[文章编号] 1003－4684(2026)01-0076-08

基于增强性 Kronecker 分解的知识图谱补全模型

陈志康，陈建峡，张姝曦

（湖北工业大学计算机学院，湖北 武汉 430068）

[摘 要] Kronecker 分解法是一种基于矢量的分解方法，有助于减少知识图谱补全模型的参数数量和所需的硬件空间，但是会降低模型相应的精度。为解决这一问题，提出了一种新颖的数据增强性 Kronecker 分解的知识图谱补全模型，通过数据增强方法提高数据的质量，不仅能够提高模型的精度，也能在一定程度上缓解数据的稀缺性。经过实验分析，该模型在保证准确度仍然优于许多先进的模型的同时，能够实现 2%的参数减少。同时，该模型可以广泛用于基于嵌入的知识图谱补全模型，具有良好的实际应用价值。

[关键词] Kronecker 分解；知识图谱；知识图谱嵌入；知识图谱补全；数据增强

[中图分类号] TP183 [文献标识码] A

作为知识库的一种主要表达形式，知识图谱(knowledge graph，KG)以图的形式存储现实世界的知识[1]，也正成为许多自然语言处理(natural language processing，NLP)任务的基石[2]。为了提高知识图谱的表示范围，研究人员使用了自动知识提取技术或依靠协作编辑来生成 KG[3]，然而这些方法仍然难以满足现实世界中不断更新的知识。如何自动更新已有的知识图谱中的现有知识，是知识图谱补全(knowledge graph complement，KGC)研究的目标。

KGC 研究方法主要包含了张量分解法、几何方法和深度学习法等三种方法。其中，张量分解方法假设一个三元组的分数可以分解成几个张量，这种方法可以捕捉实体之间的多重关系，比如实体之间的时间、空间和属性等[4-5]；几何方法将关系视为头实体到尾实体的转换[6-7]，这种方法适用于处理各种类型的知识图谱，包括有向图、无向图和超图等；深度学习方法使用深度神经网络来执行 KG 的嵌入，即知识图谱嵌入(knowledge graph embedding，KGE)[6-10]能够实现 KG 中实体和关系的分布式表现，是目前 KGC 主要研究方法之一。这种方法通常可以处理大规模的知识图谱，并且具有高效的推理能力，但需要大量的数据进行训练，而知识图谱补全任务的数据通常是有限的[11-12]。

KGE 将实体和关系转化为连续空间中的向量，不仅可以促进模型参数的计算，而且保留了知识图谱的结构信息。为了提高 KGE 模型的表现力，增加嵌入向量的大小是常用的一种方法[13]。然而，KGE 模型参数的数量增长与嵌入向量的增长呈正相关，因此大多数 KGE 模型需要在训练的参数和三元组的数量上尽量保持线性的复杂度[14]，这不仅使 KGE 模型更容易过度拟合，而且提高了硬件的要求[15]，对 KGE 模型的时间和空间复杂度都极具挑战[16]。为此，许多研究都致力于优化超参数，作为缓解上述问题的一种方式[17-18]。

与此同时，有一些研究工作通过分解预训练模型减少高内存的使用[19-23]。但是，目前对于 KGE 模型进行向量分解方法研究的工作却极为有限。Kronecker 分解(kronecker decomposition，KD)作为一种张量分解方法，用于将高维张量分解成若干低维张量的乘积形式[16]。KD 在嵌入矩阵上的应用隐含地减少了嵌入向量中的冗余，并鼓励特征重用。这种分解技术本质上减少了明确存储的参数数量，从而减少了分解模型的内存占用。因此，本文提出了一种基于 KD 的技术，解决了知识图谱嵌入模型在训练中的参数冗余问题，可以有效减少模型的内存占用，同时提高了模型的稳健性。

在知识图谱补全的研究中，数据稀疏是 KGC 必须克服的另一个重要问题[24]。这是因为在现有的 KG 数据库中，只有一小部分包含真实的事实，可以作为正面例子用于指导训练，其他的事实都至少存在一个缺失值或者零值[24]。例如，WN18RR[11]中只有 0.00047%的真实事实，FB15k-237[25]中只有 0.00054%。这种极其稀少的正向训练数据可能会

[收稿日期] 2023－05－19

[第一作者] 陈志康(1999－)，男，河南开封人，湖北工业大学硕士研究生，研究方向为知识图谱、智能问答。

导致正确推断缺失三元组的次优泛化性能[24]。因此,解决数据稀疏性问题将是提高 KGE 模型准确性的关键方法之一。针对这个问题,PUDA[26]通过合成数据,一定程度上解决了知识图谱实体和关系的缺失问题;JOIE[27]通过联合嵌入实例,用于解决同样的问题。另外,对抗性训练也很好地缓解数据稀疏性问题,提高 KGC 模型的预测精度。例如,PUDA 模型就用到了这方面的原理。

综上所述,本文提出了一种新颖的数据增强性 Kronecker 分解的知识图谱补全模型(kronecker decomposition with adversarial data augmentation, KDADA)。其中,利用 KD 分解方法,减少模型训练的参数数量;同时通过对抗性数据增强方法提高数据的质量,在一定程度上缓解数据的稀缺性,最终能有效提高模型的精度。

1)采用了 KD 分解方法,减少在三元组实体和关系中的参数数量,降低了模型训练的时间和空间复杂度。

2)设计了一种数据增强策略,以减轻对抗性训练下 KD 导致的模型精度下降。

3)实验证明了所提出的 KDADA 方法的有效性和兼容性,能够与许多 KGE 模型结合使用。

1　相关技术原理

1.1　知识图谱嵌入

KGE 模型的性能好坏直接影响到后续应用模型的性能。近年来 KGE 的研究倾向于得到图的可扩展向量表示,其中有基于因式分解的方法,随机游走法以及深度学习的方法。

基于因式分解的方法是通过获得基于向量的乘法相互作用来学习实体和关系的嵌入。例如,RESCAL 模型[28]使用三向分解方法表示 KG 的三阶二元张量。然而,它在分解等级中具有二次复杂性[29,30],导致方法的可扩展性受到限制。Yang 等[14]提出的 Distmult 模型使用对角矩阵而不是紧凑张量,它在处理具有对称关系的三元组方面表现出色,但对于反对称关系却难以获得较好的表现。为了解决这一缺陷,Trouillon 等[10]通过 ComplEx 模型学习复数向量空间的表征,嵌入输入向量的共轭转置来推断对称和反对称关系。在 ComplEx 的激励下,RotatE 模型[6]被设计为在复数空间中旋转被建模为三元组的关系,这种方法能更好地表达三元组的复杂关系。另一方面,QuatE 模型[18]通过应用四元数乘法和内积乘法将互补性扩展到四元数上。

基于随机游走的方法,最近的研究包括 DeepWalk[31],node2vec[32],DeepWalk 模型受到词嵌入的启发,与 word2vec 的计算方式相同,只是需要通过随机游走的方式构造出节点序列。DeepWalk 由于其简单易用性,得到了广泛的应用,并且能取得一定的效果,但是对于有权图来说,DeepWalk 的方法就会失效。Node2vec 在 DeepWalk 模型的基础上,在有权图的处理上提出了概率游走的方法,实现了更全面的特征挖掘。

随着深度学习技术的不断发展,基于深度学习的方法已经成了目前最流行的得到图的特征嵌入的方法,而这主要是受益于深度学习强大的编码能力。例如 ConvE[13]和 ConvKB[33]等模型,这类方法能够捕捉实体和关系之间的复杂交互关系,具有较强的表达能力,但计算复杂度较高,对大规模知识图谱的处理可能存在挑战[33-34]。

虽然比上面这些方法在预测精度等方面表现好,但增加嵌入向量的大小仍然是提高其表现力的最有效方法。例如,Dettmers 等[13]、Nguyen 等[33]和 Balazevic 等[35]已经表明,通过应用卷积操作,同时增加少量的参数,有可能提高 KGE 模型的可扩展性。ConvE 应用二维卷积操作来模拟实体和关系之间的互动[13]。还有 ConvKB[33]通过改进卷积操作来扩展 ConvE。同样地,HypER[28]通过应用特定于关系的一维卷积来扩展 ConvE。而 Demir 和 Ngomo[9]通过将二维卷积操作与 Hermitian 内积、四元数和八元数积相结合来扩展 ConvE。

本文选取 distmult 模型作为框架的评分函数,并以此进行框架的扩展,最终提出一个通用的 KGE 框架。

1.2　Kronecker 分解

Kronecker 分解的基本思想是将高维张量表示为一系列低维张量的乘积形式,其中每个低维张量对应原始张量的某个模态。KD 用于将一个大矩阵分解成多个小矩阵的乘积形式,其主要功能是对大规模矩阵进行降维和压缩,从而提高计算效率和存储效率[16]。KD 的定义详细介绍如下:

对任何矩阵 $\boldsymbol{A}\in\mathbb{C}^{m\times n}$,$\boldsymbol{B}\in\mathbb{C}^{m\times n}$,Hadamard 乘积 $\boldsymbol{A}\circ\boldsymbol{B}$ 被定义为

$$\begin{bmatrix}\boldsymbol{A}_{11}\boldsymbol{B}_{11} & \cdots & \boldsymbol{A}_{1n}\boldsymbol{B}_{1n}\\ \vdots & & \vdots\\ \boldsymbol{A}_{m1}\boldsymbol{B}_{m1} & \cdots & \boldsymbol{A}_{mn}\boldsymbol{B}_{mn}\end{bmatrix}\in\mathbb{C}^{m\times n} \tag{1}$$

其中,$\boldsymbol{A}_{ij}$ 仅与 $\boldsymbol{B}_{ij}$ 交互,对任何矩阵 $\boldsymbol{C}\in\mathbb{C}^{p\times q}$,Kronecker 积 $\boldsymbol{A}\otimes\boldsymbol{C}$ 是一个方块矩阵:

$$\begin{bmatrix}\boldsymbol{A}_{11}\boldsymbol{C} & \cdots & \boldsymbol{A}_{1n}\boldsymbol{C}\\ \vdots & & \vdots\\ \boldsymbol{A}_{m1}\boldsymbol{C} & \cdots & \boldsymbol{A}_{mn}\boldsymbol{C}\end{bmatrix}\in\mathbb{C}^{mp\times nq} \tag{2}$$

其中，$\boldsymbol{A}$ 的每个元素都与 $\boldsymbol{C}$ 的每个元素相互作用，与 Hadamard 积相比，Kronecker 积不遵守交换律，如 $\boldsymbol{A}\otimes\boldsymbol{C}\neq\boldsymbol{C}\otimes\boldsymbol{A}$。大量研究表明，公式(2)中定义的 Kronecker 积可以有效地将一个大的矩阵分解为两个较小的矩阵[38]。

在 KD 中，权值矩阵 $W\in\mathbb{C}^{pm\times qn}$ 可以分解为任意矩阵 $\boldsymbol{A}\in\mathbb{C}^{m1\times n1}$ 和 $\boldsymbol{C}\in\mathbb{C}^{\frac{pm}{m1}\times\frac{qn}{n1}}$。不同的压缩因素可以通过不同的较小矩阵的形状配置来实现。Tahaei 等人[38]已经表明，公式(3)可以有效地计算通过 KD 编码在权重矩阵 W 中的线性变换：

$$(\boldsymbol{A}\otimes\boldsymbol{B})x=\Gamma(\boldsymbol{C}\mathrm{H}_{\frac{nq}{n1}\times n1(x)}X^{\mathrm{T}}) \tag{3}$$

其中，$\boldsymbol{x}\in\mathbb{C}^{nq}$ 表示输入特征向量，Γ 表示将一个输入矩阵平铺为一个向量，$\mathrm{H}_{\frac{nq}{n1}\times n1}$ 将 x 转换为 $\frac{nq}{n_1}$ 乘 n_1 的矩阵，方法是将向量划分为大小为 $\frac{nq}{n_1}$ 的列，并将得到的列串联起来[38]，减少了执行 KD 所需的浮点运算的数量。

1.3 生成对抗网络

生成式对抗网(generative adversarial nets，GAN)被作为一种训练生成式模型的新方法引入[36]。它们由两个 "对抗性 "模型组成[37]：一个是捕捉数据分布的生成模型(generator，G)，另一个是估计样本来自训练数据而不是 G 的概率的判别模型(discriminator，D)，如：

$$\min_G\max_D V(D,G)=E_{x\sim p_{\mathrm{data}}(x)}[\log D(x)]+E_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{4}$$

其中，E 表示数学期望，p_{data} 表示生成器的分布。为了在数据 x 上学习生成器分布 p_{data}，生成器建立了一个从先验噪声分布 $h=p_z(z)$ 到数据空间的映射函数为 $G(z;\theta g)$。而鉴别器 $D(x;\theta d)$ 则输出一个标量，代表 x 来自训练数据而不是 p_{data} 的概率。G 和 D 都是同时训练的：调整 G 的参数以最小化 $\log(1-D(G(z)))$，调整 D 的参数以最小化 $\log D(X)$，使生成器 G 和鉴别器 D 遵循价值函数 $V(G,D)$ 的极大极小博弈，GAN 模型如图 1 所示。

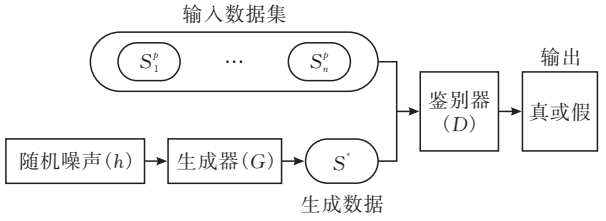

图 1 GAN 模型

2 KDADA 模型

如图 2 所示，KDADA 模型框架包括两个主要部分。一个是使用 KD 来减少参数的数量，另一个是针对数据稀少问题的对抗性数据增强，同时也提高了模型的鲁棒性。最后通过对抗性训练获得一个新的正三元组，以达到数据增强的目的。

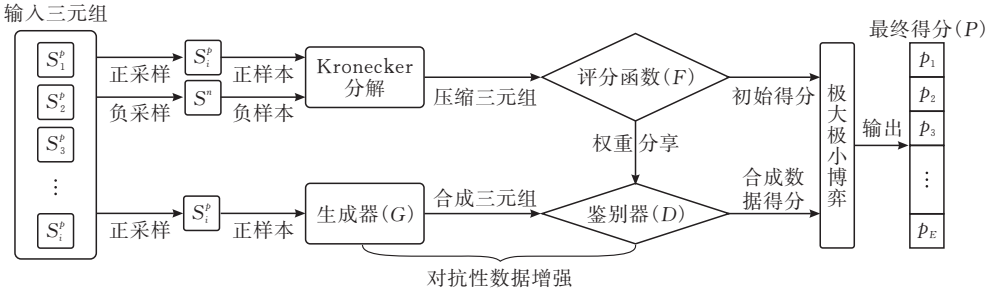

图 2 KDADA 框架

2.1 基于 Kronecker 分解的参数优化

2.1.1 知识图谱嵌入中的 Kronecker 分解 KDADA 模型参数化评分函数的计算如公式(5)所示：

$$\varphi_{(\Theta)}:\{\varepsilon\}\times\{R\}\times\{\varepsilon\}\rightarrow\mathbb{C} \tag{5}$$

该函数将输入的三元组 $(h,r,t)\in\{K\}$ 映射为一个标量值，并使用 sigmoid/logistic 函数将该标量值映射到单位区间，归一化的标量值表示三元组 (h,r,t) 为真的可能性。Θ 是 φ 的参数，由一个实体嵌入矩阵 $\boldsymbol{E}\in\mathbb{C}^{|\varepsilon|\times d}$、一个关系嵌入矩阵 $\boldsymbol{R}=\mathbb{C}^{|R|\times d}$，和其他可训练参数组成，例如对输入嵌入的仿射转换、卷积、批处理规范化或实例规范化。

假设 $\varphi=:distMult$，$\Theta:=\{E,R\}$，对于一个给定的 $(h,r,t)\in K$，该三元组的分数计算为：

$$\mathrm{Distmult}(h,r,t)=\boldsymbol{h}\circ\boldsymbol{r}\cdot\boldsymbol{t} \tag{6}$$

公式(6)表示三个 d 维实数值的嵌入向量之间的元素相互作用的结果。例如，给定 $d=2$，$x_1,x_2,y_1,y_2,z_1,z_2\in\mathbb{C}$，则该三元组的分数为：

$$[x_1\quad x_2]\circ[y_1\quad y_2]\cdot[z_1\quad z_2]=x_1y_1z_1+x_2y_2z_2 \tag{7}$$

式中，尽管 x_1 和 x_2 共同构成了 h，但 (x_1,x_2) 之间以及 (x_1,x_1) 之间的相互作用被忽略了。对与元素级之间的操作，这些被忽略的交互可以通过 KD 分解对嵌入向量来弥补，如下所示：

$$\text{KD-Distmult}(h,r,t)=[x_1x_1\quad x_1x_2\quad x_1x_2\quad x_2x_2]\circ$$

$$[y_1y_1 \quad y_1y_2 \quad y_1y_2 \quad y_2y_2]\cdot[z_1z_1 \quad z_1z_2 \quad z_1z_2 \quad z_2z_2]=x_1x_1y_1y_1z_1z_1+2(x_1x_2y_1y_2z_1z_2)+x_2x_2y_2y_2z_2z_2 \tag{8}$$

通过方程(8)的中项,合并了 (x_1,x_1) 和 (x_1,x_2) 之间的相互作用,而不需要额外的参数。因此,本论文采用了 KD-Distmult 关系嵌入。

2.1.2　关系嵌入中的 Kronecker 分解　关于 DistMult 关系嵌入的 KD 应用如公式(9)所示:

$$\text{KD-Rel-Distmult}(h,r,t)=\boldsymbol{h}\circ(\boldsymbol{r}\otimes\boldsymbol{r})\cdot\boldsymbol{t} \tag{9}$$

其中,$h,t\in E:E\in\mathbb{C}^d$,$r\in R:R\in\mathbb{C}^{\sqrt{d}}$,本文使用实体和关系嵌入大小的改变近似代表参数的收益,并最终在消融实验中呈现。

值得注意的是,三元组的分数仍然是基于三维嵌入向量计算的。且随着输入知识图谱的大小和关系的数量的增加,参数的增益变得更加可观。

2.2　基于对抗数据增强的数据稀疏处理

2.2.1　数据扩充处理　KGE 模型训练最常见的数据类型是正三元组和负三元组,但除了正负三元组外,本文还生成了合成的三元组用于数据扩充。具体来说,对于一个正的三元组 $s_i^p=\langle h_i^p,r_i^p,t_i^p\rangle$,通过对抗性训练生成 M 个合成实体 $\{t_{i1}^*,\ldots,t_{iM}^*\}$,并用每个合成实体 t_{iM}^* 替换 t_i^p,从而生成 M 个合成三元组 $S_i^*=\{h_i^p,r_i^p,t_{im}^*\}^M$,其中 $t_{im}^*\notin\varepsilon$。同理,也可以用同样的方式替换头部实体。

与给定的正三元组不同,一个合成的三元组 $s_{im}^*\in S_i^*$ 使用多层感知器(MLP)给定一个随机噪声 h_{im} 生成:

$$\begin{aligned}G(h;\Omega)&=\tanh(\boldsymbol{W}_2\cdot\mathrm{ReLU}(\boldsymbol{W}_1\cdot h+\boldsymbol{b}_1)+\boldsymbol{b}_2)\\h_{im}&\backsim N(V,\delta\boldsymbol{I})\\t_{im}^*&=G(h_{im},\Omega)\\s_{im}&=\langle h_i^p,r_i^p,t_{im}^*\rangle\end{aligned} \tag{10}$$

在公式(10)中,V 是噪声输入的平均值,其大小与嵌入维度 d 相同,$I\in\mathbb{C}^{d\times d}$ 是由噪声输入 δ 的偏差控制的单位矩阵。$\Omega=\{W_1,W_2,b_1,b_1\}$ 包括两层 MLP 的所有可学习的权重和偏置。

2.2.2　数据增强处理　给定一个生成的合成三元组 $s_{im}^*\in S_i^*$,鉴别器 D 试图预测 s_{im}^* 的合理性,以区分它是真实的还是合成的。因为评分函数 $\varphi(s;\Theta)$ 预测了每个三元组的合理性。因此,将三元组评分函数定义为鉴别器:

$$G(s;\Theta)=\varphi(s;\Theta) \tag{11}$$

本文训练 G 和 D 的对抗目标可以设计为真实正三元组和合成三元组进行极大极小博弈,如:

$$\min_D\max_G:\frac{1}{|K|}\sum_{i=1}^{|K|}\overbrace{-\frac{1}{M}\sum_{m=1}^{M}\ln\sigma\underbrace{(\varphi(s_i^p;\Theta)-\varphi(s_{ij}^*;\Theta))}_{\text{Pairwise Ordering}}}^{\text{Sorting}} \tag{12}$$

通过使用一个固定的 G,通过最小化公式(12),D 被训练优化给定的正三元组 s_i^p 和生成的合成三元组 s_{ij}^* 的相对排序,使 s_i^p 排名高于 s_{ij}^*。相反,在固定的 D 条件下,G 通过最大化等式进行训练欺骗 D,使 s_{ij}^* 排名高于或接近 s_i^p,最后将得到结果进行排序,从而得到最接近真实数据的合成三元组。

由于正三元组只占所有可能三元组的一小部分,本文遵循对抗性训练的原则来生成合成三元组 S_i^* 用于数据增强,而不是只使用真实的三元组:$\{\langle h,r,t\rangle\mid h\in\varepsilon\wedge r\in R\wedge t\in\varepsilon\}$。

与普通的对抗性训练方法不同[30],本文的最终目标是在训练后使用生成器,即利用经过训练的鉴别器来预测缺失三元组的合理性。

3　实验结果与分析

本文的实验环境为 Pytorch 1.8.1,Python 3.8,CUDA 版本为 11.1。

3.1　数据集

本文使用了 WN18RR[11] 和 FB15k-237[25] 两个数据集进行实验,具体统计数据如表 1 所示。

1)WN18RR:是基于 WordNet 语料库构建的知识图谱补全数据集,包含 10304 个实体和 11 个关系。

2)FB15k-237:是基于 Freebase 语料库构建的大规模知识图谱补全数据集,包含 14541 个实体和 237 个关系。

表 1　数据集的特征

数据集	$\lvert\varepsilon\rvert$	$\lvert R\rvert$	Train	Valid	Test
FB15K-237	14,541	237	272,115	17,535	20,466
WN18RR	40,943	11	86,835	3,034	3,134

其中,$\lvert\varepsilon\rvert$ 表示实体数量,$\lvert R\rvert$ 表示关系的数量,Train、Valid 和 Test 分别代表训练集、验证集和测试集的三元组数量。

3.2　基线

本文将 KDADA 与三类方法进行比较,如下所述。

3.2.1　基于张量分解的方法

1)DistMult[14]:采用对角矩阵表示实体和关系之间的交互。

2)ComplEx[4]:将实体和关系表示为复数向量,通过欧几里得点积计算实体和关系的交互。

3)SimplE[5]:通过简化交叉项来减少模型参数,提高计算效率。

3.2.2　基于几何的方法

1)TransE[8]:通过将实体和关系表示为向量,并计算他们之间的距离。

2)CrossE[7]：通过将实体和关系表示为向量，并通过叉积操作计算实体和关系之间的交互。

3.2.3 基于深度学习的方法 ConvE[13]：使用卷积神经网络对实体和关系进行嵌入学习。

3.2.4 基于负抽样的方法

1)KBGAN[40]：使用生成对抗网络对负例进行采样，并通过模型的判别器对正例和负例进行区分。

2)NSCaching[39]：使用负采样技术对负例进行采样，并通过缓存机制提高计算效率。

3.3 实验设置

本文提出的KGC模型通过实体排序任务进行评估：对于每个测试三元组 (h,r,t)，尾部实体预测对所有实体进行排序，给定 h 和 r，头部实体预测类似。使用四个自动评估指标：平均倒数秩(mean reciprocal rank，MRR)和 Hits@k($k\in\{1,3,10\}$)(简称 H@k)。

MRR是一种常用的衡量知识图谱补全算法性能的指标，如：

$$\mathrm{MRR}=\frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{\mathrm{rank}_i} \tag{13}$$

其中：$|Q|$ 表示查询集合中查询的数量，rank_i 表示第 i 个查询的正确答案在所有答案中的排名。MRR的取值范围为[0,1]，值越大表示算法性能越好。

Hits@k 是一种基于排名的评价指标，表示前 k 个推荐结果中有多少个与正确答案匹配。Hits@k 的具体表示如公式(14)所示：

$$\mathrm{Hits@}k=\frac{1}{|Q|}\sum_{i=1}^{|Q|}[\mathrm{rank}_i\leqslant k] \tag{14}$$

其中，$[\mathrm{rank}_i\leqslant k]$ 表示一个指示函数，如果第 i 个查询的正确答案在前 k 个推荐结果中，该指示函数的值为1，否则为0。Hits@k 的取值范围为[0,1]，值越大表示算法性能越好。

FB15K-237和WN18RR两个数据集的具体超参数设置如表2所示。

表2 超参数设置

超参数	描述	FB15K-237	WN18RR
num_ng	负样本与训练样本之比	8	8
num_ng_gen	生成样本与训练样本之比	8	8
emb_dim	嵌入维度	1024	1024
$\mathrm{lr_d}$	鉴别器的学习率	1×10^{-5}	3×10^{-6}
$\mathrm{lr_g}$	生成器的学习率	1×10^{-5}	1×10^{-6}
seed	随机种子的大小	42	42
max_epochs	训练迭代次数	3000	4000

3.4 实验结果

本文将KDADA和多种类型的基线模型进行对比。因为KDADA是在Distmult模型的基础上进行改进，所以本文首先将KDADA和其做比较，结果表明KDADA在所有的评分指标上都优于Distmult。这个表现主要归功于数据增强带来的链路预测任务精度的提升。与大多数的基线模型相比，KDADA在各方面也都有较大的提升。实验结果如表3所示。

表3 模型比较实验结果

方法	模型	WN18RR				FB15k-237			
		MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
Geometric	TransE	0.206	0.279	—	0.495	0.310	0.217	—	0.497
	CrossE	0.405	0.381	—	0.450	0.298	0.212	—	0.471
Tensor Decomposition	Distmult	0.433	0.397	—	0.502	0.313	0.224	—	0.490
	ComplEx	0.440	0.410	0.460	0.510	0.247	0.158	0.275	0.428
	SimplE	0.398	0.383	—	0.427	0.179	0.100	—	0.344
Deep Learning	ConvE	0.430	0.400	0.440	0.520	0.244	0.237	0.356	0.501
Negative Sampling	KBGAN	0.214	—	—	0.472	0.278	—	—	0.458
	NSCaching	0.443	—	—	0.518	0.302	—	—	0.481
KDADA		**0.471**	**0.433**	**0.491**	**0.542**	**0.391**	**0.295**	**0.433**	**0.585**

粗体数字为最好结果。

3.5 消融实验

本文进行了消融实验，以证明KDADA对于模型过度参数化以及KD带来的精度下降问题有效性。在表4中，本文使用了标准Distmult模型的结果[13]，KD表示将KD应用于标准Distmult模型的关系嵌入，Emb表示实体和关系的嵌入大小(KB)。

本文可以从表 4 的结果中得出结论，简单地将 KD 应用于 KGE 模型的关系嵌入会降低链接预测结果，而本文的 KDADA 框架对链接预测性能的优化是明显的，显示了数据增强的有效性。从表 4 的结果可以看出，只对关系嵌入应用 KD 可以明显减少参数的数量，最多可减少 2%的参数。

表 4　链路预测消融研究

模型	WN18RR					FB15k-237				
	MRR	H@1	H@3	H@10	Emb	MRR	H@1	H@3	H@10	Emb_
Distmult	0.433	0.397	—	0.502	168886.7	0.313	0.224	—	0.490	59439.6
KD	0.350	0.273	0.396	0.483	**167898.1**	0.277	0.182	0.309	0.471	58575.1
KDADA	**0.471**	**0.433**	**0.491**	**0.542**	**167898.1**	**0.391**	**0.295**	**0.433**	**0.585**	**58575.1**

4　结束语

本文主要针对 KGC 模型的参数数量和精度问题进行模型的设计和调整，这里提出的 KDADA 模型可以很好解决这些问题。本文首先使用 KD 对 KG 嵌入进行分解，从而达到在保留表达性的同时尽可能减少参数的目的，然后再使用一个精简的对抗网络进行数据增强，在少量增减模型大小的同时提高链路预测的结果，经过反复实验，本文证明了 KDADA 模型的有效性。

[参 考 文 献]

[1] 王硕，杜志娟，孟小峰. 大规模知识图谱补全技术的研究进展[J]. 中国科学:信息科学，2020，50(04):25.

[2] 丁建辉，贾维嘉，Ding，等. 知识图谱补全算法综述[J]. 信息通信技术，2018，12(01):7.

[3] 王瑞，李智杰，李昌华，等. 面向链接预测的知识图谱嵌入研究综述[J]. 计算机测量与控制，2022(09):30.

[4] TROUILLON T，DANCE C R，WELBL J，et al. Knowledge graph completion via complex tensor factorization[J]. Journal of Machine Learning Research，2017，18(01)：4735-4776.

[5] KAZEMI S M，POOLE D. Simple embedding for link prediction in knowledge graphs[J]. Advances in neural information processing systems，2018，31：4289-4300.

[6] SUN Z，DENG Z H，NIE J Y，et al. Rotate：Knowledge graph embedding by relational rotation in complex space[C]. 7th International Conference on Learning Representations，2019.

[7] ZHANG W，PAUDEL B，ZHANG W，et al. Interaction embeddings for prediction and explanation in knowledge graphs [C]. Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining，2019：96-104.

[8] BORDES A，USUNIER N，GARCIA-DURAN A，et al. Translating embeddings for modeling multi-relational data[C]. Advances in Neural Information Processing Systems 26，2013：2787-2795.

[9] DEMIR C，NGOMO A C N. Convolutional complex knowledge graph embeddings [C] // The Semantic Web：18th International Conference，ESWC 2021，Virtual Event，June 6-10，2021，Proceedings 18. Springer International Publishing，2021：409-424.

[10] TROUILLON T，WELBL J，RIEDEL S，et al. Complex embeddings for simple link prediction[C].International conference on machine learning. PMLR，2016：2071-2080.

[11] VASHISHTH S，SANYAL S，NITIN V，et al. Composition-based multi-relational graph convolutional networks[C]. 8th International Conference on Learning Representations，2020.

[12] NGUYEN D Q，NGUYEN T D，NGUYEN D Q，et al. A novel embedding model for knowledge base completion based on convolutional neural network [C]. Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies，2018：327-333.

[13] DETTMERS T，MINERVINI P，STENETORP P，et al. Convolutional 2d knowledge graph embeddings[C]. Proceedings of the AAAI Conference on Artificial Intelligence，2018.

[14] YANG B，YIH W，HE X，et al. Embedding entities and relations for learning and inference in knowledge bases[C]. 3rd International Conference on Learning Representations，2015.

[15] HOOKER S. The hardware lottery[J]. Communications of the ACM，2021，64(12)：58-65.

[16] EDALATI A，TAHAEI M，RASHID A，et al. Kronecker decomposition for gpt compression[C]. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics，2022：219-226.

[17] RUFFINELLI D，BROSCHEIT S，GEMULLA R. You can teach an old dog new tricks! on training knowledge graph embeddings [C]. 8th International Conference on Learning Representations，2020.

[18] ZHANG S，TAY Y，YAO L，et al. Quaternion knowledge graph embeddings[C]. Advances in Neural Information Processing Systems 32，2019：2731-2741.

[19] SANH V，DEBUT L，CHAUMOND J，et al. Distil-

BERT, a distilled version of BERT: smaller, faster, cheaper and lighter[J]. arXiv preprint arXiv: 1910.01108, 2019,2019:1-5.

[20] JIAO X, YIN Y, SHANG L, et al. Tinybert: Distilling bert for natural language understanding[C]. Findings of the Association for Computational Linguistics: EMNLP 2020, 2020: 4163-4174.

[21] SUN Z, YU H, SONG X, et al. Mobilebert: a compact task-agnostic bert for resource-limited devices[J]. arXiv preprint arXiv:2004.02984, 2020:1-9.

[22] PASSBAN P, WU Y, REZAGHOLIZADEH M, et al. Alp-kd: Attention-based layer projection for knowledge distillation[C].Proceedings of the AAAI Conference on artificial intelligence, 2021: 13657-13665.

[23] RASHID A, LIOUTAS V, REZAGHOLIZADEH M. MATE-KD: Masked adversarial TExt, a companion to knowledge distillation[C]. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 2021: 1062-1071.

[24] PUJARA J, AUGUSTINE E, GETOOR L. Sparsity and noise: Where knowledge graph embeddings fall short[C].Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017: 1751-1756.

[25] TOUTANOVA K, CHEN D, PANTEL P, et al. Representing text for joint embedding of text and knowledge bases[C].Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 1499-1509.

[26] TANG Z, PEI S, ZHANG Z, et al. Positive-unlabeled learning with adversarial data augmentation for knowledge graph completion[C].Proceedings of the 31st International Joint Conference on Artificial Intelligence, 2022: 2248-2254.

[27] HAO J, CHEN M, YU W, et al. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts[C]. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019: 1709-1719.

[28] NICKEL M, TRESP V, KRIEGEL H P. A three-way model for collective learning on multi-relational data [C]. International Conference on Machine Learning, 2011: 3104482-3104584.

[29] NiCKEL M, ROSASCO L, POGGIO T. Holographic embeddings of knowledge graphs[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2016.

[30] 肖亚新. 面向知识图谱补全的嵌入学习研究及应用[D].镇江:江苏科技大学,2022.

[31] PEROZZI B, AL-RFOU R, SKIENA S. Deepwalk: Online learning of social representations[C].Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2014: 701-710.

[32] GROVER A, LESKOVEC J. node2vec: Scalable feature learning for networks[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 855-864.

[33] NICKEL M, MURPHY K, TRESP V, et al. A review of relational machine learning for knowledge graphs [J]. Proceedings of the IEEE, 2015, 104(01): 11-33.

[34] CAI H, ZHENG V W, CHANG K C C. A comprehensive survey of graph embedding: Problems, techniques, and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(09): 1616-1637.

[35] BALAžEVI I, ALLEN C, HOSPEDALES T M. Hypernetwork knowledge graph embeddings[C]// Artificial Neural Networks and Machine Learning-ICANN 2019: Workshop and Special Sessions: 28th International Conference on Artificial Neural Networks, Munich, Germany, September 17-19, 2019, Proceedings 28. Springer International Publishing, 2019: 553-565.

[36] MARIANI G, SCHEIDEGGER F, ISTRATE R, et al. Bagan: Data augmentation with balancing gan[J]. arXiv preprint arXiv:1803.09655, 2018:1-9.

[37] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.

[38] TAHAEI M S, CHARLAIX E, NIA V P, et al. Kroneckerbert: Learning kronecker decomposition for pretrained language models via knowledge distillation[J]. arXiv preprint arXiv:2109.06243, 2021:1-10.

[39] CAI L, WANG W Y. Kbgan: Adversarial learning for knowledge graph embeddings[J]. arXiv preprint arXiv:1711.04071, 2017:1-11.

[40] ZHANG Y, YAO Q, SHAO Y, et al. NSCaching: simple and efficient negative sampling for knowledge graph embedding[C]. 2019 IEEE 35th International Conference on Data Engineering (ICDE). IEEE, 2019: 614-625.

A Knowledge Graph Complementation Model Based on Augmented Kronecker Decomposition

CHEN Zhikang,CHEN Jianxia, ZHANG Shuxi

(*School of Computer Science, Hubei Univ. of Tech., Wuhan*, 430000, *China*)

Abstract: Kronecker decomposition is a vector-based decomposition method that can reduce the number of parameters and the required hardware space of the knowledge graph complementation model; however, it will reduce the corresponding accuracy of the model. Therefore, this paper proposes a novel knowledge graph complementation model with data-enhanced Kronecker decomposition, which can not only improve the accuracy of the model but also alleviate the data scarcity to a certain extent by improving the quality of the data through data enhancement methods. After experimental analysis, the model can achieve up to 2% parameter reduction and its accuracy is still better than many advanced models. At the same time, the model can be widely used in embedding-based knowledge graph complementation models, which has good practical.

Keywords: kronecker decomposition; knowledge graph; knowledge graph embedding; knowledge graph complementation; data augmentation

[责任编校: 闫　品]

(上接第 75 页)

Point Cloud Registration Method Based on Second-order Spatial Compatibility Metric

SHEN Zhen, JIN Huazhong

(*School of Computer Science, Hubei Univ. of Tech., Wuhan* 430068, *China*)

Abstract: A point cloud registration method based on the second-order spatial compatibility metric is proposed to address the problems of large computational effort, long time consuming and difficult to eliminate the wrong matching point pairs in the feature-based point cloud registration algorithm. The method adopts two stages: coarse registration and fine registration. In the coarse registration stage, the Harris-3D algorithm is used to obtain the key points of each point cloud and calculate its fast point feature histogram (FPFH) descriptors to complete the matching between the key points; then, the second-order spatial compatibility matrix of the matched point pairs is calculated to reject the wrong matching pairs, and the preliminary transformation matrix is obtained by combining the singular value decomposition method. In the fine matching stage, the k-dimensional tree (kd-tree) accelerated ICP algorithm is used to calculate the final transformation matrix and complete the point cloud registration. The experimental results show that the proposed method is less time-consuming and more accurate than the commonly used feature-based point cloud registration algorithms, and can effectively suppress the effect of mis-matching on the registration accuracy.

Keywords: computer vision; point cloud registration; key point extraction; spatial compatibility; iterative nearest point

[责任编校: 闫　品]