Differentiable Reasoning on Large Knowledge Bases and Natural Language

UCL Centre for Artificial Intelligence, University College London
Facebook AI Research

凌静

20201/4/13

End-to-End Differentiable Proving - NIPS2017（NTP）

- 任务

  处理KB查询，r(h,?) or r(?,t)，如:grandpaOf(ABE,?)

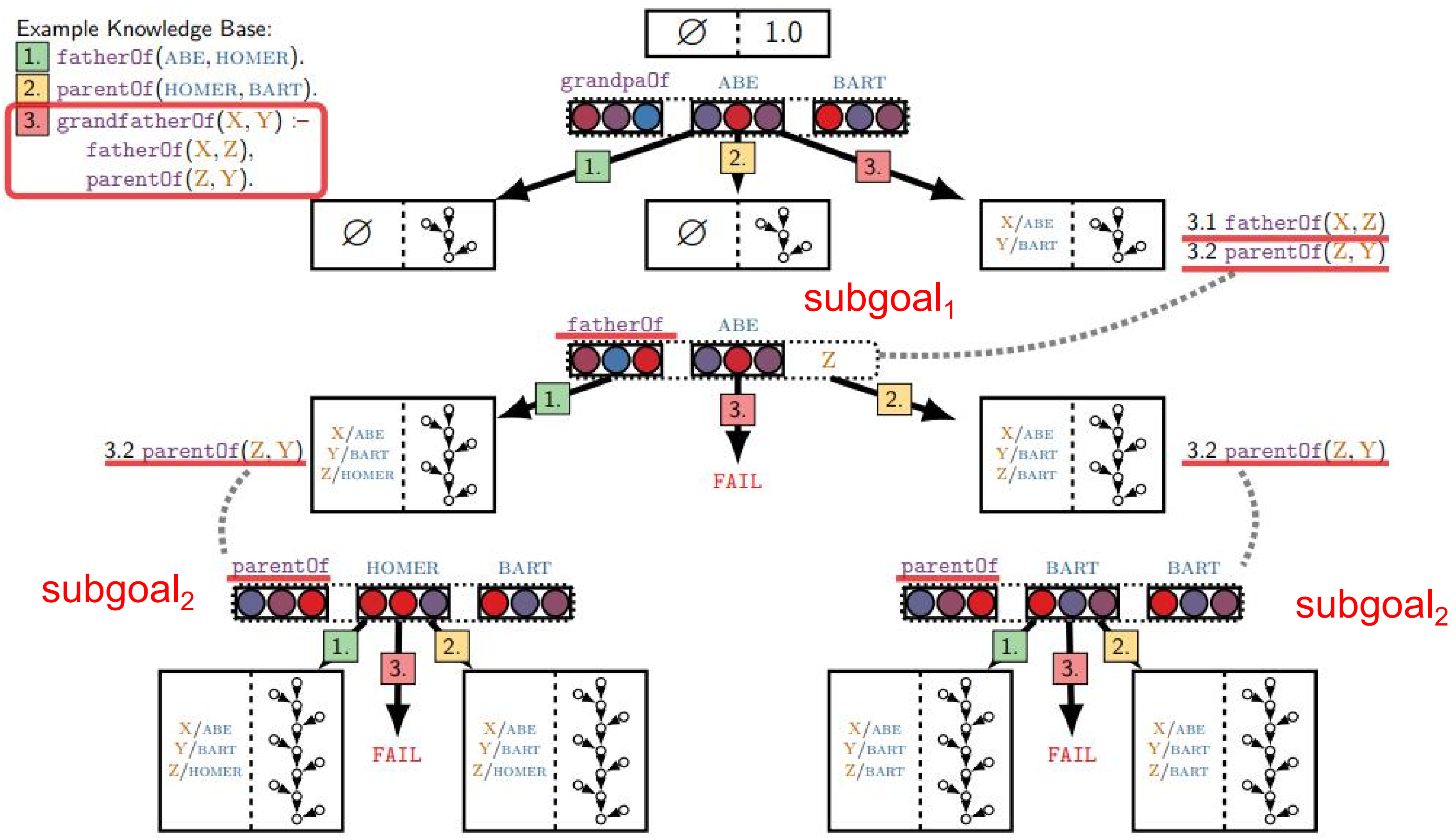
- 动机

  神经链接预测模型能建模相似性，很难处理多跳推理，没有解释性；
  符号证明太严苛，缺乏相似但不相同符号的查询的能力，容易受噪音干扰。

- 主要贡献

  可解释的可微推理，针对查询给出答案的同时给出证明“路径”
  通过训练参数反推出知识库上的一些规则，如： 0.90 locatedIn(X,Y) :– locatedIn(X,Z), locatedIn(Z,Y).

Neural Program Induction



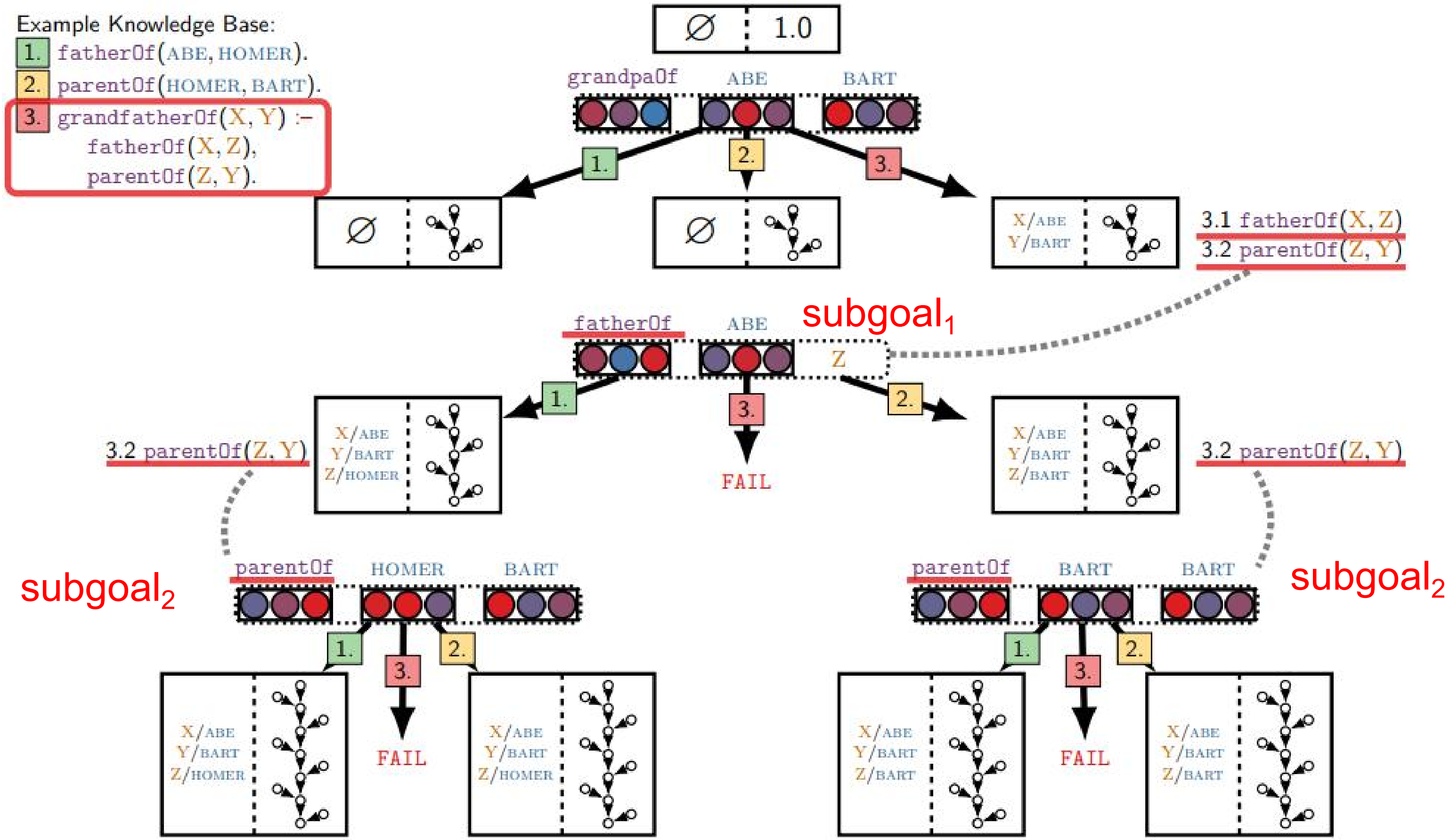
路径3-1-1表示的证明：
grandpaOf(ABE,BART) <=
fatherOf(ABE,HOMER) &
parentOf(HOMER,BART)

1. 枚举规则（OR module），利用嵌入相似性计算目标与规则的匹配度（Unification module）

$$\texttt{unify}_{\boldsymbol{\theta}}(h:\mathrm{H}, g:\mathrm{G}, S) = \texttt{unify}_{\boldsymbol{\theta}}(\mathrm{H}, \mathrm{G}, S') = (S'_\psi, S'_\rho)$$

$$S'_\psi = \left\{ \begin{array}{ll} S_\psi \cup \{h/g\} & \text{if } h \in \mathcal{V} \\ S_\psi \cup \{g/h\} & \text{if } g \in \mathcal{V}, h \notin \mathcal{V} \\ S_\psi & \text{otherwise} \end{array} \right\}, \quad S'_\rho = \min\left(S_\rho, \left\{ \begin{array}{ll} \exp\left(\frac{-\|\boldsymbol{\theta}_{h:} - \boldsymbol{\theta}_{g:}\|_2}{2\mu^2}\right) & \text{if } h, g \notin \mathcal{V} \\ 1 & \text{otherwise} \end{array} \right\} \right)$$

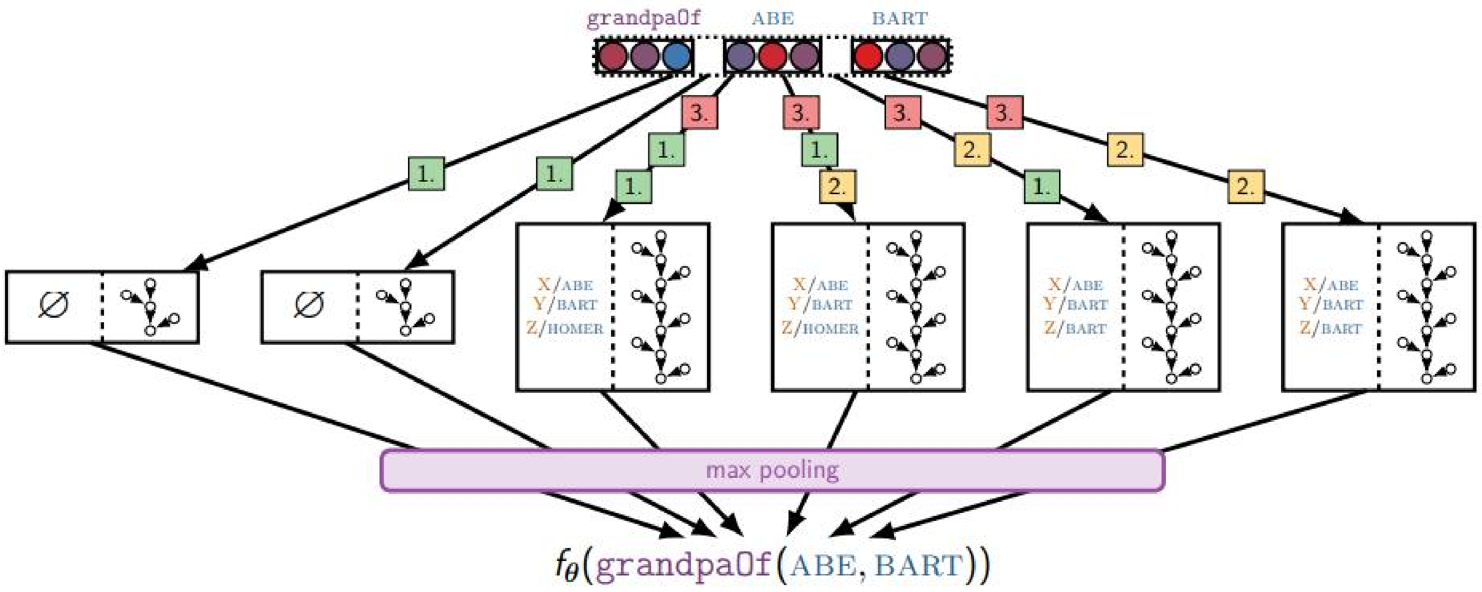
Neural Program Induction



路径3-1-1表示的证明：
grandpaOf(ABE,BART) <=
fatherOf(ABE,HOMER) &
parentOf(HOMER,BART)

1. 枚举规则（OR module），利用嵌入相似性计算目标与规则的匹配度（Unification module）
2. 处理子目标(AND module)
3. 取分数最大的路径作为结果

Training Objective



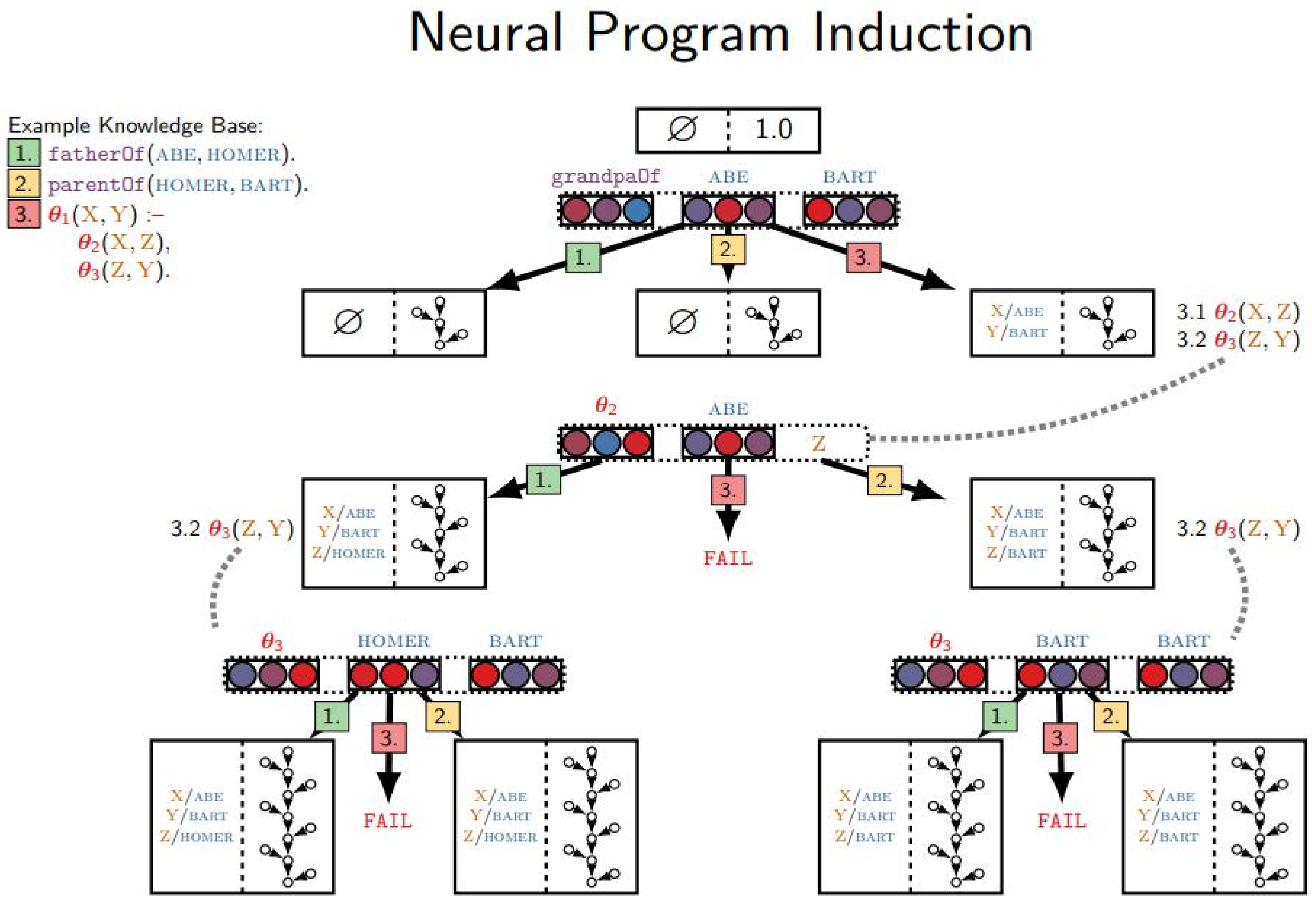
- 扰动已知事实获得negative facts
- 为了加速表示学习，与complEx联合学习

$$\mathtt{ntp}_{\boldsymbol{\theta}}^{\mathfrak{K}}(\mathrm{G}, d) = \underset{\substack{S \in \mathtt{or}_{\boldsymbol{\theta}}^{\mathfrak{K}}(\mathrm{G}, d, (\varnothing, 1)) \\ S \neq \mathtt{FAIL}}}{\arg\max} S_\rho$$

$$\mathcal{L}_{\mathtt{ntp}_{\boldsymbol{\theta}}^{\mathfrak{K}}} = \sum_{([s,i,j],y) \in \mathcal{T}} -y \log(\mathtt{ntp}_{\boldsymbol{\theta}}^{\mathfrak{K}}([s,i,j],d)_\rho) - (1-y)\log(1-\mathtt{ntp}_{\boldsymbol{\theta}}^{\mathfrak{K}}([s,i,j],d)_\rho)$$

$$\mathcal{L}_{\mathtt{ntp}\lambda_{\boldsymbol{\theta}}^{\mathfrak{K}}} = \mathcal{L}_{\mathtt{ntp}_{\boldsymbol{\theta}}^{\mathfrak{K}}} + \sum_{([s,i,j],y) \in \mathcal{T}} -y \log(\mathtt{complex}_{\boldsymbol{\theta}}(s,i,j)) - (1-y)\log(1-\mathtt{complex}_{\boldsymbol{\theta}}(s,i,j))$$

从参数中获取规则



- 指定规则模板及对应数量
- 训练结束后使用Radial Basis Function kernel获取表示与已有谓词表示相似度
- 证明路径中的最小值作为规则置信度

Differentiable Reasoning on Large Knowledge Bases and Natural Language - AAAI2020（GNTP）

- 动机

  NTP只能用在小规模知识库上

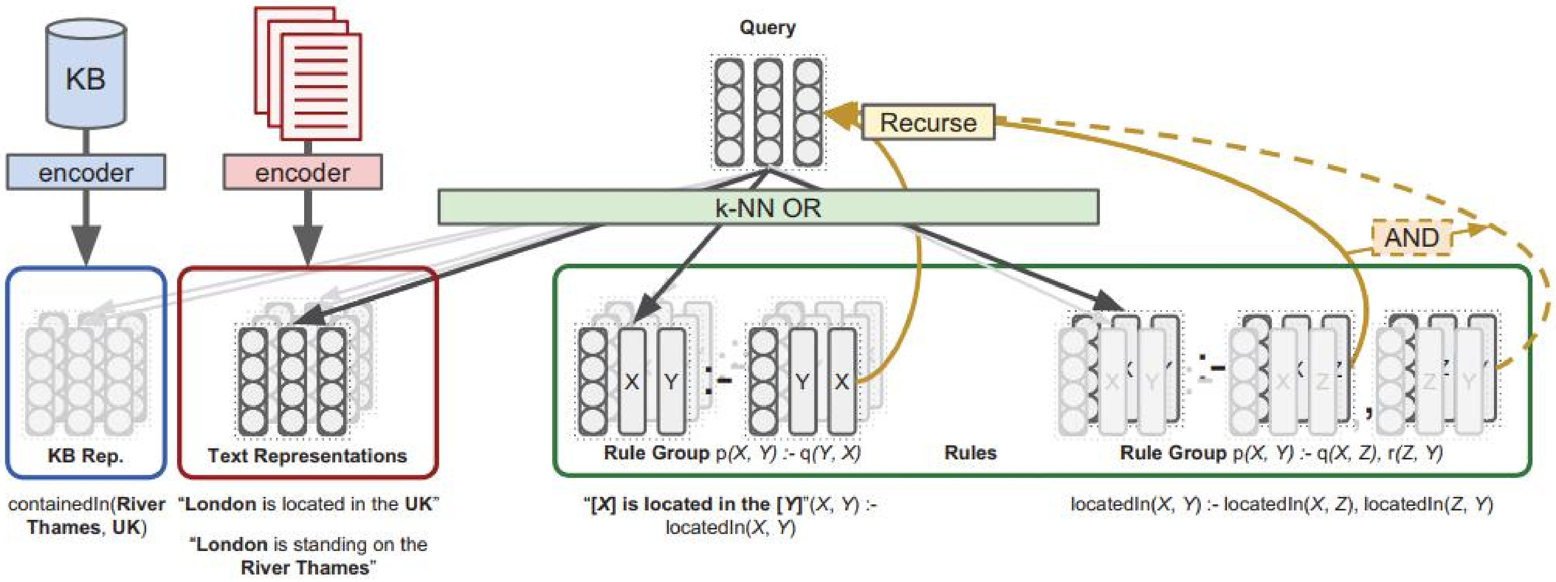
  大多数人类知识不是在KBs中获得的，而是在难以自动推理的自然语言文本中获得的

- 主要贡献

  有限损失内加速NTP，保留可解释性

  知识库和文本提及联合推理，在共享的嵌入空间中嵌入事实和自然语言句子

- 关键方法
  - L2距离最近邻，路径剪枝
  - 诱导规则的注意力机制，减少参数
  - 对自然语言扩展，联合嵌入谓词和文本



- 事实选择 (Nearest Neighbour Search (NNS))
- 规则选择(在共享相同模板的规则之间动态选择规则)
- 注意力机制

$$\mathrm{H} = [\boldsymbol{\theta}_{p:}, \mathrm{X}, \mathrm{Y}] \text{ and } \mathrm{B} = [[\boldsymbol{\theta}_{q:}, \mathrm{X}, \mathrm{Z}], [\boldsymbol{\theta}_{r:}, \mathrm{Z}, \mathrm{Y}]]$$

  把规则中k维的表示变成在已知数量为r的各个谓词上的注意力分数（适用r<k的情况）
- 联合训练

  文本编码：词嵌入求平均，当成谓词一起训练

$$t = [\texttt{[arg1]}, \texttt{borders}, \texttt{with}, \texttt{[arg2]}]$$

$$\texttt{neighborOf}(\mathrm{X},\mathrm{Y}) :\!- \text{“Y is a neighboring state to X”}(\mathrm{X},\mathrm{Y})$$
$$\texttt{locatedIn}(\mathrm{X},\mathrm{Y}) :\!- \text{“X is a neighboring state to Z”}(\mathrm{X},\mathrm{Z}), \text{“Z is located in Y”}(\mathrm{Z},\mathrm{Y})$$

Datasets		Metrics	Models					Rules Learned by GNTP
			NTP [7]	GNTP		NeuralLP	MINERVA	
				Standard	Attention			
Countries	S1	AUC-PR	90.83 ± 15.4	99.98 ± 0.05	**100.0 ± 0.0**	**100.0 ± 0.0**	**100.0 ± 0.0**	`locatedIn(X,Y) :– locatedIn(X,Z), locatedIn(Z,Y)`
	S2		87.40 ± 11.7	90.82 ± 0.88	**93.48 ± 3.29**	75.1 ± 0.3	92.36 ± 2.41	`neighborOf(X,Y) :– neighborOf(X,Z), locatedIn(Z,Y)`
	S3		56.68 ± 17.6	87.70 ± 4.79	91.27 ± 4.02	92.20 ± 0.2	**95.10 ± 1.20**	`neighborOf(X,Y) :– neighborOf(Y,X)`
Kinship		MRR	0.35	0.719	**0.759**	0.619	0.720	`term0(X, Y) :– term0(Y, X)`
		HITS@1	0.24	0.586	**0.642**	0.475	0.605	`term4(X, Y) :– term4(Y, X)`
		HITS@3	0.37	0.815	**0.850**	0.707	0.812	`term13(X,Y) :– term13(X, Z), term10(Z, Y)`
		HITS@10	0.57	0.958	**0.959**	0.912	0.924	`term2(X,Y) :– term4(X, Z), term7(Z, Y)`
Nations		MRR	0.61	**0.658**	0.645	—	—	`commonbloc1(X, Y) :– relngo(Y, X)`
		HITS@1	0.45	**0.493**	0.490	—	—	`timesincewar(X,Y) :– independence(X,Y)`
		HITS@3	0.73	**0.781**	0.736	—	—	`unweightedunvote(X,Y) :– relngo(X,Y)`
		HITS@10	0.87	**0.985**	0.975	—	—	`ngo(X, Y) :– independence(Y, X)`
UMLS		MRR	0.80	0.841	**0.857**	0.778	0.825	`isa(X,Y) :– isa(X,Z), isa(Z,Y)`
		HITS@1	0.70	0.732	**0.761**	0.643	0.728	`complicates(X,Y) :– affects(X,Y)`
		HITS@3	0.88	**0.941**	**0.947**	0.869	0.900	`affects(X, Y) :– affects(X, Z), affects(Z, Y)`
		HITS@10	0.95	**0.986**	0.983	0.962	0.968	`process_of(X,Y) :– affects(X,Y)`

规则无关三元组　　规则相关三元组

使用了FB122数据集提供的规则

		Test-I				Test-II				Test-ALL			
		Hits@N (%)			MRR	Hits@N (%)			MRR	Hits@N (%)			MRR
		3	5	10		3	5	10		3	5	10	
With Rules	KALE-Pre (Guo et al. 2016)	35.8	41.9	49.8	0.291	82.9	86.1	89.9	0.713	61.7	66.2	71.8	0.523
	KALE-Joint (Guo et al. 2016)	**38.4**	**44.7**	**52.2**	0.325	79.7	84.1	89.6	0.684	61.2	66.4	72.8	0.523
	ASR-DistMult (Minervini et al. 2017)	36.3	40.3	44.9	0.330	98.0	99.0	99.2	0.948	70.7	73.1	75.2	0.675
	ASR-ComplEx (Minervini et al. 2017)	37.3	41.0	45.9	**0.338**	**99.2**	**99.3**	**99.4**	**0.984**	71.7	73.6	75.7	0.698
	KBLR (García-Durán and Niepert 2018)	–	–	–	–	–	–	–	–	**74.0**	**77.0**	**79.7**	**0.702**
Without Rules	TransE (Bordes et al. 2013)	36.0	41.5	48.1	0.296	77.5	82.8	88.4	0.630	58.9	64.2	70.2	0.480
	DistMult (Yang et al. 2015)	36.0	40.3	45.3	0.313	**92.3**	**93.8**	**94.7**	0.874	**67.4**	**70.2**	**72.9**	0.628
	ComplEx (Trouillon et al. 2016)	**37.0**	**41.3**	**46.2**	**0.329**	91.4	91.9	92.4	**0.887**	67.3	69.5	71.9	**0.641**
	GNTPs	33.7	36.9	41.2	0.313	98.2	99.0	99.3	0.977	69.2	71.1	73.2	0.678

DistMult and ComplEx using Adversarial Sets (ASR) - a method for incorporating rules in neural link predictors via adversarial training

与KBLR和ComplEx在WN18和WN18RR上效果差不多，但是提供了可解释性

谢谢！

凌静

20201/4/13