

基于推理的知识图谱补全框架：EM-RBR

马 坤

(大连海事大学 信息科学技术学院, 辽宁省大连市 中国 116026)

摘 要 知识图完成旨在预测知识图谱中给定实体中的新链接。大多数主流的嵌入方法都集中在给定的知识图谱中包含的事实三元组, 然而忽略了由知识库隐式驱动的逻辑规则提供的丰富的背景信息。本文所介绍的是一个结合嵌入方法和基于规则推理的组合框架: EM-RBR, 以实现更好的链路预测。试图通过利用知识图谱现有三元组来预测一个新的三元组是否可能属于知识图谱。EM-RBR 的目的是利用规则中包含的关系背景知识, 在嵌入模型中进行多关系推理链路预测, 而不是表面向量三角链。通过这种方式, 我们可以在更深的上下文中探索两个实体之间的关系, 以获得更高的精度。

关键词 规则推理 知识图谱 知识图谱补全

中图法分类号 TP311

文献标识码 A

A Framework of Knowledge Graph Completion Based on Reasoning: EM-RBR

Ma kun

(School of Information Science and Technology, Dalian maritime university, Liaoning Dalian 116026 China)

Abstract Knowledge graph completion aims to predict new links in a given entity in the knowledge map. Most of the mainstream embedding methods focus on the fact triples contained in a given knowledge map, but ignore the rich background information provided by the logic rules implicitly driven by the knowledge base. This paper introduces a combination framework of embedded method and rule-based reasoning: em-rbr to achieve better link prediction. This paper attempts to predict whether a new triplet may belong to the knowledge tuple by using the existing triples of the knowledge map. The purpose of em-rbr is to use the relational background knowledge contained in the rules to predict the multi relationship reasoning links in the embedded model, rather than the surface vector triangular chain. In this way, we can explore the relationship between two entities in a deeper context for higher precision.

Key words Rule-based reasoning; Knowledge graph; Knowledge graph C

1 引言

知识图谱^[1], 具有传递世界知识的能力, 并以结构化的表示方式表达知识, 然而, 由于逻辑事实的离散性质, 知识图谱中包含的知识是不完整的。知识图谱补全目的是预测出三元组中缺失的部分, 从而使知识图谱变得更加完整。一些流行的嵌入方法旨在把实体和关系嵌入到低维潜在特征空间中。通过比较向量空间中的表示来推断实体

之间的隐式关系。但总体效果与知识图谱的密度高度相关。因为嵌入方法总是无法预测低频的弱关系和隐关系。例如, 现有的三元组 (Paul, Leader-Of, Soccer Team) 和规则 $\text{Leader-Of}(x, y) \Rightarrow \text{Member-Of}(x, y)$, 表示足球队的领导者也是运动队的成员。然后, 我们可以应用三元组上的规则来获得一个新的三元组 (Paul, Member-of, SportTeam), 即使 Member-Of 关系在知识图谱中是弱的。也有利用规则进行预测的方法: 联合模

型利用损失函数中的规则，来获得更好的实体和关系的嵌入表示。基于 ProPPR^[2]的优化模型通过嵌入规则，计算超参数。这些努力最终都是为了更好地从规则和三元组中实现嵌入，而不是通过真正的基于规则的推理来解决完成问题，这是解决前面提到的弱关系预测所必需的。与之相比，EM-RBR 可以从推理的角度进行完成，此框架结合基于规则推理的优点和先进的嵌入模型^[3]。EM-RBR 的目的是利用规则中包含的关系背景知识进行多关系推理链接预测^[4]，而不是简单的向量三角链接。通过这种方式，我们可以在更深层的语境中探索两个实体之间的关系，以获得更高的精度。

2 相关工作及贡献点

2.1 相关工作

框架 EM-RBR 结合嵌入和基于规则的推理，它本质上是一个广度优先遍历^[5]。在联合框架 EM-RBR 的开发中，我们面临两个挑战。

(1) 规则可信度

使用 AMIE^[6]自动挖掘到的大量规则不一定都完全可信。因此，有必要提出一种合理的方法来衡量规则，以便在推理时选择适当的规则。

(2) 概率表示

传统的基于推理的方法用 0 或 1 表示知识图谱拒绝或接受某三元组。这种传统的定性分析缺乏嵌入模型那样的定量信息。所以 EM-RBR 的结果需要给出一个表示三元组属于知识图谱的概率。

2.2 贡献点

(1) 有足够的灵活性和通用性，可以与许多嵌入模型相结合。

(2) 提出了一种新的三元组评级机制，结合推理过程，可以更好地区分给定的三元组和其他错误的三元组。

(3) 提出了一种新的自动判断推理规则的评级机制，每个规则都将在我们的框架中得到适当的度量。

3 框架概述

框架的核心思想是从推理的角度在更深的上下文中进行多相关路径预测，即以广度优先搜索

的形式进行。我们需要先解释相关定义。

3.1 规则定义

我们的框架中，一个规则是以 $B_1(x, z) \wedge B_2(z, y) \Rightarrow H(x, y)$ 或 $B(x, y) \Rightarrow H(x, y)$ 的形式存在的，其中一个三元组中的实体顺序是随机的，比如 $B_3(z, x) \wedge B_4(z, y) \Rightarrow R(x, y)$ 也是一条有效的规则。

3.2 相关数值定义

在搜索过程中，我们为每个状态定义了两个分数。H 是一个启发式分数，L 是用来计算 Φ 的状态分数，还需要对三元组得分进行说明。

3.2.1 三元组得分

三元组的分数 Φ 满足 $\Phi \geq 1$ 。 Φ 值较小代表三元组属于知识图谱的概率较大。三元组 (h, r, t) 的分数表示为 $\Phi_{\sim(h, r, t)}$ 。

3.2.2 启发式得分 H

$$H_{\sim O(h, r, t)} = \prod_{(B_1 \wedge B_2 \Rightarrow H) \in \Delta \text{Path}} \omega(B_1, B_2, H) \quad (1)$$

$$\omega(B_1, B_2, H) \leftarrow e^{\frac{\|B_1 + B_2 - H\|}{k}}$$

其中， $\omega(B_1, B_2, H) \leftarrow e^{\frac{\|B_1 + B_2 - H\|}{k}}$ 。将 $H \sim O(h, r, t)$ 定义为公式 (1)，表示所有规则分数的乘积。 ΔPath 表示从初始状态到当前状态的扩展中使用的规则集。 $\omega(B_1, B_2, H)$ 是

$(B_1 \wedge B_2 \Rightarrow H)$ 形状的规则得分。 k 是嵌入的维数。分数越小，可信度越高。为了使计算中的维数均匀，我们将规则的分数除以 k ，然后进行 e 指数变换，得到 H 的计算方程。规则的可视化如图 1 所示：

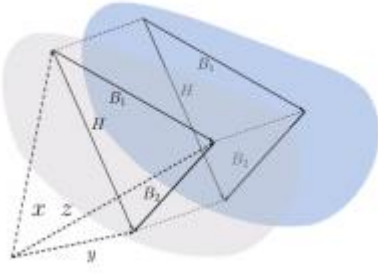


图 1: 规则的可视化

3.2.3 状态分数 O

$$L_{O \sim (h,r,t)} = H_{O \sim (h,r,t)} * \prod_{(O_h, O_r, O_t) \in O} S \sim transX(O_h, O_r, O_t) \quad (2)$$

$L_{O \sim (h,r,t)}$ 定义为公式 2, 表示 $H_{O \sim (h,r,t)}$ 的乘积和状态下所有三元组的分数。 O 表示状态, (O_h, O_r, O_t) 是属于 O 的三元组。 $S \sim transX(O_h, O_r, O_t)$ 是公式 3 中定义的这个三元组的嵌入分数^[7]。

$$S \sim transX(O_h, O_r, O_t) = \begin{cases} 1, (O_h, O_r, O_t) \in G \\ \frac{\|O_h + O_r - O_t\|}{k+1}, (O_h, O_r, O_t) \notin G \end{cases} \quad (3)$$

$\|O_h + O_r - O_t\|$ 是 transE 模型中三元组 (O_h, O_r, O_t) 的分数。其值越小, 三元组在 G 中的可能性越大。当 $(O_h, O_r, O_t) \in G$ 时, 假设分数为 0。同样, 对于规则的分数, 我们也对三元组的分数进行了一定的变换, 即除以 k 并加 1。

4 算法描述

4.1 算法概述

我们将知识图谱建模为事实 $G = \{(h, r, t)\}$ 三元组集合, 其中 E 和 R 分别表示知识图谱中的实体集和关系集。我们框架的步骤如下:

① 矢量化: 调用嵌入模型^[8] (本文采用 transE^[9] 嵌

入模型) 得到一个包含 G 中实体和关系的 k 维嵌入集合。

② 在 G 上应用 AMIE 得到推理规则集 Ω , 其中每个规则都满足规则定义。

③ 推理规则的嵌入: 是通过规则中包含的关系的嵌入来衡量的。

④ 对给定的三元组 (h, r, t) 进行推理。

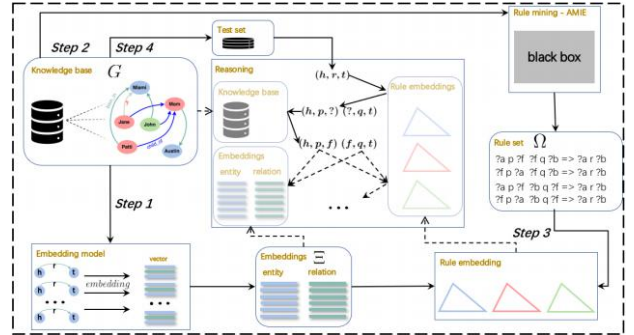


图 2: 框架概述图

4.2 推理思想

我们的方法是基于 BFS 的思想。我们使用优先级队列 Q 以 H 的升序存储状态。 H 是一个启发式分数, L 是用来计算 Φ 的状态分数。初始状态是目标三元组本身, 其 H 为 1, L 为嵌入模型 (transE) 下的分数。将初始状态推入 Q 。在搜索过程中, 弹出 Q 的顶部作为当前状态。如果 $H \sim scur \geq \Phi$, 它将被扩展, 否则我们将通过匹配规则来扩展它以获得新的状态。对于每个新状态, 计算其分数 $H \sim snew$ 和 $L \sim snew$ 。如果 $H \sim snew < L \sim scur$, 则状态将被推入 Q 。重复上述过程, 直到 Q 为空。最后, 我们从 Q 中选择所有状态的最小 L 作为 Φ 。

4.3 匹配与扩展过程

状态是一组三元组的集合, 初始状态是目标三元组本身。本质上, 状态的扩展是状态中三元组的扩展。一组三元组对应一个搜索节点, 如果所有的三元组都能满足条件, 那么该节点就可以返回。对于三元组 (h, r, t) 匹配和扩展的过程大致如下:

① 以 $B_1(x, z) \wedge B_2(z, y) \Rightarrow H(x, y)$ 的形式找到 $\omega \in \Omega$ 的规则, 其中 $H=r$ 。

② 将实体分配给规则中的变量, 如 $x=h, y=t$ 。

③ 找到所有满足 $(x, B_1, z_0) \in G$ 或 $(z_0, B_2, y) \in G$ 的 z_0 , 其中 $x=h, y=t$ 。

④ (h, r, t) 扩展到 $\{(h, B_1, z_0), (z_0, B_2, t)\}$ 。

state	triplets				
s_1	$(h, B_1, m_1)^*$	(m_1, B_2, t)			
s_2	(h, B_3, m_2)	(m_2, B_4, t)			
s_3	(h, B_5, m_3)	$(m_3, B_6, m_1)^*$	(m_1, B_2, t)		
s_4	$(h, B_5, m_4)^*$	(m_4, B_6, m_1)	(m_1, B_2, t)		
s_5	$(h, B_7, m_5)^*$	(m_5, B_8, m_1)	(m_1, B_2, t)		
s_6	(h, B_9, m_6)	$(m_6, B_{10}, m_4)^*$	(m_4, B_6, m_1)	(m_1, B_2, t)	
s_7	(h, B_{11}, m_7)	$(m_7, B_{12}, m_4)^*$	(m_4, B_6, m_1)	(m_1, B_2, t)	
s_8	(h, B_9, m_6)	$(m_6, B_{13}, m_8)^*$	(m_8, B_{14}, m_4)	(m_4, B_6, m_1)	(m_1, B_2, t)
s_9	(h, B_9, m_6)	(m_6, B_{13}, m_9)	$(m_9, B_{14}, m_4)^*$	(m_4, B_6, m_1)	(m_1, B_2, t)

图 3: 各个状态所包含三元组图

注意：一个三元组总是有多个扩展三元组。

例如，我们在初始状态下展开目标三元组。子态中有两个三元组，其中任何一个都必须在知识图谱中。当子状态进一步展开时，知识图谱中的三元组无需展开。因此，在扩展 m 次后，每个子态中应该有 $m+1$ 个三元组。其中至少有 m 属于知识图谱。

4.4 算法

首先进行数据的初始化，框架伪码代码如下：

Initialize Q as an empty priority queue

Initialize the first state:

$s_0 \leftarrow \{(h, r, t)\}$,

$H \sim s_0(h, r, t) \leftarrow 1$,

$L \sim s_0(h, r, t) \leftarrow 1$,

Initialize the score:

$\Phi \sim (h, r, t) \leftarrow (s \sim \text{transX}(h, r, t)) / k + 1$

$Q.\text{push}(s_0)$

while !Q.empty() do

$s_{\text{cur}} \leftarrow Q.\text{pop}()$

$\Phi \sim (h, r, t) \leftarrow \min\{\Phi \sim (h, r, t), L \sim s_{\text{cur}}(h, r, t)\}$

if $H \sim s_{\text{cur}}(h, r, t) < \Phi \sim (h, r, t)$ *then*

$S_{\text{ne}} \leftarrow \text{extend}(s_{\text{cur}})$

calculate $(H \sim s_{\text{ne}}(h, r, t))$, $s_{\text{ne}} \in S_{\text{ne}}$

calculate $(L \sim s_{\text{ne}}(h, r, t))$, $s_{\text{ne}} \in S_{\text{ne}}$

for $s_{\text{ne}} \in S_{\text{ne}}$ *do*

if $H \sim s_{\text{ne}}(h, r, t) < L \sim s_{\text{cur}}(h, r, t)$ *then*

$Q.\text{push}(s_{\text{ne}})$

end if

end for

end if

end while

5 实验示例

在本节中，我们使用一个示例来说明我们的算法。初始状态 S_0 只包含一个三元组 (h, r, t) ，其状态得分和启发式得分均为 1。在开始时，优先级队列 Q 只有一个元素，即初始状态及其分数。搜索过程我们在后边展开详细说明，必要的消息在假设 5.1 中定义

5.1 示例数据

首先我们假设，使用相同的符号 r_i 来表示规则的符号和规则的分数。并将一些数据关系定义为公式 4 和 5。其他的相关信息如图。

$$r_1 r_3 r_5 > L \sim s_3(h, r, t) \quad (4)$$

$$H \sim s_7(h, r, t) > \Phi \sim (h, r, t) \quad (5)$$

图 3 表示每个状态中的三元组，带星号*的三元组是不在知识图谱中的三元组，其余三元组均在知识图谱中。

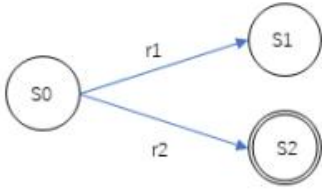
rule	score
$B_1(x, z) \wedge B_2(z, y) \Rightarrow r(x, y)$	r1
$B_3(x, z) \wedge B_4(z, y) \Rightarrow r(x, y)$	r2
$B_5(x, z) \wedge B_6(z, y) \Rightarrow B_1(x, y)$	r3
$B_7(x, z) \wedge B_8(z, y) \Rightarrow B_1(x, y)$	r4
$B_9(x, z) \wedge B_{10}(z, y) \Rightarrow B_5(x, y)$	r5
$B_{11}(x, z) \wedge B_{12}(z, y) \Rightarrow B_5(x, y)$	r6
$B_{13}(x, z) \wedge B_{14}(z, y) \Rightarrow B_{10}(x, y)$	r7

图 4: 规则得分图

图 4 表示每个规则的扩展和规则得分 r_1, r_2 表示规则得分。

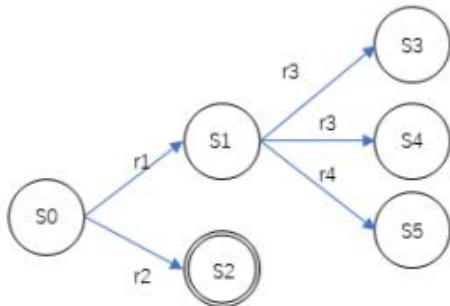
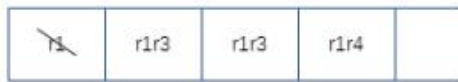
5.2 搜索过程

搜索过程分为六个阶段，每个阶段包含在一个子图中，每个子图包含三个部分。子图的顶部显示优先级队列的当前状态，中间部分是搜索的可视化，底部给出了每个阶段更新 $\Phi \sim (h, r, t)$ 的公式。



$$\begin{aligned} & \Phi_{\sim}(h, r, t) \\ &= \min \{ \Phi_{\sim}(h, r, t), \mathcal{L}_{\sim s_2}(h, r, t) \} \\ &= \mathcal{L}_{\sim s_2}(h, r, t) \end{aligned} \quad \text{I}$$

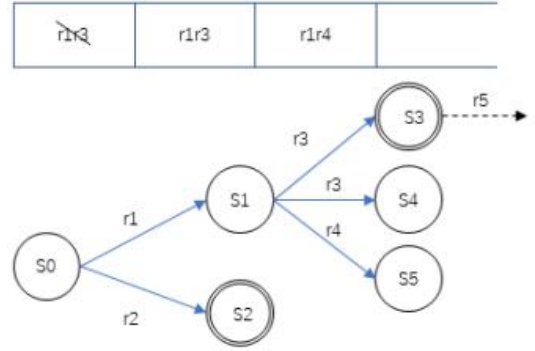
① S_0 分别匹配 R_1 和 R_2 ，并扩展到 S_1 和 S_2 。在 S_2 中， S_2 是三元组的终止状态，都在 G 中。我们使用 $\mathcal{L}_{\sim s_2}(h, r, t)$ 更新 $\Phi \sim (h, r, t)$ ，并将 S_1 推入 Q 。



$$\begin{aligned} & \Phi_{\sim}(h, r, t) \\ &= \min \{ \Phi_{\sim}(h, r, t), \mathcal{L}_{\sim s_1}(h, r, t) \} \\ &= \Phi_{\sim}(h, r, t) \end{aligned} \quad \text{II}$$

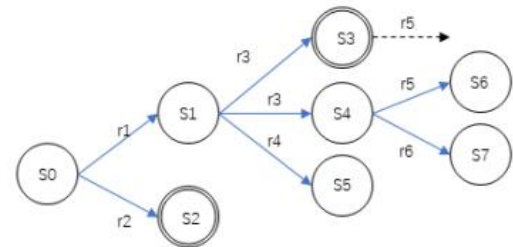
②弹出队列 s_1 的顶部数据。使用它来更新

$\Phi \sim (h, r, t)$ ，然后将其扩展到三个新的状态，这些状态将入队。



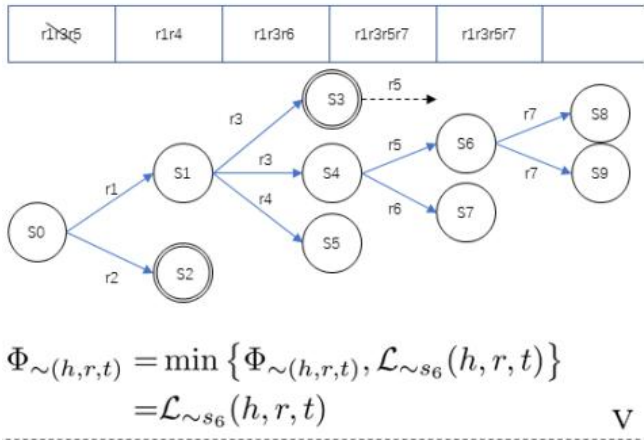
$$\begin{aligned} & \Phi_{\sim}(h, r, t) \\ &= \min \{ \Phi_{\sim}(h, r, t), \mathcal{L}_{\sim s_3}(h, r, t) \} \\ &= \mathcal{L}_{\sim s_3}(h, r, t) \end{aligned} \quad \text{III}$$

③弹出队列 s_3 的顶部数据以更新 $\Phi \sim (h, r, t)$ ，并通过匹配规则 R_5 来扩展它。由于 $r_1 r_3 r_5 > \mathcal{L}_{\sim s_3}(h, r, t)$ ，也就是说这条路径产生的解决方案将不是全局最小值。因此，这种状态不再扩展。

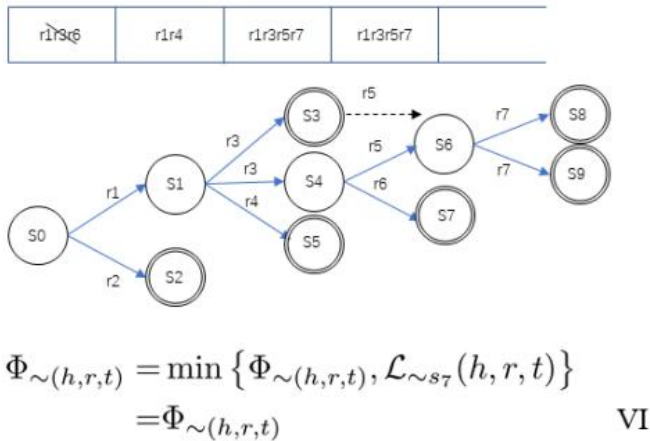


$$\begin{aligned} & \Phi_{\sim}(h, r, t) \\ &= \min \{ \Phi_{\sim}(h, r, t), \mathcal{L}_{\sim s_4}(h, r, t) \} \\ &= \mathcal{L}_{\sim s_4}(h, r, t) \end{aligned} \quad \text{IV}$$

④弹出队列 s_4 的顶部以更新 $\Phi \sim (h, r, t)$ 并扩展以获得两个新状态 S_6, S_7 。



⑤弹出队列 S6 的顶部以更新 $\Phi \sim (h, r, t)$ ，并在规则 R7 之后扩展到 S8, S9。



⑥弹出队列 S7 的顶部，现在 $H \sim S7(h, r, t) > \Phi \sim (h, r, t)$ 。所以 S7 和 Q 中的其余状态不需要扩展。因此，Q 中的所有剩余状态都成为终止状态。搜索停止了。

6 数据应用分析

6.1 分析

对于三个实数 $a, b, c(a, b, c > 1)$ ，通常存在 $c > a * b * 1$ 的可能性。考虑三元组 (h, r, t) 和规则 $B_1(x, z) \wedge B_2(z, y) \Rightarrow H(x, y)$ 。如果 c 表示三元组 (h, r, t) 的分数， a 表示规则的分数， b 表示不在知识图中的扩展新三元组的分数， 1 表示知识图中的扩展新三元组的分数。然后 (h, r, t) 的分数将降为 $a*b$ ，即。我们使用新的扩展三元组和规则来评估 (h, r, t) 。当然，这种优化只会对正确的三元组有效。对于错误的三元组，规则匹配后将得到另一个得分较

大的错误三元组。所以 $c > a * b$ 在这个结果是一个非常不可能的事件。因此，正确的三元组被优化，错误的三元组一般不会被优化。因此，从宏观的角度来看，正确的三元组的排名会增加。

状态的启发式分数 H 是沿着推理路径的所有规则的分数的乘积。规则的分数都大于 1，所以当一条搜索路径足够长时， H 必须大于 Φ 。所以搜索必须能够停止。

由上可知，我们要求规则和三元组的分数大于 1，并接近 1。给定翻译模型得到的嵌入的每个维数都是小于 1 且接近 0 的数，我们将分数除以相应的维数 k ，然后加上 1，以满足我们的要求。此外，为了突出规则在计算中的重要性，我们对规则使用 e 指数变化而不是加 1。

6.2 结果

Model	FB15k			
	MR	MRR	H@1	H@10
TransE	70.3	45.77	29.98	74.27
TransH	72.56	45.81	30.37	74.01
TransR	55.98	47.88	31.1	77.04
TransD	56.41	47.88	32.48	75.99
†E-R(E)	68.36	50.01	34.44	76.23
E-R(H)	70.72	52.39	38.82	76.52
E-R(R)	55.47	51.93	35.86	78.35
E-R(D)	55.21	53.02	38.25	78.33
†RotateE	--	--	--	--
RUGE	--	--	--	--
ComplEx	--	--	--	--
DistMult	--	--	--	--

图 5 FB15k 上的实验结果

数据集的测试集中的三元组大致可以分为两部分：一是规则可以匹配，由我们的模型进行优化；二是没有规则可以匹配。在 FB15k^[10] 中，这两部分的比例约为 1: 5。从最终的结果来看，虽然与组合前的模型相比，各种度量已经得到了改进。

事实上，只有一小部分三元组得到了优化。也正是由于这个原因，我们的模型的能力可以 FB15k-R 上充分展示，因为这个集合中的每个三元组都有许多规则可以匹配，以获得良好的优化效果。

FB15k-R			
MR	MRR	H@1	H@10
71.33	26.11	14.9	48.1
50.65	30.43	18.25	54.95
29.64	18.51	7.65	37.6
28.24	26.16	14	50.45
3.12	79.88	65.1	96.4
3.52	85.61	75.45	97.8
1.73	86.01	74.3	99.2
3.245	82.04	70.45	96.15
26.66	33.29	20.05	59.75
53.63	49.14	33.05	78.2
51.43	51.1	35.8	79.0
62.02	46.2	30.2	77.1

图 6 FB15k-R 上的实验结果

为了更好地理解每个三元组上正在优化的具体情况。我们分别分析了在平移模型和 EM-RBR 模型下，当进行头部实体替换和尾部实体替换时，每个三元组的相应排序。结果如图 6 所示。表中的数据项是 $s \sim trans$ 和 $s \sim E-R$ 从最大值排序到最小值的结果，其中 $s \sim trans$ 是对应平移模型下的排序， $s \sim E-R$ 是对应 EM-RBR 模型下的排序。

Rank	EM-RBR(E)			
	+id	*E-R	‡trans	◊L/R
1	47722	2	14141	R
2	47722	2	13900	L
3	18355	2	7525	L
4	36133	2	6884	L
5	33004	1	6253	L
6	33243	2	5883	R
7	30883	2	5674	R
8	14035	2	4862	L
9	18355	2	4525	R
10	24243	1	3655	L
...				
19372	52886	4	2	R
19373	52707	13	11	L
19374	52529	9	7	R
19375	51447	3	1	R
19376	50932	4	2	R
...				

图 7 实验优化结果

7 总结

本文介绍了一种将嵌入和基于规则的推理相结合的创新框架 EM-RBR，它可以很容易地与任何基于平移的嵌入模型集成在一起。与以往试图从规则和三重奏中获得更好的嵌入结果的联合模型不同，我们的模型允许通过进行多相关路径预测，即从推理的角度求解完成。即：广度优先遍历。我们还证明了 EM-RBR 可以有效地提高 KGC 嵌入方法的性能。这使得现有的基于平移的嵌入方法更适合和可靠地用于真实和大规模的知识推理任务。

未来有两个可研究的方向。一方面，我们将把我们的模型与更多的嵌入模型结合起来，而不仅仅是基于平移的嵌入模型。另一方面，我们将提取更多更可靠的关联规则来优化我们的工作。如上所述，FB15k 上评估时，只有部分三元组被优化。其余的根本原因是没有相应的匹配规则。如果这两个问题得到解决，EM-RBR 可以得到更好的改进。

参考文献

[1]Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. Learning structured embeddings of knowledge bases. In Twenty-Fifth AAAI Conference on Artificial Intelligence, 2011.

[2]Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In Advances in neural information processing systems, pp. 2787–2795, 2013.

[3]Antoine Bordes, Xavier Glorot, Jason Weston, and Yoshua Bengio. A semantic matching energy function for learning with multi-relational data. Machine Learning, 94(2):233–259, 2014.

[4]Rajarshi Das, Shehzaad Dhuliawala, Manzil Zaheer, Luke Vilnis, Ishan Durugkar, Akshay Krishnamurthy, Alexander J. Smola, and Andrew McCallum. Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. CoRR, abs/1711.05851, 2017.

[5]Xin Dong, Evgeniy Gabrilovich, Jeremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. Knowledge vault: A

web-scale approach to probabilistic knowledge fusion. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 601–610, 2014.

[6]Luis Antonio Galarraga, Christina Teflioudi, Katja Hose, and Fabian Suchanek. Amie: association rule mining under incomplete evidence in ontological knowledge bases. In Proceedings of the 22nd international conference on World Wide Web, pp. 413–422, 2013.

[7]Matt Gardner, Partha Talukdar, Jayant Krishnamurthy, and Tom Mitchell. Incorporating vector space similarity in random walk inference over knowledge bases. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp. 397–406, 2014.

[8]Jens Graupmann, Ralf Schenkel, and Gerhard Weikum. The spheresearch engine for unified ranked retrieval of heterogeneous xml and web documents. volume 2, pp. 529–540, 01 2005.

[9]Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Jointly embedding knowledge graphs and logical rules. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 192–202, 2016.

[10]Shu Guo, Quan Wang, Lihong Wang, Bin Wang, and Li Guo. Knowledge graph embedding with iterative guidance from soft rules. arXiv preprint arXiv:1711.11231, 2017.

[11]Frank L Hitchcock. The expression of a tensor or a polyadic as a sum of products. Journal of Mathematics and Physics, 6(1-4):164–189, 1927.

[12]Raphael Hoffmann, Congle Zhang, Xiao Ling, Luke Zettlemoyer, and Daniel S. Weld. Knowledge-based weak supervision for information extraction of overlapping relations. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies -Volume 1, HLT '11, pp. 541–550, USA, 2011. Association for Computational Linguistics. ISBN 9781932432879.

[13]赵军, 大规模知识图谱构建关键技术与应用. 北京市,中国科学院自动化研究所,2019-09-01.