知识推理近期工作综述

汇报人：凌静

2021/4/2

目录

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

- 任务

  处理KB查询，r(h,?) or r(?,t)

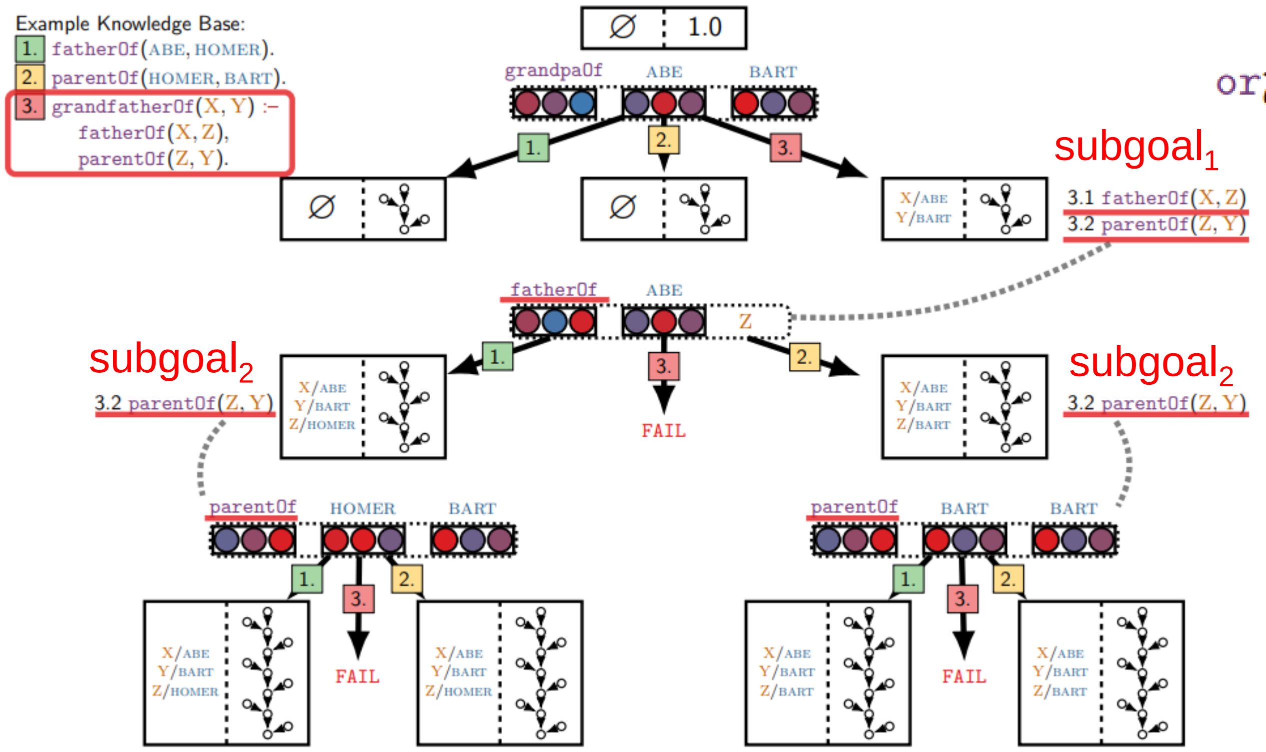
- 动机

  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

Neural Program Induction



$$\texttt{or}_{\boldsymbol{\theta}}^{\mathfrak{K}}(\mathrm{G}, d, S) = [S' \mid S' \in \texttt{and}_{\boldsymbol{\theta}}^{\mathfrak{K}}(\mathbb{B}, d, \texttt{unify}_{\boldsymbol{\theta}}(\mathrm{H}, \mathrm{G}, S)) \text{ for } \mathrm{H} :\!- \mathbb{B} \in \mathfrak{K}]$$

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

$$\texttt{and}_{\boldsymbol{\theta}}^{\mathfrak{K}}(\mathrm{G} : \mathbb{G}, d, S) = [S'' \mid S'' \in \texttt{and}_{\boldsymbol{\theta}}^{\mathfrak{K}}(\mathbb{G}, d, S')$$

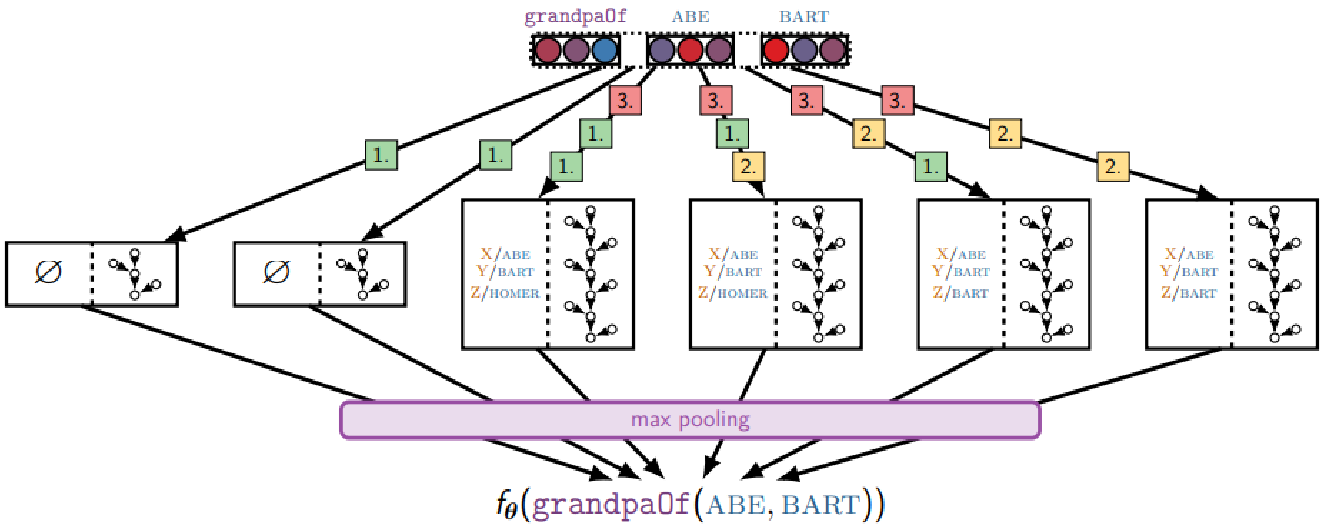
$$\text{for } S' \in \texttt{or}_{\boldsymbol{\theta}}^{\mathfrak{K}}(\text{substitute}(\mathrm{G}, S_{\psi}), d-1, S)]$$

$$\text{substitute}(g : \mathrm{G}, \psi) = \left\{ \begin{array}{ll} x & \text{if } g/x \in \psi \\ g & \text{otherwise} \end{array} \right\} : \text{substitute}(\mathrm{G}, \psi)$$

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

1. 枚举事实、规则（or module），匹配（unify module）
2. 规则变量ground到常量(and)
3. 根据规则处理子目标(and)
4. 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

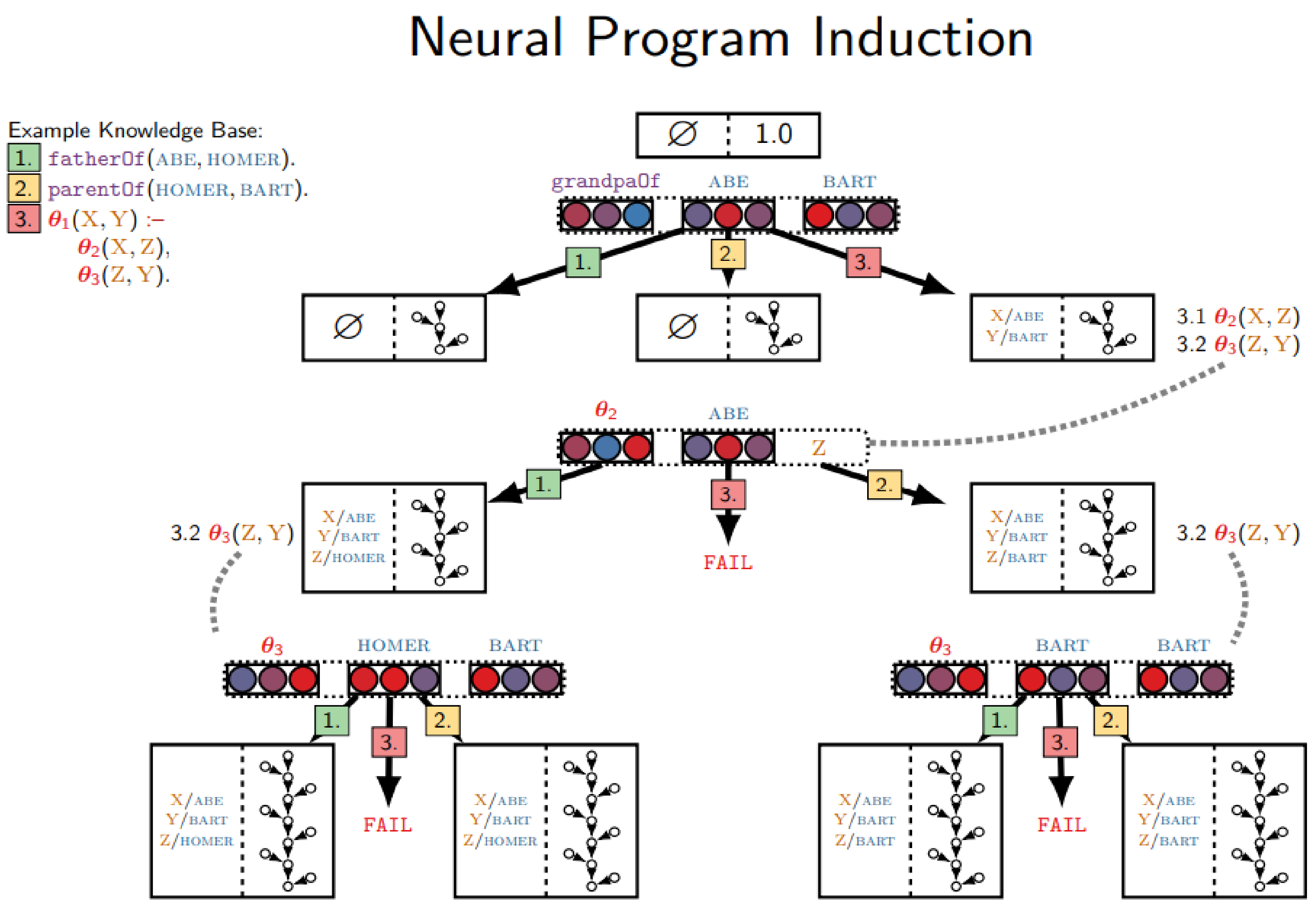
$$\mathtt{ntp}_{\boldsymbol{\theta}}^{\mathfrak{K}}(\mathrm{G}, d) = \underset{\substack{S \in \mathtt{or}_{\boldsymbol{\theta}}^{\mathfrak{K}}(\mathrm{G}, d, (\varnothing, 1)) \\ S \neq \mathtt{FAIL}}}{\arg\max} S_{\rho}$$

- Loss: negative log-likelihood w.r.t. target proof success
- Trained end-to-end using stochastic gradient descent
- Vectors are **learned such that proof success is high for known facts** and low for sampled negative facts

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

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

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

实验结果

Corpus		Metric	ComplEx	NTP	NTPλ	Examples of induced rules and their confidence
Countries	S1	AUC-PR	99.37 ± 0.4	90.83 ± 15.4	**100.00** ± 0.0	0.90 locatedIn(X,Y) :– locatedIn(X,Z), locatedIn(Z,Y).
	S2	AUC-PR	87.95 ± 2.8	87.40 ± 11.7	**93.04** ± 0.4	0.63 locatedIn(X,Y) :– neighborOf(X,Z), locatedIn(Z,Y).
	S3	AUC-PR	48.44 ± 6.3	56.68 ± 17.6	**77.26** ± 17.0	0.32 locatedIn(X,Y) :– neighborOf(X,Z), neighborOf(Z,W), locatedIn(W,Y).
Kinship		MRR	**0.81**	0.60	0.80	0.98 term15(X,Y) :– term5(Y,X)
		HITS@1	0.70	0.48	**0.76**	0.97 term18(X,Y) :– term18(Y,X)
		HITS@3	**0.89**	0.70	0.82	0.86 term4(X,Y) :– term4(Y,X)
		HITS@10	**0.98**	0.78	0.89	0.73 term12(X,Y) :– term10(X, Z), term12(Z, Y).
Nations		MRR	**0.75**	**0.75**	0.74	0.68 blockpositionindex(X,Y) :– blockpositionindex(Y,X).
		HITS@1	**0.62**	**0.62**	0.59	0.46 expeldiplomats(X,Y) :– negativebehavior(X,Y).
		HITS@3	0.84	0.86	**0.89**	0.38 negativecomm(X,Y) :– commonbloc0(X,Y).
		HITS@10	**0.99**	**0.99**	**0.99**	0.38 intergovorgs3(X,Y) :– intergovorgs(Y,X).
UMLS		MRR	0.89	0.88	**0.93**	0.88 interacts_with(X,Y) :– interacts_with(X,Z), interacts_with(Z,Y).
		HITS@1	0.82	0.82	**0.87**	0.77 isa(X,Y) :– isa(X,Z), isa(Z,Y).
		HITS@3	0.96	0.92	**0.98**	0.71 derivative_of(X,Y) :– derivative_of(X,Z), derivative_of(Z,Y).
		HITS@10	**1.00**	0.97	**1.00**	

数据集的规模都非常小
AUC-PR：area under the Precision-Recall-curve

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

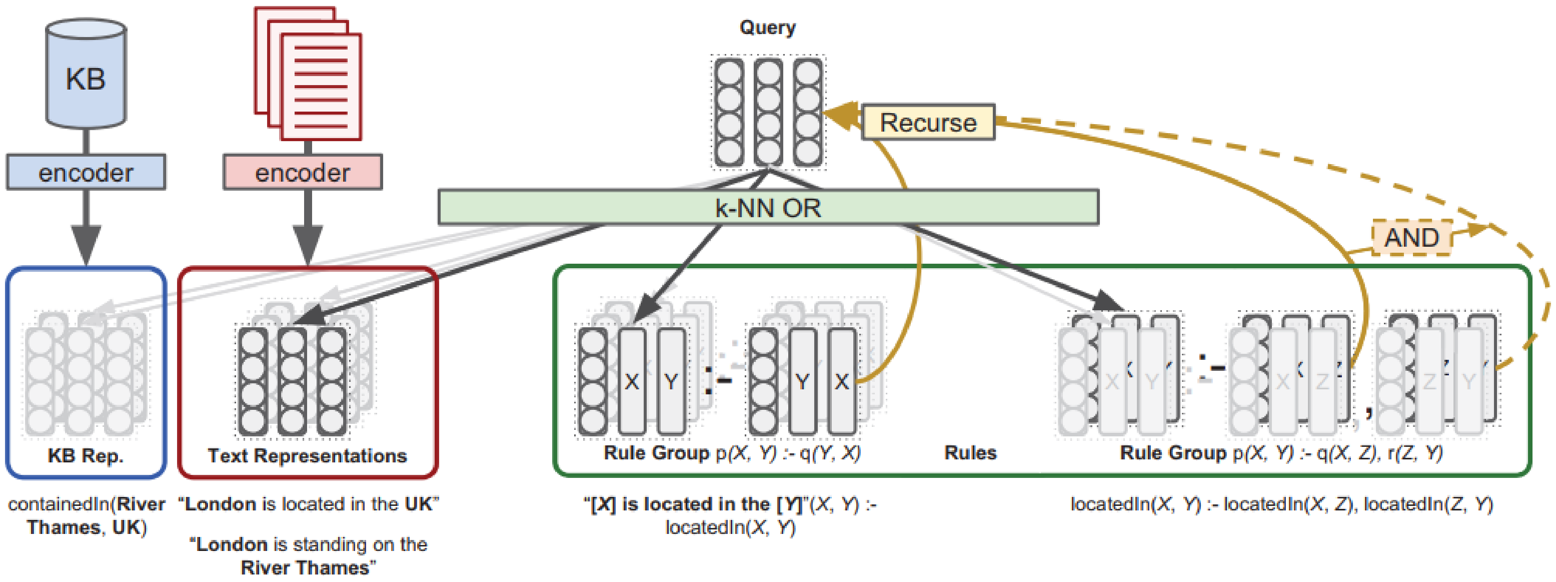
  大多数人类知识不是在KBs中获得的，而是在难以自动推理的自然语言文本中获得的

- 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

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

  $\texttt{neighborOf}(\mathrm{X}, \mathrm{Y})$:– "Y is a neighboring state to X"(X, Y)
  $\texttt{locatedIn}(\mathrm{X}, \mathrm{Y})$:– "X is a neighboring state to Z"(X, Z), "Z is located in Y"(Z, Y)

Datasets		Metrics	Models: NTP [7]	GNTP Standard	GNTP Attention	NeuralLP	MINERVA	Rules Learned by GNTP
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)

使用了FB122数据集提供的规则

		规则无关三元组 Test-I Hits@N (%) 3	5	10	MRR	规则相关三元组 Test-II Hits@N (%) 3	5	10	MRR	Test-ALL Hits@N (%) 3	5	10	MRR
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

Differentiable Learning of Logical Rules for Knowledge Base Reasoning - NIPS2017(Neural LP)

- 任务：研究了知识库推理中概率一阶逻辑规则的学习问题

- 动机

  通过为逻辑规则配备概率，可以更好地建模统计复杂和噪声数据，但这个学习问题相当困难——它需要学习结构(即模型中包含的特定规则集)和参数(即与每个规则相关联的置信度)

  许多过去的学习系统都使用了优化方法，将离散结构空间中的移动与参数空间中的移动交织起来

- 主要贡献

  第一种端到端可微&同时学习参数和逻辑规则的结构

  归纳逻辑编程方法学到的规则有相对广泛的适用性

  举例说明将神经LP模型可视化为逻辑规则的方法（从参数中导出规则）

$$\alpha \quad \mathtt{query}(\mathtt{Y},\mathtt{X}) \leftarrow \mathtt{R_n}(\mathtt{Y},\mathtt{Z_n}) \wedge \cdots \wedge \mathtt{R_1}(\mathtt{Z_1},\mathtt{X})$$

$$\mathtt{R}(\mathtt{Y},\mathtt{X}) \leftarrow \mathtt{P}(\mathtt{Y},\mathtt{Z}) \wedge \mathtt{Q}(\mathtt{Z},\mathtt{X})$$

$$\mathbf{M}_{\mathtt{P}} \cdot \mathbf{M}_{\mathtt{Q}} \cdot \mathbf{v}_{\mathtt{x}} \doteq \boxed{\mathbf{s}}$$

x实体应用到以上规则中之后，所有Y实体的得分

$$\mathbf{s} = \sum \left(\alpha_l \left(\Pi_{\mathrm{k}\in\beta_l} \mathbf{M}_{\mathrm{R_k}} \mathbf{v}_{\mathrm{x}}\right)\right), \quad \mathrm{score}(\mathrm{y} \mid \mathrm{x}) = \mathbf{v}_{\mathrm{y}}^{T} \mathbf{s}$$

$$\max_{\{\alpha_l,\beta_l\}} \sum_{\{\mathrm{x},\mathrm{y}\}} \mathrm{score}(\mathrm{y} \mid \mathrm{x}) = \max_{\{\alpha_l,\beta_l\}} \sum_{\{\mathrm{x},\mathrm{y}\}} \mathbf{v}_{\mathrm{y}}^{T} \left(\underline{\sum_{l} \left(\alpha_l \left(\Pi_{\mathrm{k}\in\beta_l} \mathbf{M}_{\mathrm{R_k}} \mathbf{v}_{\mathrm{x}}\right)\right)} \right)$$

Tensorlog: A differentiable deductive database

训练目标，β是某条规则的关系id

枚举规则操作不可微

$$\prod_{t=1}^{T} \sum_{\mathrm{k}}^{|\mathbf{R}|} a_t^{\mathrm{k}} \mathbf{M}_{\mathrm{R_k}}$$

//给每个关系分配权重参数 a_t^k

$$\mathbf{u_0} = \mathbf{v}_{\mathrm{x}}$$

$$\mathbf{u_t} = \sum_{\mathrm{k}}^{|\mathbf{R}|} a_t^{\mathrm{k}} \mathbf{M}_{\mathrm{R_k}} \left(\sum_{\tau=0}^{t-1} b_t^{\tau} \mathbf{u}_{\tau} \right) \quad \text{for } 1 \leq t \leq T$$

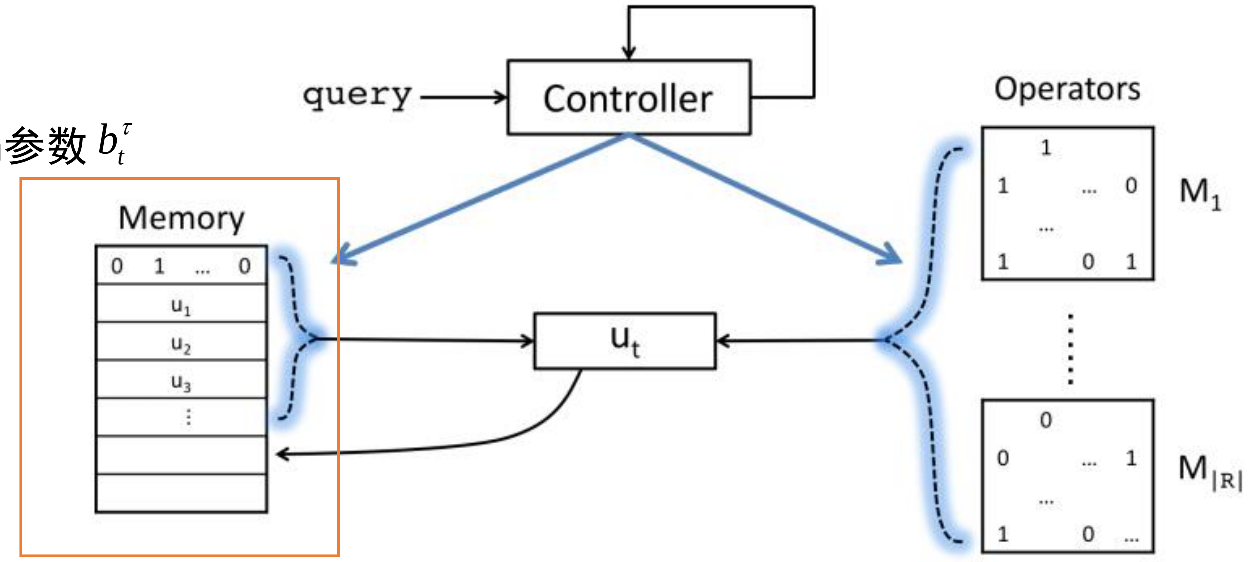
//记忆attention参数 b_t^{τ}

$$\mathbf{u_{T+1}} = \sum_{\tau=0}^{T} b_{T+1}^{\tau} \mathbf{u}_{\tau}$$

$$\mathbf{h_t} = \mathrm{update}\left(\mathbf{h_{t-1}}, \mathrm{input}\right)$$ （**RNN**）

$$\mathbf{a_t} = \mathrm{softmax}\left(W \mathbf{h_t} + b\right)$$

$$\mathbf{b_t} = \mathrm{softmax}\left([\mathbf{h_0}, \ldots, \mathbf{h_{t-1}}]^T \mathbf{h_t}\right)$$



解决规则长度限制问题

从参数中获取规则

Algorithm 1 Recover logical rules from attention vectors

Input: attention vectors $\{\mathbf{a_t} \mid t = 1, \ldots, T\}$ and $\{\mathbf{b_t} \mid t = 1, \ldots, T+1\}$
Notation: Let $R_t = \{r_1, \ldots, r_l\}$ be the set of partial rules at step t. Each rule r_l is represented by a pair of $(\alpha, \ \beta)$ as described in Equation $\boxed{1}$, where α is the confidence and β is an ordered list of relation indexes.
Initialize: $R_0 = \{r_0\}$ where $r_0 = (1, \ (\,))$.
for $t \leftarrow 1$ to $T+1$ **do**
  Initialize: $\widehat{R_t} = \emptyset$, a placeholder for storing intermediate results.
  for $\tau \leftarrow 0$ to $t-1$ **do** //枚举之前所有长度的规则状态，更新置信度
    for rule $(\alpha, \ \beta)$ in R_τ **do**
      Update $\alpha' \leftarrow \alpha \cdot b_t^\tau$. Store the updated rule $(\alpha', \ \beta)$ in $\widehat{R_t}$.
  if $t \leq T$ **then**
    Initialize: $R_t = \emptyset$
    for rule $(\alpha, \ \beta)$ in $\widehat{R_t}$ **do**
      for $k \leftarrow 1$ to $|\mathbf{R}|$ **do** //枚举更新过置信度的规则，加入新关系k并更新信息
        Update $\alpha' \leftarrow \alpha \cdot a_t^k$, $\beta' \leftarrow \beta$ append k. Add the updated rule $(\alpha', \ \beta')$ to R_t.
  else
    $R_t = \widehat{R_t}$
return R_{T+1}

	WN18		FB15K		FB15KSelected	
	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10
Neural Tensor Network	0.53	66.1	0.25	41.4	-	-
TransE	0.38	90.9	0.32	53.9	-	-
DISTMULT [29]	0.83	94.2	0.35	57.7	**0.25**	**40.8**
Node+LinkFeat [25]	0.94	94.3	**0.82**	87.0	0.23	34.7
Implicit ReasoNets [23]	-	**95.3**	-	**92.7**	-	-
Neural LP	**0.94**	94.5	0.76	83.7	0.24	36.2

Table 8: Performance comparison. Memory Network is from [28]. QA system is from [4].

Model	Accuracy
Memory Network	78.5
QA system	93.5
Key-Value Memory Network [16]	93.9
Neural LP	**94.6**

Table 6: Inductive knowledge base completion. The metric is Hits@10.

	WN18	FB15K	FB15KSelected
TransE	0.01	0.48	0.53
Neural LP	**94.49**	**73.28**	**27.97**

训练集与测试集实体不相交

DIFFERENTIABLE LEARNING OF NUMERICAL RULES IN KNOWLEDGE GRAPHS - ICLR2020（Neural Num LP）

- 任务

  学习知识图谱上包含数值比较的规则

  “People younger than 18 typically live with their parents”

- 主要贡献

  为规则加入数值比较的形式，增强表达能力

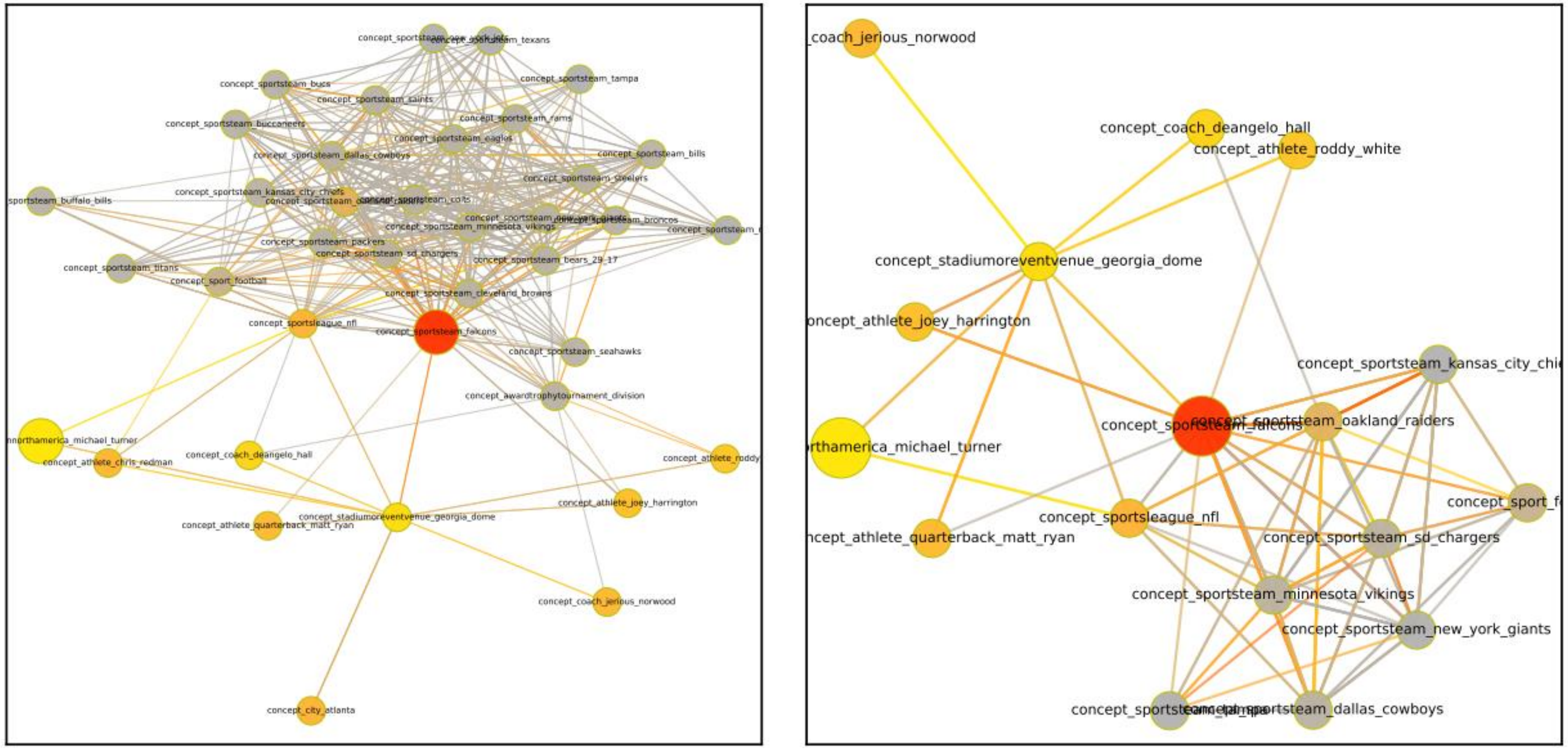
  数值规则会导致在NeuralLP框架中密集的矩阵运算，文章使用动态规划和累积和运算加速

Dataset	FB15K-237-num		DBP15K-num		Numerical1		Numerical2	
	Hit@10	MRR	Hit@10	MRR	Hit@10	MRR	Hit@10	MRR
AnyBurl	0.4262	0.2438	0.5217	0.3711	0.0306	0.0085	0.6850	0.5087
Neural-LP	0.362*	0.240*	0.4363	-	0.2398	-	0.2950	-
Neural-LP-N	0.415	**0.259**	**0.682**	**0.451**	**1.000**	**0.941**	**1.000**	**0.837**

DYNAMICALLY PRUNED MESSAGE PASSING NETWORKS FOR LARGE-SCALE KNOWLEDGE GRAPH REASONING - ICLR2020

- Dynamically Pruned Message Passing Networks (DPMPN)
- 基于输入的查询中应用图注意力机制,动态构造和扩展子图
- 构建了图结构的解释

	FB15K-237				WN18RR			
Metric (%)	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR
TransE [♠]	-	-	46.5	29.4	-	-	50.1	22.6
DistMult [♣]	15.5	26.3	41.9	24.1	39	44	49	43
DistMult [♡]	20.6 (.4)	31.8 (.2)	-	29.0 (.2)	38.4 (.4)	42.4 (.3)	-	41.3 (.3)
ComplEx [♣]	15.8	27.5	42.8	24.7	41	46	51	44
ComplEx [♡]	20.8 (.2)	32.6 (.5)	-	29.6 (.2)	38.5 (.3)	43.9 (.3)	-	42.2 (.2)
ConvE [♣]	23.7	35.6	50.1	32.5	40	44	52	43
ConvE [♡]	23.3 (.4)	33.8 (.3)	-	30.8 (.2)	39.6 (.3)	44.7 (.2)	-	43.3 (.2)
RotatE [♢]	24.1	37.5	53.3	33.8	42.8	49.2	**57.1**	47.6
ComplEx-N3[✠]	-	-	**56**	**37**	-	-	57	48
NeuralLP [♡]	18.2 (.6)	27.2 (.3)	-	24.9 (.2)	37.2 (.1)	43.4 (.1)	-	43.5 (.1)
MINERVA [♡]	14.1 (.2)	23.2 (.4)	-	20.5 (.3)	35.1 (.1)	44.5 (.4)	-	40.9 (.1)
MINERVA [△]	-	-	45.6	-	41.3	45.6	51.3	-
M-Walk [♡]	16.5 (.3)	24.3 (.2)	-	23.2 (.2)	41.4 (.1)	44.5 (.2)	-	43.7 (.1)
DPMPN	**28.6 (.1)**	**40.3 (.1)**	53.0 (.3)	36.9 (.1)	**44.4 (.4)**	**49.7 (.8)**	55.8 (.5)	**48.2 (.5)**



Query: (concept_personnorthamerica_michael_turner, concept:athleteplaysforteam, concept_sportsteam_falcons)

Selected key edges:
concept_personnorthamerica_michael_turner, concept:agentbelongstoorganization, concept_sportsleague_nfl
concept_personnorthamerica_michael_turner, concept:athletehomestadium, concept_stadiumoreventvenue_georgia_dome
concept_sportsleague_nfl, concept:agentcompeteswithagent, concept_sportsleague_nfl
concept_sportsleague_nfl, concept:agentcompeteswithagent_inv, concept_sportsleague_nfl
concept_sportsleague_nfl, concept:teamplaysinleague_inv, concept_sportsteam_sd_chargers
concept_sportsleague_nfl, concept:leaguestadiums, concept_stadiumoreventvenue_georgia_dome
concept_sportsleague_nfl, concept:teamplaysinleague_inv, concept_sportsteam_falcons
concept_sportsleague_nfl, concept:agentbelongstoorganization_inv, concept_personnorthamerica_michael_turner
concept_stadiumoreventvenue_georgia_dome, concept:leaguestadiums_inv, concept_sportsleague_nfl
concept_stadiumoreventvenue_georgia_dome, concept:teamhomestadium_inv, concept_sportsteam_falcons
concept_stadiumoreventvenue_georgia_dome, concept:athletehomestadium_inv, concept_athlete_joey_harrington
concept_stadiumoreventvenue_georgia_dome, concept:athletehomestadium_inv, concept_athlete_roddy_white
concept_stadiumoreventvenue_georgia_dome, concept:athletehomestadium_inv, concept_coach_deangelo_hall
concept_stadiumoreventvenue_georgia_dome, concept:athletehomestadium_inv, concept_personnorthamerica_michael_turner
concept_sportsleague_nfl, concept:subpartoforganization_inv, concept_sportsteam_oakland_raiders

QUERY2BOX: REASONING OVER KNOWLEDGE GRAPHS IN VECTOR SPACE USING BOX EMBEDDINGS - ICLR2020

- 任务：Existential Positive First-order (EPFO) logical queries（∧，∨，∃）

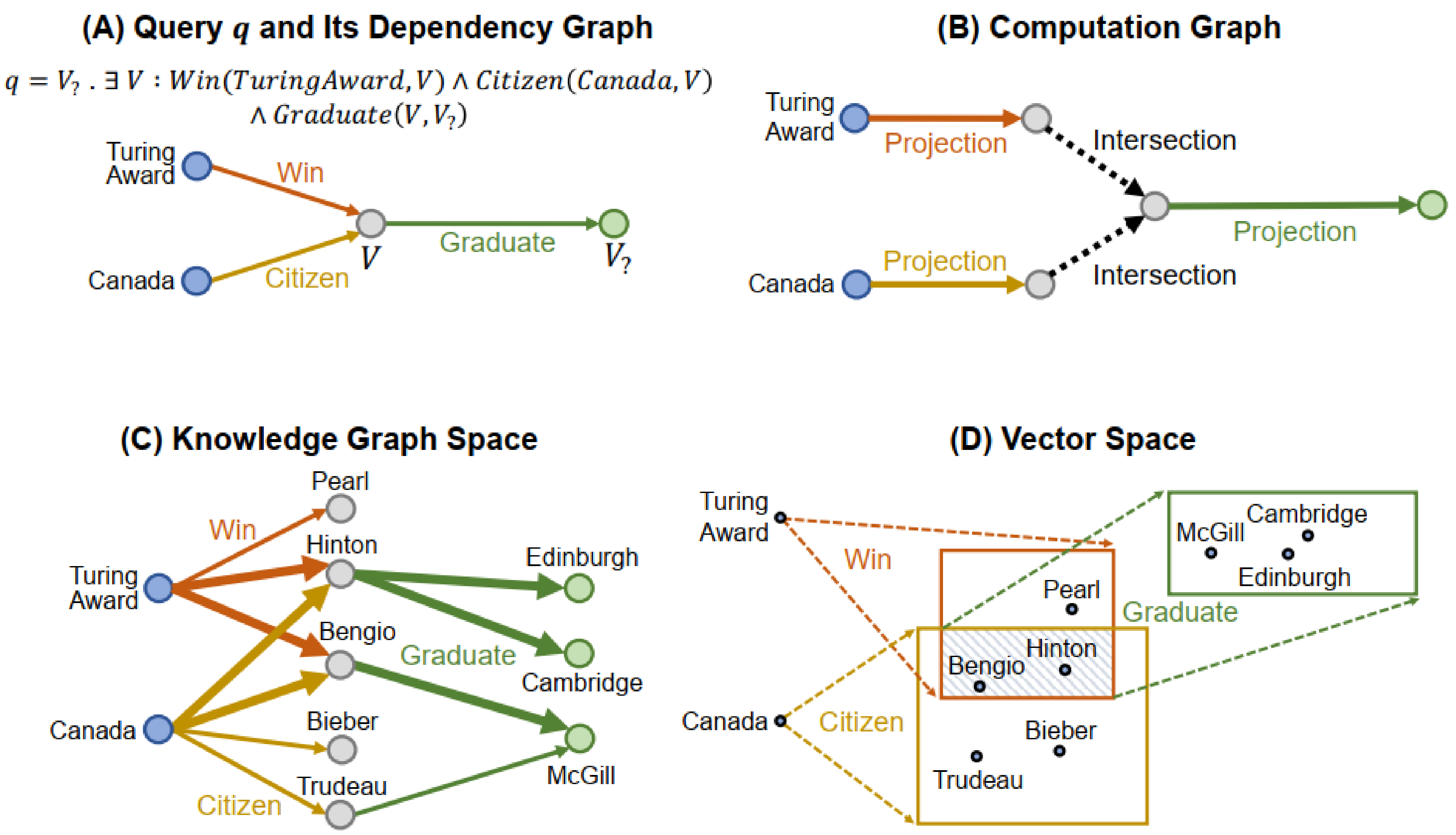
- 动机

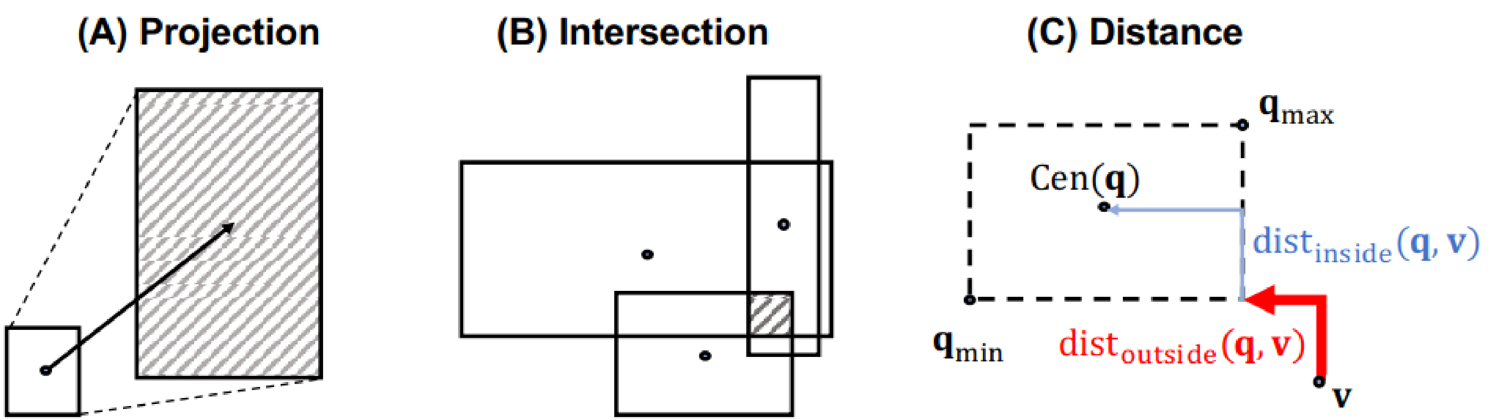
  之前的工作将查询建模为向量空间中的单个点，最近的实体作为答案，但一个查询可能有多个答案
  处理带有逻辑析取(∨)的查询仍然是一个有待解决的问题

- 主要贡献

  利用Disjunctive Normal Form (DNF)加入析取逻辑，对答案是实体集的更复杂的情况进行建模
  可以处理一些训练时没见过的结构，对缺失有一定鲁棒性

- where did Canadian citizens with *Turing Award* graduate?





define a box in $\mathbb{R}^d$ by $\mathbf{p} = (\text{Cen}(\mathbf{p}), \text{Off}(\mathbf{p})) \in \mathbb{R}^{2d}$

$$\text{Box}_{\mathbf{p}} \equiv \{\mathbf{v} \in \mathbb{R}^d : \ \text{Cen}(\mathbf{p}) - \text{Off}(\mathbf{p}) \preceq \mathbf{v} \preceq \text{Cen}(\mathbf{p}) + \text{Off}(\mathbf{p})\}$$

- **Geometric projection operator**:

  $\mathbf{p} + \mathbf{r}$ $\quad (\text{Cen}(\mathbf{r}), \text{Off}(\mathbf{r})) \in \mathbb{R}^{2d}$ with $\text{Off}(\mathbf{r}) \succeq \mathbf{0}$

- **Geometric intersection operator**:

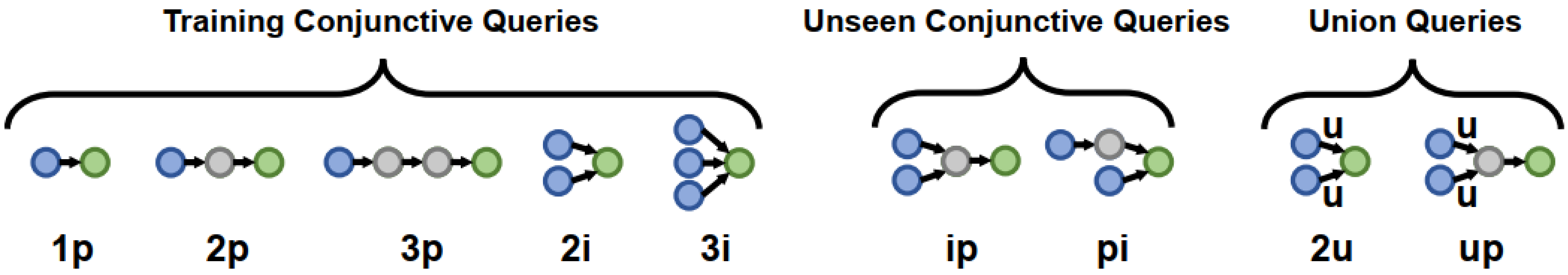
  中心上attention,sigmoid函数缩小偏移量

  $\{\mathbf{p_1}, \ldots, \mathbf{p_n}\}$ as $\mathbf{p}_{\text{inter}} = (\text{Cen}(\mathbf{p}_{\text{inter}}), \text{Off}(\mathbf{p}_{\text{inter}}))$

  $$\text{Cen}(\mathbf{p}_{\text{inter}}) = \sum_i \mathbf{a}_i \odot \text{Cen}(\mathbf{p_i}), \quad \mathbf{a}_i = \frac{\exp(\text{MLP}(\mathbf{p_i}))}{\sum_j \exp(\text{MLP}(\mathbf{p_j}))},$$

  $$\text{Off}(\mathbf{p}_{\text{inter}}) = \text{Min}(\{\text{Off}(\mathbf{p_1}), \ldots, \text{Off}(\mathbf{p_n})\}) \odot \sigma(\text{DeepSets}(\{\mathbf{p_1}, \ldots, \mathbf{p_n}\})),$$

- **Entity-to-box distance**：实体到query box中心的曼哈顿距离
- **Disjunction**（DNF下总在最后一步进行）： $\text{dist}_{\text{agg}}(\mathbf{v}; q) = \text{Min}(\{\text{dist}_{\text{box}}(\mathbf{v}; \mathbf{q}^{(1)}), \ldots, \text{dist}_{\text{box}}(\mathbf{v}; \mathbf{q}^{(N)})\})$



Method	Avg	1p	2p	3p	2i	3i	ip	pi	2u	up
FB15k										
Q2B	**0.484**	**0.786**	**0.413**	**0.303**	**0.593**	**0.712**	**0.211**	**0.397**	**0.608**	**0.33**
GQE	0.386	0.636	0.345	0.248	0.515	0.624	0.151	0.310	0.376	0.273
GQE-DOUBLE	0.384	0.630	0.346	0.250	0.515	0.611	0.153	0.320	0.362	0.271
FB15k-237										
Q2B	**0.268**	**0.467**	**0.24**	**0.186**	**0.324**	**0.453**	**0.108**	**0.205**	**0.239**	**0.193**
GQE	0.228	0.402	0.213	0.155	0.292	0.406	0.083	0.17	0.169	0.163
GQE-DOUBLE	0.23	0.405	0.213	0.153	0.298	0.411	0.085	0.182	0.167	0.16
NELL995										
Q2B	**0.306**	**0.555**	**0.266**	**0.233**	**0.343**	**0.48**	**0.132**	**0.212**	**0.369**	**0.163**
GQE	0.247	0.418	0.228	0.205	0.316	0.447	0.081	0.186	0.199	0.139
GQE-DOUBLE	0.248	0.417	0.231	0.203	0.318	0.454	0.081	0.188	0.2	0.139

两倍嵌入维度

在不能直接得到答案的复杂QA上实验，此处指标为Hits@3

SCALABLE NEURAL METHODS FOR REASONING WITH A SYMBOLIC KNOWLEDGE BASE - ICLR2020

➢ 背景/动机

- learning neural semantic parsers from denotations，使用神经方法将自然语言问题翻译成结构化查询，使用符号KB查询引擎执行。问题：难以获得数据；一个答案可能与许多可能的结构化查询相关联，从而引入噪声

- learning semantic parsers from denotations，训练数据由成对(Q,A)，其中Q是一个自然语言的问题，A是期望的答案。问题：要学习的端到端过程包括一个不可微的操作——即使用包含答案的符号知识库进行推理。一些系统也“神经化”了知识库推理，但到目前为止涉及小的知识库

➢ 主要贡献

- 可扩展性：可以分布在多个gpu上，扩展到数千万个实体和事实
- 完全端到端的神经解析器
- 新的方法架构，可以扩展到其他任务上，在两类任务上用简单的端到端架构取得了有竞争力的表现（KB completion & learning semantic parsers from denotations）

SCALABLE NEURAL METHODS FOR REASONING WITH A SYMBOLIC KNOWLEDGE BASE - ICLR2020

- weighted set as k-hot,用一个向量表示一个集合，元素=0表示不在集合里，<1表示概率，>1表示多重集
- 关系矩阵使用稀疏的矩阵数据结构来实现，COO（sparse coordinate pair）编码

$$\textit{R-neighbors}(X) \equiv \{x_j : \exists r \in R, x_i \in X \text{ so that } (x_i, x_j) \in r\}$$

$$\textit{follow}(\mathbf{x}, \mathbf{r}) \equiv \mathbf{x}\mathbf{M}_R = \mathbf{x}(\sum_{k=1}^{N_R} \mathbf{r}[k] \cdot \mathbf{M}_{r_k})$$

$$\textit{follow}(\mathbf{x}, \mathbf{r}) = \sum_{k=1}^{N_R} (\mathbf{r}[k] \cdot \mathbf{x}\mathbf{M}_{r_k})$$

$$\textit{follow}(\mathbf{x}, \mathbf{r}) = (\mathbf{x}\mathbf{M}_{subj}^T \odot \mathbf{r}\mathbf{M}_{rel}^T)\mathbf{M}_{obj}$$

subject是x的三元组　　关系是r的三元组

Strategy	Definition	Batch?	Space complexity
naive mixing	Eq 1-2	no	$O(N_T + N_E + N_R)$
late mixing	Eq 3	yes	$O(N_T + bN_E + bN_R)$
reified KB	Eq 4	yes	$O(bN_T + bN_E)$

N_T,N_E,N_R分别是三元组、实体、关系数量

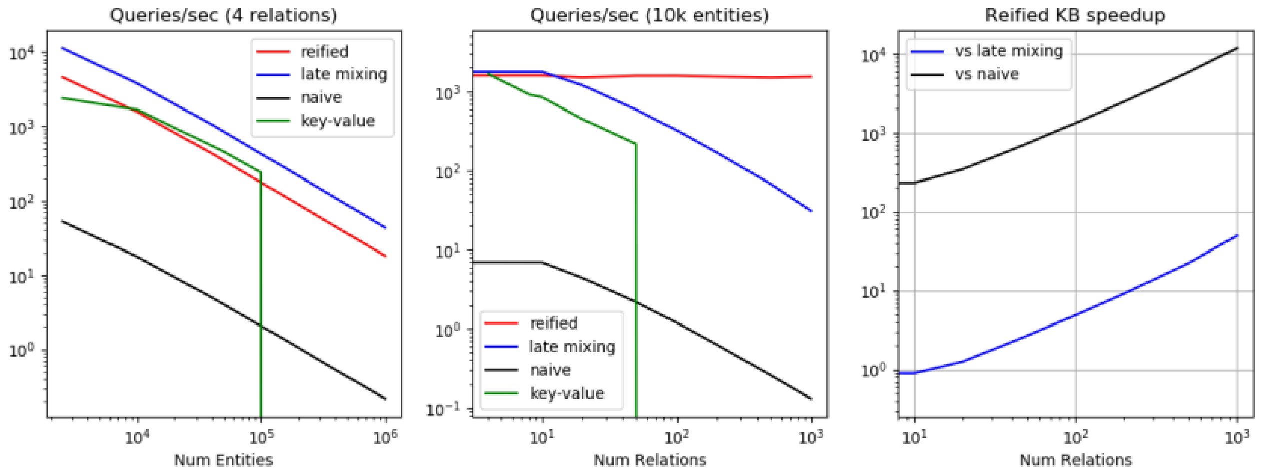


Figure 1: Left and middle: inference time in queries/sec on a synthetic KB as size and number of relations is varied. Queries/sec is given as zero when GPU memory of 12Gb is exceeded. Right: speedups of reified KBs over the baseline implementations.

	ReifKB (ours)	ReifKB + mask	KV-Mem (baseline)	VRN	GRAFT-Net	PullNet	non-differentiable components of architectures	
WebQSP	52.7	—	46.7	—	**67.8**	**68.1**	KV-Mem	initial memory
MetaQA								retrieval
1-hop	96.2	—	95.8	**97.5**	97.0	97.0		
2-hop	81.1	95.4	25.1	89.9	94.8	**99.9**	VRN	question-specific
3-hop	72.3	79.7	10.1	62.5	77.2	**91.4**	GRAFTNet	subgraph retrieval
Grid							PullNet	all iterative retrievals
5-hop	98.4	—	—	—	—	–		
10-hop	89.7	—	—	—	—	–	ReifKB(ours)	*none*

KBQA数据集上的Hits@1

The reified KB representation is quite compact, using only six integers and three floats for each KB triple.

只能产生单个实体作为回答

	NELL-995 H@1	NELL-995 H@10
ReifKB (Ours)	64.1	82.4
DistMult*	61.0	79.5
ComplEx*	61.2	82.7
ConvE*	**67.2**	**86.4**

	ReifKB (Ours)	MINERVA
NELL-995	64.1	**66.3**
Grid with seed entity		
10-hop NSEW	98.9	**99.3**
10-hop NSEW-VH	**73.6**	34.4
MetaQA 3-hop	**72.3**	41.7

Table 4: Left: Hits@1 and Hits@10 for KB completion on NELL 995. Starred KB completion methods are transductive, and do not generalize to entities not seen in training. Right: Comparison to MINERVA on several tasks for Hits@1.

	NELL-995	MetaQA-3hop	WebQuestionsSP
# Facts	154,213	196,453	43,724,175
# Entities	75,492	43,230	12,942,798
# Relations	200	9	616
Time (seconds)	44.3	72.6	1820

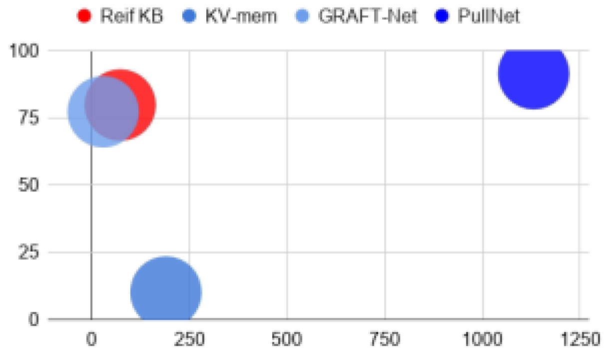


Table 5: Left, time to run 10K examples for KBs of different size. Right, time for 10k examples vs Hits@1 performance for ReifKB compared to three baselines on MetaQA-3hop questions.

Query Graph Generation for Answering Multi-hop Complex Questions from Knowledge Bases - ACL2020

- 任务

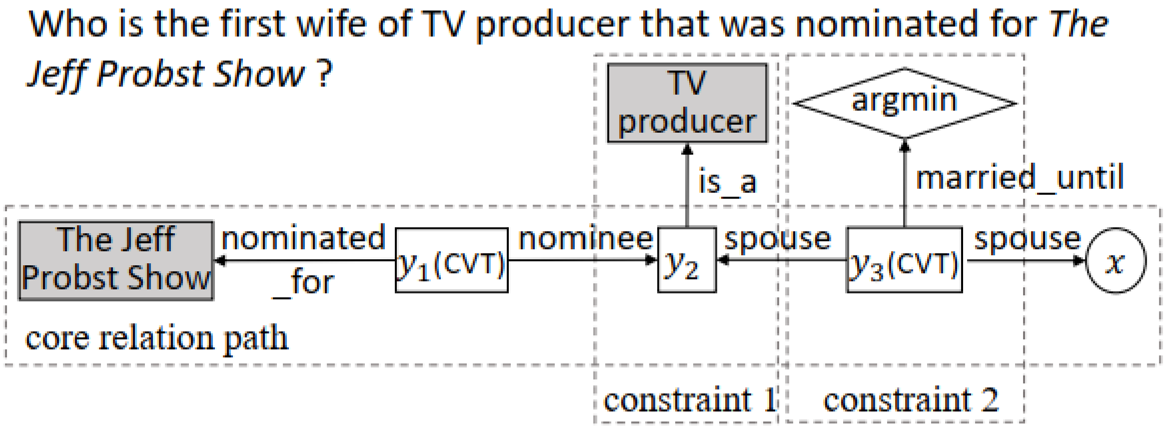
  复杂KBQA，带限制多跳推理。(自然语言查询Q,实体答案A)

- 背景/动机：

  之前复杂KBQA的工作分为带限制的单跳推理（谁是美国第一任总统？） 和多跳推理（谁是Facebook创始人的妻子?）

- 主要贡献

  处理了一种新型的更加复杂的推理任务



- 查询图生成：利用约束和beam search减少搜索空间
- 查询图排序：利用查询图涉及的实体类型数量、grounded entities的数量等特征，输入全连接层后softmax算概率

CWQ 训练、验证、测试集合问题数量：27623，3518，3531

QType	CWQ	WQSP
1-hop w/o CONS	0.1%	71.3%
1-hop w/ CONS	35.9%	28.2%
2-hop w/o CONS	33.5%	0.0%
2-hop w/ CONS	30.5%	0.5%

(a)

Method	CWQ Prec@1/F1	WQSP F1	CQ F1
Yih et al. (2015)	−/−	69.0	−
Bao et al. (2016)	−/−	−	40.9
Luo et al. (2018)	−/−	−	42.8
Lan et al. (2019a)	39.3/36.5	67.9	−
Chen et al. (2019)†	30.5/29.8	68.5	35.3
Bhutani et al. (2019)	40.8/33.9	60.3	−
Ansari et al. (2019)	−/−	72.6	−
Our Method	**44.1/40.4**	**74.0**	**43.3**

(b)

Method	CWQ Prec@1/F1
SOTA	40.8/36.5
w/ BERT	44.1/40.4
w/ LSTM	42.1/38.7
w/o extend	25.2/22.8
w/o connect	33.2/31.3
w/o aggregate	42.4/39.6

(c)

Improving Multi-hop Question Answering over Knowledge Graphs using Knowledge Base Embeddings - ACL2020

- 任务：KGQA（自然语言询问Q&关键实体e_h，答案实体A）

- 动机

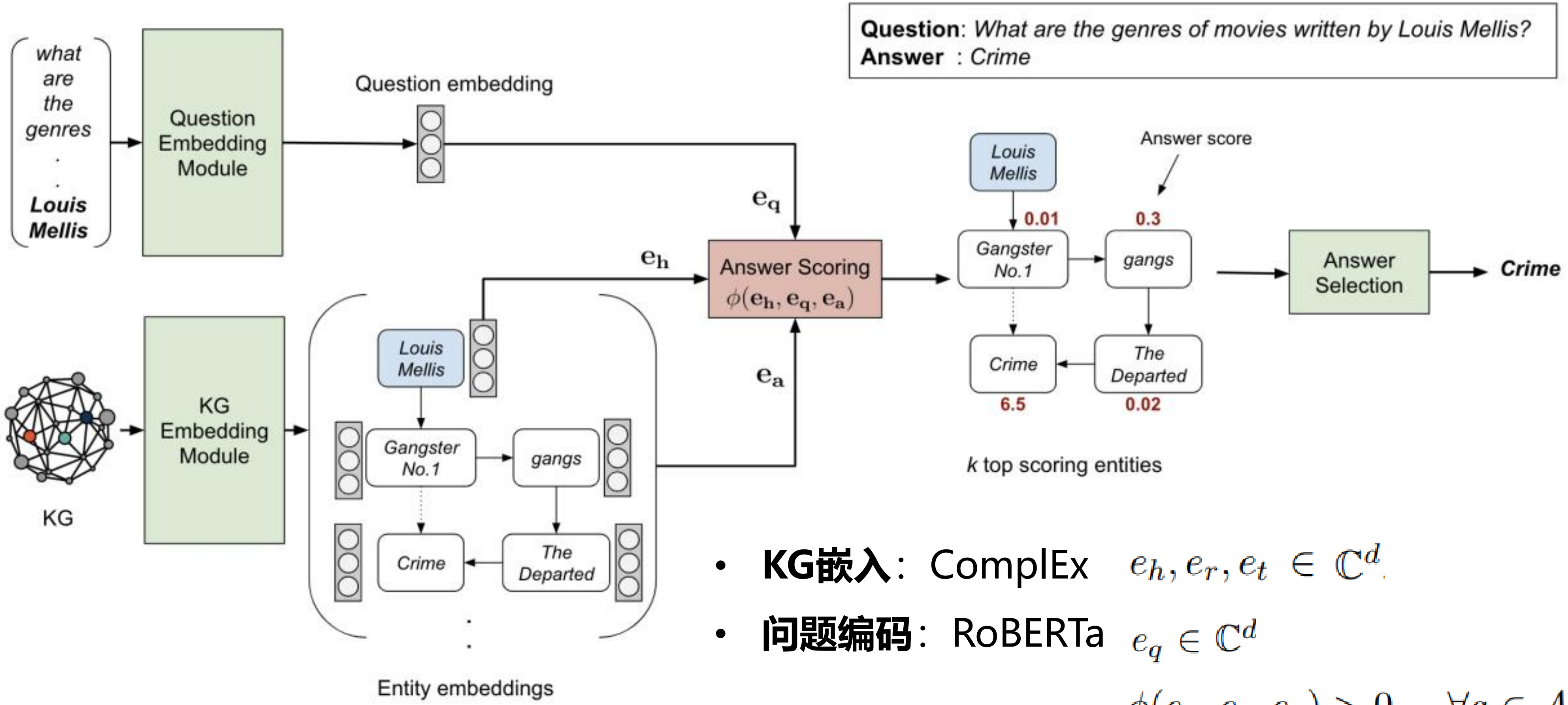
  最近关于多跳KGQA的研究试图使用相关的外部文本来处理KG稀疏性，但相关文本语料库的可用性和识别本身就是一个挑战

  KG嵌入方法虽然相关性很高，但对于多跳KGQA，目前还没有研究

- 主要贡献

  稀疏KGs上有效处理多跳KGQA

  由于**EmbedKGQA**将所有实体都视为候选答案，答案范围不像之前限定k-hop的工作一样受限



- **KG嵌入**：ComplEx $e_h, e_r, e_t \in \mathbb{C}^d$
- **问题编码**：RoBERTa $e_q \in \mathbb{C}^d$

$\phi(e_h, e_q, e_a) > 0 \quad \forall a \in \mathcal{A}$ ⟵ 如果a是答案

$\phi(e_h, e_q, e_{\bar{a}}) < 0 \quad \forall \bar{a} \notin \mathcal{A}$ ⟵ 如果a不是答案

- **答案选择**：

对于小的KGs简单地选择得分最高的实体 $e_{ans} = \underset{a' \in \mathcal{E}}{\arg\max}\, \phi(e_h, e_q, e_{a'})$

ComplEx的打分函数，a是是实体集中的实体

对于大的KG加了剪枝 $e_{ans} = \underset{a' \in \mathcal{N}_h}{\arg\max}\, \phi(e_h, e_q, e_{a'}) + \gamma * \mathrm{RelScore}_{a'}$

$$h_q = \mathrm{RoBERTa}(q')$$

$S(r, q) = \mathrm{sigmoid}(h_q^T h_r)$ $\quad \mathrm{RelScore}_{a'} = |\mathcal{R}_a \cap \mathcal{R}_{a'}|$

Ra是所有S(r,q)>0.5的关系集合，Ra'是候选答案到eh路径上的r

DIFFERENTIABLE REASONING OVER A VIRTUAL KNOWLEDGE BASE - ICLR2020

处理基于文本的QA问题中的多跳推理

谢谢！

汇报人：凌静

2021/4/2