 对于每个无类型实体， 将其视为一个实体节点
- 对于每个实体-实体对， 将其视为一个关系节点

节点向量



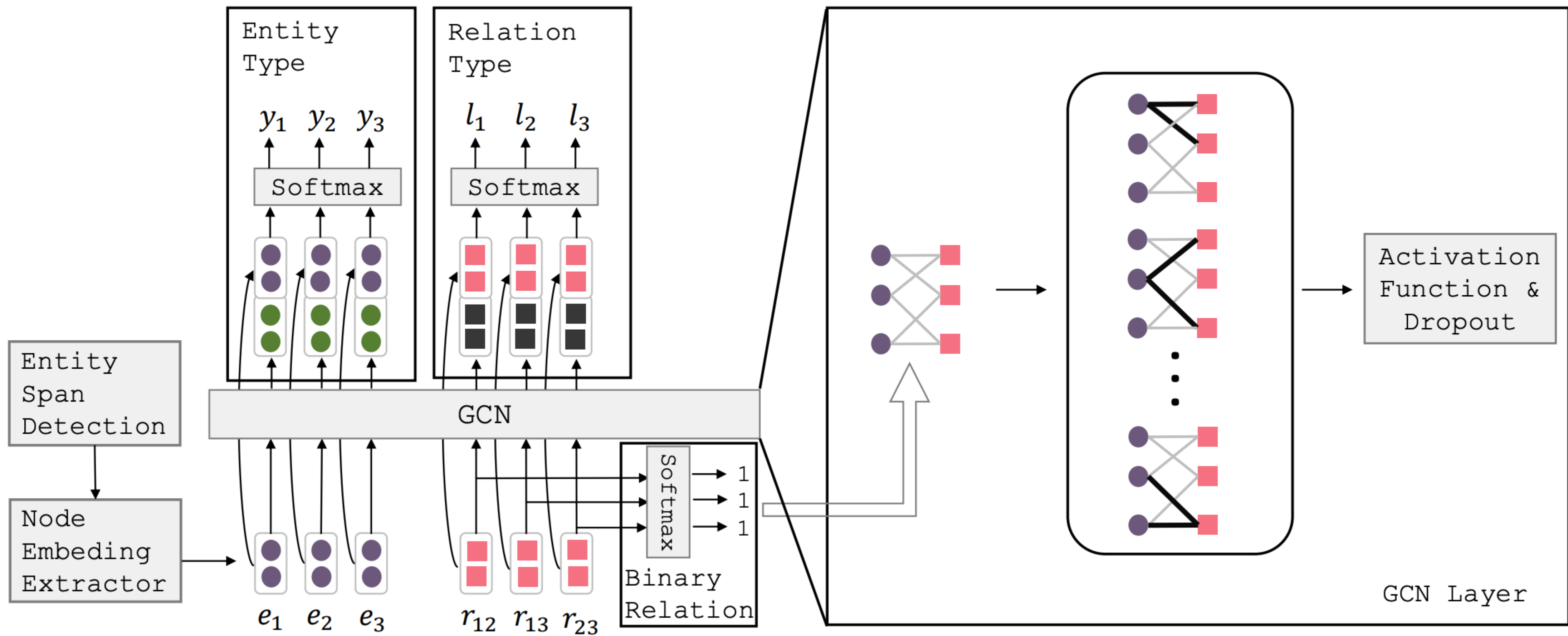
实体-关系二分图

- 静态图
 - 关系节点和对应的两个实体节点之间连边
 - 实体节点之间不连边，关系节点之间不连边
- 动态图：如果两个实体存在关系，则连两条边；如果没有关系，则不连任何一条边
 - 去除冗余的边
 - 二元关系分类

动态图

- Hard $\mathbf{A}$: $\mathbf{A}$ 是图中的邻接矩阵
 - 如果 $P(\hat{b} = 1|r_{ij}, s) > 0.5$, 则 $\mathbf{A}_{e_i, r_{ij}} = \mathbf{A}_{r_{ij}, e_i} = \mathbf{A}_{e_j, r_{ij}} = \mathbf{A}_{r_{ij}, e_j} = 1.0$
 - $\mathbf{A}$ 中对角线元素都为 1.0
 - 其他位置都为 0.0
- Soft $\mathbf{A}$
 - $\mathbf{A}_{e_i, r_{ij}} = \mathbf{A}_{r_{ij}, e_i} = \mathbf{A}_{e_j, r_{ij}} = \mathbf{A}_{r_{ij}, e_j} = P(\hat{b} = 1|r_{ij}, s)$
 - $\mathbf{A}$ 中对角线元素都为 1.0

基于GCN的联合实体关系抽取



实验

Model	Entity			Relation		
	P	R	F	P	R	F
L&J (2014)	85.2	76.9	80.8	65.4	39.8	49.5
Zhang (2017)	-	-	83.5	-	-	57.5
Sun (2018)	83.9	83.2	83.6	64.9	55.1	59.6
M&B (2016)	82.9	83.9	83.4	57.2	54.0	55.6
K&C (2017)	84.0	81.3	82.6	55.5	51.8	53.6
NN	85.7	82.1	83.9	65.6	50.7	57.2
GCN	86.1	82.4	84.2	68.1	52.3	59.1

- GCN 取得最好的实体识别性能，相当的关系抽取性能

实验

Model	Entity			Relation			Entity Span			Binary Relation		
	P	R	F	P	R	F	P	R	F	P	R	F
Sun (NN) (2018)	84.0	82.9	83.4	59.5	56.3	57.8	-	-	-	-	-	-
NN	85.7	82.1	83.9	65.6	50.7	57.2	91.2	89.6	90.4	-	-	-
GCN (static)	85.0	82.6	83.8	66.6	51.3	57.8	90.8	90.2	90.5	-	-	-
GCN (dynamic + soft)	85.3	82.3	83.8	67.3	51.6	58.5	90.8	90.2	90.5	77.3	56.4	65.2
GCN (dynamic + hard)	86.1	82.4	84.2	68.1	52.3	59.1	91.2	89.5	90.4	78.2	56.3	65.4

- 动态图的有效性和可解释性

总结

- 数据
 - 利用语言学规则进行远程监督
 - 利用高质量人工标注数据帮助远程监督
- 联合模型
 - 利用MRT进行联合实体关系抽取
 - 利用GCN进行联合实体关系抽取
- 未来工作
 - 预训练； 负样本

谢谢 Q&A