SCALABLE NEURAL METHODS FOR REASONING WITH A SYMBOLIC KNOWLEDGE BASE

William W. Cohen & Haitian Sun & R. Alex Hofer & Matthew Siegler
Google, Inc
{wcohen,haitiansun,rofer,msiegler}@google.com

ICLR2020

凌静

2021/5/18

sparse-matrix reified KB

- 目标/任务
  - 使用较少的参数表示符号知识库，建模推理过程并整合到神经网络（可微）
  - KBQA,KBC等

- 背景
  - 使用神经方法将自然语言问题翻译成结构化查询，使用符号知识库查询引擎执行
    - 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

  - 端到端QA
    - 要学习的端到端过程包括一个不可微的操作——使用包含答案的符号知识库进行推理
    - 一些工作也“神经化”了知识库推理过程，但仅涉及小的知识库

sparse-matrix reified KB

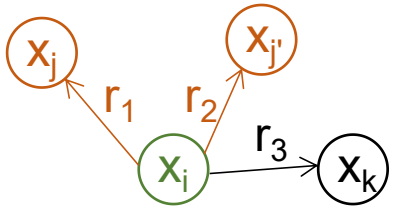
- 描述了一种表示符号知识库的新方法——sparse-matrix reified KB，完全可微，遵循原始KB语义，可建模多跳推理

- 这种有效的表示方式允许将巨大的KG直接存储在GPU内存中，例如，包含1300万实体和4300万事实的KB放到三个12G 的 GPU中。而且，在进行QA时可以对整个图谱进行推理，而不是生成候选对象（通常这是外部不可微操作）

- reified KB在KBQA和KBC任务上用简单的端到端架构与SOTA方法比较，取得了有竞争力的表现

KB上的推理

$$r\text{-}neighbors(x) \equiv \{x_j : (x_i, x_j) \in r\}$$



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

//一次推理，嵌套起来即多跳推理

$$q = \text{“}what\ movies\ were\ produced\ or\ directed\ by\ Quentin\ Tarantino\text{”}$$

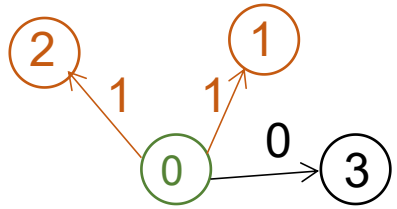
//昆汀·塔伦蒂诺制作或导演的电影？

$$X = \{Quentin\_Tarantino\} \quad R = \{producer\_of,\ writer\_of\}$$

推理起点集合表示　　涉及的关系集合表示

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

//结果是与x中实体有r中关系的实体集合



x=1,0,0,0　r=1,1,0

M_{r1}=0 0 1 0　M_{r2}=0 1 0 0　M_{r3}=0 0 0 1　M_R=0 1 1 0

....

xM_R=0 1 1 0

涉及的矩阵基本使用COO（sparse coordinate pair）编码，每一个元素用一个三元组来表示，分别是(行号，列号，数值)

实现（baseline1/baseline2/reified KB）

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

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

$$follow(\mathbf{X}, \mathbf{R}) = \sum_{k=1}^{N_R} (\mathbf{R}[:, k] \cdot \mathbf{X}\mathbf{M}_k)$$

广播

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

头实体属于x的三元组　　关系属于r的三元组

头实体属于x,关系属于r的三元组

$$follow(\mathbf{X}, \mathbf{R}) = (\mathbf{X}\mathbf{M}_{subj}^T \odot \mathbf{R}\mathbf{M}_{rel}^T)\mathbf{M}_{obj}$$

Strategy	Definition	Batch? 即x,r向量变成矩阵	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	$O(bN_T + bN_E)$

# Operations		
sp-dense matmul	dense + or $\odot$	sparse +
1	0	N_R
N_R	N_R	0
3	1	0

N_T,N_E,N_R分别是三元组、实体、关系数量，一般来说 $N_R < N_E < N_T \ll N_E^2$

$$\ell = 1, \ldots, N_T$$

represent each KB assertion $r_k(x_i, x_j)$ as a tuple (i, j, k)

$$\mathbf{M}_{subj}[\ell, m] \equiv \begin{cases} 1 & \text{if } m = i_\ell \\ 0 & \text{else} \end{cases}$$

$N_T \times N_E$

$$\mathbf{M}_{rel}[\ell, m] \equiv \begin{cases} 1 & \text{if } m = k_\ell \\ 0 & \text{else} \end{cases}$$

$N_T \times N_R$

$$\mathbf{M}_{obj}[\ell, m] \equiv \begin{cases} 1 & \text{if } m = j_\ell \\ 0 & \text{else} \end{cases}$$

$N_T \times N_E$

例子

$$\ell = 1, \ldots, N_T$$

represent each KB assertion $r_k(x_i, x_j)$ as a tuple (i, j, k)

关系集合表示：　　实体集合表示：

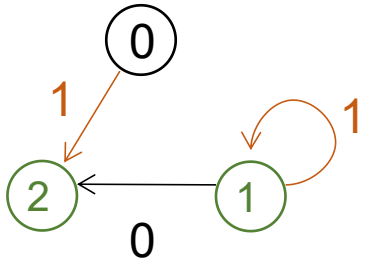
r=0,1　　x=0,1,1

id 0 1　　id 0 1 2

知识库中三元组的表示：

t=[(1,2,0),(0,2,1),(1,1,1)]

id　0　　1　　2



Msubj=0 1 0　Mobj=0 0 1　Mrel= 1 0
1 0 0　　0 0 1　　0 1
0 1 0　　0 1 0　　0 1

$$\mathbf{M}_{subj}[\ell, m] \equiv \begin{cases} 1 & \text{if } m = i_\ell \\ 0 & \text{else} \end{cases}$$

Msubjt=0 1 0　　Mrelt=1 0 0
1 0 1　　0 1 1
0 0 0

$$\mathbf{M}_{obj}[\ell, m] \equiv \begin{cases} 1 & \text{if } m = j_\ell \\ 0 & \text{else} \end{cases}$$

xMsubjt=1 0 1　rMrelt=0 1 1

$$\mathbf{M}_{rel}[\ell, m] \equiv \begin{cases} 1 & \text{if } m = k_\ell \\ 0 & \text{else} \end{cases}$$

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

subject是x的三元组　　关系是r的三元组

xMsubjt ⊙ xMsubjt=0 0 1　//说明id=2的三元组头实体的id是1or2，关系的id是1

（...）Mobj = 0 1 0　//说明id为{1,2}的实体的{1}-neighbor为1（实体id）

实验

$$follow(follow(\mathbf{x}, \mathbf{r}), \mathbf{r})$$

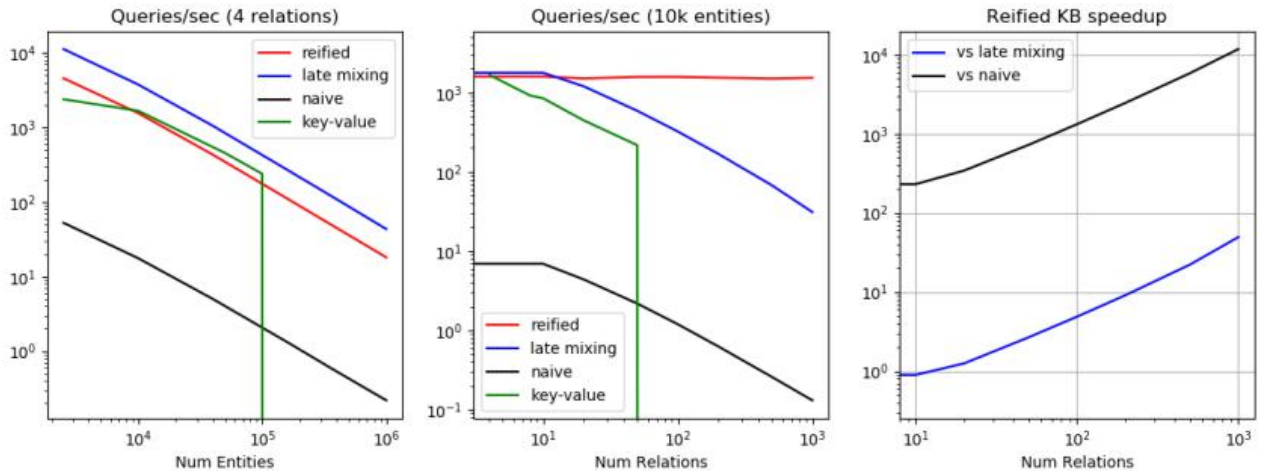


Figure 1: Left and middle: inference time in queries/sec on a synthetic KB as size and number of relations is varied. Queries/sec is given as zero when GPU memory of 12Gb is exceeded. Right: speedups of reified KBs over the baseline implementations.

任务：网格图中方位相关的两跳推理

实验

	ReifKB (ours)	ReifKB + mask	KV-Mem (baseline)	VRN	GRAFT-Net	PullNet
WebQSP	52.7	—	46.7	—	**67.8**	**68.1**
MetaQA						
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**
3-hop	72.3	79.7	10.1	62.5	77.2	**91.4**
Grid						
5-hop	98.4	—	—	—	—	–
10-hop	89.7	—	—	—	—	–

non-differentiable components of architectures	
KV-Mem	initial memory retrieval
VRN	question-specific subgraph retrieval
GRAFTNet	
PullNet	all iterative retrievals
ReifKB(ours)	*none*

KBQA数据集上的Hits@1

使用具体化知识库进行推理意味着神经模型只需要预测推理过程中各个阶段所使用的关系

$$\mathbf{r}^t = f^t(q); \quad \mathbf{x}^t = follow(\mathbf{x}^{t-1}, \mathbf{r}^t)$$

x^0即问题中涉及到的实体

实验

只能产生单个实体作为回答

reifKB参数少很多
泛化性更好

	NELL-995	
	H@1	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

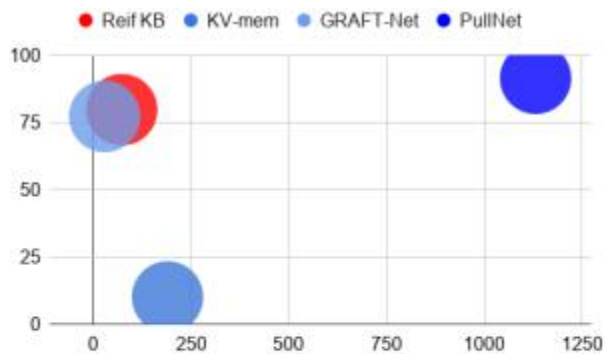


Table 5: Left, time to run 10K examples for KBs of different size. Right, time for 10k examples vs Hits@1 performance for ReifKB compared to three baselines on MetaQA-3hop questions.

感 谢!

凌静

2021/5/18