

## 用于实体集合拓展的端到端自举网络

End-to-End Bootstrapping Neural Network for Entity Set Expansion. In: AAAI 2020.

颜令勇, 韩先培, 何笨, 孙乐

联系人: 韩先培 (xianpei@iscas.ac.cn)

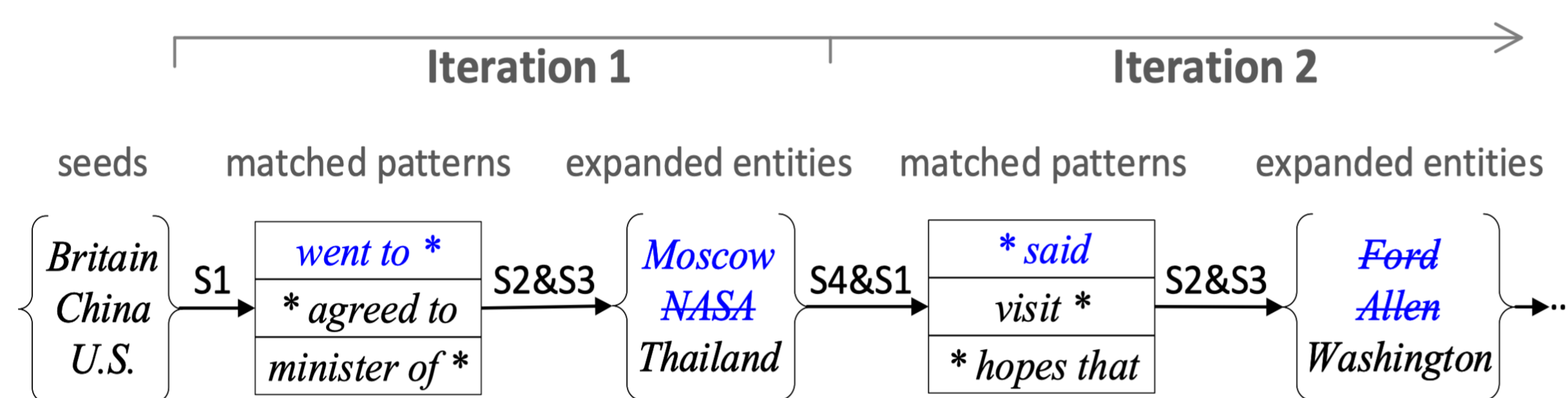
### 实体集合拓展及自举方法

- 给定某语义类下的少量实体为种子, 拓展出同属该语义类的其他实体

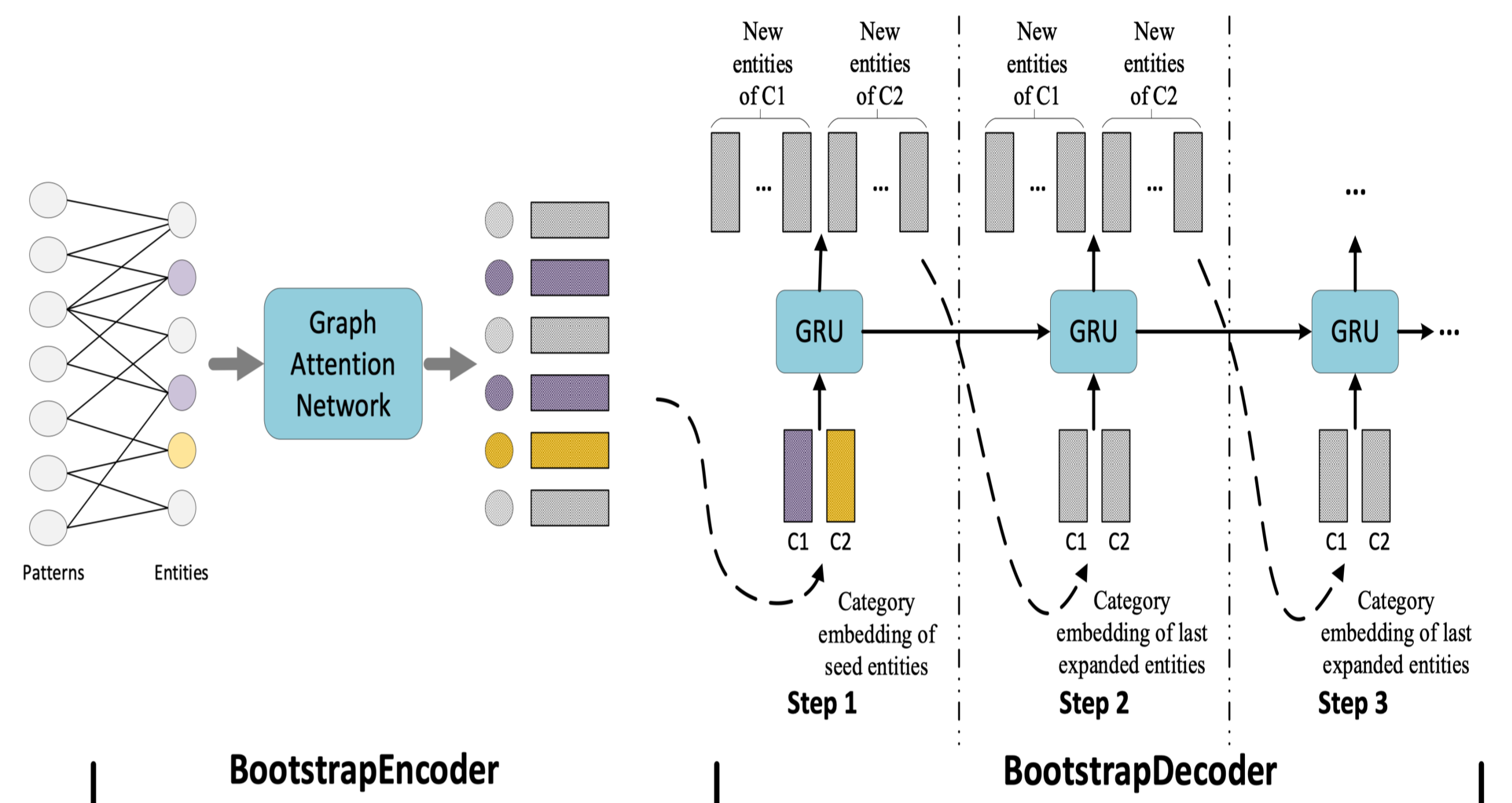
{Britain, China, U.S.}

{Britain, China, U.S., Japan, German, Thailand, Egypt, France, Russia, ...}

- 基于自举的方法: 迭代利用种子集合评估上下文模板和实体, 并添加少量新实体到种子集合



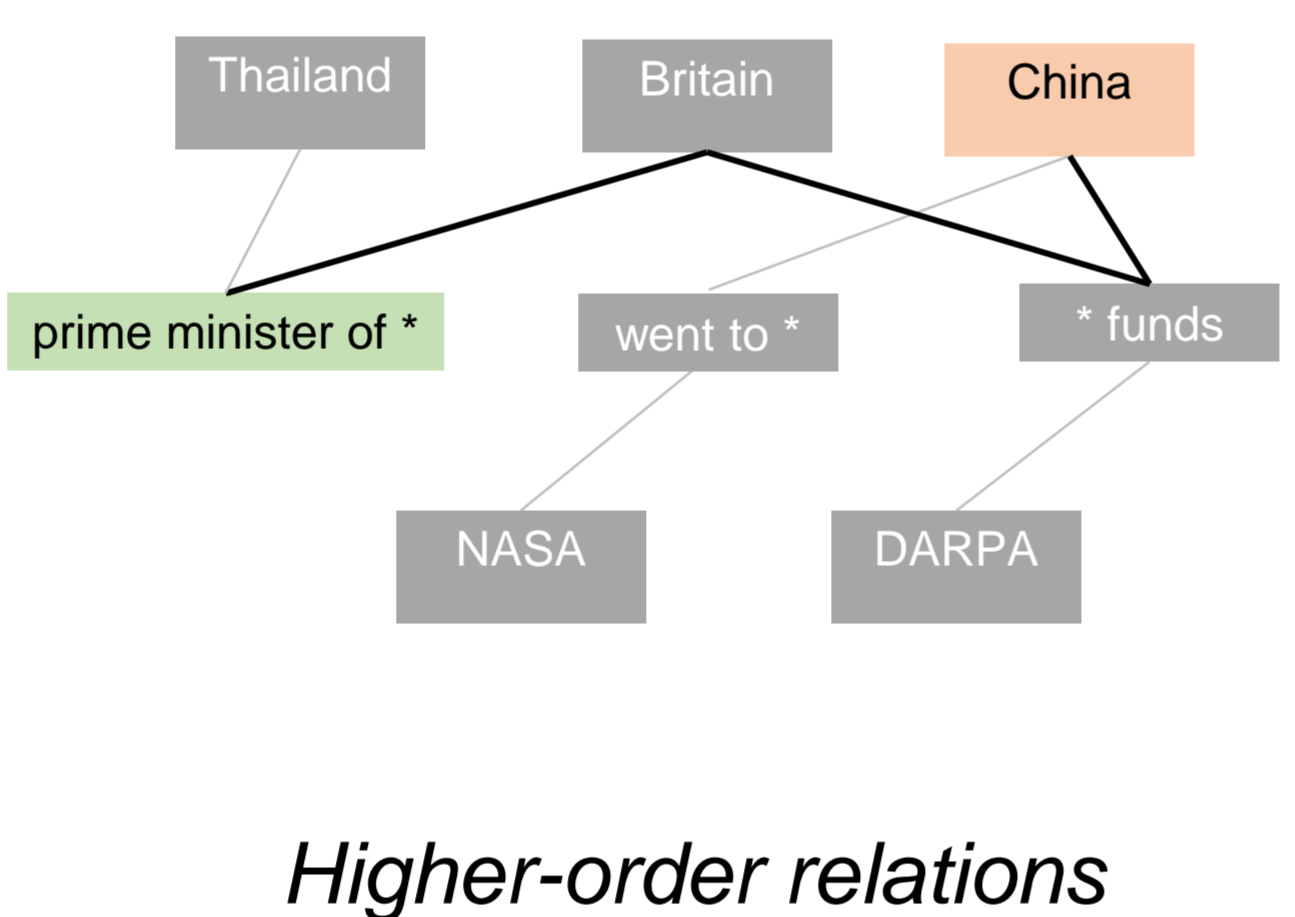
### 端到端的自举网络—BootstrapNet



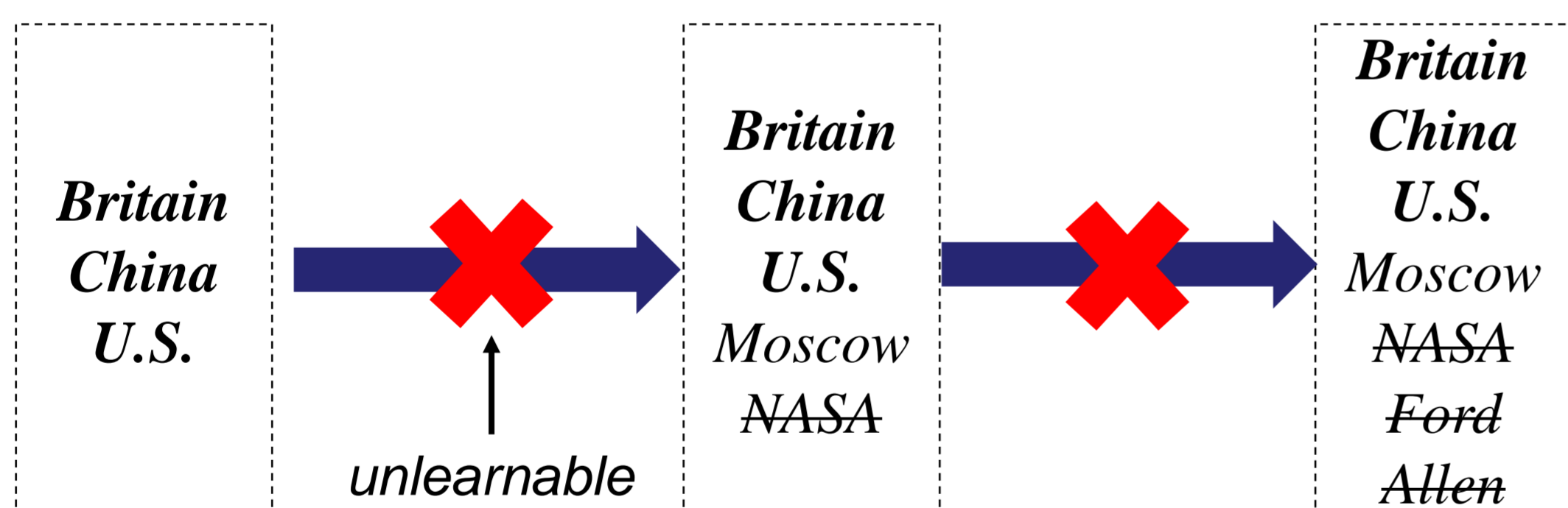
### 主要挑战

- 局部视野:** 评估实体/模板时只考虑实体-模板的直接关联 (局部视野), 忽略了高阶关联 (全局视野)

First-order relations  
 <China, went to \*>  
 <China, \* agreed to \*>  
 <China, \* funds \*>  
 <Thailand, prime minister of \*>  
 <NASA, went to \*>  
 <Britain, prime minister of \*>  
 <Britain, funds \*>  
 <DARPA, \* funds \*>  
 <DARPA, \* is trying to \*>

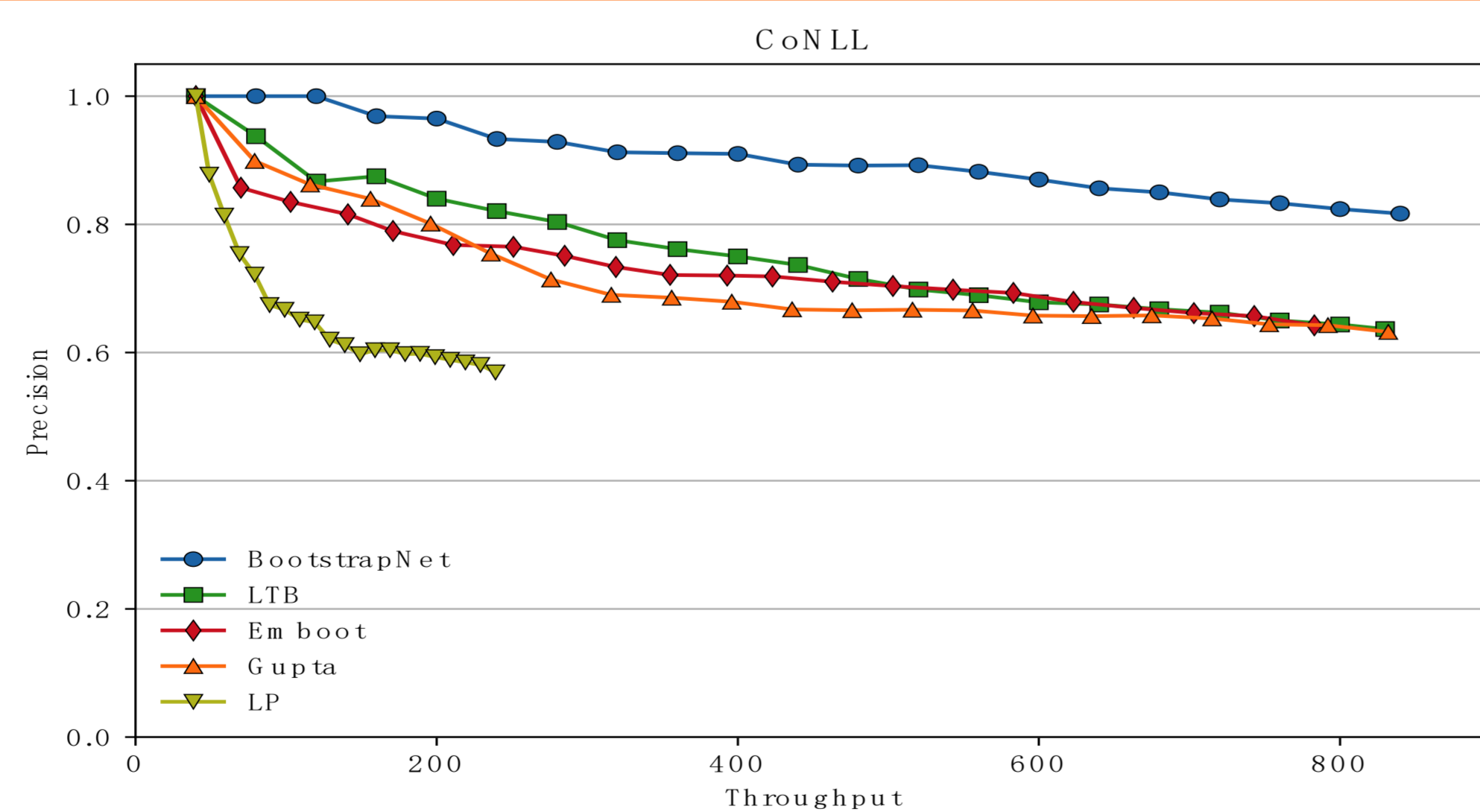


- 建模/学习困难:** 缺少对于拓展过程的统一建模, 且不可学习

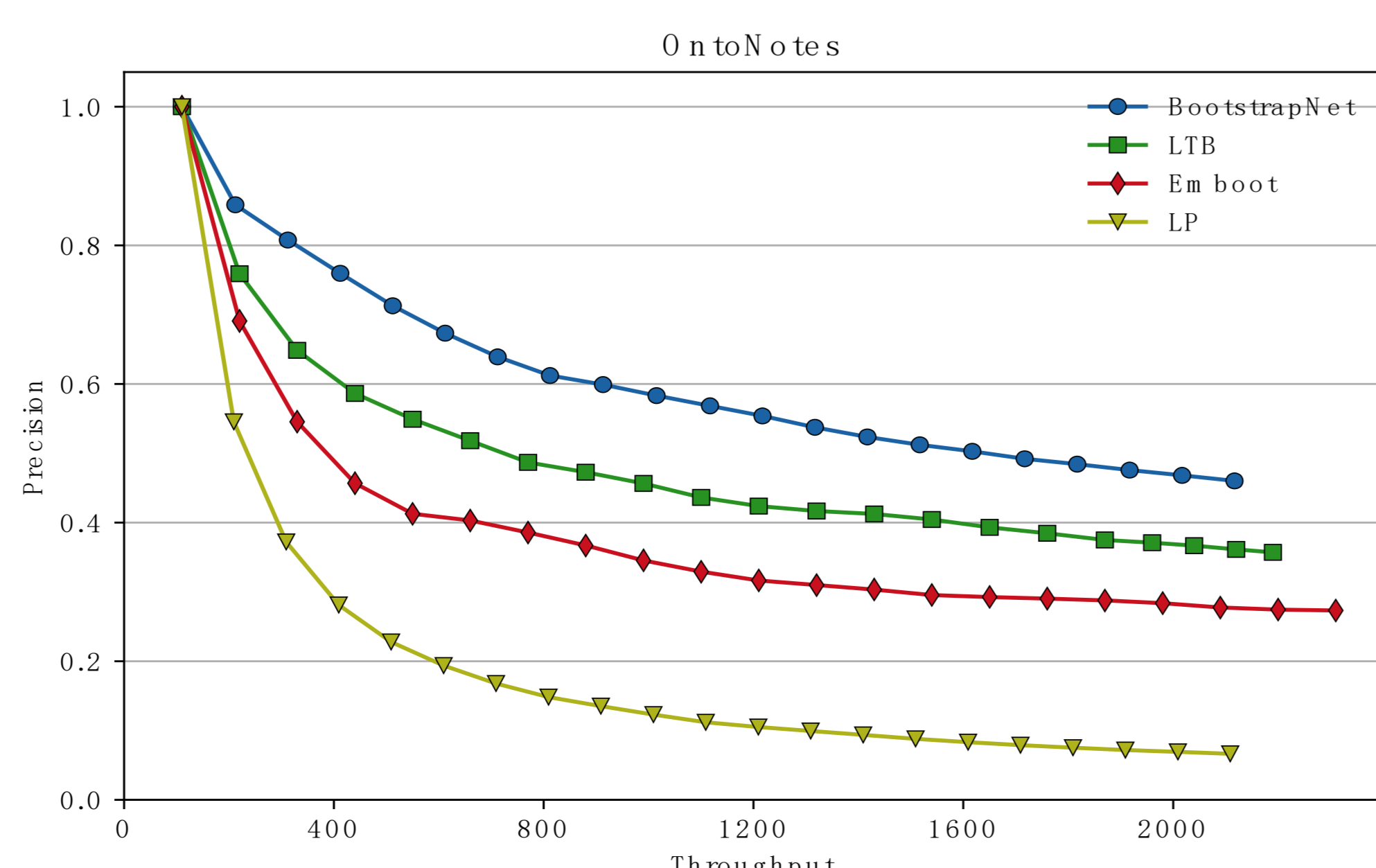


### 实验结果

- 在稠密数据集上的性能



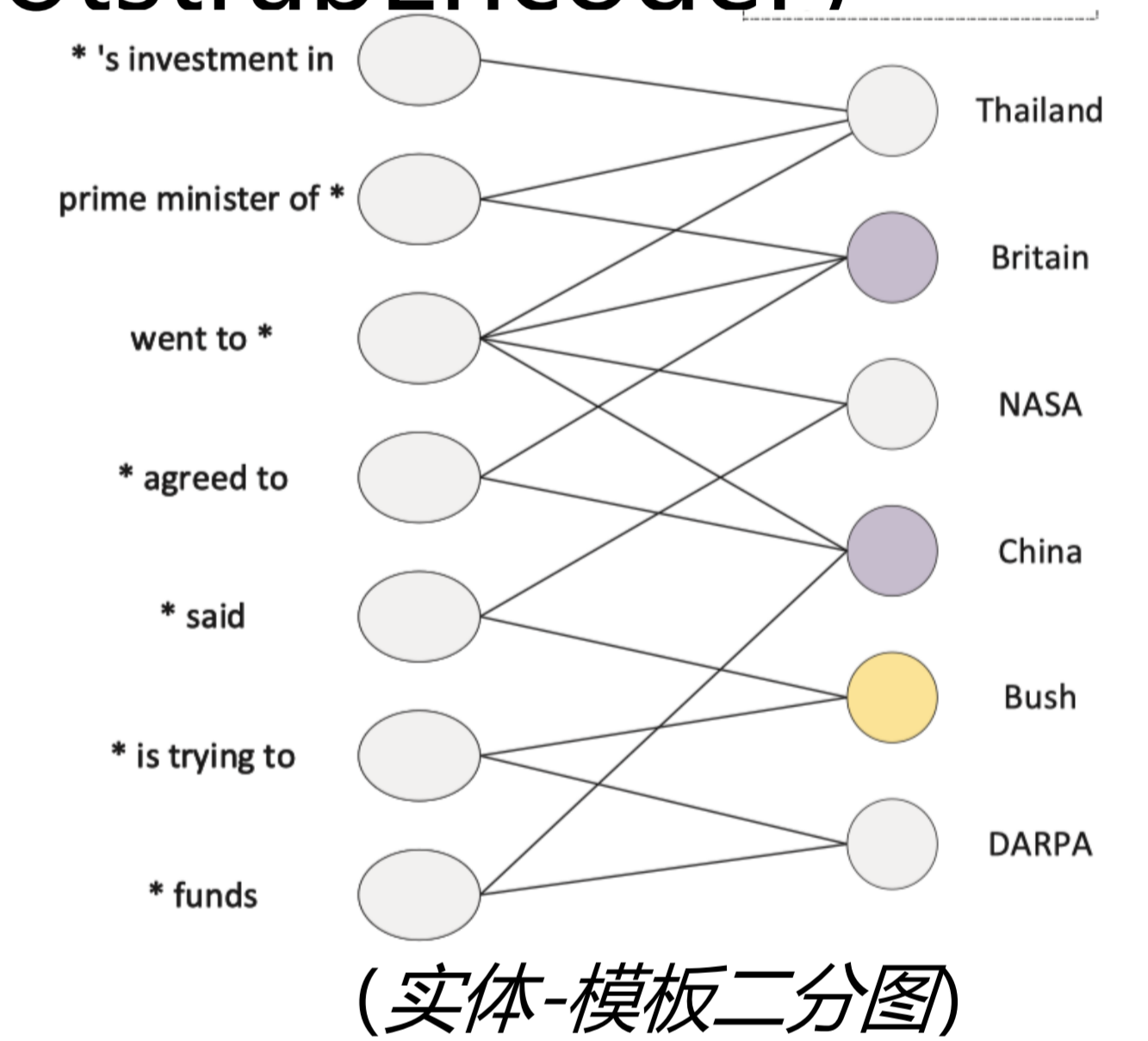
- 在稀疏数据集上的性能



- BootstrapNet: “编码器-解码器” 架构

- 捕捉高阶关联的编码器 (BootstrapEncoder)

- 以实体-模板二分图为输入
- 使用图神经网络模型捕捉高阶的实体-模板关联
- 将高阶信息编码为实体 embedding



- 建模拓展过程的解码器 (BootstrapDecoder)

- 解码器的隐状态表示语义类的语义
- 与语义类语义最相似的实体被拓展
- 拓展实体会更新语义类的语义表示

### 端到端的模型训练—多视角学习

- 多视角的模型学习:
  - 视角1: 序列化的生成 (BootstrapNet)
  - 视角2: 非序列化的分类 (BootstrapTeacher, 使用BootstrapEncoder加多层感知机做分类器)

