文献Li et al. Path reasoning over knowledge graph: A multi-agent and reinforcement learning based method. IEEE ICDMW 2018

1.摘要部分：

现在路径推理方法注重关系选择，低估实体选择。多智能体和强化学习路径推理（*MARLPaR*），用2个智能体分别以迭代形式进行关系选择和实体选择来进行复杂推理（发现隐形知识），然后用2个数据集做实验验证方法正确。

2.Introduction

目前：路径排序算法（PRA）再KG中寻找推理路径，强化学习（RL）框架与KG相关的任务也是有效的。路径推理更适合被表述为RL问题，目标：对一系列关于选择合适关系边的决策然后找到答案。（预测任务）

上面的缺点：模型/算法在给定2个实体的情况下预测缺失的关系（entity y1,?,entity2），或者只评估三元组的真值((entity y1,relation,entity y2)?)。

*MINERVA*类似RL处理问答（entity，relation，?）推理，从head entity到目标entity学习，能在图上遍历并找到目标entity。

实体选择比事实预测更重要，事实预测为找到两个实体之间的路径，实体选择更重要因为在问答中目标实体（答案实体）是未知的。

实体选择作为单独的模块，路径被分为2个步骤：1）逻辑关系路径查找，为了找到最合适的关系。2）答案实体推理，找到最合适的实体，知道找到最有效的答案实体

设计了*MARLPaR*方法（2个智能体分别为关系选择智能体和实体选择智能体）

3.Related works

A。基于知识图谱嵌入方法*TransE*：通过head entity的嵌入和relation的嵌入相加翻译成tail entity的嵌入。

语义匹配模型*RESCAL*将每个关系作为一个矩阵，对潜在因子之间的成对交互建模。DistMult通过将关系表示限制为对角矩阵来简化*RESCAL*。

ConvE……

上面对多跳关系路径建模难以解释

B。Multi-Hop Link Prediction Approaches

PRA在知识图谱上优先搜索，找到实体对之间的relation，监督学习挑选最有希望的relation，最后预测。还有用LSTM建模….

4. *MARLPaR*方法

A。强化学习系统：MDP（五元组：状态/动作空间，奖励函数，策略集，转移函数）过程为智能体-环境接口对顺序决策过程进行建模

B。关系选择智能体：状态空间，动作空间，策略集

，为当前实体，同时编码查询relation和head entity为知识

结合LSTM（长短期记忆）编码为连续向量

Noop动作：允许智能体在当前时间步不采取任何行动，相当于早起终止机制，找到答案就停止。

C。实体选择智能体：

状态空间：，同上

动作空间：，通过关系连接到当前实体的所有tail entity

策略集：

因为答案实体位置，所以模型选择与推理relation中选择已经有的实体语义相近的entity，提高正确率。

D。交替训练策略：

1.先单独训练关系选择智能体，实体选择采用随机采样

2.交替固定一个智能体，训练另外一个智能体

优化目标：，

奖励设计：智能体在t=T时接受奖励，如果实体是正确答案奖励为1，否则为0.用移动平均累计折扣奖励。

5.实验

数据集：WN18RR和NELL-995

数据预处理：在每个三元组添加逆三元组，让模型回溯。

处理后WN18RR有22个relations和NELL-995有400个relations

将平均尾实体集合基数大于1.5的关系分类为1-N/N-N关系。WN18RR中32%的关系为1-N/N-N关系，NELL-995中为27%。

用Hits@N指标评估（正确entity在all entity中排名前N的百分比）

结果：WN18RR：MARLPaR在所有Hits@N指标上都优于基线方法

NELL-995: 在9个关系任务中的大多数任务上，MARLPaR在Hits@1和Hits@3指标上表现更好

6.结论：

联合训练了一个关系选择智能体和一个实体选择智能体。关系选择智能体用于为特定查询关系找到共同的逻辑路径。实体选择智能体试图从关系的尾部实体集中选择最合适的实体