arXiv:2302.04088v1 [cs.LG] 7 Feb 2023

FFHR：用于知识图谱补全的完全灵活双曲表示

Wentao Shi、Junkang Wu、Xuezhi Cao、Jiawei Chen、Wenqiang Lei、Wei Wu、Xiangnan He，
IEEE 高级会员

摘要——由于双曲嵌入在捕获层级结构方面具有优势，针对知识图谱（KG）的双曲嵌入学习日益受到关注。然而，双曲空间中的一些重要操作仍缺乏良好的定义，导致现有方法无法充分发挥双曲空间的优势。具体而言，它们存在两项主要局限：1）现有的双曲空间图卷积网络（GCN）方法依赖切空间近似，这会在表示学习中引入近似误差；2）由于缺少双曲空间中内积操作的定义，现有方法只能通过双曲距离衡量事实（链接）的可信度，因此难以捕获复杂数据模式。

在这项工作中，我们的贡献包括：1）提出一种完整的 Poincar´e 多关系图卷积网络，无需任何近似即可在双曲空间中实现图信息传播；2）将欧几里得内积推广到双曲空间，有助于捕捉层次化和复杂模式。在此基础上，我们进一步开发了一种**完**整且**灵**活的**双**曲**表**示框架（**FFHR**），可将近期基于欧几里得空间的进展迁移至双曲空间。我们通过将 FFHR 实例化为四种具有代表性的知识图谱补全方法来验证该框架。在基准数据集上进行的大量实验表明，与其欧几里得对应模型以及最先进的双曲嵌入方法相比，我们的 FFHR 均具有优越性。

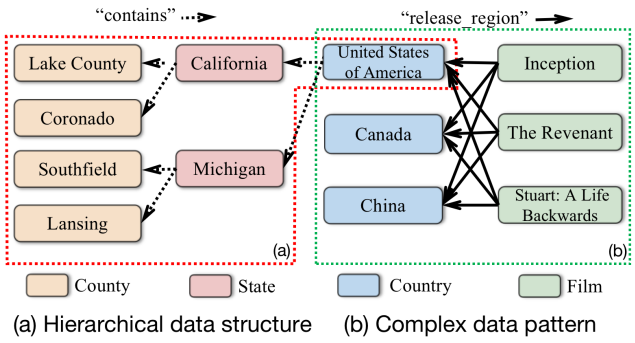
索引词——图神经网络、双曲表示、知识图谱补全

1 引言

KNOWLEDGE 识图谱已助力许多下游任务，例如对话生成 [1] [2]，、推荐系统 [3] [4] 和问答 [5]。尽管知识图谱中已包含大量实体和事实，但仍远未完整。这一不足催生了一个热门研究课题——知识图谱补全，旨在发现给定知识图谱中缺失的事实。

嵌入方法在知识图谱补全中十分普遍，它们将实体和关系映射到潜在空间，并使用评分函数衡量事实的可信度。近期研究主要聚焦于在欧几里得空间中嵌入知识图谱。然而，由于知识图谱通常具有层级结构（图 1 给出了一个示例），而欧几里得空间并未对数据内在的几何模式作出假设，因此难以保留此类层级结构。在欧几里得空间中描绘层级结构需要极高维的嵌入，这不可避免地会带来高昂的时间和内存成本 [6] [7]。

近年来，双曲空间已被用于知识图谱补全 [8] [9]，，为学习低维嵌入同时保留潜在层次结构提供了令人振奋的机会。双曲空间可以被视为一种



(a) Hierarchical data structure (b) Complex data pattern

图1. 一个同时包含层次化数据结构和复杂数据模式的知识图谱玩具示例。

离散树的连续对应物，具有强大的层次模式捕捉能力——例如，它仅用两个维度就能以任意低的失真嵌入任意树 [10]。尽管取得了令人瞩目的成果，我们认为现有方法尚未充分利用双曲空间。双曲空间中的许多重要运算仍缺乏良好的定义，导致模型无法继承近期先进的基于欧几里得空间的模型的优点。具体而言，它们存在两个主要局限：

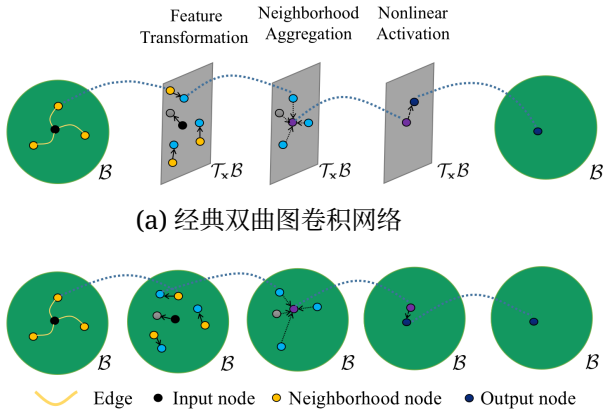
(1) 图卷积网络在知识图谱补全中的有效性已得到验证 [11] [12]。为了将图卷积网络转换到双曲空间，现有方法 [13] [14] 通常依赖切空间进行特征变换和邻域聚合。然而，由于切空间只是双曲空间的局部近似，通过切空间执行图卷积网络

• *W. Shi*、*J. Wu*、*J. Chen* 和 *X. He* 曾就职于中国科学技术大学信息科学与技术学院。邮箱：{*shiwentao123*、*jkwu0909*}*@mail.ustc.edu.cn*、*cjwustc@ustc.edu.cn*、*xiangnanhe@gmail.com*。

• *X. Cao* 和 *W. Wu* 曾就职于美团。邮箱：*caoxuezhi@meituan.com*、*wuwei19850318@gmail.com*。

• *W. Lei*曾就职于四川大学。电子邮箱：*wenqianglei@gmail.com M*

稿件于*2005*年*4*月*19*日收到；于*2015*年*8*月*26*日修订，*2015*年。



(a) 经典双曲图卷积网络

(b) 我们提出的全庞加莱多关系图卷积网络

图2。(a) 经典双曲图卷积网络将所有运算定义在切空间中，由于近似处理，无法保持双曲结构。(b) Pur提出的FPM-GCN完全基于庞加莱模型，不进行近似。

不可避免地会产生近似误差，进而扭曲双曲流形的全局结构。

(2) 现有的双曲方法总是使用双曲距离作为评分函数，这通常不足以捕获复杂关系（图1b）。事实上，对于欧几里得空间中的方法，最先进的结果往往由采用内积评分函数的方法取得。内积更加灵活有效，但在双曲空间中仍缺少明确的对应定义。

为解决这些局限，本文提出两种重要的双曲运算：

(1) 全双曲庞加莱多关系图卷积网络（**FPM-GCN**）：该方法基于庞加莱球模型，无需进行切空间近似（如图 2b 所示）。我们通过两项改进实现这一点：在特征变换模块中利用双曲等距群变换，并直接在双曲空间中采用陀螺中点，而非在切空间中采用。通过这两项设计，FPM-GCN 能够保持双曲结构，并更充分地利用双曲空间。

(2) 双曲内积（**HIN**）：它是欧几里得内积的双曲推广，相较于双曲距离具有更大的灵活性。HIN 的推导在理论上是严谨的：我们转而采用庞加莱球模型的莫比乌斯陀螺向量空间，仔细类比欧几里得内积的定义，并参考双曲余弦定律 [15]。

基于这两项重要提议，我们构建了一个**Fully and FlexibleHyperbolicRepresentation**框架（**FFHR**），能够将近期基于欧几里得的方法迁移到双曲空间，并更有效地发挥双曲空间的优势。FFHR并非一个具体模型，而是一套高层次的解决方案蓝图。我们通过将FFHR实例化为四种具有代表性的知识图谱补全方法来展示其应用，包括DistMult[16],ComplEx [17], RESCAL [18]和DualE [19]。我们进一步通过实验证明了它们相较于欧几里得对应模型和双曲空间领域最先进方法的优越性。

简而言之，本文的主要贡献总结如下：

- 提出了 FPM-GCN，其将所有运算定义在庞加莱球模型中。这些运算无需进行切空间近似，因此能够更好地保留双曲结构。
- 提出双曲内积（HIN）——欧几里得内积的双曲推广形式；在建模复杂数据模式方面，它比双曲距离评分函数更加灵活。
- 构建FFHR框架，将近期先进的基于欧几里得的方法迁移到双曲空间，并将其实例化为四种具有代表性的知识图谱补全方法，以验证其有效性。

2 相关工作

我们总结了双曲表示用于知识图谱补全的两个主要局限。本节将详细回顾经典的欧氏知识图谱补全方法和双曲嵌入方法。

2.1 欧氏嵌入模型

根据不同的相似度度量，知识图谱补全方法大致可分为两类。i）平移方法，例如 TransE [20], TransH [21], RotatE [6], , 通过计算实体之间的距离来衡量事实的可信度。ii）语义匹配方法，例如 TuckER [22], QuatE [23],DihEdral [24], 基于内积评分函数来衡量相似度。然而，这两类方法都需要高维空间才能达到最先进水平，这会导致较高的内存成本。

近年来，基于图卷积网络的方法也吸引了大量研究关注。R-GCN [11] 首次将图卷积网络引入多关系图，并利用图结构信息来提升预测性能。CompGCN [12] 缓解了过参数化问题，并提出了一个用于知识图谱补全的通用框架。得益于强大的信息聚合能力，图卷积网络对于建模复杂数据模式至关重要。

2.2 双曲嵌入

双曲几何是具有恒定负曲率的非欧几何。 [25] [26] 首先将同质图嵌入双曲空间；与嵌入欧几里得空间相比，在嵌入层次数据时表现出显著提升。近年来，MuRP [8] 和 AttH [9] 将双曲嵌入推广到多关系图。然而，它们都使用双曲距离作为评分函数，缺乏灵活性，难以捕捉多关系图中的复杂数据模式。

此外，还有一项工作 [27] 尝试将双曲嵌入映射到切空间，并计算它们之间的欧几里得内积，将其作为双曲内积。然而，由于切空间只是双曲空间的局部近似，这不可避免地会导致近似误差。在实践中，它最终会退化为使用欧几里得内积的等价形式。

2.3 双曲图神经网络

HNN [28] 首先推广了基本的双曲运算工具（例如 MLR、GRU），并形式化了双曲神经

网络层，所依据的是 [29] [30]提出的理论。随后，HNN ++ [31]提出了一种改进的双曲 MLR 层，并将其推广至卷积层。然而，通过全连接层获得的特征彼此独立，因此缺乏明确的几何含义，也无法保持双曲结构。与之不同，HGCN [32]和HGNN [33]同时提出了双曲图卷积网络，其核心思想相同。它们将大多数运算（特征变换、邻域聚合）形式化于切空间中，这会导致表示学习中的失真。

最近，H2HGCN [34]提出了一种改进的双曲到双曲图卷积网络，其运算定义在不同的等距双曲模型中。该方法避免了切空间近似，但引入了不同双曲模型之间不必要的变换，这会导致数值不稳定性。此外，我们指出，上述所有方法均针对同质图定义，不适用于多关系知识图谱。

有少数工作（如 M^{2}GNN [13] 和 H2E [35]）将 HGNN 推广到多关系图，以便更好地捕捉层次化数据结构，但它们将所有运算都定义在切空间中。更全面的分析可参见近期的一篇综述 [14]。与这些方法相比，我们的 FPM-GCN 定义在庞加莱模型中，不存在任何近似误差。此外，我们的 FPM-GCN 无需像 H2HGCN [34],那样在不同的等距模型之间进行变换，因此更加稳定且具有表达性。

3 预备知识

本节首先给出知识图谱补全任务的形式化定义，然后介绍一些双曲几何的背景知识。

3.1 任务定义。

设 $\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{L}\}$ 是知识图谱的一个实例，其中 $\mathcal{E}$、$\mathcal{R}$ 和 $\mathcal{L}$ 分别表示实体、关系和事实的集合。每个事实 $f \in \mathcal{L}$ 都表示为一个三元组 $(h,\ r,\ t) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$，描述从头实体 h 到尾实体 t 存在关系 $r \in \mathcal{R}$。知识图谱补全任务旨在推断知识图谱中的缺失边。其目标是学习一个函数 $s:\ \mathcal{E} \times \mathcal{R} \times \mathcal{E} \to \mathbb{R}$，为有效三元组赋予比无效三元组更高的分数。

3.2 黎曼流形。

n 维流形 $\mathcal{M}$ 是一个拓扑空间，其中每个点 $\mathbf{x} \in \mathcal{M}$ 的邻域在局部都可以用切空间 $\mathcal{T}_{\mathbf{x}}\mathcal{M}$近似。配备度量张量 $g_{\mathbf{x}}$ 的流形 $\mathcal{M}$称为黎曼流形，记作 $(\mathcal{M}, g)$。度量张量 $g_{\mathbf{x}}$：$\mathcal{T}_{\mathbf{x}}\mathcal{M} \times \mathcal{T}_{\mathbf{x}}\mathcal{M} \to \mathbb{R}$提供了关于 $\mathcal{T}_{\mathbf{x}}\mathcal{M}$中切向量的角度和长度的局部信息，通过积分可得到全局距离。全局距离定义为$d(\mathbf{x}, \mathbf{y}) = \int_0^1 \sqrt{g_{\gamma(t)}(\dot{\gamma}(t)\ \dot{\gamma}(t))}\mathrm{d}t$，其中光滑路径 γ是测地线，满足 $\gamma(0) = \mathbf{x}$ 且 ~~$\gamma(1) = \mathbf{y}$~~。指数映射 $\exp_{\mathbf{x}}(\mathbf{v})$：$\mathcal{T}_{\mathbf{x}}\mathcal{M} \to \mathcal{M}$ 将切向量 $\mathbf{v} \in \mathcal{T}_{\mathbf{x}}\mathcal{M}$ 沿测地线映射到流形上。反之，对数映射 $\log_{\mathbf{x}}(\mathbf{y}):\ \mathcal{M} \to \mathcal{T}_{\mathbf{x}}\mathcal{M}$将 $\mathbf{y} \in \mathcal{M}$ 投影回切空间，自然满足对数映射$_{\mathbf{x}}$(指数映射$_{\mathbf{x}}(\mathbf{v})) = \mathbf{v}$。

3.3 双曲空间与庞加莱球模型。

双曲空间是一种具有恒定负曲率的特定黎曼流形。在五种不同的双曲空间等距模型中，我们选择庞加莱球模型进行推导，因为它支持一些重要的向量运算 [36] [37],，例如莫比乌斯加法、莫比乌斯矩阵-向量乘法。这些向量运算有助于在双曲空间中构建图神经网络。形式上，设$(\mathcal{B}_c^n, g^c)$ 为曲率恒定为 $-c$ 的庞加莱球模型。其黎曼度量为：

$$g_{\mathbf{x}}^c = (\lambda_{\mathbf{x}}^c)^2 \mathbf{I}_n, \quad \lambda_{\mathbf{x}}^c = \frac{2}{1 - c||\mathbf{x}||^2} \tag{1}$$

莫比乌斯加法和莫比乌斯矩阵-向量乘法如下所示：

$$\mathbf{x} \oplus_c \mathbf{y} = \frac{(1 + 2c\left\langle \mathbf{x}, \mathbf{y} \right\rangle + c||\mathbf{y}||^2)\mathbf{x} + (1 - c||\mathbf{x}||^2)\mathbf{y}}{1 + 2c\left\langle \mathbf{x}, \mathbf{y} \right\rangle + c^2||\mathbf{x}||^2||\mathbf{y}||^2}, \tag{2}$$

$$\mathbf{W} \otimes_c \mathbf{x} = exp_{\mathbf{0}}^c(\mathbf{W} \cdot log_{\mathbf{0}}^c(\mathbf{x})), \tag{3}$$

其中，$exp_{\mathbf{0}}^c$ 和 $log_{\mathbf{0}}^c$ 分别是$(\mathcal{B}_c^n, g^c)$上的指数映射和对数映射，而$\mathbf{0}$ 是庞加莱球模型的原点。利用莫比乌斯加法，我们可以用更清晰的几何解释重写一些重要公式。连接点$\mathbf{x}$,$\mathbf{y} \in \mathcal{B}_c$ 的测地线为：

$$\gamma_{\mathbf{x} \to \mathbf{y}}(t) := \mathbf{x} \oplus_c (-\mathbf{x} \oplus_c \mathbf{y}) \otimes_c t,$$
$$s.t. \gamma_{\mathbf{x} \to \mathbf{y}}(0) = \mathbf{x}\ and\ \gamma_{\mathbf{x} \to \mathbf{y}}(1) = \mathbf{y}.$$

指数映射和对数映射的闭式推导如下：

$$exp_{\mathbf{x}}^c(\mathbf{v}) = \mathbf{x} \oplus_c \left(tanh\left(\sqrt{c}\frac{\lambda_{\mathbf{x}}^c||\mathbf{v}||}{2} \right) \frac{\mathbf{v}}{\sqrt{c}||\mathbf{v}||} \right),$$

$$log_{\mathbf{x}}^c(\mathbf{y}) = \frac{2}{\sqrt{c}\lambda_{\mathbf{x}}^c} artanh(\sqrt{c}||-\mathbf{x} \oplus_c \mathbf{y}||)\frac{-\mathbf{x} \oplus_c \mathbf{y}}{||-\mathbf{x} \oplus_c \mathbf{y}||},$$

其中 $\mathbf{x} \in \mathcal{B}_c$、$\mathbf{v} \in \mathcal{T}_{\mathbf{x}}\mathcal{B}_c$，且 $\lambda_{\mathbf{x}}^c$ 是 $\mathcal{B}_c$的共形因子。仔细观察这些公式，可以发现它们与欧几里得对应模型非常相似。与欧几里得加法不同，M¨obius 加法是一种非交换且非结合的运算。

此外，双曲距离函数定义如下：

$$d^{\mathcal{B}_c}(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{c}} arctanh(\sqrt{c}||-\mathbf{x} \oplus_c \mathbf{y}||). \tag{4}$$

借助 exp 和 log 映射函数，先前的工作提出了经典的特征变换和邻域聚合模块，它们定义在原点的切空间中。

如果您想进一步了解庞加莱´球模型，可以参考 [30]。

3.4 图卷积网络

给定图 $\mathcal{G} = (\mathcal{E}, \mathcal{R})$，其实体集为 $\mathcal{E}$，关系集为$\mathcal{R}$，我们可以将一般的欧氏图卷积网络概括为两个主要步骤。令 x_i^l 表示第 l层中第 i 个实体的欧氏表示。

3.4.1 特征变换

$$\mathbf{m}_i^{l,E} = \mathbf{W}_r^l \mathbf{x}_i^{l-1,E} + \mathbf{b}_r^l \tag{5}$$

其中，$\mathbf{W}_r^l$ 表示关系 r 的变换矩阵，$\mathbf{b}_r$表示关系 r 的偏置向量。

3.4.2 邻域关系 *gg g*

$$\mathbf{x}_i^{l,E} = \sigma(\mathbf{m}_i^{l,E} + \sum_{j \in N(i)} \mathbf{m}_j^{l,E}) \tag{6}$$

其中，$N(i)$ 表示实体 i 的出边所对应的邻域集合。$\sigma(\cdot)$ 表示激活函数。

随后，通过堆叠多个图卷积层可以接收邻域消息，这对下游任务很有用。

4 我们的方案

本节首先详细介绍 FPM-GCN 和 HIN 的定义。然后，我们提出 FFHR框架，并展示其如何将基于欧几里得空间的进展推广到双曲空间。最后，我们使用四种具有代表性的知识图谱补全方法实例化 FFHR。

4.1 全庞加莱多关系图卷积网络（FPM-GCN）

本节提出一种新型双曲图卷积网络，无需通过切空间进行近似，也无需在不同等距模型之间进行转换。如图 2 所示，FPM-GCN 完全在双曲空间上运行，包含三个重要模块：

4.1.1 特征变换模块。

根据文献 [38] [39]，，我们了解到，等距群变换（欧几里得空间中的正交变换）具有更强的表达性，并且能够提升神经网络的稳定性。因此，我们在特征变换模块中采用等距群变换。具体而言，我们将特征变换模块形式化为左陀螺平移和正交变换的复合，如下所示：

$$\mathbf{m}_i^{l,\mathcal{B}_c} = \mathbf{b}_r^l \oplus_c \mathbf{W}_r^l \mathbf{x}_i^{l-1,\mathcal{B}_c}, \quad s.t. (\mathbf{W}_r^l)^T \mathbf{W}_r^l = \mathbf{I}, \tag{7}$$

其中，$\mathbf{W}_r^l$ 是关系感知变换矩阵，$\mathbf{b}_r^l$ 是关系感知变换向量。我们证明，所提出的模块满足以下两个理想性质：

命题 1. 流形保持。

特征变换模块具有流形保持性。变换后的向量仍然位于庞加莱球模型中。

证明。 首先，我们将特征变换模块分为两部分。正交变换部分：

$$\Phi(\mathbf{x}) = \mathbf{W} \cdot \mathbf{x}, \ s.t. \ \mathbf{W}^T \mathbf{W} = \mathbf{I}$$

莫比乌斯变换部分：

$$\mathbf{L}_{\mathbf{u}}(\mathbf{x}) = \mathbf{u} \oplus_c \mathbf{x}$$

对于第一部分，容易得到 $||\Phi(\mathbf{x})|| = ||\sqrt{\mathbf{x}^T \mathbf{W}^T \mathbf{W} \mathbf{x}}|| = ||\mathbf{x}|| < \frac{1}{c}$。对于 M¨obius 变换部分，我们希望证明：

$$||\mathbf{u} \oplus_c \mathbf{x}|| < \frac{1}{\sqrt{c}}$$

这等于：

$$(\mathbf{u} \oplus \mathbf{x})^T (\mathbf{u} \oplus \mathbf{x}) < \frac{1}{c}.$$

利用莫比乌斯加法公式，我们只需证明：

$$c(||\mathbf{u}||^2 - \frac{1}{c})(||\mathbf{x}||^2 - \frac{1}{c})(1 + 2c\langle \mathbf{u}, \mathbf{x} \rangle + c^2 ||\mathbf{u}||^2 ||\mathbf{x}||^2) > 0$$

我们已经得到 $||\mathbf{u}||^2 < \frac{1}{c}$，$||\mathbf{x}||^2 < \frac{1}{c}$。利用柯西-施瓦茨不等式，容易得到 $1 + 2c\langle \mathbf{u}, \mathbf{x} \rangle + c^2 ||\mathbf{u}||^2 ||\mathbf{x}||^2 > 0$。因此，上述不等式成立，且这两个变换部分都保持流形结构。证明完毕。

□

注 该命题保证我们的特征变换完全在 Poincar´e 球内进行，无需近似。

定理 1. 表达性。

我们的特征变换定义了一个关于 $d_{\mathcal{B}_c}$的 $\mathcal{B}_c$ 等距变换：

$$d_{\mathcal{B}_c}(Transformed(\mathbf{x}), Transformed(\mathbf{y})) = d_{\mathcal{B}_c}(\mathbf{x}, \mathbf{y}),$$

而且，$(\mathcal{B}_c, d_{\mathcal{B}_c})$ 的每个等距变换都可以表示为与我们的特征变换相同的形式。

证明。首先，我们证明我们的特征变换是关于 $d_{\mathcal{B}_c}$的 $\mathcal{B}_c$ 等距变换。距离函数为：

$$d^{\mathcal{B}_c}(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{c}} arctanh(\sqrt{c}|| - \mathbf{x} \oplus_c \mathbf{y}||)$$

因此，我们只需证明 $|| - \mathbf{L_u}(\mathbf{x}) \oplus \mathbf{L_u}(\mathbf{y})|| = || - \mathbf{x} \oplus \mathbf{y}||$, $|| - \Phi(\mathbf{x}) \oplus \Phi(\mathbf{y})|| = || - \mathbf{x} \oplus \mathbf{y}||$, $\forall \mathbf{x}, \mathbf{y} \in \mathcal{B}$。对于第一个恒等式，我们有

$$\begin{aligned}
&|| - \mathbf{L}_{\mathbf{u}}(\mathbf{x}) \oplus \mathbf{L}_{\mathbf{u}}(\mathbf{y})|| \\
=&|| - (\mathbf{u} \oplus \mathbf{x}) \oplus (\mathbf{u} \oplus \mathbf{y})|| \\
=&|| - (\mathbf{u} \oplus \mathbf{x}) \oplus (\mathbf{u} \oplus (\mathbf{x} \oplus (-\mathbf{x} \oplus \mathbf{y})))|| \\
=&|| - (\mathbf{u} \oplus \mathbf{x}) \oplus (\mathbf{u} \oplus \mathbf{x}) \oplus gyr[\mathbf{u}, \mathbf{x}](-\mathbf{x} \oplus \mathbf{y})|| \\
=&||gyr[\mathbf{u}, \mathbf{x}](-\mathbf{x} \oplus \mathbf{y})|| \\
=&|| - \mathbf{x} \oplus \mathbf{y}||
\end{aligned}$$

在推导过程中，我们使用了引理1中提到的结论。

对于第二个恒等式，我们有

$$\begin{aligned}
&|| - \Phi(x) \oplus \Phi(y)|| \\
=&||\mathbf{W} \frac{(1 + 2c\langle -\mathbf{x}, \mathbf{y} \rangle + c||\mathbf{y}||^2)(-\mathbf{x}) + (1 - c||\mathbf{x}||^2)\mathbf{y}}{1 + 2c\langle -\mathbf{x}, \mathbf{y} \rangle + c^2 ||\mathbf{x}||^2 ||\mathbf{y}||^2}|| \\
=&|| - \mathbf{x} \oplus \mathbf{y}||.
\end{aligned}$$

其次，我们证明，$(\mathcal{B}_c, d_{\mathcal{B}_c})$ 的每个等距变换都可以表示为与我们的特征变换相同的形式。我们将$(\mathcal{B}_c, d_{\mathcal{B}_c})$的等距变换群记为 $Iso(\mathcal{B}_c, d_{\mathcal{B}_c})$。

假设 $\mathbf{R} \in Iso(\mathcal{B}_c, d_{\mathcal{B}_c})$，根据定义，$\mathbf{R}$ 是 $\mathcal{B}$ 上的双射自映射。利用 [40], 中所示的定理11，我们有 $\mathbf{R} = \mathbf{L}^{\mathbf{R}(\mathbf{0})} \circ \tau$，其中 τ 是 $\mathcal{B}$ 上固定 $\mathbf{0}$ 的双射自映射。按照 [41], 中定理18(2)的证明，$\mathbf{L}^{-1}{}_{\mathbf{R}(\mathbf{0})} = \mathbf{L}^{-\mathbf{R}(\mathbf{0})}$。因此，我们有 $\tau = \mathbf{L}^{-\mathbf{R}(\mathbf{0})} \circ \mathbf{R}$，且 τ 是 $(\mathcal{B}_c, d_{\mathcal{B}_c})$ 的等距变换。我们可以发现，τ 保持 M¨obius 陀螺度量。利用 [42], 中的定理3.2，我们

注意到 τ 是 $\mathcal{R}^n$ 上的双射正交变换。我们的证明至此完成。□

注 我们注意到，先前 HGCN 中的其他特征变换模块只是等距群变换的特例，因此表达能力和性能较差。

对于多关系图，通用正交矩阵的内存开销很大，优化过程也非常耗时 [38]。因此在实践中，我们将该矩阵形式化为一个分块对角矩阵：

$$\mathbf{W}_r = diag(A(\theta_{1,r}), A(\theta_{2,r}), \cdots, A(\theta_{\frac{n}{2},r})), \tag{8}$$

$$A(\theta_{i,r}) = \begin{bmatrix} cos(\theta_{i,r}) & -sin(\theta_{i,r}) \\ sin(\theta_{i,r}) & cos(\theta_{i,r}) \end{bmatrix}. \tag{9}$$

4.1.2 邻域聚合模块。

这里我们提出使用 M¨obius 陀螺中点 [37] [31]计算邻域的加权中心点，该中心点可直接基于庞加莱球模型计算，无需任何近似或空间变换。中心点可表征为校准后距离平方加权和的最小化点。

$$\mathbf{x}_i^{k,l,\mathcal{B}_c} = \frac{1}{2} \otimes_c \left(\frac{\sum_{j \in N(i)} v_{i,j}^k \cdot \lambda^c_{\mathbf{m}_j^{l,\mathcal{B}_c}} \cdot \mathbf{m}_j^{l,\mathcal{B}_c}}{\sum_{j \in N(i)} |v_{i,j}^k| \cdot (\lambda^c_{\mathbf{m}_j^{l,\mathcal{B}_c}} - 1)} \right), \tag{10}$$

其中 $v_{i,\ j}^k$ 是第 k 个注意力权重，$N\ (i)$ 是 i 的一跳邻域集合。$\lambda^c\ _{\mathbf{m}_j^{\mathcal{B}_c}}$是公式 1 中定义的共形因子。

为获得注意力权重，我们采用如下注意力模块：

$$v_{i,j}^k = LeakyReLU((\mathbf{a}_h^k)^T \mathbf{m}_i^{l,\mathcal{B}_c} + (\mathbf{a}_t^k)^T \mathbf{m}_j^{l,\mathcal{B}_c}), \tag{11}$$

其中，$\mathbf{a}_h$ 和$\mathbf{a}_t$ 是注意力向量。值得注意的是，注意力权重不包含结构信息，因此可以直接获得。

此外，我们可以采用多头注意力：将它们视为权重相等的点，并再次使用 M¨obius 陀螺中点进行聚合。

4.1.3 非线性激活模块。

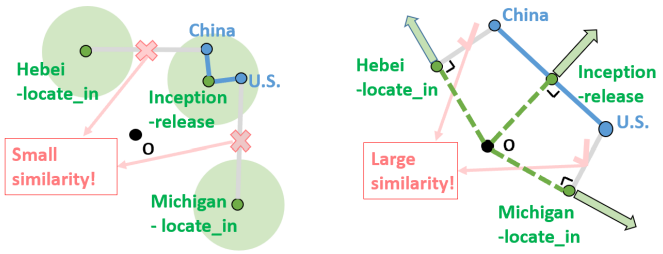
有趣的是，在庞加莱球模型上执行常见的非线性激活函数（例如，ReLU、Softmax）能够保持流形。这里我们将 FPM-GCN 中的非线性激活模块形式化为：

$$\mathbf{x}_i^{l,\mathcal{B}_c} = \frac{1}{\sqrt{c}} \cdot \sigma(\sqrt{c} \cdot \mathbf{x}_i^{l,\mathcal{B}_c}), \tag{12}$$

我们在实现中简单地选择了 LeakyReLU。

4.2 双曲内积（HIN）

在深入探讨我们的 HIN 之前，我们先总结欧几里得空间领域的最新进展，并简要分析评分函数；这两者都能为了解其特性提供有益的见解。



(a) Distance score function (b) Inner product score function

图 3. 《盗梦空间》在美国和中国上映。内积只约束嵌入向量在特定方向上的投影，因此能够保留美国与中国之间的差异。相比之下，距离函数会约束向量的所有维度，难以保留差异，也缺乏灵活性。

4.2.1 两种评分函数的比较。

形式上，这两类评分函数可写为：

$$s_{dist}(h,r,t) = -||\phi_r(\mathbf{e}_h) - \mathbf{e}_t||^2, \tag{13}$$

$$s_{inner}(h,r,t) = \langle \phi_r(\mathbf{e}_h), \mathbf{e}_t \rangle, \tag{14}$$

其中，$\mathbf{e}(\cdot)$ 表示每个实体学习到的嵌入，$\phi_r(\mathbf{e}_h)$ 是头实体h 经过关系感知变换 $\phi_r(\cdot)$后得到的变换嵌入。评分函数通过计算 $\phi_r(\mathbf{e}_h)$ 与$\mathbf{e}_t$ 之间的相似度（内积或距离），来估计给定事实的可信度。

复杂关系模式在知识图谱中很常见。以图 1 为例，知识图谱中存在复杂的多对多关系，即，一个头实体可以通过同一关系连接多个尾实体。例如，《盗梦空间》在美国和中国上映。在学习表示以捕捉此类复杂关系模式时，基于距离的评分函数往往难以胜任。为更好地理解这一点，我们给出一个玩具示例，如图 3 所示，展示了使用两种评分函数学习到的表示。可以看出，基于距离的函数施加了相当严格的约束——如果多个尾实体与同一个头实体相连，它就会迫使这些尾实体的嵌入彼此绝对接近。这种做法缺乏灵活性，还可能导致矛盾。例如，“中国”和“美国”也必须分别接近各自独有的其他尾实体，如“河北—位于”和“密歇根—位于”。这些条件无法同时满足：让具有区别性的“中国”和“美国”彼此更接近，必然会破坏它们之间固有的差异。对于内积，这种负面影响可以得到缓解，因为内积是一种相对温和的度量——它只约束实体在特定方向上的相似度。内积具有灵活性，能够保留复杂数据模式中的差异。例如，“中国”可以同时与“《盗梦空间》—上映”和“河北—位于”具有较高的相似度。

虽然我们的分析直观但不够严谨，却有助于理解内积为何优于距离。接下来，我们还将进行实证研究来验证这一观点。

4.2.2 近期欧几里得方法进展总结

为了比较两种评分函数，我们对近期基于欧几里得距离的

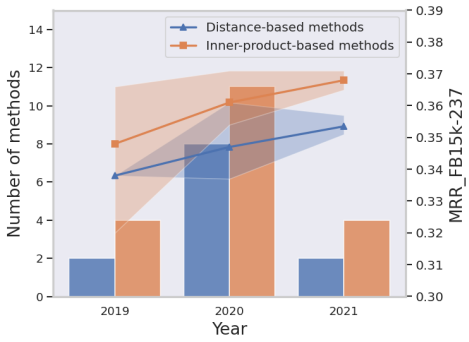


图4. 柱状图和折线图分别展示了近年来基准测试中知识图谱补全方法的数量及其平均倒数排名结果。（阴影区域表示不同年份平均倒数排名的范围。）

发表于 SIGKDD、NIPS、WWW 等顶级会议的知识图谱补全方法进行了全面综述。图 4 展示了采用基于距离的评分函数或基于内积的函数的方法数量及其性能对比。我们发现，近年来内积评分函数受到越来越多的关注。在嵌入复杂数据模式（例如FB15K-237）时，基于内积的方法平均表现更好。此外，最先进水平的结果也由基于内积的方法取得。这些统计结果或许表明，内积评分函数在嵌入复杂数据模式方面具有优势。

上述理论分析和统计结果促使我们探索双曲空间中的内积。

4.2.3 双曲空间中的内积

尽管内积在知识图谱补全中表现出色，但在双曲空间中仍缺乏良好的定义。为了计算双曲空间中的内积，现有工作 [27] 通常将双曲嵌入映射到切空间，并计算它们之间的欧几里得内积。具体而言，这可以表示为：

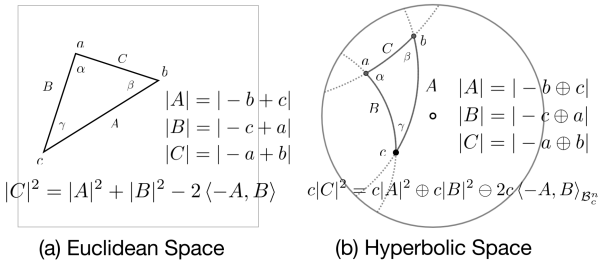
$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{B}_c^n} = \langle log_{\mathbf{0}}^c(\mathbf{x}), log_{\mathbf{0}}^c(\mathbf{y}) \rangle \tag{15}$$

然而，这种表述有两个主要缺点：1）切空间只是双曲空间的局部近似，这不可避免地会导致近似误差。采用 [9], 中流行的优化方法时，它很遗憾会退化为等价的欧几里得内积形式。2）这种表述不符合双曲余弦定律，因此无法保持数据的双曲结构。

为解决这一问题，本小节将欧几里得内积推广到双曲空间，提出一种有意义且实用的推广形式。我们参考欧几里得内积和双曲余弦定律的定义来实现这一点 [15]。形式化地，对于任意双曲三角形$\triangle \mathbf{0xy}$，$\mathbf{x}$，$\mathbf{y} \in \mathbb{B}_c^n$，有：

$$c||\mathbf{C}||^2 = c||\mathbf{A}||^2 \oplus_c c||\mathbf{B}||^2 \ominus_c 2c\Gamma(\mathbf{A}, \mathbf{B}, \psi),$$

其中$\mathbf{A} = -\mathbf{0} \oplus_c \mathbf{x}$、$\mathbf{B} = -\mathbf{0} \oplus_c \mathbf{y}$、$\mathbf{C} = -\mathbf{x} \oplus_c \mathbf{y}$，$\psi$ 是原点处的双曲角，$\Gamma(\mathbf{A},\mathbf{B}\ \psi)$ 表示$\mathbf{A}$ 与$\mathbf{B}$之间特定的相似度函数。这里，双曲距离 $||\mathbf{C}||^2$ 可以分解为两个双曲范数 $||\mathbf{A}||^2$ 和 $||\mathbf{B}||^2$ 的莫比乌斯差，再减去相似度



(a) Euclidean Space (b) Hyperbolic Space

图 5. Poincar´e 球模型中 (a) 欧几里得余弦定律和 (b) 双曲余弦定律的示意图。

函数 $\Gamma(\mathbf{A},\mathbf{B}\ \psi)$。如图5所示，类比欧几里得内积的定义（根据余弦定理），我们将相似度函数 $\Gamma(\mathbf{A},\mathbf{B}\ \psi)$ 称为“双曲内积”，其定义如下：

$$\begin{aligned}\langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{B}_c^n} &= \Gamma(\mathbf{A}, \mathbf{B}, \psi) \\ &= \frac{||\mathbf{A}|| \cdot ||\mathbf{B}|| \cdot cos\psi}{(1 + c^2||\mathbf{A}||^2)(1 + c^2||\mathbf{B}||^2) - 2c^2||\mathbf{A}|| \cdot ||\mathbf{B}|| \cdot cos\psi},\end{aligned}$$

其中，$\mathbf{A} = -\mathbf{0} \oplus_c \mathbf{x}$、$\mathbf{B} = -\mathbf{0} \oplus_c \mathbf{y}$，且 ψ 是原点处的双曲角。显然，我们可以观察到$\lim_{c\to 0} \langle \mathbf{x}, \mathbf{y} \rangle_{\mathcal{B}_c^n} = \langle \mathbf{x}, \mathbf{y} \rangle$，这表明我们的双曲内积是欧几里得内积的一种推广形式，因此继承了基于内积的评分函数的灵活性。我们的 HIN 在理论上也是合理的，因为我们只是类比欧几里得内积，并借助双曲余弦定律，而没有作出任何额外假设。

4.3 全能灵活双曲表示框架（FFHR）

基于上述两项重要的双曲改进，我们构建了一个完全灵活的双曲表示框架 FFHR。FFHR 并非一个具体模型，而是一套高层次的解决方案蓝图，能够将近期基于欧几里得空间的进展迁移到双曲空间。如表1所示，我们展示了 FFHR框架如何将基于欧几里得空间的内积方法推广到双曲空间。

对于知识图谱补全方法，使用一般的三元评分函数 $\mathcal{F}(h, r, t)$ 既耗时又低效。因此，大多数方法都可以表示为 $\mathcal{F}(\phi_r(h)\ t)$。根据不同的评分函数 $\mathcal{F}$，这些方法可分为两类：基于距离的方法和基于内积的方法。由于将基于距离的方法转换到双曲空间相对容易，且已有相关工作进行研究，因此我们主要讨论 FFHR框架如何将所有基于内积的方法迁移到双曲空间。

基于内积的方法之间的区别在于它们的变换函数 ϕ_r。为了迁移这些模型，同时保留其变换函数 ϕ，我们用我们的 HIN 函数替换内积评分函数，并用我们的 FPM-GCN 替换它们的欧几里得图卷积网络。本文中，我们将 FFHR 实例化为四种具有代表性的知识图谱补全方法，包括 DistMult [16], ComplEx [17],

表1 FFHR 将欧几里得方法推广至双曲空间的概要。

模块	欧几里得空间	FFHR
特征变换	$\mathbf{m}_i^{l,E} = \mathbf{W}_r^l \mathbf{x}_i^{l-1,\ E} + \mathbf{b}_r^l$	$\mathbf{m}_i^{l,\ \mathcal{B}_c} = \mathbf{b}_r^l \oplus_c \mathbf{W}_r^l \mathbf{x}_i^{l-1,\ \mathcal{B}_c}$
邻域聚合	$\mathbf{x}_i^{l,E} = \sum_{j \in N\ (i)} v_{i,\ j}^k \cdot \mathbf{m}_j^{l,E}$	$\mathbf{x}_i^{k,\ l,\ \mathcal{B}_c} = \frac{1}{2} \otimes_c \left(\frac{\sum_{j \in N(i)} v_{i,\ j}^k \cdot \lambda^c \mathbf{m}_j^{l\ \mathcal{B}_c}}{\left\lvert \sum_{j \in N(i)} v_{i,\ j}^k \right\rvert \left(\lambda^c \mathbf{m}_j^{l\ \mathcal{B}_c} \right)^{-1}} \right)$
非线性激活	$\mathbf{x}_i^{l,E} = \sigma(\mathbf{x}_i^{l\ E}\ ,\ )$	$\mathbf{x}_i^{l,\ \mathcal{B}_c} = \frac{1}{\sqrt{c}} \cdot \sigma(\sqrt{c} \cdot \mathbf{x}_i^{l\ \mathcal{B}_c},)$
内积评分函数	$\langle \phi_r(\mathbf{e}_h),\ \mathbf{e}_t \rangle$	$\langle \phi_r(\mathbf{e}_h),\ \mathbf{e}_t \rangle_{\mathcal{B}_c}$

表 2 基准数据集的统计信息。

数据集	#实体	#关系	#三元组
WN18RR	40,943	11	93,003
FB15k-237	14,541	237	310,116
WN18	40,943	18	151,442

RESCAL [18]和 DualE [19]，并在实验中测试它们的性能。

$$s(h,r,t) = \langle \mathbf{M}_r \otimes_c \mathbf{e}_h, \mathbf{e}_t \rangle_{\mathcal{B}_c^n}, \tag{16}$$

其中，$\mathbf{e}_{(.)} \in \mathcal{B}_c^n$表示实体嵌入，$\mathbf{M}_r$ 表示关系变换矩阵。

- **DistMult-FFHR**：$\mathbf{M}_r$ 是一个对角矩阵；
- **ComplEx-FFHR**：$\mathbf{M}_r$ 是分块对角矩阵；
- **DualE-FFHR**：$\mathbf{M}_r$ 为分块对角矩阵；
- **RESCAL-FFHR**：$\mathbf{M}_r$ 为一般矩阵。

我们还测试了不使用 FPM-GCN 的相应变体，分别记为 DistMult-H、ComplEx-H、RESCAL-H 和 DualE-H。

4.3.1 训练与优化。
在本文中，我们指 [7] 并使用全多类对数损失作为目标函数：

$$L((h,r,t)) = -s(h,r,t) + log\left[\sum_{t' \in \mathcal{E}} exp(s(h,r,t'))\right].$$

此外，我们参考了先前关于 [9] [13]双曲模型优化的工作。

5 实验与分析

在本节中，我们通过实验评估所提出方法的性能。我们的实验旨在回答以下主要问题：

- **（研究问题1）**FFHR 与现有方法相比表现如何，尤其是与基于欧几里得的方法和基于双曲距离的方法相比?
- **（研究问题2）**双曲空间和 HIN 是否有助于捕获层次数据?
- **（研究问题3）**在捕获复杂关系模式方面，HIN 是否优于双曲距离?
- **（研究问题4）**所提出的组件 FPM-GCN 是否确实能够减少近似误差并提升性能?
- **（RQ5）**不同的超参数设置（例如维度、正则化系数）会如何影响性能?

5.1 实验设置

5.1.1 数据集。
为公平比较，我们采用三个标准基准数据集 WN18 [20], WN18RR[43] 和 FB15k-237 [44] 。WN18 是 WordNet 的一个子集，包含丰富的层级关系，例如“上位词”“部分关系”以及许多逆关系。WN18RR 是 WN18 的一个子集，其中删除了逆关系。FB15k-237 是 Freebase 的一个子集，包含一般的世界知识。其大多数关系都不是层级关系，例如“出生地或国籍”，但也包含一些层级关系，例如“部分关系”。数据集的统计信息见表 2。

5.1.2 基线方法。
接下来，我们将本文方法与不同类别的最先进方法进行比较：

基于距离的模型:
- **RotatE** [6]是一种经典的基于欧几里得距离的方法，它将各种模式的关系表示为从头实体到尾实体的二维旋转。
- **Rotate3D** [45]是 RotatE 的扩展版本，它将二维旋转推广到三维旋转，有利于多跳推理。

基于内积的模型:
- **ConvE** [43] 是一种典型的基于内积的方法，它利用二维卷积和全连接神经网络来建模实体与关系之间复杂的交互。
- **TuckER** [22] 是一种简单而强大的线性模型，基于知识图谱三元组二元张量表示的 Tucker分解。
- **QuatE** [23] 采用具有三个虚部的超复数嵌入来表示实体和关系，是 ComplEx 的扩展版本。
- **LowFER** [46] 提出了一种因子化双线性池化模型，以更好地融合实体和关系，并自然地推广了 TuckER 模型。
- **AutoSF** [47] 利用自动化机器学习技术建立搜索空间，在其中自动为给定的知识图谱搜索嵌入模型。

基于距离的双曲模型和变体:
- **MuRP**、**MuRE** [8] 是首个利用双曲嵌入对知识图谱中的层次数据进行建模的模型。评分函数基于双曲距离函数。
- **AttH**、**AttE** [9]将关系表示为双曲空间上的参数化运算（旋转和反射），从而更好地建模关系的复杂模式。

表3 低维情形（d=32）下的KGC性能比较，其中“IP”“HS”“图卷积网络”分别表示是否使用内积、双曲空间和图卷积网络。最佳结果以粗体标示，次佳结果带下划线。

	特征			WN18RR				FB15K-237				WN18			
模型	IP?	HS?	图卷积网络?	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
RotatE	×	×	×	0.387	0.330	0.417	0.491	0.290	0.208	0.316	0.458	0.830	0.802	0.852	0.868
MuRE	×	×	×	0.458	0.421	0.471	0.525	0.313	0.226	0.340	0.489	0.940	0.938	0.941	0.943
MuRP	×	√	×	0.465	0.420	0.484	0.544	0.323	0.235	0.353	0.501	0.941	0.939	0.942	0.944
AttE	×	×	×	0.456	0.419	0.471	0.526	0.311	0.223	0.339	0.488	0.939	0.937	0.941	0.942
AttH	×	√	×	0.466	0.419	0.484	0.551	0.324	0.236	0.354	0.501	0.941	**0.940**	0.942	0.945
CompGCN	√	×	√	0.434	0.406	0.445	0.485	0.313	0.227	0.342	0.487	0.934	0.930	0.936	0.939
DistMult	√	×	×	0.420	0.393	0.429	0.472	0.315	0.230	0.344	0.485	0.935	0.927	0.943	0.945
DistMult-H	√	√	×	0.433	0.401	0.441	0.497	0.322	0.235	0.354	0.496	0.936	0.928	0.943	0.946
DistMult-FFHR	√	√	√	0.438	0.395	0.457	0.519	0.322	0.235	0.352	0.496	0.938	0.932	0.943	0.948
ComplEx	√	×	×	0.429	0.406	0.436	0.473	0.319	0.233	0.350	0.493	0.939	0.936	0.942	0.944
ComplEx-H	√	√	×	0.442	0.408	0.456	0.507	0.325	0.237	0.356	0.500	0.941	0.939	0.942	0.944
ComplEx-FFHR	√	√	√	0.459	0.420	0.476	0.530	0.326	0.236	0.359	0.503	**0.943**	**0.940**	**0.944**	0.946
DualE	√	×	×	0.385	0.369	0.388	0.413	0.290	0.205	0.318	0.462	0.935	0.932	0.937	0.939
DualE-H	√	√	×	0.422	0.371	0.451	0.512	0.314	0.227	0.345	0.486	0.935	0.931	0.938	0.941
DualE-FFHR	√	√	√	0.440	0.383	0.474	0.536	0.314	0.225	0.348	0.490	0.937	0.930	0.943	0.945
RESCAL	√	×	×	0.449	0.420	0.461	0.503	0.337	0.247	0.369	0.514	0.927	0.912	0.939	**0.951**
RESCAL-H	√	√	×	0.459	0.419	0.480	0.532	0.344	0.255	0.378	**0.522**	0.939	0.936	0.942	0.944
RESCAL-FFHR	√	√	√	**0.468**	**0.422**	**0.490**	**0.552**	**0.345**	**0.256**	**0.379**	0.521	0.939	0.936	0.942	0.944

•**ConE** [48] 将实体和查询表示为二维锥的笛卡尔积，在知识图谱上的多跳推理中展现出强大的能力。

基于图神经网络的模型:

• **CompGCN** [12] 是一个通用框架，通过利用多种实体关系组合操作，将图卷积网络与知识图谱嵌入相结合。

• **M^2图神经网络** [13] 将多关系知识图谱嵌入混合曲率空间，以实现知识图谱补全。

5.1.3 评估指标。

我们遵循标准评估设置，在测试时过滤掉已有三元组 [20] , 并采用链接预测文献中两种常用的评估指标:

- 平均倒数排名（MRR）：正确实体的平均倒数排名。
- H@K，K∈{1,3,10}：衡量真实三元组排名处于前 k 个候选三元组中的比例。

5.1.4 实现细节。

我们完全遵循先前工作的做法 [9] 进行数据集划分和数据增强。在低维情形下，为了公平比较，所有方法（包括基线方法）的嵌入维度均设为 32。由于我们的实验设置与 [9], 完全相同，部分基线方法的结果直接取自 [9]。我们还复现了其中一部分重要方法，并得到与 [9]相近的结果。在高维情形下，我们将所提方法及其变体的维度设为 512。对于其他基线方法，我们不限制其维度，而是报告其在 [200, 2000] 范围内任意维度下的最佳性能，具体参考其原始论文。

关于超参数，对于我们的 FFHR 及其变体，我们使用网格搜索来寻找最佳设置：批次大小为 $\in \{100, 200, 500, 1000, 2000\}$，学习率为$\in \{0.005, 0.01, 0.05, 0.1, 0.5\}$，正则化系数为 $\in \{0.005, 0.01, 0.05, 0.1, 0.5\}$，注意力头数量为 $\in \{1, 2, 4, 8\}$，图卷积网络层数为 $\in \{1, 2\}$。基线方法的超参数参考了原论文，或在我们的实验中进行了验证。

对于所有方法及其欧几里得对应模型，我们均采用 Adagrad 作为优化器，并使用默认的 Xavier均匀初始化器初始化嵌入。值得注意的是，基于内积的方法还有一个优势: 可以使用各种正则化器来提升性能，而基于距离的方法则不能。本文采用 DURA [49] 作为正则化器。为保证公平比较，我们重新实现了 DualE、DualE-H 和 DualE-FFHR，并移除了原论文中使用的类型约束 [19]。我们使用 PyTorch 实现所提出的方法，并在 NVIDIA Tesla V100 GPU 上进行所有实验。

5.2 性能比较（研究问题1）

5.2.1 低维结果。

表3展示了低维（d=32）下的结果。我们观察到：（1）我们的方法在所有基准数据集上始终取得最佳表现，证明了所提出的 HIN 和 FPM-GCN 的有效性。（2）与欧几里得对应模型相比，不使用 FPM-GCN 的双曲模型（*-H）在 WN18RR、FB15k-237 和 WN18 上的平均倒数排名分别提升了 2.8、2.0 和 0.5 个百分点。这与我们的预期一致：双曲空间更适合表示层次数据，尤其是在低维情形下。值得注意的是，WN18RR 包含更多层次数据结构，因此我们的双曲模型在该数据集上的表现相比其欧几里得对应模型有更大提升。（3）我们的方法在 FB15k-237 上也显著优于以往基于距离的双曲方法（例如 AttH、MuRP），因为灵活的内积函数有助于捕捉复杂关系。（4）除了少数例外，采用 FPM-GCN 都能提升模型表现。

表4 高维情况下（d=512）的知识图谱补全性能比较，其中“IP”、“HS”、“GCN”分别表示是否使用内积、双曲空间和图卷积网络。最佳结果以粗体显示，次佳结果加下划线。

	特征			WN18RR				FB15k-237				WN18			
模型	IP?	H?	图卷积网络?	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
ConvE	√	×	×	0.430	0.400	0.440	0.520	0.325	0.237	0.356	0.501	0.943	0.935	0.946	0.956
TuckER	√	×	×	0.470	0.443	0.482	0.526	0.358	0.266	0.394	0.544	**0.953**	**0.949**	0.955	0.958
QuatE	√	×	×	0.488	0.438	0.508	0.582	0.348	0.248	0.382	0.550	0.950	0.945	0.954	0.959
AutoSF	√	×	×	0.490	0.451	-	0.567	0.360	0.267	-	0.552	0.952	0.947	-	0.961
RotatE	×	×	×	0.476	0.428	0.492	0.571	0.338	0.241	0.375	0.533	0.947	0.938	0.953	0.961
Rotate3D	×	×	×	0.489	0.442	0.505	0.579	0.347	0.250	0.385	0.543	0.951	0.945	0.953	0.961
LowFER	√	×	×	0.465	0.434	0.479	0.526	0.359	0.266	0.396	0.544	0.950	0.946	0.952	0.958
MuRP	×	√	×	0.481	0.440	0.495	0.566	0.335	0.243	0.367	0.518	0.947	0.942	0.951	0.954
AttH	×	√	×	0.486	0.443	0.499	0.573	0.348	0.252	0.384	0.540	0.948	0.944	0.950	0.953
ConE	×	√	×	0.496	0.453	0.515	0.579	0.345	0.247	0.381	0.540	-	-	-	-
CompGCN	×	×	√	0.479	0.443	0.494	0.546	0.355	0.264	0.390	0.535	0.945	0.941	0.948	0.951
M^2GNN	×	√	√	0.485	0.444	0.498	0.572	0.362	0.275	0.398	**0.565**	-	-	-	-
DistMult-FFHR	√	√	√	0.459	0.418	0.473	0.544	0.360	0.266	0.395	0.547	0.947	0.942	0.949	0.953
ComplEx-FFHR	√	√	√	0.488	0.446	0.504	0.568	0.362	0.268	0.399	0.551	**0.953**	0.947	**0.956**	**0.963**
DualE-FFHR	√	√	√	0.486	0.446	0.504	0.564	0.353	0.259	0.390	0.541	0.951	0.946	0.955	0.961
RESCAL-FFHR	√	√	√	**0.500**	**0.461**	**0.515**	**0.577**	**0.370**	**0.277**	**0.407**	0.552	0.951	0.945	0.955	0.961

表 5 WN18RR 上 d=32 时的 Hits@10。Khs_G 越高且 ξ_G 越低，表示层次结构越明显。

关系	Khs_G	ξ_G	RESCAL	RESCAL-H	关系改进
成员部分词	1.00	-2.90	0.330	0.368	+11.5%
上位词	1.00	-2.46	0.175	**0.211**	**+20.6%**
具有部分	1.00	-1.43	0.279	0.308	+10.4%
实例上位词	1.00	-0.82	0.516	0.570	+10.5%
领域区域成员	1.00	-0.78	0.288	**0.346**	**+20.1%**
领域用法成员	1.00	-0.74	0.375	0.438	+16.8%
同义词集 主题域	0.99	-0.69	0.395	0.439	+11.1%
另见	0.36	-2.09	0.625	0.616	-1.4%
派生相关形式	0.07	-3.84	0.959	0.963	+0.4%
相似于	0.07	-1.00	1.000	1.000	+0.0%
动词组	0.07	-1.00	0.974	0.974	+0.0%

表6 在不同关系类别（按 [21] 分类）上的结果，数据集为 FB15k-237，d=32。

		MuRP		AttH		RESCAL-H-dist		RESCAL-H	
		平均倒数排名	Hits@10	平均倒数排名	Hits@10	MRR	H@10	平均倒数排名	Hits@10
头实体预测	1-1	**0.477**	0.568	0.467	**0.573**	0.458	0.531	0.437	**0.573**
	1-N	0.088	0.170	0.073	0.150	0.094	0.178	**0.103**	**0.193**
	N-1	0.445	0.643	0.439	0.630	0.447	0.630	**0.449**	**0.645**
	N-N	0.246	0.449	0.243	0.439	0.255	0.456	**0.272**	**0.471**
尾部预测	1-1	**0.474**	0.557	0.453	**0.573**	0.470	0.552	0.439	0.557
	1-N	0.733	0.858	0.730	0.852	0.733	0.851	**0.760**	**0.865**
	N-1	0.059	0.131	0.060	0.111	0.070	0.136	**0.079**	**0.156**
	N-N	0.352	0.581	0.350	0.577	0.361	0.589	**0.382**	**0.601**

5.2.2 高维结果。

表4展示了我们的双曲模型及其变体在 $d = 512$ 下的表现。对于其他基线方法，我们不限制其维度，而是参考其原始论文，报告其在任意维度范围 [200, 2000] 内的最佳表现。缺失的基线方法是因为缺少开源代码。除了少数例外，我们可以发现，我们的方法在所有基准数据集上仍达到最先进水平，并且在处理复杂数据模式（例如 FB15K-237）方面有了很大提升，而其他方法在这类情况下表现不佳。这可能是由于它们的评分函数缺乏灵活性以及近似误差所致。

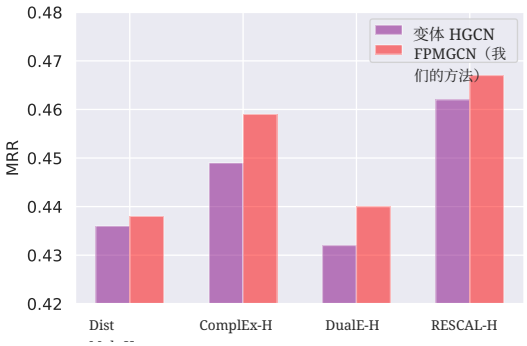


图6. 在 WN18RR 上比较 FPM-GCN 与 HGCN 的变体，d=32。

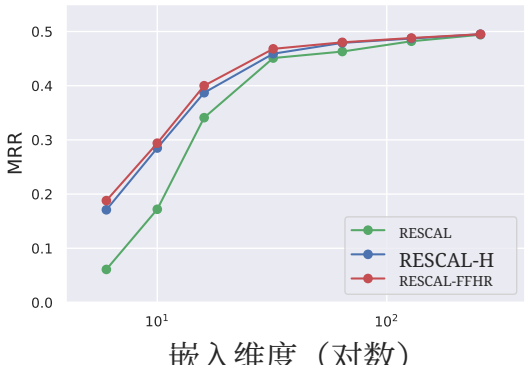


图 7. MRR 随嵌入维度 d $\in$ {6, 10, 16, 32, 64, 128, 256} 的变化。

5.3 案例研究（研究问题2、研究问题3）

5.3.1 捕捉层次化数据的案例研究（研究问题*2*）

为了更深入地理解双曲空间在建模层次数据方面的优势，我们在 WN18RR 上按不同关系类型衡量性能。遵循 [9]，，我们使用 Krackhardt 层级得分（Khs_G）和估计曲率（ξ_G）来衡量每种关系。较低的 ξ_G 和较高的 Khs_G 分别反映全局和局部树结构的程度。我们计算低维情形（d=32）下每种关系在 10 次运行中的平均 Hit@10。如表 5 所示，类似树结构的关系（例如 “上位词”、“具有部分”）通过 FFHR 获得了更大的提升。相反，对于非层级关系（例如 “另见”、“相似于”），如表 5 下半部分所示，双曲模型的提升幅度很小，甚至出现负向提升。

5.3.2 捕捉复杂关系模式的案例研究（研究问题*3*）

按照 [21]，我们将 FB15k-237 中的关系分为四类：一对一（1-1）、一对多（1-N）、多对一（N-1）和多对多（N-N）。我们将模型 RESCAL-H 与基于双曲距离的模型 AttH、MuRP 和 RESCAL-H-dist（RESCAL-H 的一种变体，仅将 HIN 替换为双曲距离函数 [8]）进行比较。表6显示，与基于双曲距离的方法（AttH、MuRP 和 RESCAL-H-dist）相比，RESCAL-H 在复杂关系（1-N、N-1、N-N）上的表现显著提升。对于简单关系（1-1），双曲距离函数的表现更好。这一结果清楚地验证了双曲内积在捕捉复杂数据模式方面的优势。在现实世界和我们的基准数据集中，复杂关系（1-N、N-1、N-N）占多数，这进一步表明采用双曲内积函数的优势。

5.4 消融研究（RQ4、RQ5）

5.4.1 FPM-GCN 消融研究（*RQ4*）

为了公平比较我们提出的 FPM-GCN 与以往采用切空间近似的双曲图卷积网络，本节实现了四种变体：DistMult-V-HGCN、ComplEx-V-HGCN、DualE-V-HGCN 和 RESCAL-V-HGCN。这些变体仅通过切空间近似执行特征变换模块和聚合模块。所有其他组成部分均与 FPM-GCN 相同。具体而言，基于切空间的模块定义如下：

HGCN 的特征变换模块：

$$\mathbf{m}_i^{l,\mathcal{B}_c} = \mathbf{W}_r^l \otimes_c \mathbf{x}_i^{l-1,\mathcal{B}_c}$$

其中$\mathbf{W}_r^l$ 为一般矩阵。

HGCN 的邻域聚合模块：

$$\mathbf{x}_i^{k,l,\mathcal{B}_c} = exp_0\left(\sum_{j\in N(i)} \frac{v_{i,j}^k \cdot log_0(\mathbf{m}_j^{l,\mathcal{B}_c})}{\sum_{j\in N(i)} v_{i,j}^k}\right)$$

其中，$v_{i,\ j}^k$ 是第 k 个注意力权重，N（i）是实体 i的一跳邻域集合。

由于这些操作是在原点的切空间中定义的，因此它们在双曲空间中没有明确的含义。此外，使用这些操作也会破坏实体之间的双曲结构。

这里我们选择 HGCN 而非 H2HGNN [34], 进行比较，因为后者是为洛伦兹模型上的单关系图而设计的，无法适用于我们在庞加莱球模型上定义的评分函数。

图6展示了在维度为 d=32 时，WN18RR 上的平均倒数排名结果。对于所有基础模型，我们的 FPM-GCN 始终优于变体 HGCN，平均提升 1.5%，这清楚地验证了 FPM-GCN 的优越性。关键原因在于，我们定义的操作完全基于双曲空间，且无需近似。

5.4.2 维度消融研究（*RQ5*）

为探究嵌入维度的作用，我们报告了不同维度下各模型（RESCAL、RESCAL-H、RESCAL-FFHR）在 WN18RR 上的平均倒数排名，如图 7 所示。结果表明，RESCAL-H 和 RESCAL-FFHR 始终优于其欧几里得对应模型 RESCAL，并且在低维情形下提升显著。这证实了我们的预期：得益于几何特性，双曲嵌入在低维情形下具有更强的表达性。此外，与 RESCAL-H 相比，FPM-GCN 作为编码器在广泛的维度范围内均能带来性能提升。

6 结论

本文通过严谨的数学推理，提出了双曲空间中的两项重要操作：（1）FPM-GCN，无需进行切空间近似；（2）HIN，即欧几里得内积的双曲推广，在建模复杂数据模式方面比双曲距离更加灵活。我们进一步提出了完全灵活双曲表示嵌入框架 FFHR，它能够将近期基于欧几里得空间的进展迁移到双曲空间。我们通过将四种表示学习知识图谱补全方法实例化为 FFHR，并开展了大量实验，以验证这些方案的有效性。

后续工作中，严格证明内积的优越性，并进一步探讨这两种评分函数，是很有前景的方向。与此同时，混合空间具有更强的表示能力，近年来也受到了广泛关注。将我们的图卷积网络模块和内积函数推广到混合空间，将会很有价值。

参考文献

[1] H. He、A. Balakrishnan、M. Eric 和 P. Liang，“利用动态图谱嵌入学习对称协作式对话代理”，发表于*ACL*，2017年7月。[2] Y.-L. Tuan、Y.-N. Chen 和 H. yi Lee，“Dykgchat：动态知识图谱上的对话生成落地基准”，2019年。[3] X. Wang、X. He、Y. Cao、M. Liu 和 T.-S. Chua，“用于推荐的知识图谱注意力网络”，发表于*SIGKDD*，2019年，第950–958页。[4] Q. Guo、F. Zhuang、C. Qin、H. Zhu、X. Xie、H. Xiong 和 Q. He，“ 基于知识图谱的推荐系统综述 ”，发表于 *IEEE* 知识与数据工程汇刊，第01期，第1–1页，2020年10月。[5] X. Huang、J. Zhang、D. Li 和 P. Li，“基于知识图谱嵌入的问答”，发表于*WSDM*，2019年，第105–113页。[6] 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”，发表于*ICLR*，2019年。[7] T. Lacroix、N. Usunier 和 G. Obozinski，“用于知识库补全的规范张量分解”，发表于*ICML*，第80卷，2018年10月15日，第2863–2872页，7月。[8] I. Balažević、C. Allen 和 T. Hospedales，“多关系庞加莱图嵌入”，发表于*NIPS*，2019年。[9] I. Chami、A. Wolf、D.-C. Juan、F. Sala、S. Ravi 和 C. Ré，“低维双曲知识图谱嵌入”，发表于*ACL*，2020年，第6901–6914页。[10] F. Sala、C. De Sa、A. Gu 和 C. Re，“双曲嵌入的表示权衡”，发表于*ICML*，第80卷，2018年7月10–15日，第4460–4469页。[11] M. S. 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建模”，发表于*ESWC*，第10843卷，2018年，第593–607页。[12] S. Vashishth、S. Sanyal、V. Nitin 和 P. Talukdar，“基于组合的多关系图卷积网络”，发表于*ICLR*，2020年。[13] S. Wang、X. Wei、C. N. Nogueira dos Santos、Z. Wang、R. Nallapati、A. Arnold、B. Xiang、P. S. Yu 和 I. F. Cruz，“用于知识图谱补全的混合曲率多关系图神经网络”，发表于*WWW*，2021年，第1761–1771页。[14] M. Yang、M. Zhou、Z. Li、J. Liu、L. Pan、H. Xiong 和 I. King，“双曲图神经网络：方法与应用综述”，2022年。 [在线]。可访问：https://arxiv.org/abs/2202.13852 。[15] A. Ungar，“双曲三角学及其在双曲几何庞加莱球模型中的应用”，发表于*Computers & Mathematics with Applications*，第41卷第1期，第135–147页，2001年。[16] B. Yang、W. 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”，发表于*ICLR*，2015年。[17] 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 ICML，2016年，第2071–2080页。[18] M. Nickel、V. Tresp 和 H.-P. Kriegel，《用于多关系数据集体学习的三元模型》，载于 ICML，2011年，第809–816页。[19] Z. Cao、Q. Xu、Z. Yang、X. Cao 和 Q. Huang，《双四元数知识图谱嵌入》，AAAI，第35卷，第8期，第6894–6902页，2021年5月。[20] A. Bordes、N. Usunier、A. Garcia-Durán、J. Weston 和 O. Yakhnenko，《用于多关系数据建模的嵌入平移》，载于 NIPS，2013年，第2787–2795页。[21] 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 AAAI，2014年，第1112–1119页。[22] I. Balažević、C. Allen 和 T. M. Hospedales，《TuckER：用于知识图谱补全的张量分解》，载于 EMNLP，2019年。[23] S. Zhang、Y. Tay、L. Yao 和 Q. Liu，《四元数知识图谱嵌入》，载于 NIPS，第32卷，2019年。[24] C. Xu 和 R. Li，《知识图谱中基于二面体群的关系嵌入》，载于 ACL，2019年7月。[25] M. Nickel 和 D. Kiela，《用于学习层次表示的庞加莱嵌入》，载于 NIPS，2017年，第6341–6350页。[26] ——，《在双曲几何的洛伦兹模型中学习连续层次结构》，载于 ICML，第80卷，2018年，第3776–3785页。[27] B. P. Chamberlain、J. Clough 和 M. P. Deisenroth，《图在双曲空间中的神经嵌入》，2017年。[28] O. Ganea、G. Bécigneul 和 T. Hofmann，《双曲神经网络》，载于 NIPS，2018年，第5345–5355页。[29] A. A. Ungar，《解析双曲几何与阿尔伯特·爱因斯坦狭义相对论》。WORLD SCIENTIFIC，2008年。[30] ——，《双曲几何的陀螺向量空间方法》，《数学与统计学综合讲座》，第1卷，第1期，第1–194页，2008年。[31] R. Shimizu、Y. Mukuta 和 T. Harada，《双曲神经网络》，载于 ICLR，2021年。[32] I. Chami、Z. Ying、C. Ré 和 J. Leskovec，《双曲图卷积神经网络》，载于 NIPS，第32卷，2019年。[33] Q. Liu、M. Nickel 和 D. Kiela，《双曲图神经网络》，载于 NIPS，第32卷，2019年。[34] J. Dai、Y. Wu、Z. Gao 和 Y. Jia，《双曲到双曲图卷积网络》，载于 CVPR，2021年6月，第154–163页。[35] S. Wang、X. Wei、C. N. Dos Santos、Z. Wang、R. Nallapati、A. Arnold 和 P. S. Yu，《通过层次双曲神经图嵌入进行知识图谱表示》，载于2021年IEEE国际大数据会议（大数据），2021年，第540–549页。[36] A. Ungar，《解析双曲几何与阿尔伯特·爱因斯坦狭义相对论》，2008年2月。[37] ——，《双曲几何的陀螺向量空间方法》，2009年1月，第1卷。[38] L. Huang、X. Liu、B. Lang、A. Yu、Y. Wang 和 B. Li，《正交权重归一化：深度神经网络中多个相关斯蒂费尔流形优化问题的解决方案》，载于 AAAI，2018年，第3271–3278页。[39] Y. Tang、J. Huang、G. Wang、X. He 和 B. Zhou，《用于知识图谱嵌入的图上下文建模正交关系变换》，载于 ACL，2020年7月，第2713–2722页。[40] T. Suksumran 和 K. Wiboonton，《陀螺群和 l-子陀螺群的同构定理（《几何与物理对称性杂志》37卷（2015），67–83页）》，《几何与物理对称性杂志》，第37卷，第67–83页，2015年2月。[41] T. Rassias 和 P. Pardalos，《数学及其应用论文集：谨献给 Vladimir Arnold》，2016年1月。[42] T. Abe，《爱因斯坦陀螺群、莫比乌斯陀螺群和固有速度陀螺群上的陀螺度量保持映射》，《非线性泛函分析与应用》，第19卷，2014年1月。[43] T. Dettmers、M. Pasquale、S. Pontus 和 S. Riedel，《卷积二维知识图谱嵌入》，载于 AAAI，2018年2月，第1811–1818页。[44] K. Toutanova 和 D. Chen，《知识库与文本推理中的观测特征与潜在特征》，载于第三届连续向量空间模型及其组合性研讨会，2015年7月。[45] 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 CIKM，美国纽约州纽约市，2020年，第385–394页。

[46] S. Amin、S. Varanasi、K. A. Dunfield 和 G. Neumann，《LowFER：用于链接预测的低秩双线性池化》，载于 *ICML*，第119卷，2020年7月13–18日，第257–268页。[47] Y. Zhang、Q. Yao、W. Dai 和 L. Chen，《AutoSF：搜索知识图谱嵌入的评分函数》，载于 *ICDE*，2020年，第433–444页。[48] Y. Bai、R. Ying、H. Ren 和 J. Leskovec，《利用关系特定双曲锥建模异构层次结构》，载于 *NIPS*，2021年。[49] Z. Zhang、J. Cai 和 J. Wang，《用于基于张量分解的知识图谱补全的对偶性诱导正则项》，载于 *NIPS*，2020年。



WentaoShi 目前正在中国科学技术大学（USTC）攻读硕士学位，该校位于中国合肥。他的研究兴趣包括数据挖掘和推荐系统。



Junkang Wu 目前正在中国科学技术大学（USTC）攻读硕士学位，该校位于中国合肥。他的研究兴趣包括数据挖掘和推荐系统。



XuezhiCao 是美团自然语言处理团队的高级研究员。他于2018年获得上海交通大学博士学位。他的研究兴趣包括知识图谱、推荐系统和社交网络。他已在SIGIR、WWW、AAAI等顶级会议上发表超过15篇论文。



Jiawei Chen 是浙江大学教授。他于2020年获得浙江大学计算机科学博士学位。他的研究兴趣包括信息检索、数据挖掘和因果推断。他已在WWW、AAAI、SIGIR、CIKM、ICDE和TOIS等国际会议和期刊上发表十余篇学术论文。此外，他还担任SIGIR、WWW、WSDM、ACMMM、AAAI和IJCAI等顶级会议的程序委员会／高级程序委员会委员，以及知名会议和期刊的特邀审稿人

TNNLS、TKDE、TOIS 等期刊。



WeiWu 现任美团技术负责人，领导团队专注于知识图谱、自然语言处理和信息检索等人工智能技术。在此之前，张博士曾任微软亚洲研究院研究员。他获得了计算机科学博士学位，由中国科学技术大学和微软亚洲研究院联合指导。他已在KDD、WWW、AAAI 和 IJCAI 等国际顶级会议和期刊上发表 50 多篇论文。他曾获得

他曾获得ICDM2013和CIKM2020的最佳论文奖。长期以来，他一直担任KDD、WWW和TKDE等国际顶级会议和期刊的审稿人。



Weiqiang Lei 是四川大学教授。他于2019年获得新加坡国立大学博士学位。他的研究兴趣集中于对话式人工智能，包括对话式推荐、对话与问答系统以及用户反馈建模。他在KDD、WSDM、TOIS、ACL、EMNLP等顶级会议和期刊上发表了相关论文，并获得了ACM MM 2020最佳论文奖。他还积极在多个关于对话式推荐主题的会议上开展教程，包括

RecSys 2021、SIGIR 2020、CCL 2020和CCIR 2020



Xiangnan He 是中国科学技术大学（USTC）教授。他在新加坡国立大学（NUS）获得计算机科学博士学位。他的研究兴趣涵盖信息检索、数据挖掘和多媒体分析。他已发表100多篇论文，发表于SIGIR、WWW和KDD等顶级会议，以及TKDE、TOIS和TNNLS等期刊。他的工作曾获得SIGIR（2021年、2016年）和WWW最佳论文奖荣誉提名

（2018年）。他担任多个期刊的编委，包括 ACM信息系统汇刊（TOIS）、IEEE大数据汇刊（TBD）、AI Open 等。此外，他曾担任 IEEE CCIS 2019 的程序委员会主席，并担任 SIGIR、WWW、KDD、MM、WSDM、ICML 等多个顶级会议的程序委员会委员（高级委员）。