基于三维复合几何变换的知识图谱嵌入

Xiou Ge、Yun-Cheng Wang，*IEEE* 学生会员，Bin Wang，*IEEE* 会员，
C.-C. Jay Kuo，*IEEE* 和 *ACM* 会士，

arXiv:2304.00378v1 [cs.AI] 1 Apr 2023

摘要——**本文利用二维几何变换级联来建模知识图谱（KG）中实体之间的关系，从而提出了一种有效的知识图谱嵌入（KGE）模型 CompoundE。此外，本文利用三维空间中旋转的非交换性质，提出了一种新的 KGE 模型 Rotate3D。受 CompoundE 和 Rotate3D 启发，本文采用包括平移、旋转、缩放、反射和错切在内的三维复合几何变换，并提出一系列名为 CompoundE3D 的 KGE 模型。CompoundE3D 提供多种设计变体，以适应知识图谱丰富的内在特性。由于每种变体在部分关系上各有优势，将多种变体集成可取得更优性能。本文在四个常用链接预测数据集上的实验验证了 CompoundE3D 的有效性和灵活性。**

索引词—**知识图谱嵌入、链接预测、几何变换。**

I. 引言

知NOWLEDGE识图谱（KG）在知识管理与发现、推荐系统 [4], [5], 欺诈检测、 [6], [7], 聊天机器人[8], [9], 等领域有着广泛应用。知识图谱是有向关系图，由一系列形式为（h, r, t）的三元组构成，其中h、 r和 t 分别表示头实体、关系和尾实体。头实体和尾实体称为实体，以节点表示；关系则是知识图谱中的链接。知识图谱往往并不完整。知识图谱管理中的一项关键任务是“缺失链接预测”。近年来，知识图谱嵌入（KGE）方法因其在缺失链接预测中的有效性而受到广泛关注。许多知识图谱（例如 DBpedia [10], 、YAGO [11],、Freebase [12], 、NELL [13],、Wikidata [14], 和 ConceptNet [15]）已被创建并公开提供，用于知识图谱嵌入模型的开发与评估。

一类知识图谱嵌入模型会构建一个高维嵌入空间，其中每个实体都是一个向量。关系通过某种几何变换来建模，例如平移和旋转。为评估候选三元组的可能性，先将与关系相关的几何变换应用于头实体，然后测量变换后的头实体与尾实体之间的距离。距离越短，三元组成立的可能性就越高。因此，这些知识图谱嵌入模型被称为基于距离的知识图谱嵌入。基于距离的知识图谱嵌入模型包括 TransE [16], RotatE [17],以及 PairRE [18]。它们各自使用单一的几何变换来表示实体之间的关系。具体而言，TransE、RotatE 和 PairRE 分别采用平移、旋转和缩放操作。

Xiou Ge、Yun-Cheng Wang 和 C.-C. Jay Kuo 任职于美国加利福尼亚州洛杉矶市 90089 邮编的南加州大学明锡电气与计算机工程系。Bin Wang 任职于新加坡国立大学，新加坡。

上述知识图谱嵌入模型仅通过一种几何变换，就能在链接预测中取得相当不错的性能。多个二维几何变换的级联为图像处理提供了强大的工具 [19]。在 [20] 中，研究者利用这一思路开发了一种名为 CompoundE 的新型知识图谱嵌入模型。TransE、RotatE 和 PairRE 都是 CompoundE 的退化情形。因此，CompoundE 在链接预测方面的性能优于这些模型。CompoundE 在统一框架下整合了平移、旋转和缩放操作。它具有若干可通过数学证明的性质，有助于对知识图谱中的不同复杂关系类型进行建模。在 [20] 中，通过大量实验以及在实体类型判定和多跳查询回答等下游任务中的应用，成功验证了这些复合算子的有效性。此外，借鉴三维空间中的旋转概念，Rotate3D [21]–[23] 实现了更有效的参数化；与基于二维旋转的 RotatE 相比，它赋予模型更强的建模能力。也就是说，Rotate3D 比 RotatE 更擅长建模非交换关系。

受 CompoundE 和 Rotate3D 成功的启发，我们思考，在知识图谱嵌入模型的设计中，探索三维空间中的复合几何变换是否有益。在此，我们沿着三个方向扩展 CompoundE 工作 [20]。首先，除平移、旋转和缩放之外，我们还纳入了更多仿射运算，例如反射和错切。其次，我们将这些几何变换从二维空间扩展到三维空间，并提出了一系列知识图谱嵌入模型 CompoundE3D。第三，CompoundE3D 支持多种设计变体，以匹配知识图谱丰富的潜在特征。由于每种变体在部分关系上各具优势，将多个变体进行集成可以带来更优的性能。我们在四个广泛使用的链接预测数据集上通过实验验证了 CompoundE3D 的有效性。

值得强调的是，我们通过解决两个关键问题来增强 CompoundE。首先，复合运算会产生许多模型变体，而如何确定在给定数据集上表现最佳的评分函数尚不明确。为此，我们提出一种改进的束搜索算法，逐步从简单而有效的评分函数构建出更复杂的评分函数。其次，尽管集成学习是一种广受欢迎的策略，但它在构建知识图谱嵌入模型方面仍未得到充分探索，

在这项工作中，我们探索了两种有望提升链接预测性能的集成策略，使不同的 CompoundE3D 变体能够协同工作、互相补充。首先，我们对不同评分函数进行加权求和，以用于链接预测。其次，我们应用无监督排序聚合函数，整合各个模型变体的排序预测。这两种策略都有助于提升有效候选实体的排名，并减轻异常值的影响。

本文的主要贡献总结如下。

- 我们考察三维空间中的仿射运算，而非二维空间中的仿射运算，以实现更加灵活多样的关系表示。除了 CompoundE 中使用的平移、旋转和缩放，我们还引入了反射和剪切变换，从而进一步拓展了设计空间。
- 我们提出了一种改进的束搜索算法，用于发现更优的模型变体。该方法避免对较差的变体进行不必要的探索，而是聚焦于更有效的变体，从而在模型复杂度与预测性能之间取得良好平衡。
- 我们分析了各项运算的性质及其在建模不同关系时的优势。我们的分析得到了四个数据集上的实证结果支持。
- 为减少单个模型变体的误差并提升整体链接预测性能，我们采用两种方法汇总不同变体的决策，即加权距离求和与排名融合。

本文其余部分安排如下。第 II 节简要回顾相关工作。第 III 节阐述 CompoundE3D 的设计方法，并详细介绍多个模型变体的决策集成。第 IV 节展示实验结果，并与以往工作进行性能基准比较。最后，第 V 节给出总结性评述和未来研究方向。

II. 相关工作

A. 知识图谱嵌入（*KGE*）模型

许多基于距离的知识图谱嵌入模型通过将关系视为特定变换而构建。下面对它们进行简要回顾。

1) 二维几何变换：不少知识图谱嵌入模型受到二维几何变换的启发，例如二维平面上的平移、旋转和缩放。TransE [16]将关系建模为头实体与尾实体之间的平移。这种简单模型无法有效地建模对称关系。RotatE [17]将关系视为复数空间中的特定旋转，因此能够很好地处理对称关系。此外，RotatE 引入了自对抗负采样损失，显著提升了基于距离的知识图谱嵌入模型的性能。PairRE [18]通过缩放操作对关系进行建模，从而支持可变间隔，这有助于编码复杂关系。PairRE 对实体嵌入施加的酉约束在实践中也很有效。CompoundE [20] 采用包括平移、旋转和缩放在内的复合几何变换来建模不同关系。它在几乎不增加整体复杂度的情况下，提供了性能更优的知识图谱嵌入模型。

2) 高级变换：NagE [22] 将通用群论引入知识图谱嵌入模型的设计，并为其构建提供了一套通用方法。QuatE [24] 将知识图谱嵌入模型的设计扩展到四元数空间，使实体与关系之间的交互更加紧凑，同时提供了更多自由度。为更有效地建模关系组合中的非交换性，RotatE3D [21]和 DensE [23]都利用了四元数旋转，但采用的形式不同。ROTH [25] 采用双曲曲率来捕捉知识图谱中的层次结构。另一方面，文献 [26] 对在知识图谱嵌入中引入双曲几何是否必要提出了质疑。

B. 基于分类的模型

另一类模型通过将未见过的三元组划分为“有效”（或正类）和“无效”（或负类）两类，并使用软决策来衡量三元组的可能性。

1）简单神经网络：采用多层感知机（MLP）网络 [27]来衡量未见过的三元组用于链接预测的可能性。神经张量网络（NTN） [28] 采用双线性张量神经层，对三元组中实体与关系之间的交互进行建模。ConvE[29] 将头实体与关系堆叠起来，将其重塑为二维数组，并使用卷积神经网络（CNN）从中提取信息。随后，所得特征图通过点积与尾实体进行交互。R-GCN [30]使用具有关系特定权重的图卷积网络（GCN）来获取实体表示，随后将其输入 DistMult [31] 进行链接预测。尽管它具有处理归纳设置的潜力，但其性能仍不及基于嵌入的方法。

2）高级神经网络：KG-BERT [32]使用预训练语言模型 BERT [33], ，从文本描述中获取实体表示（而非从知识图谱链接中获取）。然而，与基于嵌入的模型相比，其推理时间要长得多。SimKGC [34] 通过构造对比样本对改进了基于Transformer的分类方法。它使用 BERT 估计语义相似度，并将相似度得分较高的三元组视为正样本对，反之亦然。然而，其性能对语言模型质量较为敏感，且所需的计算资源较高。

3）轻量级分类模型：KGBoost [35]提出了一种新颖的负采样方案，并使用 XG-Boost [36] 分类器进行链接预测。受判别特征学习（DFT） [37], [38]的启发，该方法从训练得到的嵌入中提取最具判别性的特征；GreenKGC[39] 是一种轻量、模块化的分类方法，通过训练二分类器对未见过的三元组进行分类。

C. 高级关系建模

为了建模复杂关系，人们开发了专门的技术。例如，为了有效建模1对N、N对1和N对N等关系，TransH [40]将嵌入实体空间投影到关系特定超平面上。TransR [41]学习一种关系特定投影，将

实体向量映射到特定的关系空间。TransD [42] 基于关系和实体投影向量推导出动态映射。TranSparse [43] 使关系投影矩阵保持稀疏。近年来，包括X+AT [44],SFBR [45], 和 STaR [46] 在内的许多知识图谱嵌入模型，将平移和缩放运算应用于基于距离和基于语义匹配的[47]模型，以提升性能。事实证明，引入平移有助于提升四元数空间中的知识图谱嵌入模型性能，例如DualE [48], 和BiQUE [49]。ReflectE [50] 将每种关系建模为超平面的法向量，该超平面用于反射实体向量。它能够很好地建模对称关系和逆关系。迄今为止，将各种仿射运算级联起来是一个自然而尚未得到探索的思路。

D. 模型集成

尽管集成学习是机器学习中的一种常见策略，但它在知识图谱补全领域仍未得到充分探索。链接预测评估本质上是一个排序问题。我们希望优化集成决策，使有效三元组在所有候选项中的排名高于无效三元组。排序聚合是信息检索中的经典问题。监督方法 [36], [51] 和无监督方法 [52], [53] 均已得到研究。由于知识图谱链接预测中的真实排序不可得（除排名第一的三元组外），无监督设定更为相关。然而，利用排序聚合提升链接预测性能这一方法尚未得到充分关注。下面给出若干示例。KEnS [54] 通过集成推理结合多个基于语言的知识图谱嵌入模型的预测，用于多语言知识图谱补全。AutoSF [55]提出了一种算法，用于从多个语义匹配模型中搜索最佳评分函数。在[56] 中，采用多个相同的低维知识图谱嵌入模型构成的集成，以提升链接预测性能。最近，DuEL[57] 将链接预测视为分类问题，并使用无监督技术聚合多个不同分类器给出的二元决策。

III. 所提方法

A. CompoundE3D

在本文中，我们使用三维仿射变换，包括平移、缩放、旋转、反射和错切，如图 1 所示，以建模知识图谱中的不同关系。这组规模庞大的变换算子为针对知识图谱数据集不同特征的知识图谱嵌入设计提供了极大的灵活性。下面，我们在齐次坐标中正式定义每个三维仿射算子。

1) 平移： 分量 $\mathbf{T} \in \mathbf{SE(3)}$，如图 1a 所示，定义为

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & v_x \\ 0 & 1 & 0 & v_y \\ 0 & 0 & 1 & v_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

2) 缩放： 分量 $\mathbf{S} \in \mathbf{Aff(3)}$，如图 1b 所示，定义为

$$\mathbf{S} = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{2}$$

3）旋转：如图 1c 所示，分量 $\mathbf{R} \in \mathbf{SO(3)}$ 定义为

$$\mathbf{R} = \mathbf{R}_z(\alpha)\mathbf{R}_y(\beta)\mathbf{R}_x(\gamma) = \begin{bmatrix} a & b & c & 0 \\ d & e & f & 0 \\ g & h & i & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

其中

$$\begin{aligned} a &= \cos(\alpha)\cos(\beta), \\ b &= \cos(\alpha)\sin(\beta)\sin(\gamma) - \sin(\alpha)\cos(\gamma), \\ c &= \cos(\alpha)\sin(\beta)\cos(\gamma) + \sin(\alpha)\sin(\gamma), \\ d &= \sin(\alpha)\cos(\beta), \\ e &= \sin(\alpha)\sin(\beta)\sin(\gamma) + \cos(\alpha)\cos(\gamma), \\ f &= \sin(\alpha)\sin(\beta)\cos(\gamma) - \cos(\alpha)\sin(\gamma), \\ g &= -\sin(\beta), \\ h &= \cos(\beta)\sin(\gamma), \\ i &= \cos(\beta)\cos(\gamma). \end{aligned} \tag{4}$$

这一通用三维旋转算子是对偏航、俯仰和滚转旋转进行复合的结果。它们分别定义为

- 偏航旋转分量:

$$\mathbf{R}_z(\alpha) = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) & 0 & 0 \\ \sin(\alpha) & \cos(\alpha) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{5}$$

- 俯仰旋转分量:

$$\mathbf{R}_y(\beta) = \begin{bmatrix} \cos(\beta) & 0 & -\sin(\beta) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(\beta) & 0 & \cos(\beta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{6}$$

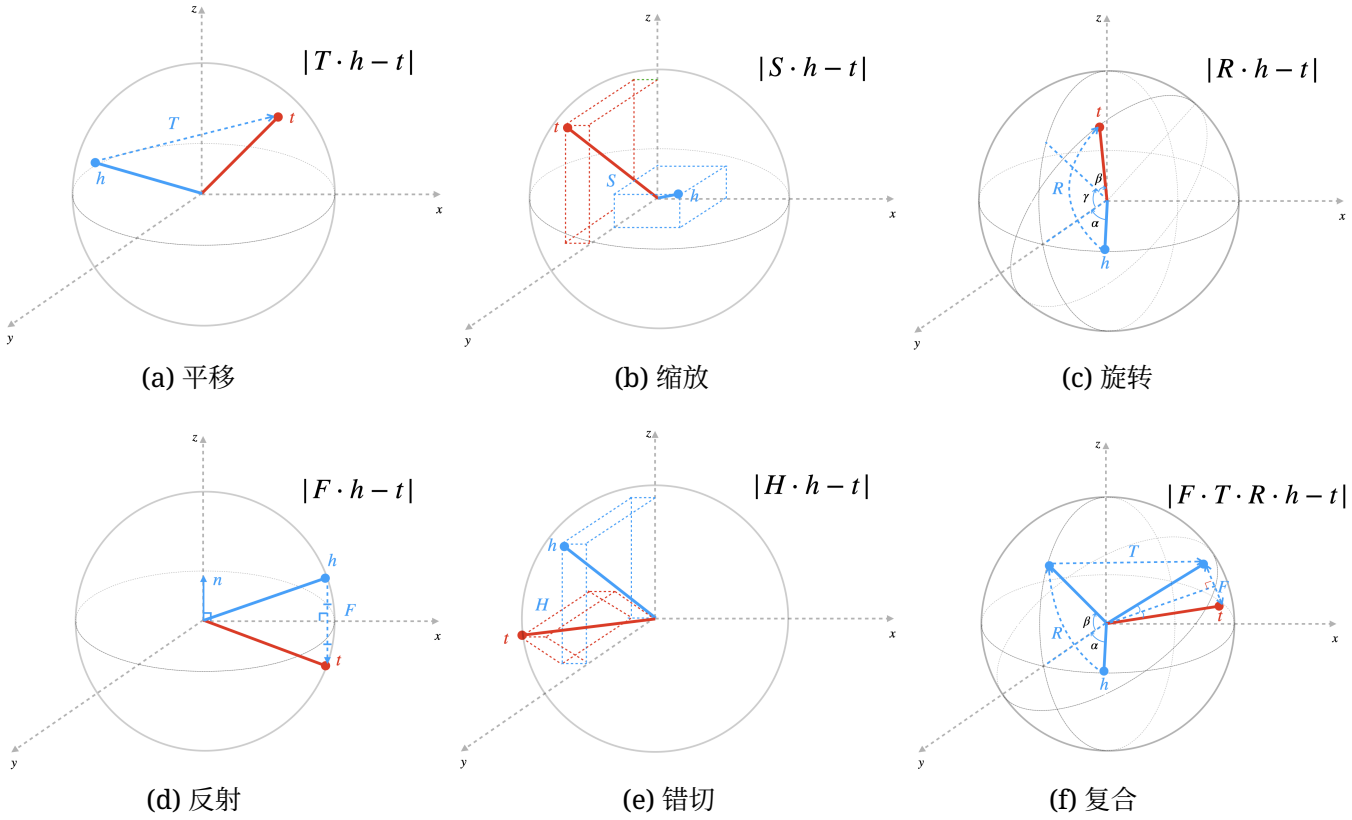
- 横滚旋转分量:

$$\mathbf{R}_x(\gamma) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\gamma) & -\sin(\gamma) & 0 \\ 0 & \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{7}$$

4) 反射：分量 $\mathbf{F} \in \mathbf{SO(3)}$，如图 1d 所示，定义为

$$\mathbf{F} = \begin{bmatrix} 1-2n_x^2 & -2n_xn_y & -2n_xn_z & 0 \\ -2n_xn_y & 1-2n_y^2 & -2n_yn_z & 0 \\ -2n_xn_z & -2n_yn_z & 1-2n_z^2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{8}$$

上述表达式源自豪斯霍尔德反射，$\mathbf{F} = \mathbf{I} - \mathbf{2nn^T}$。在三维空间中，$\mathbf{n}$ 是垂直于反射超平面$\mathbf{n} = [n_x, n_y, n_z]$的三维单位向量。



(a) 平移 (b) 缩放 (c) 旋转

(d) 反射 (e) 错切 (f) 复合

图 1：在三维子空间中组合不同的几何运算。

5）错切： 分量$\mathbf{H}$ $\in$ **Aff**(3)，如图1e所示，定义为

$$\mathbf{H} = \mathbf{H}_{yz}\mathbf{H}_{xz}\mathbf{H}_{xy} = \begin{bmatrix} 1 & \mathrm{Sh}_x^y & \mathrm{Sh}_x^z & 0 \\ \mathrm{Sh}_y^x & 1 & \mathrm{Sh}_y^z & 0 \\ \mathrm{Sh}_z^x & \mathrm{Sh}_z^y & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{9}$$

剪切算子是由 3 个算子复合而成的结果：$\mathbf{H}_{yz}$、$\mathbf{H}_{xz}$ 和$\mathbf{H}_{xy}$ 。其数学定义如下：

$$\mathbf{H}_{yz} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \mathrm{Sh}_y^x & 1 & 0 & 0 \\ \mathrm{Sh}_z^x & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{10}$$

$$\mathbf{H}_{xz} = \begin{bmatrix} 1 & \mathrm{Sh}_x^y & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \mathrm{Sh}_z^y & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{11}$$

$$\mathbf{H}_{xy} = \begin{bmatrix} 1 & 0 & \mathrm{Sh}_x^z & 0 \\ 0 & 1 & \mathrm{Sh}_y^z & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{12}$$

矩阵 $\mathbf{H}_{yz}$ 具有物理意义——它表示一种剪切变换，按 x 分量的倍数对 y 和 z 分量进行错切。类似的物理解释也适用于 $\mathbf{H}_{xz}$ 和 $\mathbf{H}_{xy}$。

上述变换可以级联，从而得到一个复合算子；例如，

$$\mathbf{O} = \mathbf{T} \cdot \mathbf{S} \cdot \mathbf{R} \cdot \mathbf{F} \cdot \mathbf{H}, \tag{13}$$

在实际实现中，我们使用该算子在常规笛卡尔坐标系中的表示，而非齐次坐标。此外，高维关系算子可以表示为如下形式的分块对角矩阵：

$$\mathbf{M_r} = \mathbf{diag}(\mathbf{O_{r,1}}, \mathbf{O_{r,2}}, \ldots, \mathbf{O_{r,n}}), \tag{14}$$

其中，$\mathbf{O_{r,\ i}}$ 是第 i 阶段的复合算子。

我们可以为 CompoundE3D 定义以下三个评分函数：

- CompoundE3D-Head

$$f_r^{(h)}(h,t) = \|\mathbf{M_r} \cdot \mathbf{h} - \mathbf{t}\|, \tag{15}$$

- CompoundE3D-Tail

$$f_r^{(t)}(h,t) = \|\mathbf{h} - \hat{\mathbf{M}}_\mathbf{r} \cdot \mathbf{t}\|, \tag{16}$$

- CompoundE3D-Complete

$$f_r^{(h,t)}(h,t) = \|\mathbf{M_r} \cdot \mathbf{h} - \hat{\mathbf{M}}_\mathbf{r} \cdot \mathbf{t}\|, \tag{17}$$

其中 $\mathbf{h}$ 和 $\mathbf{t}$ 分别表示头实体嵌入和尾实体嵌入，而$\hat{}\mathbf{M_r}$ 和 $\mathbf{M_r}$ 分别表示作用于头实体和尾实体的关系特定算子。

一般来说，我们有五种可用的不同仿射变换，即平移、缩放、旋转、反射，

以及错切。每个算子可以应用于 1）头实体、2）尾实体，或 3 ）头实体和尾实体。因此，在每个阶段总共有 15 种不同的算子应用方式。所有这些可能的选择都称为 CompoundE3D 变体。对于给定的知识图谱数据集，要找到最优的 CompoundE3D 变体，搜索空间非常庞大。开发一种简单而有效的机制，以便在特定复杂度约束下找到性能最佳的变体，至关重要。

B. 搜索最佳 *CompoundE3D* 变体的束搜索 *ant*

在本小节中，我们提出一种束搜索算法，用于寻找最优的 CompoundE3D 变体。对于第 i 阶段，在某一步可应用的所有算子对集合为

$$\begin{aligned}\mathbf{P} \in \{&(\mathbf{T},\mathbf{I}),(\mathbf{S},\mathbf{I}),(\mathbf{R},\mathbf{I}),(\mathbf{F},\mathbf{I}),(\mathbf{H},\mathbf{I}),\\ &(\mathbf{I},\hat{\mathbf{T}}),(\mathbf{I},\hat{\mathbf{S}}),(\mathbf{I},\hat{\mathbf{R}}),(\mathbf{I},\hat{\mathbf{F}}),(\mathbf{I},\hat{\mathbf{H}}),\\ &(\mathbf{T},\hat{\mathbf{T}}),(\mathbf{S},\hat{\mathbf{S}}),(\mathbf{R},\hat{\mathbf{R}}),(\mathbf{F},\hat{\mathbf{F}}),(\mathbf{H},\hat{\mathbf{H}})\},\end{aligned} \tag{18}$$

其中 $\mathbf{I}$ 是恒等算子。首先，我们应用 $\mathbf{P}$ 中的所有算子对，并计算所有中间变体的评分函数。每个变体使用训练集进行 l 次迭代优化，并在验证数据集上评估其性能。然后，我们选择表现最佳的前 k 个变体，作为下一步进一步探索的起点。重复相同的过程，直到触发某个终止条件。之后，我们进入第（$i+1$）阶段。到达最后一个阶段后，整个搜索过程即告完成。阶段总数是由用户选定的超参数。

通过由简单模型构建更复杂知识图谱嵌入模型的束搜索过程如算法 1 所述。补充说明如下。

- 我们通过设置一个循环来初始化算法，该循环遍历由 $\mathbf{P}$表示的所有可能算子组合的集合，对这些组合进行训练和评估，并找出排名前k 的变体作为起始点。
- 在下一个循环中，我们使用两个停止准则来终止束搜索：1）算子数量 $>\lambda$，即当算子数量超过上限 λ时停止搜索；2） $\Delta_{\text{平均倒数排名}}\Delta_{\text{参数}} < \gamma$，即平均倒数排名的增幅与自由参数增幅之比低于阈值 γ，此时继续增加模型复杂度已不值得，因为模型性能只能获得边际提升。
- P $\times$W 表示算子对集合 P 与上一步得到的排名前k 的变体集合 W 之间的笛卡尔积，而 $\dot{f}_r^{i-1}(h,t) \triangleleft (\mathbf{M}^i, \hat{\mathbf{M}}^i)$ 表示将算子对 $(\mathbf{M}^i, \hat{\mathbf{M}}^i)$ 应用于之前的最优评分函数 $\dot{f}_r^{i-1}(h, t)$。•例如，如果 $\dot{f}_r^{i-1}(h,t) = \|\mathbf{R}\cdot\mathbf{h} - \mathbf{t}\|$，且 $(\mathbf{M}^i, \hat{\mathbf{M}}^i) = (\mathbf{S}, \hat{\mathbf{S}})$，那么 $\tilde{f}_r^i(h,t) = \|\mathbf{S}\cdot\mathbf{R}\cdot\mathbf{h} - \hat{\mathbf{S}}\cdot\mathbf{t}\|$。
- 循环因触发任一终止条件而结束后，我们从已探索的变体集合$\mathbf{W}$中选择表现最佳的前 1 个变体作为最佳选择。

C. 模型集成

1) 加权距离和（*WDS*）策略：我们选择表现最优的前-k 个CompoundE3D变体，并进行

算法 1 用于选择最佳 CompoundE3D 变体的束搜索

初始化 $i \leftarrow 1$，$\mathbf{U} \leftarrow \{\}$
对于$(\mathbf{M}^i, \hat{\mathbf{M}}^i) \in \mathbf{P}$，**执行**
  $\tilde{f}_r^i(h,t) \leftarrow \|\mathbf{M}^i\cdot\mathbf{h} - \hat{\mathbf{M}}^i\cdot\mathbf{t}\|$;
  训练 $\tilde{f}_r^i(h, t)$ 迭代 l 次;
  平均倒数排名 $\leftarrow$评估 $\tilde{f}_r^i(h, t)$ 使用验证集;
  U.insert({平均倒数排名, $\tilde{f}_r^i(h,t)$});
结束循环
$\mathbf{W} \leftarrow$ top-k 变体来自 $\mathbf{U}$
$i \leftarrow i+1$
Δ平均倒数排名 $\leftarrow \gamma$，Δ参数 $\leftarrow 1$
当 # 算子 $< \lambda$ **和** 最大 $\Delta_{\text{平均倒数排名}}\Delta_{\text{参数}} \geq \gamma$ **时执行**
  初始化 $\mathbf{V} \leftarrow \{\}$
  对于 $\{(\mathbf{M}^i, \hat{\mathbf{M}}^i)\ \dot{f}_r^{i-1}(h, t)\} \in \mathbf{P} \times \mathbf{W}$ **执行**
    $\tilde{f}_r^i(h, t) \leftarrow \dot{f}_r^{i-1}(h, t) \triangleleft (\mathbf{M}^i, \hat{\mathbf{M}}^i)$;
    训练 $\tilde{f}_r^i(h, t)$ 进行 l 次迭代;
    评估 Δ平均倒数排名 $\leftarrow f_r^i(h, t)$, 使用验证集; $f_r^i(h, t)$ 平均倒数排名; $\dot{f}_r^{i-1}(h, t)$ 平均倒数排名;
    Δ参数 $\leftarrow \tilde{f}_r^i(h, t)$ 参数 $- \dot{f}_r^{i-1}(h, t)$ 参数;
    V.insert(平均倒数排名, Δ平均倒数排名, Δ参数, $\tilde{f}_r^i(h,t)$));
  结束 for
  $\mathbf{W} \leftarrow$ 来自 $\mathbf{V}$ 的前-k 个变体;
结束 while
$f_r^*(h,t) \leftarrow$ 来自 $\mathbf{W}$ 的最佳变体;

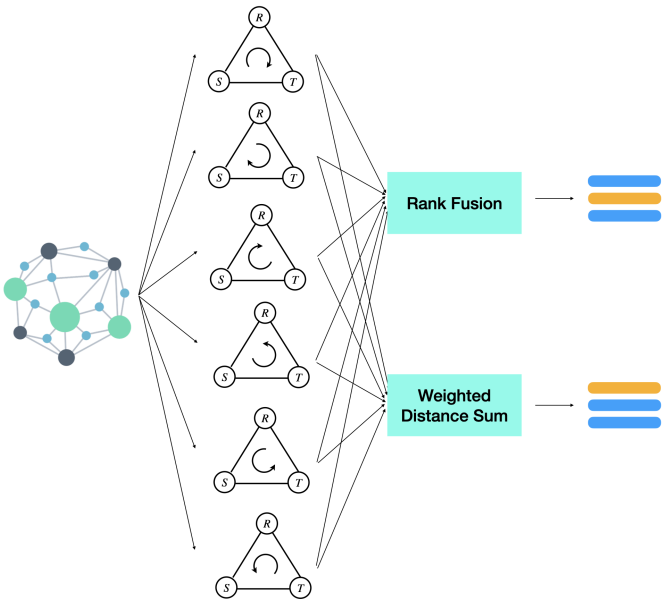


图 2：多个 CompoundE3D 变体的集成。

其预测分数的加权平均。考虑以下三种加权方案。

- **均匀权重**。该方案对所选 k 变体赋予相同的权重，作为

$$\hat{f}_r(h,t) = \frac{1}{k}\sum_{i=1}^{k} f_r^i(h,t), \tag{19}$$

其中， $f_r^i(h, t)$ 是第 i 个变体的评分函数。

• **几何权重**。 该方案根据各变体在验证数据集上的MRR性能按降序排序，并为第 k 个变体赋予权重 λ^k，$0 < \lambda < 1$。也就是说，

$$\hat{f}_r(h,t) = \frac{1}{\sum_{i=1}^{k} \lambda^k} \sum_{i=1}^{k} \lambda^k f_r^i(h,t). \tag{20}$$

由于 $\lambda^k > \lambda^{k+1}$，在计算聚合距离时，我们会为表现更好的模型赋予更高的权重。

• **可学习权重**。 该方案基于训练数据集训练一组可学习权重 $w_i > 0$，以最小化以下加权分数：

$$\hat{f}_r(h,t) = \frac{1}{\sum_{i=1}^{k} w_i} \sum_{i=1}^{k} w_i f_r^i(h,t). \tag{21}$$

在寻找最佳变体的优化过程中，可学习权重被作为参数实现，并使用相同的学习率和优化器。

对于每种关系，我们比较这三种权重方案，并选择性能最佳的一种。

2）排名融合策略：链接预测是一个列表排序问题。排名融合可用于提升性能。基于评分的知识图谱嵌入方法可以采用一些简单的排名融合方法。例如，可以取候选项排名的最大值、最小值、中位数、总和和欧几里得距离。在表 I 中，它们分别记为最大值组合、Comb-MIN、中位数组合、求和组合和欧几里得距离。表中还考虑并列出了三种高级排名融合方法，分别是：波达计数法 [58],倒数排名融合（RRF） [53],和 RBC（排名偏置质心） [59]。波达计数法根据候选项在单个偏好列表中的位置为其分配分数，排名最靠前的候选项得分最高，最后一名候选项得分最低。RRF 汇总倒数排名，以降低排名靠后候选项的重要性。表中的因子 k 可减轻异常值带来的高排名影响。RBC 使用几何分布降低排名靠后候选项的权重。所有排名融合函数的数学公式均列于表 I 的第二列，其中 R_i 是第 i 个基模型（或变体）的排名，$1 \le i \le n$ $e \in E$ 表示实体集中的一个实体，而 k和 ϕ 是超参数。

表 I：所考虑的排名融合函数列表。

Name	函数
最大值组合	$\max\{R_1(e) \cdots, R_n(e)\}$
最小值组合	$\min\{R_1(e) \cdots, R_n(e)\}$
中位数组合	中位数$\{R_1(e) \cdots, R_n(e)\}$
求和组合	$\sum_{i=1}^{n} R_i(e)$
欧几里得距离	$\sqrt{R_1(e)^2 + \cdots + R_n(e)^2}$
波达计数法	$\sum_{i=1}^{n} \frac{\lvert E\rvert - R_i(e) + 1}{\lvert E\rvert}$
倒数排名融合 [53]	$\sum_{i=1}^{n} \frac{1}{k + R_i(e)}$
RBC [59]	$\sum_{i=1}^{n} (1-\phi)\phi^{R_i(e)-1}$

D. 优化

遵循 RotatE 的负采样损失和自对抗训练策略，我们选择 CompoundE3D 的以下损失函数

$$\begin{aligned} L_{\text{KGE}} = & -\log \sigma(\zeta_1 - f_r(h,t)) \\ & - \sum_{i=1}^{n} p(h'_i, r, t'_i) \log \sigma(f_r(h'_i, t'_i) - \zeta_1), \end{aligned} \tag{22}$$

其中，σ 是Sigmoid函数，ζ_1 是预设的间隔超参数，(h'_i, r, t'_i) 是第 i个负三元组，而 $p(h'_i, r, t'_i)$是抽取负三元组(h'_i, r, t'_i)的概率。给定一个正三元组(h_i, r, t_i)，负采样分布为

$$p(h'_j, r, t'_j | \{(h_i, r, t_i)\}) = \frac{\exp \alpha_1 f_r(h'_j, t'_j)}{\sum_i \exp \alpha_1 f_r(h'_i, t'_i)}, \tag{23}$$

其中 α_1 是 softmax 函数中的温度。

IV. 实验

A. 实验设置

1）数据集：我们在以下四个知识图谱数据集上评估 CompoundE3D 的链接预测性能，并将其与几种基准方法进行比较。

• **DB100K [60].** 它是 DBpedia 知识图谱的一个子集。该数据集包含与音乐内容相关的信息，例如流派、乐队和音乐艺术家。它是一个相对稠密的知识图谱，因为每个实体至少出现在 20 种不同的关系中。

• **YAGO3-10 [61].** 它是 YAGO3 的一个子集，用于描述人物的公民身份、性别和职业。YGGO3-10 包含至少涉及 10 种不同关系的实体。

• **WN18RR [16], [29].**它是 WordNet 词汇数据库的一个子集。为避免测试泄漏，WN18RR 中移除了逆关系。

• **Ogbl-Wikikg2 [62].**它从维基百科中提取，包含 250 万个实体，是所选四个数据集中规模最大的一个。

四个知识图谱数据集的统计信息见表 II。

2）评估协议：下面介绍链接预测任务中常用的评估协议。对于测试集中的每个三元组（h，r，t），我们通过替换头实体 h或尾实体 t来生成测试样例（？ r，t）和（h，r，？）。然后，对于构成三元组（$\hat{h}$，r，t）的每个头实体候选项，以及构成三元组（h，r，$\hat{t}$）的每个尾实体候选项，我们分别计算基于距离的评分函数 $f_r(\hat{h}, t)$ 和 $f_r(h, \hat{t})$。分数越低，生成的三元组越有可能为真。然后，我们按升序排列所有候选三元组的分数，并确定真实三元组的排名。此外，我们在过滤排名设置 [16] 下评估链接预测性能，该设置

表 I：四个链接预测数据集的统计信息。

数据集	#实体	#关系	#训练集	#验证集	#测试集	平均度
DB100K	99,604	470	597,572	50,000	50,000	12
ogbl-wikikg2	2,500,604	535	16,109,182	429,456	598,543	12.2
YAGO3-10	123,182	37	1,079,040	5,000	5,000	9.6
WN18RR	40,943	11	86,835	3,034	3,134	2.19

表 III：六个超参数的搜索空间。

数据集	DB100K	ogbl-wikikg2	YAGO3-10	WN18RR
Dim	$\{150, 300, 450, 600\}$	$\{90, 150, 180, 240, 300\}$	$\{450, 600, 750, 900\}$	$\{180, 240, 360, 480, 600\}$
lr	$\{2, 3, 4, 5, 6, 7, 8, 9\} \times 10^{-5}$	$\{0.0005, 0.001, 0.005, 0.01\}$	$\{3, 4, 5, 6, 7\} \times 10^{-4}$	$\{4, 5, 6, 7, 8\} \times 10^{-4}$
B	$\{256, 512, 1024, 2048\}$	$\{2048, 4096, 8192\}$	$\{512, 1024, 2048, 4096\}$	$\{512, 1024, 2048, 4096\}$
N	$\{256, 512, 1024, 2048\}$	$\{125, 250, 500\}$	$\{256, 512, 1024, 2048\}$	$\{256, 512, 1024, 2048\}$
ζ	$\{4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}$	$\{5, 6, 7, 8, 9\}$	$\{11, 12, 13, 13.1, 13.3, 13.5\}$	$\{5, 6, 7, 8, 9\}$
α	$\{0.5, 0.7, 0.9, 1.0, 1.2\}$	$\{0.5, 1.0\}$	$\{0.8, 0.9, 1.0, 1.1, 1.2\}$	$\{0.5, 0.7, 0.9, 1.0, 1.2\}$

表 IV：链接预测任务的最优配置，其中 B 和 N 分别表示批次大小和负样本大小。

数据集	CompoundE3D 变体	#Dim	*lr*	*B*	*N*	ζ	α
DB100K	$\|\|\mathbf{S}\cdot\mathbf{h} - \hat{\mathbf{T}}\cdot\hat{\mathbf{R}}\cdot\hat{\mathbf{S}}\cdot\mathbf{t}\|\|$	600	0.00005	1024	512	9	1
ogbl-wikikg2	$\|\|\mathbf{T}\cdot\mathbf{h} - \hat{\mathbf{H}}\cdot\mathbf{t}\|\|$	300	0.001	8192	125	8	1
YAGO3-10	$\|\|\mathbf{T}\cdot\mathbf{S}\cdot\mathbf{R}\cdot\mathbf{h} - \mathbf{t}\|\|$	600	0.0005	1024	1024	13.3	1.1
WN18RR	$\|\|\mathbf{R}\cdot\mathbf{S}\cdot\mathbf{T}\cdot$小时 $-$吨$\|\|$	480	0.00005	512	256	6	1

能够突出未见三元组预测的重要性，因为嵌入模型往往会给已观测到的三元组更高的排名。我们采用 Hits@k 和平均倒数排名（MRR）作为评估指标，以比较知识图谱嵌入模型的质量。

3）超参数搜索：我们针对不同的知识图谱数据集，对 CompoundE3D的六个超参数进行了广泛搜索，分别是：1）嵌入空间的维度（**维度**），2）学习率（***lr***），3）批次大小（B），4）负样本大小（N），5）间隔超参数（ζ），以及 6）采样温度（α）。它们的搜索取值列于表 III。

在搜索过程中，我们首先在特定的超参数设置下计算评分函数，使若干变体取得尚可的性能，其中每个变体的训练迭代次数设为 $l = 30000$。找到最优变体后，我们在该最优变体下微调超参数。最优配置如表 IV 所示。所有参数调优均采用 Adam优化器 [63] 。对于集成实验，我们为每个基础变体模型采用相同的最优配置。

4）其他实现细节：我们根据任务的显存需求，在多种图形处理器上开展实验并进行超参数调优，包括 Nvidia P100（16G）、V100（32G）、A100（40G）和 A40（48G）。通常，每项任务申请 8 个 CPU 核心和少于 70G 的内存。表 IV 中每种最优配置的结果均可在单块 V100 上针对所有数据集复现。对于 WN18RR 数据集，我们采用 Rotate3D [21] 中的旋转实现。

B. 实验结果

1) 性能评估：我们将几种基准知识图谱嵌入方法的链接预测性能与

表 V：DB100K 在过滤排名设置下的链接预测性能比较。

数据集	DB100K			
模型	**MRR**	**H@1**	**H@3**	**H@10**
TransE [16]	0.111	0.016	0.164	0.27
DistMult [31]	0.233	0.115	0.301	0.448
HolE [64]	0.26	0.182	0.309	0.411
ComplEx [65]	0.242	0.126	0.312	0.44
Analogy [66]	0.252	0.142	0.323	0.427
RUGE [67]	0.246	0.129	0.325	0.433
ComplEx-NNE+AER [60]	0.306	0.244	0.334	0.418
SEEK [68]	0.338	0.268	0.37	0.467
AcrE（并行） [69]	0.413	0.314	0.472	0.588
PairRE [18]	0.412	0.309	0.472	0.600
TransSHER [70]	0.431	0.345	0.476	0.589
CompoundE [20]	0.405	0.306	0.461	0.588
CompoundE3D	0.450	0.373	0.488	0.594
CompoundE3D RRF	0.457	0.376	0.497	0.607
CompoundE3D WDS	**0.462**	**0.378**	**0.506**	**0.616**

采用表 IV 所示最优配置的 CompoundE3D 进行比较。DB100K、ogbl- wikikg2 和 YAGO3-10 数据集的性能评估结果分别如表 V、表 VI 和表 VII 所示。此外，每列中的最佳结果和次优结果分别以粗体和下划线标示。与 CompoundE 及其他近期模型相比，CompoundE3D 实现了显著的性能提升。我们可以清楚地看到，通过在新框架中引入更多仿射算子，并将仿射变换从二维扩展到三维，CompoundE3D 展现出了明显优势，

为验证模型集成的有效性，我们针对 DB100K 和 YAGO3-10 数据集考察了两种不同的集成策略。

•对于 DB100K 数据集，我们选取表现最佳的两个

表 VI：在 ogbl-wikikg2 的过滤排名设置下的链接预测性能比较。

数据集	ogbl-wikikg2		
指标	Dim	验证集 MRR	Test MRR
AutoSF+NodePiece	100	0.5806	0.5703
ComplEx-N3-RP	100	0.6701	0.6481
TransE [16]	500	0.4272	0.4256
DistMult [31]	500	0.3506	0.3729
ComplEx [65]	250	0.3759	0.4027
RotatE [17]	250	0.4353	0.4353
Rotate3D [21]	100	0.5685	0.5568
PairRE [18]	200	0.5423	0.5208
TripleRE [71]	200	0.6045	0.5794
CompoundE [20]	100	0.6704	0.6515
CompoundE3D	90	0.6994	0.6826
	180	0.7146	0.6962
	300	**0.7175**	**0.7006**

表 VII：YAGO3-10 在过滤排名设置下的链接预测性能比较。

数据集	YAGO3-10			
指标	MRR	命中率@1	命中率@3	Hit@10
DistMult [31]	0.34	0.24	0.38	0.54
ComplEx [65]	0.36	0.26	0.4	0.55
DihEdral [72]	0.472	0.381	0.523	0.643
ConvE [29]	0.44	0.35	0.49	0.62
RotatE [17]	0.495	0.402	0.55	0.67
InteractE [73]	0.541	0.462	-	0.687
HAKE [74]	0.545	0.462	0.596	0.694
DensE [23]	0.541	**0.465**	0.585	0.678
Rot-Pro [75]	0.542	0.443	0.596	0.699
CompoundE [20]	0.477	0.376	0.538	0.664
CompoundE3D	0.542	0.450	0.602	0.701
CompoundE3D RRF	0.541	0.446	0.607	**0.707**
CompoundE3D WDS	**0.551**	0.463	**0.608**	0.703

变体。它们分别是 $\|\mathbf{S}\cdot\mathbf{h}-\hat{}\mathbf{R}\cdot\hat{}\mathbf{S}\cdot\mathbf{t}\|$ 和 $\|\mathbf{S}\cdot\mathbf{h}-\hat{}\mathbf{T}\cdot\hat{}\mathbf{R}\cdot\hat{}\mathbf{S}\cdot\mathbf{t}\|$

- 对于 YAGO3-10 数据集，我们选取表现最好的三个变体。它们分别是 $\|\mathbf{T}\cdot\mathbf{S}\cdot\mathbf{R}\cdot\mathbf{h}-\mathbf{t}\|$、$\|\mathbf{S}\cdot\mathbf{R}\cdot\mathbf{T}\cdot\mathbf{h}-\mathbf{t}\|$，以及 $\|\mathbf{S}\cdot\mathbf{T}\cdot\mathbf{R}\cdot\mathbf{h}-\mathbf{t}\|$。

2）模型集成：如第 III-C 节所述，我们采用两种策略进行模型集成：加权距离求和（WDS）和排名融合。在三种 WDS 策略中，可学习权重策略在 DB100K 上效果最好，而均匀权重在 YAGO3-10 上表现最佳。我们在表 V 和表 VII 中用 CompoundE3D WDS 表示最佳 WDS 方案，并在表 VIII 中列出其他权重策略的性能。在全部八种排名融合策略中，我们发现倒数排名融合（RRF）在 DB100K 和 YAGO3-10 上均最为有效。因此，我们在表 V 和表 VII 中用 CompoundE3D RRF 表示最佳排名融合方案，并在表 IX 中列出其他排名融合策略的性能。

3）束搜索的有效性：我们在 DB100K 和 YAGO3-10 数据集上开展消融研究，以深入了解不同变换算子对模型性能的影响。我们从最简单配置的变体开始

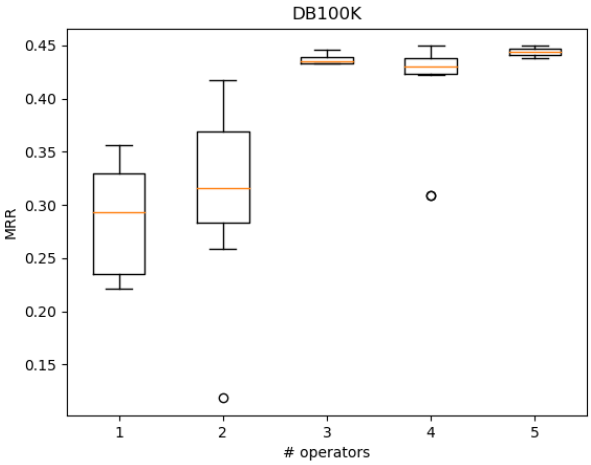


图3：DB100K数据集中不同模型变体的平均倒数排名性能与算子数量的分布。

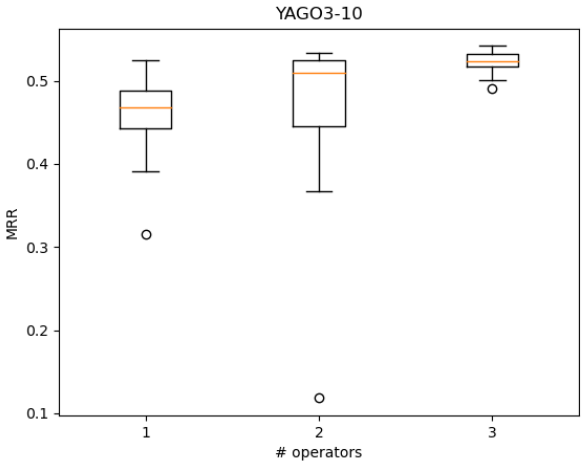


图4：YAGO3-10数据集中不同模型变体的MRR性能与算子数量的分布。

并在每个阶段添加额外的算子。效果良好的简单模型、由此得到的最优变体及其性能数据分别列于表 X 和表 XI。此外，我们分别在图 3 和图 4 中展示了随着算子数量增加，DB100K 和 YAGO-3 上平均倒数排名性能的分布。箱线图中，黄色横线表示中位数，箱体表示四分位距，两端的横线表示下须线和上须线，圆点则表示异常值。这些结果均表明，所提出的束搜索算法是有效的。

4）对称关系的建模：旋转和反射是等距运算。正如 [17], [50], 所述，它们的二维版本在某些情况下能够很好地处理对称关系。我们猜想，相应的三维算子也具有相同的性质。为验证这一点，我们进行了消融研究，并评估了仅使用

表 VIII：DB100K 和 YAGO3-10 上不同加权距离和（WDS）策略的比较。

数据集	DB100K				YAGO3-10			
策略	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
可学习权重	**0.462**	**0.378**	**0.506**	**0.616**	0.545	0.451	0.586	0.696
均匀权重	0.460	0.376	0.503	0.614	**0.551**	**0.463**	**0.608**	**0.703**
几何权重	0.446	0.348	0.503	0.618	0.531	0.439	0.580	0.691

表 I　　X：DB100K 和 YAGO3-10 上不同排序融合方法的性能比较　　.

数据集	DB100K				YAGO3-10			
聚合函数	MRR	命中率@1	命中率@3	Hit@10	MRR	命中率@1	命中率@3	Hit@10
最大值组合	0.455	0.375	0.496	0.603	0.536	0.440	0.600	0.701
最小值组合	0.452	0.369	0.4955	0.606	0.527	0.427	0.597	0.702
中位数组合	0.456	0.376	0.497	0.606	0.541	0.445	0.605	0.705
求和组合	0.456	0.376	0.4969	0.6060	0.540	0.446	0.606	0.704
欧几里得距离	0.455	0.375	0.496	0.605	0.540	0.445	0.603	0.702
博尔达计数法	0.456	0.376	0.497	0.606	0.540	0.446	0.606	0.704
RRF	**0.457**	**0.376**	**0.497**	**0.607**	**0.541**	**0.446**	**0.607**	**0.707**
RBC	0.456	0.376	0.497	0.606	0.540	0.445	0.604	0.703

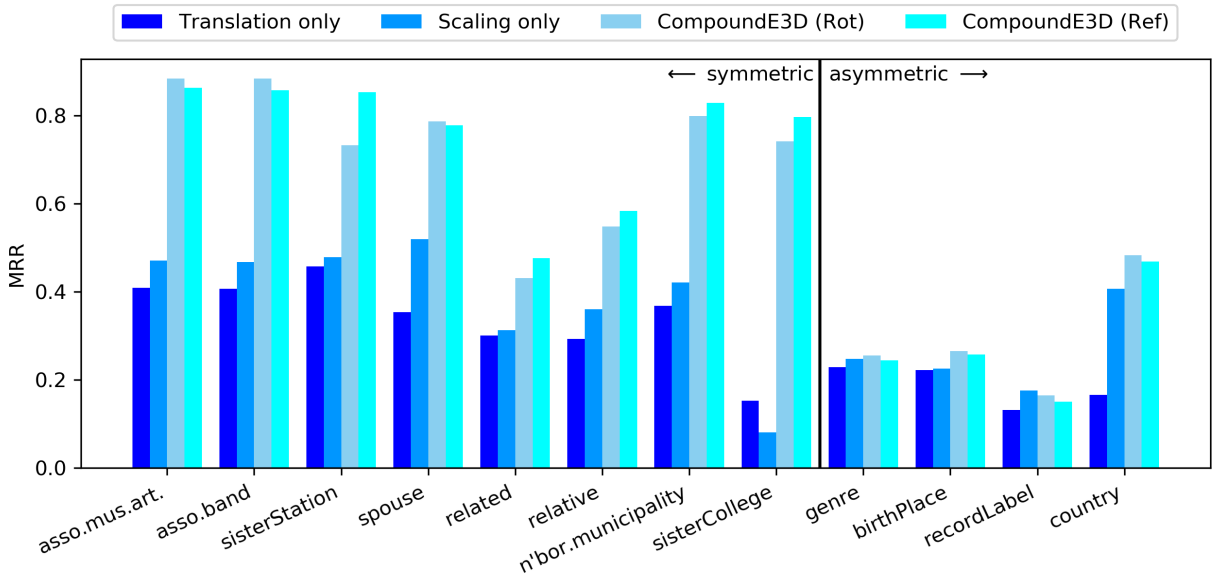


图 5：旋转和反射算子对对称关系的影响。

表 X：CompoundE3D 在 DB100K 上的消融研究。

数据集	DB100K			
变体	MRR	命中率@1	命中率@3	Hit@10
$\|\mathbf{S}\cdot\mathbf{h}-\hat{\mathbf{S}}\cdot\mathbf{t}\|$	0.417	0.323	0.471	0.590
$\|\mathbf{S}\cdot\mathbf{h}-\hat{\mathbf{R}}\cdot\hat{\mathbf{S}}\cdot\mathbf{t}\|$	0.447	0.364	0.492	0.600
$\|\mathbf{S}\cdot\mathbf{h}-\hat{\mathbf{T}}\cdot\hat{\mathbf{R}}\cdot\hat{\mathbf{S}}\cdot\mathbf{t}\|$	0.450	0.373	0.488	0.594

表XI：YAGO3-10 数据集上 CompoundE3D 的消融研究。

数据集	YAGO3-10			
指标	MRR	命中率@1	命中率@3	Hit@10
$\|\mathbf{R}\cdot\mathbf{h}-\mathbf{t}\|$	0.496	0.402	0.547	0.676
$\|\mathbf{S}\cdot\mathbf{R}\cdot\mathbf{h}-\mathbf{t}\|$	0.501	0.404	0.554	0.680
$\|\mathbf{T}\cdot\mathbf{S}\cdot\mathbf{R}\cdot\mathbf{h}-\mathbf{t}\|$	0.542	0.450	0.602	0.701

平移和缩放的模型与同时使用旋转和反射的模型的基础评分函数。图 5 对比了 DB100K 中不同模型变体在对称关系和非对称关系上的平均倒数排名性能。在该图中，我们选取了出现频率最高的关系类型，以便进行有意义的比较。不出所料，旋转和反射算子

确实能在对称关系上带来比非对称关系更显著的性能提升。这支持了我们的猜想：旋转和反射算子在建模对称关系方面具有内在优势。

5）多重性建模：多重性是指两个实体之间存在多种关系的情形；也就是说，

表 XII：WN18RR 各关系类型的过滤后 MRR 性能比较。

关系	类型	Khs_{G_r}	ξ_{G_r}	TransE	RotatE	CompoundE	CompoundE3D
相似于	一对一	0.07	-1.00	0.294	**1.000**	**1.000**	**1.000**
动词组	一对一	0.07	-0.50	0.363	0.961	**0.974**	0.898
成员部分词	1对N	1.00	-0.50	0.179	**0.259**	0.230	0.246
包含部分	1对N	1.00	-1.43	0.117	0.200	0.190	**0.202**
属于使用领域	1对N	1.00	-0.74	0.113	0.297	0.332	**0.378**
属于地域领域	1对N	1.00	-0.78	0.114	0.217	0.280	**0.413**
上位词	N对1	1.00	-2.64	0.059	0.156	0.155	**0.182**
实例上位词	N对1	1.00	-0.82	0.289	0.322	0.337	**0.356**
同义词集所属主题领域	N对1	0.99	-0.69	0.149	0.339	0.367	**0.396**
另请参见	N对N	0.36	-2.09	0.227	0.625	**0.629**	0.622
派生相关形式	N对N	0.07	-3.84	0.440	0.957	0.956	**0.959**

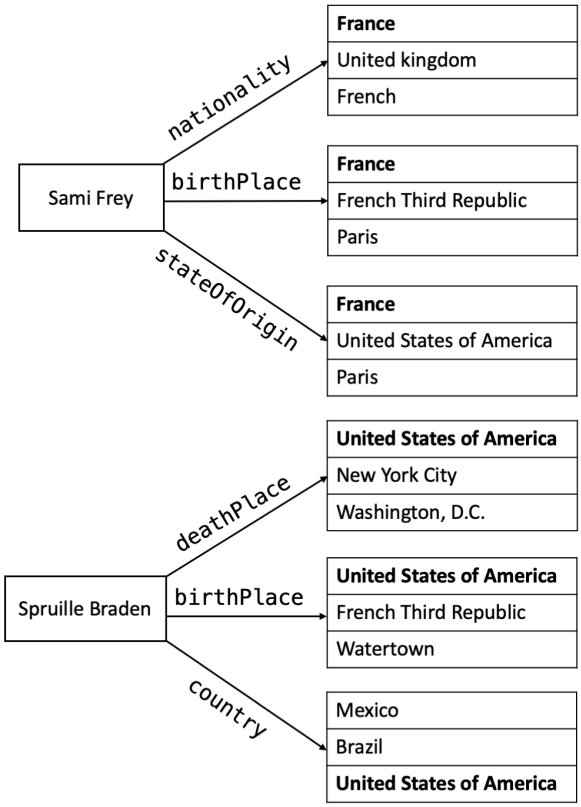


图6：CompoundE3D对多重性建模能力的示意图。

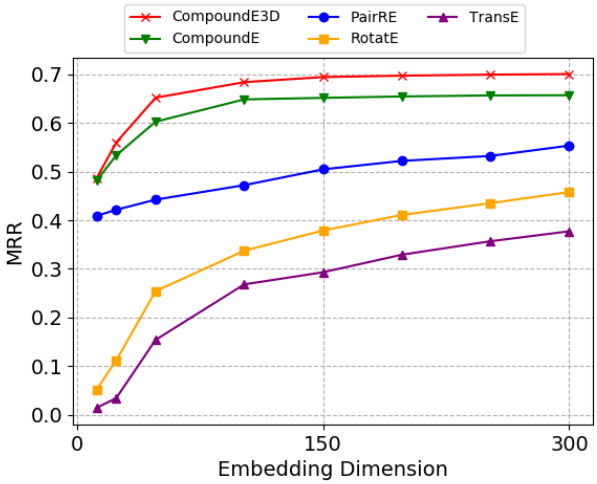


图 7：比较不同模型在不同维度下的平均倒数排名性能指标。

表 XIII：在相近测试集平均倒数排名下，ogbl-wikikg2 上知识图谱嵌入模型的复杂度比较。

模型	参数数量
TransE [16]	1,251M
DistMult [31]	1,251M
ComplEx [65]	1,251M
ComplEx-RP [76]	250.1M
RotatE [17]	500M
RotatE3D [21]	750.4M
PairRE [18]	500M
CompoundE [20]	250.1M
CompoundE3D	**225.2M**

三元组 $(h, r_1, t) \ldots, (h, r_n, t)$同时成立。通常，由于传统知识图谱嵌入模型在关系建模方面能力有限，因此难以对多重性进行建模。相比之下，CompoundE3D能够很好地建模多重关系模式，因为它可以使用多组不同的变换将头实体映射到尾实体。我们在图6中给出两个示例，以说明CompoundE3D对多重关系进行建模的能力。这些示例取自DB100K中的实际链接预测案例。对于一个固定的（头实体，尾实体）对，存在三种不同的关系。图中展示了这两个示例中每种关系对应的前三个尾实体预测结果。我们看到，CompoundE3D能够处理由于其丰富的变体集合，多重性表现良好。

6）层级关系建模：我们希望研究 CompoundE3D 对层级关系的建模能力。WN18RR 提供了一个包含层级关系的代表性数据集。可使用两个指标来衡量关系的层级行为 [25]：1）记为 Khs_{G_r} 的克拉克哈特分数；2）记为 ξ_{G_r} 的曲率估计。如果关系 r 的 Khs_{G_r} 分数较高且 ξ_{G_r} 分数较低，则其层级行为更强，反之亦然。我们在表 XII 中比较了不同基线模型的过滤后 MRR 性能，包括 TransE、RotatE、CompoundE（二维

版本）和 CompoundE3D。在同一张表中，我们还列出了每种关系的 Khs$_{G_r}$ 和 ξ_{G_r} 值，以考察其层级行为是否更强。从表中可以看出，CompoundE 和 CompoundE3D 在几乎所有关系上的表现都优于 TransE 和 RotatE。此外，除“成员部分词”外，CompoundE3D 在所有层级关系上的表现都优于 CompoundE。这一结果表明，通过纳入更多样的三维变换，CompoundE3D 能够比 CompoundE 更有效地建模层级关系。

7）模型效率：研究模型性能与模型维度之间的关系十分重要。模型维度反映了内存和计算复杂度。为展示 CompoundE3D 相较于先前模型在各种嵌入维度下的优势，我们在图 7 中绘制了 Wikikg2 数据集上的链接预测平均倒数排名性能，其中维度值设为 12、24、48、102、150、198、252 和 300。从图中可以看出，CompoundE3D 在所有维度下均持续优于所有基准模型。此外，我们从自由参数数量的角度分析了不同知识图谱嵌入模型的复杂度。表 XIII 比较了不同知识图谱嵌入模型在 ogbl-wikikg2 数据集上的自由参数数量。

报告基线模型所使用的自由参数数量时，我们参考 ogbl-wikikg2 排行榜。CompoundE3D 报告的参数数量对应的嵌入维度为 90。如图 7 和表 XIII 所示，CompoundE3D 在所有基准模型中性能最佳，同时自由参数数量最少。

V. 结论与未来工作

本文提出了一种基于三维空间中复合仿射变换的新颖且有效的知识图谱嵌入（KGE）模型，称为 CompoundE3D。我们设计了一种束搜索过程，从最简单的配置逐步构建出所需的 KGE。我们还探索了对排名前 k 的模型变体进行集成，以进一步提升链接预测性能。我们通过大量实验结果展示了 CompoundE3D 的优越性能。我们进行了消融研究，以评估各个算子的作用，并开展了案例研究，以阐明 CompoundE3D 在多重性、对称关系和层级关系等若干关系类型上的建模能力。

关于未来的研究方向，探索 CompoundE3D 在实体类型判定 [77] 和实体对齐[78]等其他重要知识图谱任务中的有效性，将会很有意义。此外，在低维嵌入空间中提升性能的研究对于实际应用具有重要价值，值得进一步探讨。

致谢

作者感谢南加州大学高级计算研究中心（CARC）提供计算资源，这些资源为本文报告的研究成果作出了贡献。网址：https: //carc.usc.edu。

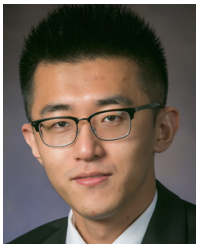
参考文献

[1] T. Yu、J. Li、Q. Yu、Y. Tian、X. Shun、L. Xu、L. Zhu 和 H. Gao，“中医养生知识图谱：设计、构建与应用”，Artif. Intell. Med.，第 77 卷，第 48–52 页，2017年。[2] S. Sang、Z. Yang、L. Wang、X. Liu、H. Lin 和 J. Wang，“Sematyp：一种基于知识图谱的药物发现文献挖掘方法”，BMC Bioinform.，第 19 卷，第 1–11 页，2018年。[3] X. Zeng、X. Tu、Y. Liu、X. Fu 和 Y. Su，“借助知识图谱推动更好的药物发现”，Curr. Opin. Struct. Biol.，第 72 卷，126，2022年 114页。– 。[4] H. Wang、F. Zhang、J. Wang、M. Zhao、W. Li、X. Xie 和 M. Guo，“RippleNet：在知识图谱上传播用户偏好以用于推荐系统”，载于第 27 届 ACM 国际信息与知识管理会议（CIKM’18）论文集，2018年，第 417–426 页。[5] K. Zhou、W. X. Zhao、S. Bian、Y. Zhou、J.-R. Wen 和 J. Yu，“通过基于知识图谱的语义融合改进对话式推荐系统”，载于第 26 届 ACM SIGKDD 国际知识发现与数据挖掘会议（KDD ’20）论文集，2020年，第 1006–1014 页。[6] Q. Zhan 和 H. Yin，“一种基于知识图谱和神经网络的贷款申请欺诈检测方法”，载于第 2 届国际人工智能创新会议论文集，2018年，第 111–115 页。[7] X. Zhu、X. Ao、Z. Qin、Y. Chang、Y. Liu、Q. He 和 J. Li，“后疫情时代的智能金融欺诈检测实践”，The Innov.，第 2 卷，第 4 期，第 100176 页，2021年。[8] H. He、A. Balakrishnan、M. Eric 和 P. Liang，“利用动态知识图谱嵌入学习对称协作式对话代理”，载于第 55 届计算语言学协会年会（ACL 2017）论文集，2017年，第 1766–1776 页。[9] A. Ait-Mlouk 和 L. Jiang，“Kbot：一种用于关联数据自然语言理解的基于知识图谱的聊天机器人”，IEEE Access，第 8 卷，第 149 220–149 230 页，2020年。[10] S. Auer、C. Bizer、G. Kobilarov、J. Lehmann、R. Cyganiak 和 Z. Ives，“DBpedia：开放数据网络的核心”，载于《语义网》。Springer，2007年，第 722–735 页。[11] F. M. Suchanek、G. Kasneci 和 G. Weikum，“YAGO：语义知识的核心”，载于第 16 届国际万维网会议（WWW’07）论文集，2007年，第 697–706 页。[12] K. Bollacker、C. Evans、P. Paritosh、T. Sturge 和 J. Taylor，“Freebase：一个协作创建的图数据库，用于组织人类知识”，载于 2008 年 ACM SIGMOD 国际数据管理会议（SIGMOD ’08）论文集，2008年，第 1247–1250 页。[13] A. Carlson、J. Betteridge、B. Kisiel、B. Settles、E. R. Hruschka 和 T. M. Mitchell，“构建永不停歇的语言学习架构”，载于第 24 届 AAAI 人工智能会议论文集，2010年，第 1306–1313 页。[14] D. Vrande˘ci´c 和 M. Kr¨otzsch，“Wikidata：一个免费的协作式知识库”，Commun. ACM，第 57 卷，第 10 期，第 78–85 页，2014年。[15] R. Speer、J. Chin 和 C. Havasi，“ConceptNet 5.5：一个开放的多语言常识图”，载于第 31 届 AAAI 人工智能会议论文集，2017年，第 4444–4451 页。[16] A. Bordes、N. Usunier、A. Garcia-Duran、J. Weston 和 O. Yakhnenko，“用于多关系数据建模的嵌入平移”，《神经信息处理系统进展》第 26 卷（NeurIPS 2013），第 2787–2795 页，2013年。[17] Z. Sun、Z.-H. Deng、J.-Y. Nie 和 J. Tang，“RotatE：通过复数空间中的关系旋转进行知识图谱嵌入”，载于第 8 届国际学习表征会议（ICLR）论文集，2019年，第 1–18 页。[18] L. Chao、J. He、T. Wang 和 W. Chu，“PairRE：通过成对关系向量实现知识图谱嵌入”，载于第 59 届计算语言学协会年会（ACL 2021）论文集，2021年，第 4360–4369 页。[19] W. K. Pratt，《数字图像处理导论》。CRC 出版社，2013年。[20] X. Ge、Y.-C. Wang、B. Wang 和 C.-C. J. Kuo，“CompoundE：通过平移、旋转和缩放复合运算实现知识图谱嵌入”，arXiv 预印本 arXiv:2207.05324，2022年。[21] C. Gao、C. Sun、L. Shan、L. Lin 和 M. Wang，“Rotate3D：将关系表示为三维空间中的旋转以实现知识图谱嵌入”，载于第 29 届 ACM 国际信息与知识管理会议（CIKM’20）论文集，2020年，第 385–394 页。[22] T. Yang、L. Sha 和 P. Hong，“NagE：知识图谱的非阿贝尔群嵌入”，载于第 29 届 ACM 国际信息与知识管理会议（CIKM’20）论文集，2020年，第 1735–1742 页。[23] H. Lu、H. Hu 和 X. Lin，“DensE：一种具有自适应语义层次结构的增强型知识图谱嵌入非交换表示”，Neurocomputing，第 476 卷，第 115–125 页，2022年。[24] S. Zhang、Y. Tay、L. Yao 和 Q. Liu，“四元数知识图谱嵌入”，载于《神经信息处理系统进展》第 32 卷（NeurIPS 2019），2019年

[25] I. Chami、A. Wolf、D.-C. Juan、F. Sala、S. Ravi 和 C. R´e，《低维双曲知识图谱嵌入》，载于《第58届计算语言学协会年会论文集》（ACL 2020），2020年，第6901–6914页。[26] K. Wang、Y. Liu、D. Lin 和 Q. Z. Sheng，《双曲几何并非必需：用于低维知识图谱嵌入的轻量级欧几里得模型》，arXiv 预印本 arXiv:2103.14930，2021年。[27] X. Dong、E. Gabrilovich、G. Heitz、W. Horn、N. Lao、K. Murphy、T. Str ohmann、S. Sun 和 W. Zhang，《Knowledge Vault：一种用于概率知识融合的网络规模方法》，载于《第20届ACM SIGKDD知识发现与数据挖掘国际会议论文集》（KDD '14），2014年，第601–610页。[28] R. Socher、D. Chen、C. D. Manning 和 A. Ng，《利用神经张量网络进行知识库补全推理》，《神经信息处理系统进展》第26卷（NeurIPS 2013），第26卷，2013年。[29] T. Dettmers、P. Minervini、P. Stenetorp 和 S. Riedel，《卷积二维知识图谱嵌入》，载于《第32届AAAI人工智能会议论文集》，2018年，第1811–1818页。[30] M. Schlichtkrull、T. N. Kipf、P. Bloem、R. Van Den Berg、I. Titov 和 M. Welling，《利用图卷积网络对关系数据进行建模》，载于《第15届扩展语义网会议论文集》（ESWC'18），Springer，2018年，第593–607页。[31] B. Yang、W.-t. Yih、X. He、J. Gao 和 L. Deng，《为知识库中的学习与推理嵌入实体和关系》，载于《第4届国际学习表征会议论文集》（ICLR），2015年，第1–13页。[32] L. Yao、C. Mao 和 Y. Luo，《KG-BERT：用于知识图谱补全的BERT》，arXiv 预印本 arXiv:1909.03193，2019年。[33] J. Devlin、M.-W. Chang、K. Lee 和 K. Toutanova，《BERT：用于语言理解的深度双向变换器预训练》，arXiv 预印本 arXiv:1810.04805，2018年。[34] L. Wang、W. Zhao、Z. Wei 和 J. Liu，《SimKGC：基于预训练语言模型的简单对比知识图谱补全》，载于《第60届计算语言学协会年会论文集》（ACL 2022），2022年，第4281–4294页。[35] Y.-C. Wang、X. Ge、B. Wang 和 C.-C. J. Kuo，《KGBoost：一种采用负采样的基于分类的知识库补全方法》，《模式识别快报》，第157卷，第104–111页，2022年。[36] T. Chen 和 C. Guestrin，《XGBoost：一种可扩展的树提升系统》，载于《第22届ACM SIGKDD知识发现与数据挖掘国际会议论文集》（KDD '16），2016年，第785–794页。[37] Y. Yang、W. Wang、H. Fu、C.-C. J. Kuo 等，《关于从高维特征空间中进行监督特征选择》，《APSIPA信号与信息处理汇刊》，第11卷，第1期，2022年。[38] C.-C. J. Kuo 和 A. M. Madni，《绿色学习：介绍、示例与展望》，《视觉传播与图像表示杂志》，第103685页，2022年。[39] Y.-C. Wang、X. Ge、B. Wang 和 C.-C. J. Kuo，《GreenKGC：一种轻量级知识图谱补全方法》，arXiv 预印本 arXiv:2208.09137，2022年。[40] Z. Wang、J. Zhang、J. Feng 和 Z. Chen，《通过在超平面上平移进行知识图谱嵌入》，载于《第28届AAAI人工智能会议论文集》，2014年，第1112–1119页。[41] Y. Lin、Z. Liu、M. Sun、Y. Liu 和 X. Zhu，《学习用于知识图谱补全的实体与关系嵌入》，载于《第29届AAAI人工智能会议论文集》，2015年，第2181–2187页。[42] G. Ji、S. He、L. Xu、K. Liu 和 J. Zhao，《通过动态映射矩阵进行知识图谱嵌入》，载于《第53届计算语言学协会年会论文集》（ACL 2015），第1卷，2015年，第687–696页。[43] G. Ji、K. Liu、S. He 和 J. Zhao，《采用自适应稀疏迁移矩阵进行知识图谱补全》，载于《第30届AAAI人工智能会议论文集》，2016年，第985–991页。[44] J. Yang、Y. Shi、X. Tong、R. Wang、T. Chen 和 X. Ying，《利用与各关系对应的实体仿射变换改进知识图谱嵌入》，载于《自然语言处理实证方法会议论文集》（EMNLP'21 Findings），2021年，第508–517页。[45] Z. Liang、J. Yang、H. Liu 和 K. Huang，《一种基于关系的知识图谱补全语义过滤器》，载于《自然语言处理实证方法会议论文集》（EMNLP'21），2021年，第7920–7929页。[46] J. Li 和 Y. Yang，《STaR：通过缩放、平移和旋转进行知识图谱嵌入》，载于《第11届人工智能移动服务会议论文集》（AIMS 2022），Springer，2022年，第31–45页。[47] Q. Wang、Z. Mao、B. Wang 和 L. Guo，《知识图谱嵌入：方法与应用综述》，《IEEE知识与数据工程汇刊》，第29卷，第12期，第2724–2743页，2017年。[48] Z. Cao、Q. Xu、Z. Yang、X. Cao 和 Q. Huang，《双四元数知识图谱嵌入》，载于《第35届AAAI人工智能会议论文集》，第35卷，第8期，2021年，第6894–6902页。

[49] J. Guo 和 S. Kok，《Bique：知识图谱的双四元数嵌入》，载于《自然语言处理经验方法会议论文集》（EMNLP'21），2021年，第8338–8351页。[50] Q. Zhang、R. Wang、J. Yang 和 L. Xue，《通过反射变换进行知识图谱嵌入》，《基于知识的系统》，第238卷，107861，2022年。[51] Z. Cao、T. Qin、T.-Y. Liu、M.-F. Tsai 和 H. Li，《学习排序：从成对方法到列表式方法》，载于第24届国际机器学习会议论文集（ICML 2007），2007年，第129–136页。[52] A. Klementiev、D. Roth 和 K. Small，《基于距离的模型进行无监督排序聚合》，载于第25届国际机器学习会议论文集（ICML 2008），2008年，第472–479页。[53] G. V. Cormack、C. L. Clarke 和 S. Buettcher，《倒数排名融合优于孔多塞法和单独的排序学习方法》，载于第32届 ACM SIGIR 信息检索研究与开发年会论文集（SIGIR '09），2009年，第758–759页。[54] X. Chen、M. Chen、C. Fan、A. Uppunda、Y. Sun 和 C. Zaniolo，《通过集成知识迁移进行多语言知识图谱补全》，载于自然语言处理经验方法会议论文集：研究成果（EMNLP'20 Findings），2020年，第3227–3238页。[55] Y. Zhang、Q. Yao、W. Dai 和 L. Chen，《AutoSF：为知识图谱嵌入搜索评分函数》，载于第36届 IEEE 国际数据工程会议论文集（ICDE 2020），2020年，第433–444页。[56] C. Xu、M. Nayyeri、S. Vahdati 和 J. Lehmann，《利用低维知识图谱嵌入进行多次运行集成学习》，载于2021年国际神经网络联合会议论文集（IJCNN）。IEEE，2021年，第1–8页。[57] U. Joshi 和 J. Urbani，《基于集成的知识图谱嵌入事实分类》，载于第19届扩展语义网会议论文集（ESWC'22）。Springer，2022年，第147–164页。[58] M. Van Erp 和 L. Sc homaker，《用于组合排序分类器假设的博尔达计数法变体》，载于第7届手写识别前沿国际研讨会。International Unipen Foundation，2000年，第443–452页。[59] P. Bailey、A. Moffat、F. Scholer 和 P. Thomas，《存在查询变体时的检索一致性》，载于第40届 ACM SI GIR 信息检索研究与开发年会论文集（SIGIR '17），2017年，第395–404页。[60] B. Ding、Q. Wang、B. Wang 和 L. Guo，《利用简单约束改进知识图谱嵌入》，arXiv预印本 arXiv:1805 02408，2018年。[61] F. Mahdisoltani、J. Biega 和 F. Suchanek，《YAGO3：来自多语言维基百科的知识库》，载于第7届双年数据系统研究创新会议论文集。CIDR Conference，2014年。[62] W. Hu、M. Fey、M. Zitnik、Y. Dong、H. Ren、B. Liu、M. Catasta 和 J. Leskovec，《开放图基准：用于图机器学习的数据集》，arXiv预印本 arXiv:2005.00687，2020年。[63] D. P. Kingma 和 J. Ba，《Adam：一种随机优化方法》，arXiv预印本 arXiv:1412.6980，2014年。[64] Nickel、M. 等、L. Rosasco 和 T. Poggio，《知识图谱的全息嵌入》，载于第30届 AAAI 人工智能会议论文集，2016年。[65] T. Trouillon、J. Welbl、S. Riedel、 ´E. Gaussier 和 G. Bouchard，《用于简单链接预测的复数嵌入》，载于第33届国际机器学习会议论文集（ICML 2016），2016年，第2071–2080页。[66] H. Liu、Y. Wu 和 Y. Yang，《多关系嵌入的类比推理》，载于第34届国际机器学习会议论文集（ICML 2017），2017年，第2168–2178页。[67] S. Guo、Q. Wang、L. Wang、B. Wang 和 L. Guo，《通过软规则的迭代指导进行知识图谱嵌入》，载于第32届 AAAI 人工智能会议论文集，第32卷，第1期，2018年。[68] W. Xu、S. Zheng、L. He、B. Shao、J. Yin 和 T.-Y. Liu，《SEEK：知识图谱的分段嵌入》，arXiv预印本 arXiv:2005.00856，2020年。[69] F. Ren、J. Li、H. Zhang、S. Liu、B. Li、R. Ming 和 Y. Bai，《利用空洞卷积和残差学习进行知识图谱嵌入》，载于第58届计算语言学协会年会论文集（ACL 2020），2020年，第1532–1543页。[70] Y. Li、W. Fan、C. Liu、C. Lin 和 J. Qian，《TranSHER：通过超椭球限制转换知识图谱嵌入》，载于第60届计算语言学协会年会论文集（ACL 2022），2022年12月，第8517–8528页。[71] L. Yu、Z. Luo、H. Liu、D. Lin、H. Li 和 Y. Deng，《TripleRE：通过三重关系向量进行知识图谱嵌入》，arXiv预印本 arXiv:2209.08271，2022年。[72] C. Xu 和 R. Li，《知识图谱中的二面体群关系嵌入》，载于第57届计算语言学协会年会论文集（ACL 2019），2019年，第263–272页。

[73] S. Vashishth、S. Sanyal、V. Nitin、N. Agrawal 和 P. Talukdar，《Interacte：通过增加特征交互改进基于卷积的知识图谱嵌入》，载于第*34*届 *AAAI* 人工智能会议论文集，第34卷，2020年，第3009–3016页。[74] Z. Zhang、J. Cai、Y. Zhang 和 J. Wang，《学习用于链接预测的层级感知知识图谱嵌入》，载于第*34*届 *AAAI* 人工智能会议论文集，第34卷，2020年，第3065–3072页。[75] T. Song、J. Luo 和 L. Huang，《Rot-Pro：通过知识图谱嵌入中的投影建模传递性》，神经信息处理系统进展第*34*卷（*NeurIPS 2021*），第34卷，第24 695–24 706页，2021年。[76] Y. Chen、P. Minervini、S. Riedel 和 P. Stenetorp，《将关系预测作为辅助训练目标，以改进多关系图表示》，载于第*3*届自动知识库构建会议论文集（*AKBC2021*），2021年。[77] X. Ge、Y.-C. Wang、B. Wang 和 C. J. Kuo，《CORE：一种通过复数空间回归和嵌入进行知识图谱实体类型预测的方法》，《模式识别快报》，第157卷，第97–103页，2022年。[78] X. Ge、Y. C. Wang、B. Wang、C.-C. J. Kuo 等，《Typeea：用于知识图谱实体对齐的类型关联嵌入》，《*APSIPA* 信号与信息处理汇刊》，第12卷，第1期，2023年。



Xiou Ge（S’14）于2016年和2018年分别获得伊利诺伊大学厄巴纳-香槟分校（位于伊利诺伊州厄巴纳）的电气与计算机工程理学学士和理学硕士学位。目前，他正在南加州大学明谢电气与计算机工程系攻读电气与计算机工程哲学博士学位。他的研究兴趣包括知识图谱嵌入、节点分类、图对齐和自然语言处理。



Yun-Cheng Wang（S’19）于2018年获得国立台湾大学（位于台湾台北）的电气工程学士学位。他于2020年获得理学硕士学位，目前正在南加州大学明谢电气与计算机工程系攻读电气与计算机工程哲学博士学位。他的研究兴趣包括知识图谱嵌入和自然语言处理。



Bin Wang（S’19-M’22）于2017年6月获得中国成都电子科技大学（UESTC）电子信息工程专业工程学学士学位，并于2021年5月获得美国加利福尼亚州洛杉矶南加州大学（USC）电气工程专业哲学博士学位。他目前是新加坡新加坡国立大学（NUS）的研究员。他的研究兴趣包括表示学习、图学习和对话摘要。



C.-C. Jay Kuo（F’99）于1980年获得台湾台北国立台湾大学电气工程专业理学学士学位，并分别于1985年和1987年获得剑桥麻省理工学院电气工程专业理学硕士和哲学博士学位。他现任William M. Hogue电气与计算机工程讲席教授，同时是电气与计算机工程及计算机科学杰出教授，并担任南加州大学多媒体通信实验室（MCL）主任，南加州大学。他是约340篇期刊论文、1000篇会议论文和15本书的合著者。他的研究兴趣包括多媒体和绿色计算。他是国家发明家科学院、美国科学促进会、国际光学工程师学会以及计算机协会的会士。