

关系预测作为改进多关系图表示的辅助训练目标

Yihong 陈

YIHONG.CHEN@CS.UCL.AC.UK

伦敦大学学院, 英国伦敦 Facebook AI Research, 英国伦敦 Pasquale Minervini

P.MINERVINI@CS.UCL.AC.UK

伦敦大学学院, 伦敦, 英国

塞巴斯蒂安·里德尔

S.RIEDEL@CS.UCL.AC.UK

伦敦大学学院, 英国伦敦 Facebook AI Research, 英国伦敦

P.STENETORP@CS.UCL.AC.UK

蓬托斯 Stenetorp

伦敦大学学院, 英国伦敦

摘要

在多关系图上学习良好的表示对于知识库补全(KBC)至关重要。本文提出了一个新的多关系图表示学习的自监督训练目标, 通过简单地将关系预测纳入常用的 1vsAll 目标。新的训练目标不仅包含用于预测给定三元组的主客体的术语, 还包含用于预测关系类型的术语。我们分析了这一新的目标如何影响 KBC 中的多关系学习: 在各种数据集和模型上的实验表明, 关系预测可以显著提高实体排名, 这是 KBC 最广泛使用的评估任务, 在 FB15k-237 上产生了 6.1% 的 MRR 和 9.9% 的增长 Hits@1, 在 Aristo-v4 上产生了 3.1% 的 MRR 和 3.4% 的增长 Hits@1。此外, 我们观察到所提出的目标在高度多关系数据集(即具有大量谓词的数据集)上特别有效, 并在使用更大的嵌入尺寸时生成更好的表示。代码将在 <https://github.com/facebookresearch/ssl-relation-prediction> 上提供。

1. 介绍

知识库补全(Knowledge Base Completion, KBC), 又称链路预测, 旨在补全缺失条目, 在构建大规模知识图谱中起着至关重要的作用[Nickel 等人, 2016, Ji 等人, 2020, Li 等人, 2020]。在过去的几年里, 大部分关于 KBC 的研究都集中在知识图谱嵌入模型上, 这些模型学习知识图谱中所有实体和关系的表示, 并使用它们对是否存在边进行评分[Nickel et al., 2016]。文献中提出了许多模型和架构创新, 包括但不限于基于翻译的模型[Bordes 等人, 2013], 潜在因子分解模型[Nickel 等人, 2011, Trouillon 等人, 2016, Balazevic 等人, 2019], 以及基于神经网络的模型[detmers 等人, 2018, Schlichtkrull 等人, 2018, Xu 等人, 2020]。

其他更近期的研究一直在分析这些 KBC 模型的评估程序方面做出补充努力。例如, Sun 等人[2020]呼吁将评估协议标准化; Kadlec 等人[2017]、Ruffinelli 等人[2020]和 Jain 等人[2020]强调了训练策略的重要性, 并表明仔细的超参数调整可以产生更准确的结果

比采用更精细的模型架构,Lacroix 等人[2018]认为, 当适当选择训练目标时, 一个简单的模型可以产生最先进的结果。

从这些发现中获得灵感, 本文探索了关系预测: 一个简单的辅助训练目标, 可以显著提高几个 KBC 模型的多关系图表示学习。除了训练模型来预测知识图谱中三元组的主体和客体实体, 我们还训练它们来预测关系类型, 从而形成一个自监督的训练目标。直观地说, 这种方法类似于使用类似于掩码语言模型的训练目标[Devlin et al., 2019]而不是 KBC 常用的自回归训练目标。在我们的实验中, 我们发现新的辅助训练目标显著提高了下游链路预测的精度。

对各种模型和数据集的实证评估支持了我们新的训练目标的有效性: 在嵌入大小介于 2K 和 4K 之间的 ComplEx-N3 [Trouillon et al., 2016]和 CP-N3 [Lacroix et al., 2018]上观察到最大的改进, 在 Hits@1 上提供高达 9.9%的提升, 在 FB15k-237 上提供 6.1%的提升, 而计算开销可以忽略不计。

在具有不同数量谓词的数据集上进行了进一步的实验, 发现当数据集高度多关系时, 即包含大量谓词时, 关系预测的帮助更大。此外, 我们的定性分析表明, 一些多对多[Bordes et al., 2013]谓词的预测得到了改进, 关系表示更加多样化。

2.背景及相关工作

知识图谱 $G \subseteq E \times R \times E$ 包含一组主语-谓语-宾语 h, p, o 三元组, 其中每个三元组表示三元组的主语 $s \in E$ 和宾语 $o \in E$ 之间的类型 $p \in R$ 的关系。这里, E 和 R 分别表示所有实体和关系类型的集合。

知识图谱嵌入模型知识图谱嵌入 (KGE) 模型, 也称为神经链接预测器, 是一个可微模型, 其中 E 中的实体和 R 中的关系类型在一个连续的嵌入空间中表示, 两个实体之间链接的可能性是它们表示的函数。更正式地说, KGE 模型由一个参数评分函数 $\phi_\theta: E \times R \times E \rightarrow \mathbb{R}$ 定义, 参数 θ 在给定三个 h, p, o 的情况下, 产生实体 s 和 o 与关系 p 相关的可能性。

评分函数 KGE 模型可以通过它们的评分函数 ϕ_θ 来表征。例如, 在 TransE [Bordes et al., 2013] 中, 一个三重 h, p, o 的得分由 $\phi_\theta(s, p, o) = \|s + p - o\|_2^2$ 给出, 其中 $s, p, o \in \mathbb{R}^k$ 分别表示 s, p 和 o 的嵌入表示。Pk 在 Dist-Mult [Yang et al., 2015] 中, 评分函数定义为 $\phi_\theta(s, p, o) = \sum_{i=1}^k s_i p_i o_i$, 其中 $h \cdot, \cdot \cdot, \cdot i$ 表示三线性点积。正则张量分解[CP, Hitchcock, 1927]类似于 DistMult, 不同的是, 每个实体 x 有两种表示, $x_s \in \mathbb{R}^k$ 和 $x_o \in \mathbb{R}^k$, 取决于它是被用作主语还是被用作宾语: $\phi_\theta(s, p, o) = h s s, p, o o i$ 。在 RESCAL [Nickel et al., 2011] 中, 评分函数是由 $\phi_\theta(s, p, o) = s^T P o$ 给出的双线性模型, 其中 $s, o \in \mathbb{R}^k$ 是 s 和 p 的嵌入表示, $p \in \mathbb{R}^{k \times k}$ 是 p 的表示。注意, 如果 p 被约束为对角线, DistMult 等价于 RESCAL。该模型的另一种变体是 ComplEx [Trouillon et al., 2016], 其中 s, p 和 o 的嵌入表示是复数向量-即 $s, p, o \in \mathbb{C}^k$ -, 评分函数由 $\phi_\theta(s, p, o) = \langle (h s, p, o i) \rangle$ 给出, 其中 $\langle x \rangle$ 表示 x 的实部, \bar{x} 表示 x 的复数共轭。在 TuckER [Balazevic et al., 2019] 中, 评分函数为

定义为 $\phi^\theta(s, p, o) = W \times 1s \times 2p \times 3o$, 其中 $W \in \mathbb{R}^{k_s \times k_p \times k_o}$ 是参数的三支张量, $s \in \mathbb{R}^{k_s}$, $p \in \mathbb{R}^{k_p}$, $o \in \mathbb{R}^{k_o}$ 是 s 、 p 和 o 的嵌入表示。在这项工作中, 我们主要关注 DistMult、CP、ComplEx 和 TuckER, 因为它们在几个链路预测基准上的有效性[Ruffinelli 等人, 2020, Jain 等人, 2020]。

训练目标表征 KGE 模型的另一个维度是它们的训练目标。早期的张量分解模型, 如 RESCAL 和 CP, 被训练以最小化整个邻接张量的重建误差[Nickel 等人, 2011]。为了扩展到更大的知识图谱, 后续的方法如 Bordes 等人[2013]和 Yang 等人[2015]通过使用负采样简化了训练目标: 对于每个训练三元组, 一个腐败过程通过腐败三元组的主体和客体来生成一批负样本, 通过增加训练三元组的分数同时降低其腐败的分数来训练模型。这种方法后来被 detmers 等人[2018]扩展, 其中, 给定一个主语 s 和一个谓语 p , 预测正确对象的任务被转换为 $|\mathbb{E}|$ 维多标签分类任务, 其中每个标签对应一个不同的对象, 多个标签可以分配给 (s, p) 对。这种方法被 Ruffinelli 等人称为 KvsAll[2020]。另一种扩展是由 Lacroix 等人提出的[2018], 给定一个受试者 s 和一个对象 p , 预测训练三元组中正确对象 o 的任务被转换为 $|\mathbb{E}|$ 维多类分类任务, 其中每个类对应一个不同的对象, 只有一个类可以分配给 (s, p) 对。这被 Ruffinelli 等人称为 1vsAll[2020]。

请注意, 对于基于因子分解的模型, 如 DistMult、ComplEx 和 CP, KvsAll 和 1vsAll 目标可以在 gpu 上高效计算[Lacroix 等人, 2018, Jain 等人, 2020]。例如, 对于 DistMult, 所有具有主语 s 和谓语 p 的三元组的分数可以通过 $E(sp)$ 计算, 其中 E 表示元素乘积, $E \in \mathbb{R}^{|\mathbb{E}| \times k}$ 是实体嵌入矩阵。在这项工作中, 我们遵循 Lacroix 等人[2018]并采用 1vsAll 损失, 以便能够与他们的结果进行比较, 并且由于 Ruffinelli 等人[2020]表明, 他们在下游链路预测精度方面产生了类似的结果。

最近关于 KBC 模型标准化评估协议的工作[Sun 等人, 2020]及其系统评估[Kadlec 等人, 2017 年, Mohamed 等人, 2019 年, Jain 等人, 2020 年, Ruffinelli 等人, 2020 年]表明, 基于潜在因子分解的模型, 如 RESCAL、ComplEx 和 CP, 在适当调优超参数时非常有竞争力[Kadlec 等人, 2017 年, Ruffinelli 等人, 2020]。在这项工作中, 我们表明, 使用关系预测作为辅助训练任务可以进一步提高其下游链路预测精度。

3. 关系预测作为辅助训练目标

在所谓的 1vsAll 设置中[Ruffinelli 等人, 2020], 通过最大化受试者的条件似然($\text{resp. } s$), 使用自监督训练目标来训练 KBC 模型。对象 o 的训练三元组, 给定谓词和对象 o (resp. 主语 s), 更正式地说,

KBC 模型通过最大化以下目标进行训练:

$$\begin{aligned} \arg \max_{\theta \in \Theta} \sum_{\langle s, p, o \rangle \in \mathcal{G}} [\log P_{\theta}(s | p, o) + \log P_{\theta}(o | s, p)] \\ \log P_{\theta}(o | s, p) = \phi_{\theta}(s, p, o) - \log \sum_{o' \in \mathcal{E}} \exp [\phi_{\theta}(s, p, o')] \\ \log P_{\theta}(s | p, o) = \phi_{\theta}(s, p, o) - \log \sum_{s' \in \mathcal{E}} \exp [\phi_{\theta}(s', p, o)], \end{aligned} \quad (1)$$

其中 $\theta \in \Theta$ 是模型参数, 包括实体和关系嵌入, ϕ_{θ} 是一个以 θ 为参数的评分函数。这样的客观限制了训练目标中的预测位置为三元组的第一个(s)或最后一个(o)元素。

在这项工作中, 我们提出将关系预测作为训练 KBC 模型的辅助任务。新的训练目标不仅包含了公式(1)中 $-\log P(s | p, o)$ 和 $\log P(o | s, p)$ 的预测主客体的术语, 还包含了预测关系类型 p 的客观 $\log P(p | s, o)$:

$$\begin{aligned} \arg \max_{\theta \in \Theta} \sum_{\langle s, p, o \rangle \in \mathcal{G}} [\log P_{\theta}(s | p, o) + \log P_{\theta}(o | s, p) + \lambda \log P_{\theta}(p | s, o)] \\ \log P_{\theta}(p | s, o) = \phi_{\theta}(s, p, o) - \log \sum_{p' \in \mathcal{R}} \exp [\phi_{\theta}(s, p', o)], \end{aligned} \quad (2)$$

其中 $\lambda \in \mathbb{R}^+$ 是用户指定的超参数, 决定了关系预测目标的贡献; 除非另有规定, 否则我们假设 $\lambda = 1$ 。这个新的训练目标给训练过程增加了很少的开销, 并且可以很容易地添加到现有的 KBC 实现中: PyTorch 示例包含在附录 a 中。与传统方法相比, 关系预测可以帮助模型学习进一步区分不同谓词。

4. 实证研究

在本节中, 我们进行了几个实验来验证将关系预测作为辅助训练目标的有效性。我们对以下研究问题感兴趣: RQ1: 新的训练目标如何影响跨不同数据集的下游知识库补全任务的结果? 数据集上的关系类型数量如何影响新的训练目标?

RQ2: 新的训练目标如何影响不同的模型? 它是统一地使所有模型受益, 还是特别地帮助其中的一些模型?

RQ3: 新的训练目标是否产生了更好的实体和关系表示?

数据集我们使用了在 KBC 文献中常用的 Kok 和 Domingos[2007]、WN18RR [Dettmers 等人, 2018] 和 FB15k-237 [Toutanova 等人, 2015] 中的 Nations、UMLS 和亲属关系。由于这些数据集包含的谓词数量相对较少, 我们还对 Aristo-v4 进行了实验, 这是 Aristo 元组 KB 的第 4 版 [Mishra et al., 2017], 其中包含超过 1.6k 个谓词。由于 Aristo-v4 对 KBC 没有标准化的划分, 我们随机抽取了 20k 个三元组进行测试, 20k 个用于验证。表 1 总结了这些数据集的统计数据。

数据集	E	R	#火 车	#验 证	#测 试
国家	14	55	1592 年	One hundred.	301
uml	135	46	5216 年	652	661
亲属关系	104	25	8544 年	1068 年	1074 年
WN18RR	40943 年	11	86835 年	3034 年	3134 年
fb15k - 237	27395 年	237	272115 年	17535 年	20466 年
Aristo-v4	44950 年	1605 年	242594 年	20000 年	20000 年

表 1:数据集统计，其中|E|和|R|表示实体和谓词的数量。

实体排序是知识库补全中最常用的评估协议。对于给定的查询(s, p, ?)或(?), p, o)，所有候选实体根据模型产生的分数进行排序，并使用得到的排序来计算真实答案的排名。我们使用标准的过滤均值倒数排名(MRR)和 Hits@K(排名前 K 个结果的命中率)， $K \in \{1, 3, 10\}$ 作为指标。

模型我们使用几个具有竞争力和可重现性的模型 [Ruffinelli 等人, 2020, Sun 等人, 2020]: RESCAL [Nickel 等人, 2011], ComplEx [Trouillon 等人, 2016], CP [Lacroix 等人, 2018], 和 TuckER [Balazevic 等人, 2019]。为了确保各种比较中的公平性，我们使用验证集对超参数进行了广泛的调优，该验证集总共包含 41,316 次训练运行。对于所有数据集上的主要结果，我们使用网格搜索(grid-search)调整了 λ 。对于关于谓词数量和不同模型选择的消融实验，为了简单起见，我们将 λ 设置为 1。关于超参数扫描的详细信息可以在附录 B 中找到。

4.1 RQ1:关系 预测 对 不同数据集的影响

提出的训练目标如何影响不同数据集上的知识库补全?为了回答这个研究问题，我们在几个流行的 KBC 数据集上比较了训练与关系预测和没有关系预测的训练的性能。对于较小的数据集(, Nations 和 UMLS)，我们从 RESCAL, ComplEx, CP 和 TuckER 中选择了最好的一个。对于更大的数据集(WN18RR, FB15k-237 和 Aristo-v4)，由于计算预算有限，我们使用了 ComplEx，它在我们的初步实验中表现优于其他模型。

表 2 总结了较小数据集上的结果，其中“表示使用关系预测进行训练，而%表示不使用关系预测进行训练。我们可以观察到，关系预测带来了在 MRR 和 Hits@1 上 2% - 4%的改进，以及保持有竞争力的 Hits@3 和 Hits@10。

表 3 总结了在更大的数据集上的结果。包括关系预测作为辅助训练目标带来了 3 个数据集上所有指标的一致改进，除了 WN18RR 上的 Hits@10。特别是，在 FB15k-237 数据集上，关系预测导致了 6.1%的 MRR、9.9%的 Hits@1、6.1%的 Hits@3 以及 3.1%的 MRR、3.4%的 Hits@1、3.8%的 Hits@3 在 Aristo-v4 数据集上的增长。与 WN18RR 相比，我们观察到 FB15k-237 和 Aristo-v4 的改善更大。一个潜在的原因是，FB15k-237 上的谓词集合(|R| = 237)和 Aristo-v4 (|R| = 1605)比 WN18RR (|R| = 11)更多样化。WN18RR 上|R|的谓词数量相对较少，模型从区分不同实体中获得的收益大于区分关系。换句话说，使用更低的 λ 值(权重的

数据集	实体的预测	关系预测	MRR	Hits@1	Hits@3	Hits@10
亲属关系	%	“ % ”	0.920	0.867	0.970	0.990
	"		0.897	0.835	0.955	0.987
	"		0.916	0.866	0.964	0.988
国家	%	“ % ”	0.686	0.493	0.871	0.998
	"		0.813	0.701	0.915	1.000
	"		0.827	0.726	0.915	0.998
uml	%	“ % ”	0.863	0.795	0.914	0.979
	"		0.960	0.930	0.991	0.998
	"		0.971	0.954	0.986	0.997

表 2:亲属关系、民族和 UMLS 的测试性能比较。我们对 4 个不同的模型进行了广泛的超参数搜索，分别是 RESCAL, ComplEx, CP 和 Tucker，其中模型也被视为超参数。在这三个数据集上包括关系预测作为辅助训练目标在测试 MRR 和 Hits@1 方面有帮助，同时保持竞争性测试 Hits@3 和 Hits@10。关于超参数选择过程的更多细节可在附录 B.1.1 中找到。

数据集	实体的预测	关系预测	MR R	Hits@ 1	Hits@ 3	Hits@1 0
WN18R R	%	“ % ”	0.258	0.21	0.290	0.33
	"		0.487	2	0.501	9
	"		0.488	0.44	0.505	0.58
				1		0
				0.44		0.57
fb15k - 237	%	“ % ”	0.263	0.18	0.287	0.41
	"		0.366	7	0.401	1
	"		0.388	0.27	0.425	0.55
				1		7
				0.29		0.56
Aristo- v4	%	“ % ”	0.169	0.12	0.177	0.26
	"		0.301	0	0.324	7
	"		0.311	0.23	0.336	0.43
				2		8
				0.24		0.44
				0		7

表 3:使用 ComplEx 对 WN18RR、FB15k-237、Aristo-v4 的测试性能比较。将关系预测作为辅助训练目标，在三个数据集上除了 Hits@10 在 WN18RR 上的所有指标上都带来了一致的改进。在 FB15k-237 和 Aristo-v4 上，添加关系预测会在下游链路预测任务上产生更大的改进。关于超参数选择过程的更多细节可以在附录 B.1.2 中找到。

关系预测目标)更适合谓词较少但实体数量较多的数据集。我们在 4.1.2 节中包括了R上的消融。

数据集	MRR	Hits@1	Hits@3	Hits@10
WN18RR	(15.0, 0.03125)	(15.0, 0.03125)	(15.0, 0.03125)	(3.0, 0.76740)
fb15k - 237	(15.0, 0.03125)	(15.0, 0.03125)	(15.0, 0.03125)	(15.0, 0.03125)
Aristo-v4	(15.0, 0.03125)	(15.0, 0.03125)	(15.0, 0.03125)	(15.0, 0.03125)

表 4:复数- n3 在几个数据集上的 Wilcoxon 符号秩检验。对于每个数据集和指标，我们报告相应的统计量-即正差异的秩和-和 p 值 as (statistics, p-value)。

以下 4.4.1 意义测试

为了表明关系扰动带来的改进是显著的，我们使用 5 个随机种子进行实验，并对使用和不使用关系预测获得的指标进行 Wilcoxon 符号秩检验 [Wilcoxon, 1992]。测试执行如下。首先，我们计算了在有关系预测和没有关系预测的情况下进行复杂训练得到的结果之间的差异。零假设是差异的中位数是负的。表 4 总结了结果。我们可以观察到几乎所有的 p 值都在 0.03 左右，这意味着我们可以在约 97% 的置信水平上拒绝零假设。将关系预测作为辅助训练目标的新训练目标显著提高了 KBC 模型在 WN18RR 上的性能，除了 Hits@10。

4.1.2 对谓词数量的消融

如前所述，关系预测对 WN18RR、FB15k-237 和 Aristo-v4 带来了不同的影响。由于这些数据集最大的差异之一是不同谓词的数量 $|R|$ (Aristo-v4 为 1,605, FB15k-237 为 237，而 WN18RR 仅为 11)，因此我们想用各种 $|R|$ 来确定扰动关系的影响。为了实现这一点，我们通过从 FB15k-237 中采样包含一个谓词子集的三元组来构建一系列具有不同 $|R|$ 的数据集。例如，为了构建一个只有 5 个谓词的数据集，我们首先从 237 个谓词集合中采样 5 个谓词，然后提取包含这 5 个谓词的三元组作为新的数据集。总的来说，我们拥有了 $|R| \in [5, 25, 50, 100, 150, 200]$ 谓词的数据集。为了解决在数据集构建过程中谓词采样中引入的噪声，我们用 3 个随机种子进行了实验。为了方便起见，我们将关系预测 λ 的权重设置为 1，并在正则化和其他超参数上执行类似的网格搜索，以确保模型被正则化并使用不同数量的训练和测试数据点进行适当的训练。

结果总结在图 1 中。如图 1 的右侧部分所示，预测关系有助于拥有更多谓词的数据集，从而在 MRR、Hits@1 和 Hits@3 上获得 2%-4% 的提升。对于少于 50 个谓词的数据集，在图的左侧部分所示的相对变化中有相当大的波动——但有明显的下降趋势。这些结果验证了我们的假设，即关系预测给谓词数量较多的数据集带来好处。请注意，我们没有调整关系预测目标 λ 的权重(并将其固定为 1)，这个选择在谓词数量较少的数据集上可能是次优的。

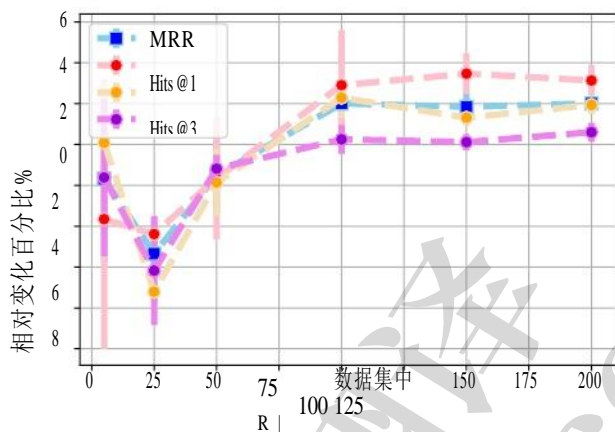


图 1:复杂训练与 w/o 关系预测在不同谓词数量的数据集 $|R|$ 之间的相对变化。我们用 3 个随机种子进行了实验。条形越大，方差越大。相对变化用 $(m^+ - m^-) / m^-$ 计算，其中 m^+ 和 m^- 表示与 w/o 关系预测的度量值。对于 $|R| < 50$ 的数据集，可以观察到明显的下降趋势，而在 $|R| > 50$ 的数据集中，MRR、Hits@1 和 Hits@3 明显增加了 2% - 4%。

模型	关系预测	MRR	Hits@1	Hits@3	Hits@10
CP	%	0.356	0.262	0.392	0.546
		-	-	-	-
		0.366	0.274	0.401	0.550
复杂的	%	0.366	0.271	0.401	0.557
		-	-	-	-
		0.382	0.289	0.419	0.568
RESCAL	%	0.356	0.266	0.390	0.532
		-	-	-	-
		0.359	0.271	0.395	0.533
塔克	%	0.351	0.260	0.386	0.532
		-	-	-	-
		0.354	0.264	0.388	0.535

表 5:FB15k-237 在 CP、ComplEx、RESCAL 和 TuckER 4 种不同模型上的测试性能比较。我们将关系预测的权重设置为 1，并对超参数进行网格搜索。更多细节可在附录中查阅。虽然关系预测似乎对所有 4 个模型都有帮助，但与 TuckER 和 RESCAL 相比，它给 CP 和 ComplEx 带来了更多的好处。

4.2 RQ2:关系预测对不同 KBC 模型的影响

为了衡量关系预测如何影响 KBC 模型的下游精度，我们在 FB15k-237 上使用几个模型进行了实验，即 ComplEx、CP、TuckER 和 RESCAL。

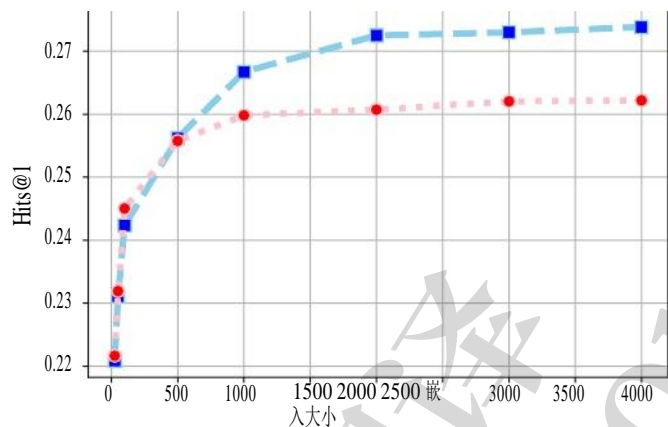


图 2: Hits@1 vs 嵌入大小对于 FB15k-237 上的 CP，每个点代表一个用一些特定嵌入大小训练的模型，具有(蓝色)/ out(红色)扰动关系。最小的嵌入大小是 25。

为了简单起见，我们将关系预测 λ 的权重设置为 1。如表 5 所示，包括关系预测作为辅助训练目标会给所有模型带来持续的改进。值得注意的是，CP 和 ComplEx 在 Hits@1 上分别可以观察到高达 4.4% 和 6.6% 的增长。对于 TuckER 和 RESCAL 来说，关系扰动带来的改进相对较小。这可能是因为我们必须为 TuckER 和 RESCAL 使用较小的嵌入尺寸，因为众所周知，这些模型在使用较大的嵌入尺寸时存在可扩展性问题。我们在 4.2.1 节中包括了对模型嵌入尺寸的消融。至于计算成本，在我们的实验中，采用新的损失只增加了每个 epoch 平均 2% 的训练时间增加，尽管它可能需要更多的 epoch 才能收敛。

4.2.1 嵌入尺寸的消融

在我们的实验中，增加模型的嵌入大小可以带来更好的性能。然而，可能存在一个饱和点，较大的嵌入大小停止提高性能。我们感兴趣的是扰动关系将如何影响饱和点以及哪些嵌入尺寸从中受益最大。图 2 显示了 FB15k-237 上 CP 的嵌入大小和 MRR 之间的关系。在较小的嵌入尺寸下，扰动关系的影响不大。然而，它确实有助于较大嵌入尺寸的 CP，并延迟饱和点。我们可以看到，蓝色曲线的斜率比红色曲线的斜率大，红色曲线在嵌入尺寸 1000 和嵌入尺寸 4000 之间弯曲很小。因此，我们可以观察到，扰动关系为通过增加嵌入尺寸来改进模型留下了更多的空间。

4.3 RQ3: 对学习到的实体和关系表示进行定性分析

在我们的实验中，我们观察到关系预测提高了多对多谓词的链接预测精度，这对于 KBC 模型来说是众所周知的困难的[Bordes 等人, 2013]。表 6 列出了从关系预测中获益最多的前 10 个谓词。我们对谓词进行排名

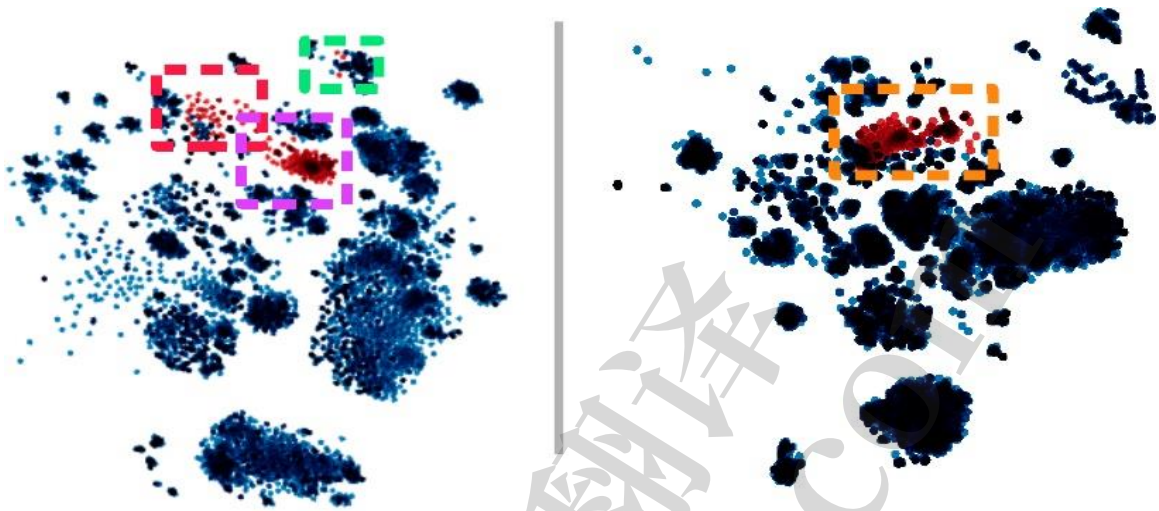


图 3: 复杂嵌入的 t-SNE 可视化，用关系预测进行训练（左）和没有关系预测（右）。红点和蓝点对应谓词分别和实体。虚线框突出不同的簇。

-/冰球/冰球队/当前名单。/体育/运动队名单/位置	-	-
/体育/运动团队名单。/棒球棒球球员位置/位置	-	-
/位置/国家/二级部门	-	-
/电视/电视制片人/程序。/电视/电视制片人术语/程序		
/奥运会奥林匹克运动/运动员。/奥运会奥林匹克运动员所属/奥运会	-	-
/获奖作品/获奖情况。荣誉/奖项/奖项/荣幸		
/音乐/仪表/家庭	-	-
/体育/奥运会奥运会		
/基地/ bibliomess /肚兜位置/状态	-	-

表 6:关系预测改进最多的前 10 个谓词。

通过平均(s, p, ?)和(?)的关联 MRR， p, o)查询。表 14 和表 15 列出了(s, p, ?)的前 20 个查询(? , p, o)，通过关系预测提高最多。我们可以看到，关系预测有助于诸如“电影《魔力麦克》在哪里上映？”，“派拉蒙影业在哪里成立？”、“电影《独裁者 2012》中出现的是哪个人？”，“哪些地方位于英国？”、“维拉·德雷克获得了哪个奖项？”

为了直观地理解为什么它对这些谓词有帮助，我们在学习到的实体和谓词表示上运行了 t-SNE。互反谓词也包含在 t-SNE 可视化中。我们将嵌入大小设置为 1000，并使用 N3 正则化。超参数是根据验证 MRR 来选择的。我们运行 t-SNE 5000 步，复杂度为 50。从图 3 中我们可以看到，相比之下，用于关系预测的 t-SNE 可视化中有更多的谓词簇

没有关系的预测。这表明关系预测有助于模型区分不同的谓词:大多数谓词从实体中分离(粉色区域),而一些具有相似语义或主-客体上下文的谓词形成一个集群(红色区域);还有一些谓词,它们与对应的谓词并不接近,而是与高度相关的实体(绿色区域)接近。表 7 列出了每个区域的 3 个示例谓词。虽然在将高维嵌入向量投影到二维空间的过程中可能会有信息丢失,但我们希望这种可视化将有助于说明关系预测如何帮助学习更多样化的谓词表示。

粉色区域
/额外基础/ schemastaging /组织/电话号码。/base/schemastaging/电话沙盒/联系人类别/位置/统计区域/导出到的地方。/位置/导入和导出/导出到 - - - /体育/运动聪明/团队 //体育/运动聪明参与的团队
红色区域
/人/人/国籍 /人/人/宗教 - - - - -
绿色区域
graduates. /教育/教育机构/学生/教育/教育/学生 /共同/主题/网页/共同/网页/类别 - - - - -

表 7:t-SNE 图中每个区域的三个示例谓词。

5.讨论和结论

在本文中,我们建议使用一个新的自监督目标来训练 KBC 模型-通过简单地将关系预测合并到常用的 1vsAll 目标中。在我们的实验中,我们表明添加这样一个简单的学习目标对各种 KBC 模型有显著的帮助。它使在 FB15k-237 上训练的复杂模型在 Hits@1 上提高了 9.9%,即使实体排名的评估任务似乎与关系预测无关。

我们的工作为设计 KBC 的关系感知的自我监督目标提供了一个有价值的方向。在本文中,我们主要关注基于因子分解的简单模型。未来的工作将考虑对更复杂的 KBC 模型,如基于图神经网络的 KBC 模型,以及在更多的数据集上分析所提出的目标。另一个有趣的未来工作方向是在链路预测之外的更下游应用上分析所提出的辅助目标,并评估它是否可以用于学习有用的多关系图表示。

确认

我们要感谢审稿人的建设性反馈。我们也要感谢 UCL NLP 的所有其他成员在 COVID - 19 期间的支持。PM 和 PS 由欧盟地平线 2020 研究和创新计划支持,资助协议编号为。875160.

参考文献

- Ivana Balazevic, Carl Allen 和 Timothy M. Hospedales。Tucker:张量分解用于知识图谱补全。在 EMNLP / IJCNLP, 2019。
- Antoine Bordes, Nicolas Usunier, Alberto García-Durán, J. Weston, Oksana Yakhnenko。转换嵌入以建模多关系数据。在少量的酒,2013。
- 蒂姆·德特默斯、帕斯夸尔·密涅瓦尼、本图斯·斯坦托普和塞巴斯蒂安·里德尔。卷积 2d 知识图谱嵌入。在 2018 年第 32 届 AAAI 人工智能会议上。
- Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。Bert:用于语言理解的深度双向 transformer 预训练。在计算语言学协会 2019 年北美分会会议论文集:人类语言技术, 第 1 卷(长短文), 第 4171-4186 页, 2019 年。
- f · 1 · 希区柯克。张量或多元作为积和的表达式。*j. 数学. 物理*, 6(1):164-189,1927。
- Prachi Jain, Sushant Rathi, Mausam, 和 Soumen Chakrabarti。知识库完成:基线反击战(再次)。ArXiv, abs / 2005.00804, 2020 年。
- 季少雄、潘世瑞、埃里克·坎布里亚、佩卡·马丁宁、余菲利普。知识图谱研究综述:表示、获取与应用。arXiv 预印本 arXiv:2002.00388, 2020。
- Rudolf Kadlec, Ondrej Bajgar, and Jan Kleindienst。知识库完成:基线反击。在 Phil Blunsom、Antoine Bordes、Kyunghyun Cho、Shay B. Cohen、Chris Dyer、Edward Grefenstette、Karl Moritz Hermann、Laura rimel、Jason Weston 和 Scott Yih, 编辑, 《第二届 NLP 表示学习研讨会会议录》Rep4NLP@ACL 2017, 加拿大温哥华, 2017 年 8 月 3 日, 第 69-74 页。Association for Computational Linguistics, 2017。doi: 10.18653 / v1 / w17 - 2609。URL <https://doi.org/10.18653/v1/w17-2609>。
- Stanley Kok 和 Pedro M. Domingos。统计谓词的发明。ICML, ACM 国际会议论文集系列第 227 卷, 第 433-440 页。ACM, 2007 年。
- Timothée Lacroix, Nicolas Usunier 和 Guillaume Obozinski。规范张量分解用于知识库补全。摘自 Jennifer G. Dy 和 Andreas Krause, 编辑, 第 35 届机器学习国际会议论文集, ICML 2018, 斯德哥尔摩, 瑞典, 2018 年 7 月 10-15 日, 《机器学习研究论文集》第 80 卷, 2869 - 2878 页。PMLR, 2018 年。URL <http://proceedings.mlr.press/v80/lacroix18a.html>。
- 李晨晨、李爱萍、王烨、涂洪奎、宋一晨。知识表示学习方法与应用综述。2020 年 IEEE 第五届网络空间数据科学国际会议(DSC), 第 312-319 页。IEEE 2020。
- B. D. Mishra, Niket Tandon 和 P. Clark。以领域为目标的高精度知识提取。*计算语言学协会汇刊*, 5:233-246,2017。
- Sameh K Mohamed, Vít Nováček, Pierre-Yves Vandenbussche 和 Emir Mušez。知识图谱嵌入模型中的损失函数。在《dl4kg2019 - 知识图谱深度学习研讨会论文集》2019 年第 1 页。

M. Nickel、Volker Tresp 和 H. Krieger。在多关系数据上进行集体学习的三支模型。2011 年 ICML。

M. 尼克尔、凯文·墨菲、沃尔克·特雷普、叶夫根尼·盖布里洛维奇。面向知识图谱的关系机器学习综述。《IEEE 学报》，104:11-33,2016。

D. Ruffinelli, Samuel Broscheit, Rainer Gemulla。你可以教老狗新把戏!关于训练知识图谱嵌入。2020 年 ICLR。

Michael Schlichtkrull、Thomas N Kipf、Peter Bloem、Rianne Van Den Berg、Ivan Titov 和 Max Welling。用图卷积网络建模关系数据。欧洲语义网会议，第 593-607 页。施普林格,2018 年。

孙志清、Shikhar Vashishth、Soumya Sanyal、Partha P. Talukdar、杨一鸣。知识图谱补全方法的再评价。在 Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, 编辑, 计算语言学协会第 58 届年会论文集, ACL 2020, 在线, 2020 年 7 月 5-10 日, 第 5516-5522 页。计算语言学协会, 2020 年。URL <https://www.aclweb.org/anthology/2020.acl-main.489/>。

Kristina Toutanova、陈丹奇、Patrick Pantel、Hoifung Poon、Pallavi Choudhury、Michael Gamon。代表文本，用于文本和知识库的联合嵌入。2015 年自然语言处理经验方法会议论文集，1499-1509 页，葡萄牙里斯本，2015 年 9 月。计算语言学协会。doi: 10.18653 / v1 / d15 - 1174。URL <https://www.aclweb.org/anthology/d15-1174>。

Théo Trouillon、Johannes Welbl、Sebastian Riedel、Eric Gaussier 和 Guillaume Bouchard。用于简单链接预测的复杂嵌入。在 Maria-Florina Balcan 和 Kilian Q. Weinberger, 编辑, 第 33 届国际机器学习会议论文集, ICML 2016, 美国纽约州纽约市, 2016 年 6 月 19-24 日, JMLR 研讨会和会议论文集第 48 卷, 2071-2080 页。JMLR.org, 2016。URL <http://proceedings.mlr.press/v48/trouillon16.html>。

弗兰克 Wilcoxon。通过排名方法进行个体比较。《统计学的突破》，第 196-202 页。施普林格,1992 年。

徐小冉、冯伟、姜云生、谢晓辉、孙志清、邓志宏。动态剪枝的消息传递网络，用于大规模知识图谱推理。国际学习表征会议，2020 年。URL <https://openreview.net/forum?id=rkeuAhVKvB>。

杨璧山，易文头，何晓东，高剑峰，邓丽。在知识库中嵌入用于学习和推理的实体和关系。在 ICLR(海报)，2015 年。

附录 a作为辅助训练目标的关系预测代码片段

图4 演示了如何将关系预测添加到现有的 ComplEx 实现中。

附录 b超参数扫描

在这一节中，我们总结了实验中使用的所有超参数。我们使用 Tesla P100 和 Tesla V100 gpu 来运行实验。我们通过 PyTorch 实现了每个模型。我们的代码库基于 <https://github.com/facebookresearch/kbc>。

B.1 跨数据集关系预测的超参数范围

B.1.1 亲属关系、国家和 umls

模型	D 或 D_r, D_s	lr	bsz	注册
RESCAL	(50、100、200)	[0.1, 0.01]	[10、50、100、500]	[0, 0.005, 0.01, 0.05, 0.1, 0.5]
复杂的	[100、200、500、1000]	[0.1, 0.01]	[10、50、100、500]	[0, 0.005, 0.01, 0.05, 0.1, 0.5]
CP	[200, 400, 1000, 2000]	[0.1, 0.01]	[10、50、100、500]	[0, 0.005, 0.01, 0.05, 0.1, 0.5]
塔克	25[(100),(200 年 25),(100 年,50),(200 年,50),(100、100),(200、100)]	[0.1, 0.01]	[10、50、100、500]	[0, 0.005, 0.01, 0.05, 0.1, 0.5]

表 8:超参数在小数据集(亲属关系，国家，UMLS)上搜索不同的 KBC 模型; D 代表嵌入大小， D_r 代表关系的单独嵌入大小， lr 是学习率， BSZ 是批量大小， reg 是正则化强度。

数据集	关系预测	实体的预测	模型	d	d_r	lr	bsz	注册	λ	Dev MRR
亲属关系	“ % ”	%	塔克	200	100	0.10	10	0.1	NA	0.919581
		"	CP	2000	NA	0.10	50	0.01	NA	0.897429
		"	CP	2000	NA	0.10	50	0.05	4.000	0.918323
国家	“ % ”	%	塔克塔	200	50	0.01	10	0.1	NA	0.686010
		"	克 CP	2000	NA	0.01	10	0.01	NA	0.855388
		"		200	25	0.01	10	0.10	0.250	0.865352
uml	“ % ”	%	CP复杂	1000	NA	0.1	500	0.01	NA	0.863008
		"	复杂	1000	NA	0.10	10	0.00	NA	0.967626
		"		1000	NA	0.01	10	0.00	0.500	0.971612

表 9:在较小的数据集上的最佳超参数配置和相应的验证 MRR。 d 代表嵌入大小， d_r 代表关系的单独嵌入大小， lr 是学习率， bsz 是批量大小， reg 是正则化强度， λ 是关系预测的权重-NA 表示“不适用”。

对于所有小数据集(, Nations, UMLS)，我们用 Adagrad 优化器和 N3 正则化训练 RESCAL, ComplEx, CP 和 TuckER，最多 400 个 epoch。倒数三元组被包括在内，因为据报道它们是有帮助的[detmers 等人，2018,Lacroix 等人，2018]。我们对超参数组合进行了网格搜索，并为每种选择了最佳配置

有道文档翻译
pdf.youdao.com


```

1  class ComplEx(KBCModel):
2      def __init__(self, sizes, rank, init_size):
3          super(ComplEx, self).__init__()
4          self.sizes = sizes
5          self.rank = rank
6
7          self.embeddings = nn.ModuleList([
8              nn.Embedding(s, 2 * rank, sparse=False)
9              for s in sizes[:2]
10         ])
11         self.embeddings[0].weight.data *= init_size
12         self.embeddings[1].weight.data *= init_size
13
14     def forward(self, x, score_rhs=True, score_rel=False, score_lhs=False, normalize_rel=False):
15         lhs = self.embeddings[0](x[:, 0])
16         rel = self.embeddings[1](x[:, 1])
17         rhs = self.embeddings[0](x[:, 2])
18
19         lhs = lhs[:, :self.rank], lhs[:, self.rank:]
20         rel = rel[:, :self.rank], rel[:, self.rank:]
21         rhs = rhs[:, :self.rank], rhs[:, self.rank:]
22
23         rhs_scores, rel_scores = None, None
24         if score_rhs:
25             to_score_entity = self.embeddings[0].weight
26             to_score_entity = to_score_entity[:, :self.rank], to_score_entity[:, self.rank:]
27             rhs_scores = (
28                 (lhs[0] * rel[0] - lhs[1] * rel[1]) @ to_score_entity[0].transpose(0, 1) +
29                 (lhs[0] * rel[1] + lhs[1] * rel[0]) @ to_score_entity[1].transpose(0, 1)
30             )
31         if score_rel:
32             to_score_rel = self.embeddings[1].weight
33             to_score_rel = to_score_rel[:, :self.rank], to_score_rel[:, self.rank:]
34             rel_scores = (
35                 (lhs[0] * rhs[0] + lhs[1] * rhs[1]) @ to_score_rel[0].transpose(0, 1) +
36                 (lhs[0] * rhs[1] - lhs[1] * rhs[0]) @ to_score_rel[1].transpose(0, 1)
37             )
38         if score_lhs:
39             to_score_lhs = self.embeddings[0].weight
40             to_score_lhs = to_score_lhs[:, :self.rank], to_score_lhs[:, self.rank:]
41             lhs_scores = (
42                 (rel[0] * rhs[0] + rel[1] * rhs[1]) @ to_score_lhs[0].transpose(0, 1) +
43                 (rel[0] * rhs[1] - rel[1] * rhs[0]) @ to_score_lhs[1].transpose(0, 1)
44             )

```

图 4:关系预测对于复杂，红色区域显示了使用关系预测作为辅助训练任务相关的线。

基于验证 MRR 的数据集。我们在表 8 中列出了超参数搜索的网格，并在表 9 中报告了搜索最佳的配置。关于关系预测和实体预测之间的平衡，我们搜索了关系预测的权重超过 {4,2,0.5,0.25,0.125}。

B.1.2 wn18rr、fb15k-237 和 aristo-v4

对于所有数据集，我们用 N3 正则化器和 Adagrad 优化器和 N3 正则化器训练 ComplEx，最多 400 个 epoch。倒数三元组被包括在内，因为据报道它们是有帮助的[detmers 等人，2018,Lacroix 等人，2018]。至于关系预测的权重，我们在不同的数据集上的不同区域进行了搜索。对于 WN18RR，我们搜索了关系预测的权重超过 [0.005,0.001,0.05,0.1,0.5,1]。对于 FB15k-237，我们搜索了 [0.125,0.25,0.5,1,2,4]。对于 Aristo-v4，我们搜索了 [0.125,0.25,0.5,1,2,4]。我们在超参数组合上进行网格搜索，并根据验证 MRR 为每个数据集选择最佳配置。我们在表 10 中报告了每个数据集的网格，在表 11 中发现了最好的配置。

数据集	d	lr	bsz	注册
WN18RR	(100、500、1000)	[0.1, 0.01]	(100、500、1000)	[0.005, 0.01, 0.05, 0.1, 0.5, 1]
fb15k - 237	(100、500、1000)	[0.1, 0.01]	(100、500、1000)	[0.0005, 0.005, 0.01, 0.05, 0.1, 0.5, 1, 0]
Aristo-v4	(500、1000、1500)	[0.1, 0.01]	(100、500、1000)	[0, 0.005, 0.01, 0.05, 0.1, 0.5, 1]

表 10:不同数据集上复杂的 vanilla 关系扰动的超参数搜索 d代表嵌入大小。Lr 是学习率。BSZ 是批量大小。Reg 是正则化强度。 λ 是关系预测的权重。

数据集	关系预测	实体的预测	d	l r	bs z	注册	λ	Dev MRR
WN18R R	“ % ”	% " "	1000	0.10	50	0.5	NA	0.25794
			1000	0.10	0	0.1	NA	5
			1000	0.10	10	0	0.050	0.48808
					0	0.1		3
					10	0		0.49005
					0			3
fb15k - 237	“ % ”	% " "	1000	0.10	100	0.000	NA	0.26288
			1000	0.10	0	5 0.05	NA	8
			1000	0.10	100	0.05	4.000	0.37230
					100			5
					0			0.39372
								2
Aristo- v4	“ % ”	” “ ”	1500	0.10	1000	0.0	NA	0.16870
			1500	0.01	500	1	NA	0
			1500	0.10	100	0.0	0.125	0.30707
						1		6
						0.0		0.31444
						5		3

表 11:带加权关系扰动的复杂跨数据集的最佳超参数配置和相应的验证 MRR 。 D 代表嵌入大小, lr 是学习率, BSZ 是批量大小, reg 是正则化强度, λ 是关系预测的权重。NA 表示不适用。

B.2 跨模型关系预测的超参数范围

我们在 FB15k-237 上对每个模型进行了实验。请注意，最初的 TucKER [Balazevic et al., 2019] 包括一些在 CP, ComplEx 和 TucKER 中未使用的训练策略，如 dropout，学习率衰减等。然而，为了公平地比较关系预测如何影响每个模型，我们使用 Adagrad 优化器和 N3 正则化在最多 400 个 epoch 的类似设置下训练了所有模型。我们进行了网格搜索，并根据验证 MRR 选择了最佳的超参数配置。在本实验中，我们将关系预测的权重设置为 1。表 12 列出了共享超参数的网格。对于 RESCAL，谓词矩阵的正则化可以在秩上进行规范化，以实现更好的结果。此外，对于 RESCAL, F2 正则化在经验上比 N3 正则化器表现更好。对于塔克来说，谓词和实体的排名是不同的。表 13 列出了我们搜索中找到的最佳超参数配置。

模型	D or (D, dr)	lr	bsz	注册
RESCAL	(128、256、512)	[0.1, 0.01]	(100、500、1000)	[0, 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1]
复杂的	(100、500、1000)	[0.1, 0.01]	(100、500、1000)	[0, 0.0005, 0.005, 0.01, 0.05, 0.1, 0.5, 1]
CP	[64、128、256、512、4000]	[0.1, 0.01]	(100、500、1000)	[0.005, 0.01, 0.05, 0.1, 0.5, 1]
塔克	[(1000, 150), (1000, 100), (400, 400), (500, 75), (300, 300), (200、200)]	[0.1, 0.01]	(100、500、1000)	[0.005, 0.01, 0.05, 0.1, 0.5, 1]

表 12:超参数在 FB15k-237 上搜索不同的 KBC 模型; D 代表嵌入大小, Dr 代表关系的单独嵌入大小, lr 是学习率, BSZ 是批量大小, reg 是正则化强度。

模型	关系预测	D or (D, dr)	lr	bsz	注册	Dev MRR
RESCAL	%	512	0.1	500	0.00	0.365384
	”	512	0.1	One hundred.	0.00	0.366789
复杂的	%	1000	0.1	One hundred.	0.05	0.372305
	”	1000	0.1	1000	0.05	0.387133
CP	%	4000	0.1	One hundred.	0.05	0.364245
	”	4000	0.1	1000	0.05	0.372408
塔克	%	(1000, 100)	0.1	One hundred.	0.10	0.358857
	”	(1000, 100)	0.1	One hundred.	0.50	0.359932

表 13:各模型在 FB15k-237 上的最佳超参数配置和相应的验证 MRR;为简单起见，我们将权重 λ 设置为 1。 D 代表嵌入大小， D_r 代表关系的单独嵌入大小， lr 是学习率， BSZ 是批量大小， reg 是正则化强度。

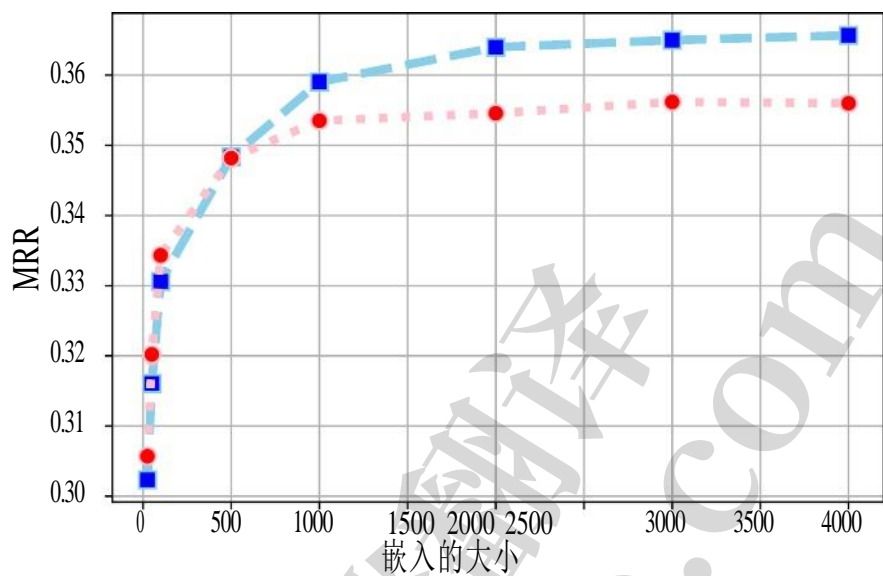


图 5:FB15k-237 上 CP 的 MRR 与 Rank 的对比。

附录 c 其他结果

C.1 更多关于 Rank 消融的指标

图 5 (MRR)、图 6 (Hits@3)和图 7 (Hits@10)显示了消融排名的实验的额外指标。蓝色表示有关系预测的训练，红色表示没有预测的训练。排名的范围是[25、50、100、500、1000、2000、3000、4000]

C.2 关系预测改进最多的前 20 个查询

表 14 显示了($?$, p , o)形式中关系预测改进最多的查询，而表 15 显示了(s , p , $?$)形式中关系预测改进最多的前 20 个查询。

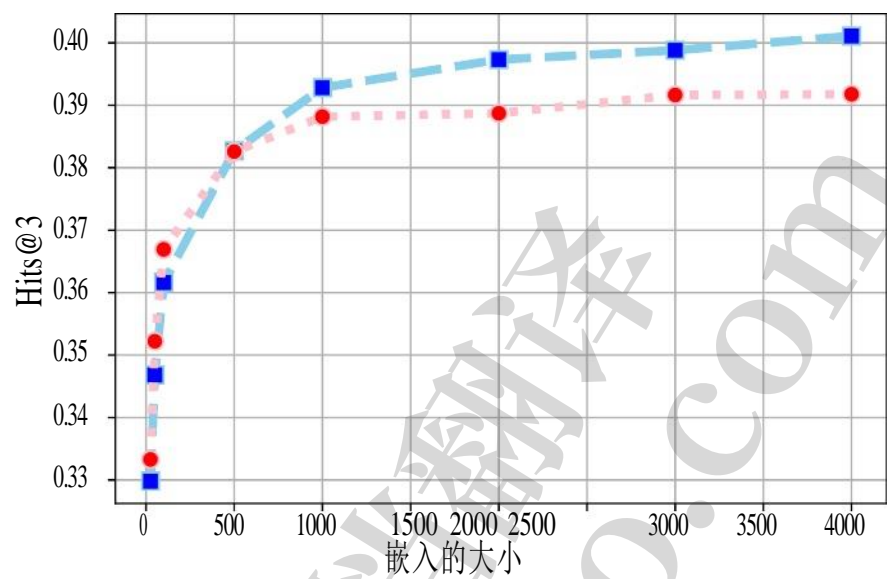


图 6:Hits @ 3 对 FB15k-237 上 CP 的排名。

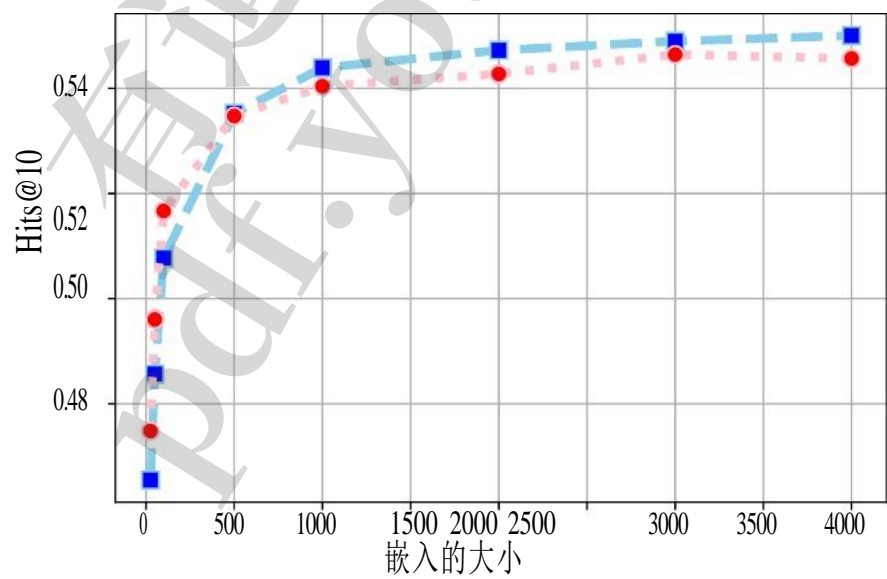


图 7:Hits @ 10 对 FB15k-237 上 CP 的排名。

主题	谓 词	对 象	Δ 1 /排 名
派拉蒙电影公司	/共同/主题/网页/共同/网页/类别	NA	1.000
中场	/运动/运动位置/球员/运动/运动队名单/球队	Gaziantepspor	0.988
独裁者(2012 年电影)	/电影/电影/个人形象/电影/个人电影形象/人	希拉里·克林顿(Hillary Clinton)	0.970
奥斯卡最佳影片	/award/奖项类别/获奖者/奖项荣誉/获奖者	莫林 Stapleton	0.963
女配角			
克里斯托弗·哥伦布	/user/tsegaran/随机/分类主题/条目	美国国会图书馆分类法 -	0.950
	/user/tsegaran/分类/随机/分类条目	阳离子	
联合王国	/位置/位置/包含	西敏寺	0.933
总统	/组织/角色/领导/组织/领导/组织	西弗吉尼亚大学	0.923
奥斯卡最佳影片奖	/award/奖项类别/获奖者/奖项荣誉/获奖者	克里斯托弗·沃肯	0.923
男配角			
总统	/组织/角色/领导/组织/领导/组织	布林莫尔学院	0.917
总统	/组织/角色/领导/组织/领导/组织	迪金森大学	0.917
全国电影学会	/award/award_category/winner /award/award_honor/award_winner	瑞茜·威瑟斯彭	0.917
影评人奖最佳女主角			
总统	/组织/角色/领导/组织/领导/组织	路易斯安那州立大学	0.917
维拉·德雷克	/award/获奖作品/获奖/奖项/奖项荣誉/奖项	洛杉矶影评人 As-	0.900
		协会奖最佳女主角	
总统	/组织/角色/领导/组织/领导/组织	奥克拉荷马大学	0.900
美国	/位置/国家/二级部门	马里恩县,印第安纳州	0.900
总统	/组织/角色/领导/组织/领导/组织	南加州大学	0.900
特拉维斯 Tritt	/电影/演员/电影/电影/性能/电影	2000 年蓝调兄弟	0.900
伦敦影评人圈	/award/award_category/winners /award/award_honor/award_winner	尼尔乔丹	0.900
年度最佳导演奖			

美国	/位置/国家/二级部门	纽约州尼亚加拉县	0.900
描述桑德斯	/人/人/住过的地方/人/住过的地方/地点	亚特兰大	0.900

表 14:前 20 个(s, p, o)测试三元组，基于它们右手边(即在给定 p 和 s 预测 o 的任务上)的倒数排名的增加，在我们引入关系预测辅助目标后。

主 题	谓 词	对 象	Δ 1 /排 名
中场	足球/足球/团队/当前名单	温布尔登足球俱 乐部	0.990
	/足球/足球球员位置/位置		
转发(英式足球)	/足球/足球队/现役阵容/运动/运动队阵容/位置	伊拉克国家足球队	0.989
美国	/电影/电影/发布日期	克利奥帕特拉(1963 片)	0.975
	/电影/电影区域上映日期/电影上映区域		
影评人选择电影奖最佳表演-	/奖项/奖提名/奖提名	马特•达蒙	0.975
荷兰国际集 团(ing)合奏	/奖项/奖提名/奖		
女	/人/人/性别	格雷迪莱尔	0.968
联合王国	/电影/电影/发布日期	《跨年》(2011 年电影)	0.967
	/电影/电影区域上映日期/电影上映区域		
美国美元	/位置/统计/ rent50 2	安克雷奇,阿拉斯 加	0.966
	/计量单位/日期货币价值/货币		
联合王国	/电影/电影/发布日期	轻轻地杀害他们	0.962
	/电影/电影区域上映日期/电影上映区域		
加州格兰岱尔市	/人/人/配偶/人/结婚/仪式的地点	简奥	0.962
联合王国	/电影/电影/发布日期	ParaNorman	0.962
	/电影/电影区域上映日期/电影上映区域		
流媒体	/film/film/分销商/film/film 电影发行商关系/电影发行媒介	《低俗小 说》	0.960
联合王国	/电影/电影/发布日期	魔力迈克	0.958
	/电影/电影区域上映日期/电影上映区域		
美国美元	/位置/统计/ rent50 2	加州纳帕县	0.952
	/计量单位/日期货币价值/货币		
联合王国	/电影/电影/发布日期	摇滚年代(2012 年电影版)	0.952
	/电影/电影区域上映日期/电影上映区域		
联合王国	/电影/电影/发布日期	接 触(1997 美国	0.950

	/电影/电影区域上映日期/电影上映区域	电 影)	
流媒体	/film/film/分销商/film/film 电影发行商关系/电影发行媒介	美国历史上 X	0.950
洛杉矶	/机构/组织/地方成立了	派拉蒙电影公司	0.947
联合王国	/电影/电影/发布日期	洛城机密》(电影)	0.941
	/电影/电影区域上映日期/电影上映区域		
心理惊悚片	/电影/电影/类 型	家庭阴谋	0.941
联合王国	/电影/电影/发布日期	《点球成金》 (电影)	0.938
	/电影/电影区域上映日期/电影上映区域		

21 表 15:前 20 个(s, p, o)测试三元组，基于它们的左手边(即在给定 p 和 o 预测 s 的任务上)的倒数排名的增加，在我们引入关系预测辅助目标后。