

# 基于双曲图注意力网络的知识图谱链路预测方法

吴铮\* 陈鸿昶 张建朋

(战略支援部队信息工程大学信息技术研究所 郑州 450002)

**摘要:** 现有的大多数知识表示学习模型孤立地看待每个知识三元组, 未能发现和利用实体周围邻域特征信息, 并且将树状层级结构的知识图谱嵌入到欧式空间, 会带来嵌入式向量高度失真的问题。为解决上述问题, 该文提出了一种基于双曲图注意力网络的知识图谱链路预测方法(HyGAT-LP)。首先将知识图谱嵌入到负常数曲率的双曲空间中, 从而更契合知识图谱的树状层级结构; 然后在所给实体领域内基于实体和关系两种层面的注意力机制聚合邻域特征信息, 将实体嵌入到低维的双曲空间; 最后利用得分函数计算每个三元组的得分值, 并以此作为判定该三元组成立的依据完成知识图谱上的链路预测任务。实验结果表明, 与基准模型相比, 所提方法可显著提高知识图谱链路预测性能。

**关键词:** 知识图谱; 链路预测; 双曲空间; 图注意力网络

中图分类号: TP181

文献标识码: A

文章编号: 1009-5896(2022)06-2184-11

DOI: [10.11999/JEIT210321](https://doi.org/10.11999/JEIT210321)

## Link Prediction in Knowledge Graphs Based on Hyperbolic Graph Attention Networks

WU Zheng CHEN Hongchang ZHANG Jianpeng

(Institute of Information Technology, PLA Strategic Support Force Information Engineering University, Zhengzhou 450002, China)

**Abstract:** Most existing knowledge representation learning models treat knowledge triples independently, it fail to cover and leverage the feature information in any given entity's neighborhood. Besides, embedding knowledge graphs with tree-like hierarchical structure in Euclidean space would incur a large distortion in embeddings. To tackle such issues, a link prediction method based on Hyperbolic Graph Attention networks for Link Prediction in knowledge graphs (HyGAT-LP) is proposed. Firstly, knowledge graphs are embedded in hyperbolic space with constant negative curvature, which is more suited for knowledge graphs' tree-like hierarchical structure. Then the proposed method aggregates feature information in the given entity's neighborhood with both entity-level and relation-level attention mechanisms, and further, embeds the given entity in low dimensional hyperbolic space. Finally, every triple's score is computed by a scoring function, and links in knowledge graphs are predicted based on the scores indicating the probabilities that predicted triples are correct. Experimental results show that, compared with baseline models, the proposed method can significantly improve the performance of link prediction in knowledge graphs.

**Key words:** Knowledge graphs; Link prediction; Hyperbolic space; Graph attention networks

### 1 引言

知识图谱是一种用图模型来描述知识和建模世

界万物之间关联关系的技术方法<sup>[1]</sup>。现如今, 大规模知识图谱已经成为推动实现信息检索<sup>[2]</sup>、自动问答<sup>[3]</sup>和推荐系统<sup>[4]</sup>等人工智能相关应用的重要基础。知识图谱将现实世界中的知识建模成 $(v_h, r, v_t)$ 三元组的形式, 其中 $v_h$ 和 $v_t$ 分别表示头、尾实体,  $r$ 表示实体之间的关系, 例如(Beijing, capital\_of, China)。虽然知识图谱的应用取得了巨大的成功, 但是目前例如FreeBase, WordNet, Yago, DBpedia以及NELL等现代著名知识图谱存储的知识是稀疏的和不完全的, 很多存在于真实世

收稿日期: 2021-04-16; 改回日期: 2022-02-28; 网络出版: 2022-02-14

\*通信作者: 吴铮 wzxuexizhuanyong@163.com

基金项目: 国家自然科学基金青年基金(62002384), 郑州市协同创新重大专项(162/32410218), 中国博士后科学基金面上项目(47698) Foundation Items: The National Natural Science Foundation for Young Scholars (62002384), The Collaborative Innovation Program of Zhengzhou (162/32410218), The General Program of China Postdoctoral Science Foundation (47698)

界中的关系在知识图谱的实体之间却是缺失的，因此需要知识图谱补全技术来提高知识图谱的完整程度和数据质量。知识图谱补全目前主要被抽象成一个链路预测问题，即预测出三元组中缺失的部分，如头实体预测 $(?, r, v_t)$ 、尾实体预测 $(v_h, r, ?)$ 和关系预测 $(v_h, ?, v_t)$ ，问号表示要预测的部分，而另外两个部分是已知的。

近年来，网络表示学习在特征提取方面展现出了强大的能力。许多研究者将网络表示学习迁移到知识图谱数据上，完成知识图谱中的各类任务，其中就包含基于知识表示学习的知识图谱链路预测。基于知识表示学习的知识图谱链路预测方法，大多将知识网络中的实体与实体之间的语义关系嵌入到欧式空间中连续的低维稠密向量空间中，当学习到知识图谱的欧式空间嵌入式向量后，实体和关系之间可以进行计算与推理，再利用线性模型或神经网络模型就可以高效地完成知识三元组中缺失元素的预测。尽管基于欧式空间的表示学习取得了成功，但是最近研究结果表明，欧几里德对称模型不能很好地反映复杂的数据模式，欧几里德空间里的嵌入表示不是最有意义的几何表示<sup>[9]</sup>。现实世界中的网络很多情况下表现为树形结构的无标度图，在这样的结构中，图的体积(以一个节点为中心，某个半径范围内的节点数量)以半径的指数级增长，而欧式空间中球的体积只以半径的多项式级增长，这就会造成嵌入式向量的高度失真。知识图谱是一个典型的树状分层多关系结构的图数据，知识图谱中的许多关系呈现出了实体之间明显的分层结构。大多数现有的知识图谱链路预测方法将知识图谱嵌入到欧式空间中进行表示学习，容易造成嵌入式向量的失真，而将知识图谱嵌入到双曲空间中则可能会有较明显的性能改进。此外，大多数方法孤立地分析每个知识三元组，忽视了对三元组中实体周围局部结构化信息的挖掘和利用。

针对以上问题，本文提出了一种基于双曲注意力网络的知识图谱链路预测方法(Hyperbolic Graph Attention networks for Link Prediction in knowledge graph, HyGAT-LP)，通过聚合双曲几何对层级结构图数据的建模能力和图注意力网络对网络结构的学习能力，共同学习知识图谱上的嵌入式表示。首先设计双曲空间中的邻居融合和非线性激活的函数表达式，在每一层中用不同的可训练的曲率将欧式输入特征转化为双曲空间中的表示向量。然后计算不同邻居关系的权重，并计算对应每种关系下不同邻居实体的权重，最后综合衡量邻居实体和关系的影响，分别通过实体和关系级别的注

意力机制聚合双曲空间中的邻居信息来学习和更新双曲嵌入式表示向量。多个知识图谱数据集上的实验结果表明，HyGAT-LP的性能优于目前最先进的知识图谱链路预测方法。

## 2 相关工作

### 2.1 知识表示学习

最近几年，知识图谱补全技术受到广泛关注并取得深入发展，在所有知识补全方法中，最有效的就是基于知识表示学习方法，此类方法可以在保留知识图谱内在结构的基础上，通过学习实体和关系低维向量表示，使用线性或者神经网络模型来预测知识三元组中缺失的信息<sup>[6]</sup>。众多算法模型大致可以被分为4类：(1)转移距离模型，典型的模型为TransE<sup>[7]</sup>。受词向量中平移不变性的启发，在TransE中，把关系的向量表示解释成头实体向量向尾实体向量的转移向量，把衡量知识图谱中三元组的合理性问题，转化成衡量头实体到尾实体的距离问题。后续的RotatE<sup>[8]</sup>模型在复数空间中建模，把关系当作头尾实体之间的旋转，从理论上证明能够解决对称/反对称、翻转、组合关系。(2)语义匹配模型，早期典型的模型为RESCAL<sup>[9]</sup>，核心思想是将整个知识图谱编码为一个3维张量，将关系表示成实体向量的线性变换，得分函数设计成头实体向量、关系矩阵以及尾实体向量的双线性乘积的形式。DistMult<sup>[10]</sup>通过限制关系矩阵为对角矩阵来简化和修正RESCAL模型容易过拟合的问题，但是DistMult限定所有关系对称，无法解决非对称关系。ComplEx<sup>[11]</sup>通过将实体和关系在复数域进行向量表示来扩展DistMult，实现了对非对称关系的预测，但是它不能推断组合模式。TuckER<sup>[12]</sup>通过对三元组向量的Tucker分解实现了不同关系之间多任务学习，以上语义匹配模型都可以看作TuckER模型的特例。相比较于转移距离模型，语义匹配模型通过挖掘向量化后的实体和关系的潜在语义来度量事实的可信性。(3)基于卷积神经网络(Convolutional Neural Networks, CNN)的模型，ConvE<sup>[13]</sup>和ConvKB<sup>[14]</sup>是最近两年新提出的基于CNN的知识图谱表示学习模型，ConvE通过一个2维卷积来预测链路，模型包括卷积层、全连接层和最后输出的内积层。ConvKB和ConvE类似，把每个三元组看作一个 $k \times 3$ 的矩阵，卷积层中尺寸为 $1 \times 3$ 的过滤器对矩阵的每一行进行遍历，提取实体之间的全局关系。相比较于其他模型，这些模型虽然提取局部特征的效率较高，但是仍将实体和关系视为独立的元素，损坏了三元组的完整性，并且忽视了对相邻的不同三元组之间关系的考量。

(4)基于图神经网络(Graph Neural Networks, GNN)的模型,典型代表为R-GCN<sup>[15]</sup>(Relational Graph Convolutional Networks),基于已知实体或关系在图中周围节点的表示,推理得到未知实体的表示,从而可以获取知识图谱中缺失实体的表示向量。此外还有RGHAT<sup>[16]</sup>模型基于图注意力网络(Graph Attention networks, GAT)来进行链接预测和实体发现等任务。GNN模型可以处理数据之间具有复杂关系和相互依赖的图结构数据<sup>[17]</sup>。GNN模型的引入丰富了知识图谱中实体和关系的表达,尤其在得到未知实体和关系的表示方面具备一定的推理能力。

## 2.2 双曲空间知识表示学习

知识图谱链路预测的性能取决于模型对图数据结构和关系模式建模的契合程度。双曲空间是最近机器学习领域中十分活跃的研究热点之一,已有研究表明,在建模多关系树形层次结构的知识图谱数据时,负常数曲率空间即双曲空间要优于零曲率空间即欧式空间<sup>[18]</sup>。近年来,许多研究尝试把各种各样的层级数据嵌入到双曲空间,在嵌入向量维度更少的条件下却在下游的机器学习任务中取得了更好的效果,并在知识问答、机器翻译、图分析等应用上取得了性能上的提升。Zhang等人<sup>[19]</sup>通过双曲层级注意力网络模型Hype-HAN生成从词向量到句向量,再到最终的文档向量,然后利用生成的这些语义向量用于自动文本分类任务。Balažević等人<sup>[20]</sup>提出的MuRP模型在双曲几何空间中学习知识图谱表示向量,通过最小化头实体向量和尾实体向量转移后的双曲距离,学习到最佳双曲嵌入式向量,但是其缺点在于它像其他转移距离模型一样无法挖掘关系中的逻辑属性,且模型曲率固定会导致准确率的丧失。针对此问题,Chami等人<sup>[21]</sup>提出ATTH模型对其进行改进,使用旋转、反射、平移变换对知识图谱中的逻辑和层次模式进行建模,取得了最佳的预测效果。

## 3 相关概念及定义

### 3.1 知识图谱链路预测

知识图谱可以看作用多关系图 $G = (V, R)$ 来表示知识三元组集合,其中 $V$ 和 $R$ 分别表示实体(节点)和关系(边)集合,知识三元组 $(v_h, r_{h,t}, v_t) \in E \subseteq V \times R \times V$ 表示包含语义信息的事实,且其中 $v_h \in V$ 是头实体, $v_t \in V$ 是尾实体, $r_{h,t} \in R$ 是两者之间的语义关系。知识表示学习的目的就是在某个空间 $U$ 中将实体 $v_h, v_t \in V$ 映射成向量 $\mathbf{v}_h, \mathbf{v}_t \in U^{d_v}$ ,将关系 $r_{h,t} \in R$ 映射成向量 $\mathbf{r}_{h,t} \in U^{d_r}$ ,学习到的向量可以保留知识图谱的结构和语义信息。由于知识

图谱经常是不完备的,所以需要链路预测来推断其他的正确事实。如果数据集被划分为训练集三元组集合 $E_{Train}$ 和测试集三元组集合 $E_{Test}$ ,知识图谱上的表示向量学习过程如下:定义一个用于衡量三元组成立可能性的得分函数 $f: V \times R \times V \rightarrow \mathbb{R}$ ,使用训练集 $E_{Train}$ 中的知识三元组训练,用学习到的表示向量计算测试集 $E_{Test}$ 中的知识三元组得分,目标是使得测试集中真实存在的三元组得分大于不存在的三元组得分。

基于以上定义,知识图谱上的链路预测问题可以描述成一个排序问题。以预测尾实体 $(v_h, r_{h,t}, ?)$ 为例,给定头实体 $v_h$ 和关系 $r_{h,t}$ ,首先,将尾实体 $v_t$ 替换成实体集合 $V$ 中的其他实体 $v_t' \in V \setminus v_t$ 来生成 $(|V| - 1)$ 个无效三元组 $(v_h, r_{h,t}, v_t')$ ,用得分函数 $f$ 计算正确三元组的得分 $f(v_h, r_{h,t}, v_t)$ 以及 $(|V| - 1)$ 个无效三元组的得分 $f(v_h, r_{h,t}, v_t')$ ,将这些得分按倒序排列,可以得到有效三元组 $(v_h, r_{h,t}, v_t)$ 的排序值,以此作为衡量链路预测效果的依据。

### 3.2 双曲几何模型

不同的双曲空间表达模型为双曲空间中进行数学运算提供了不同的视角,庞加莱球模型是5个双曲几何等距同构模型中的一种(等距同构是指从一个模型 $(X, d)$ 的度量空间到另一个模型 $(X', d')$ 的度量空间存在一对一距离保持映射,其中 $X$ 和 $X'$ 是集合, $d$ 和 $d'$ 是距离函数或者度量,它提供了一种衡量模型之间等价关系的定义。)曲率为负常数 $-1/K$  ( $K > 0$ ),半径为 $\sqrt{K}$ 的庞加莱球模型双曲空间 $(\mathbb{B}^{d,K}, g^{\mathbb{B}})$ 是一个黎曼度量为 $g^{\mathbb{B}}$ 的 $d$ 维流形 $\mathbb{B}^{d,K} := \{v \in \mathbb{R}^d : \|v\|^2 < K\}$ ,其中 $g^{\mathbb{B}}$ 与欧式度量 $g^{\mathbb{E}}$ 的共因子为 $\lambda_v^K = 2K / (K - \|v\|^2)$ ,即 $g^{\mathbb{B}} = (\lambda_v^K)^2 g^{\mathbb{E}}$ 。

因为从Poincare平面到Poincare圆盘的映射为一个等距映射,所以在庞加莱球模型双曲空间中两个点 $v_i, v_j \in \mathbb{B}^{d,K}$  ( $v_i \neq v_j$ )间的测地线距离为

$$d_{\mathbb{B}}^K(v_i, v_j) = 2\sqrt{K} \operatorname{arctanh} \left( \frac{1}{\sqrt{K}} \|\mathbf{v}_i \oplus_K \mathbf{v}_j\| \right) \quad (1)$$

其中, $\|\cdot\|$ 是欧式范数, $\oplus_K$ 是莫比乌斯加法:

$$\begin{aligned} \mathbf{v}_i \oplus_K \mathbf{v}_j &= \frac{(K^2 + 2K \langle \mathbf{v}_i, \mathbf{v}_j \rangle + \|\mathbf{v}_j\|^2) \mathbf{v}_i + (K^2 - K \|\mathbf{v}_i\|^2) \mathbf{v}_j}{K^2 + 2K \langle \mathbf{v}_i, \mathbf{v}_j \rangle + \|\mathbf{v}_i\|^2 \|\mathbf{v}_j\|^2} \end{aligned} \quad (2)$$

其中, $\langle \cdot, \cdot \rangle$ 是欧式内积。

在庞加莱球面 $\mathbb{B}^{d,K}$ 上每一点 $v$ 都有一个 $d$ 维的欧式正切平面空间为

$$\mathcal{T}_v \mathbb{B}^{d,K} := \{p \in \mathbb{R}^d : \langle v, p \rangle = 0\} \quad (3)$$

$T_{\mathbf{v}}\mathbb{B}^{d,K} = \mathbb{R}^d$ 是双曲流形 $\mathbb{B}^{d,K}$ 在点 $\mathbf{v}$ 处的局部一阶估计, 包含了所有从 $\mathbf{v}$ 点出发的所有方向的切线, 它有助于进行双曲空间中未定义的欧式运算。

庞加莱球上两个点 $\mathbf{v}_i, \mathbf{v}_j \in \mathbb{B}^{d,K}$  ( $\mathbf{v}_i \neq \mathbf{v}_j$ ), 其中在点 $\mathbf{v}_i$ 处的欧式正切平面空间 $T_{\mathbf{v}_i}\mathbb{B}^{d,K}$ 中某一方向上的向量 $\mathbf{p} \in T_{\mathbf{v}_i}\mathbb{B}^{d,K}$ , 欧式正切平面空间 $T_{\mathbf{v}_i}\mathbb{B}^{d,K}$ 和双曲空间 $\mathbb{B}^{d,K}$ 之间的双射关系如图1所示。

切空间和庞加莱球之间的切换通过指数映射和对数映射进行, 特别地, 其中沿着向量 $\mathbf{p}$ 的方向从 $T_{\mathbf{v}_i}\mathbb{B}^{d,K}$ 映射到 $\mathbb{B}^{d,K}$ 通过指数映射, 如式(4)所示,

$$\exp_{\mathbf{v}_i}^K(\mathbf{p}) = \mathbf{v}_i \oplus_K \left( \tanh \left( \frac{\lambda_{\mathbf{v}_i}^K \|\mathbf{p}\|}{2\sqrt{K}} \right) \frac{\sqrt{K}\mathbf{p}}{\|\mathbf{p}\|} \right) \quad (4)$$

$\mathbb{B}^{d,K}$ 到 $T_{\mathbf{v}_i}\mathbb{B}^{d,K}$ 通过对数映射, 如式(5)所示

$$\log_{\mathbf{v}_i}^K(\mathbf{v}_j) = \frac{2\sqrt{K}}{\lambda_{\mathbf{v}_i}^K} \operatorname{arctanh} \left( \frac{\|-\mathbf{v}_i \oplus_K \mathbf{v}_j\|}{\sqrt{K}} \right) \frac{-\mathbf{v}_i \oplus_K \mathbf{v}_j}{\|-\mathbf{v}_i \oplus_K \mathbf{v}_j\|} \quad (5)$$

## 4 算法介绍

本文的目的是学习可以保存复杂逻辑关系和层次结构的双曲空间中的知识图谱表示向量, 并用学到的向量通过得分函数在知识图谱上进行链路预测, 还原损坏三元组中丢失的信息, 从而完成知识图谱补全任务。为此, 本文提出了基于双曲图注意力网络的知识图谱链路预测算法HyGAT-LP, 算法总共包括3个部分: (1)欧式空间到双曲空间的映射; (2)双曲空间中的特征变换; (3)双曲空间中的邻居信息聚合。

### 4.1 欧式空间到双曲空间的映射

对于知识图谱中每个实体, 其经过神经网络初始化后生成的欧式特征向量假设为 $\mathbf{v}^{0,E} \in \mathbb{R}^d$ , 以 $\mathbf{o} := \{0, \dots, 0\} \in \mathbb{B}^{d,K}$ 为双曲空间中的原点, 它同时也是进行正切空间中相关运算的参考点, 且满足 $\langle \mathbf{v}^{0,E}, \mathbf{o} \rangle = 0$ , HGAT-LP根据庞加莱球模型中的指数映射, 利用式(4)将欧式正切空间中的点 $\mathbf{v} \in T_{\mathbf{o}}\mathbb{B}^{d,K}$ 映射到双曲空间中(如图1所示)

$$\mathbf{v}^{0,H} = \exp_{\mathbf{o}}^K(\mathbf{v}^{0,E}) = \tanh \left( \frac{\lambda_{\mathbf{o}}^K \|\mathbf{v}^{0,E}\|}{2\sqrt{K}} \right) \frac{\sqrt{K}\mathbf{v}^{0,E}}{\|\mathbf{v}^{0,E}\|} \quad (6)$$

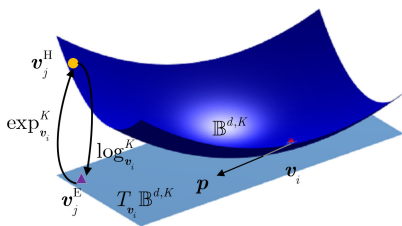


图1 欧式正切平面空间和双曲空间之间映射关系示意图

### 4.2 双曲空间中的特征变换

线性变换通过特征向量与权重矩阵的相乘和与常数偏置项的相加来实现, 是神经网络特征学习的重要组成部分, 但是双曲空间中没有直接进行线性变换的定义, 所以需要设计双曲空间中的特征变换。对于双曲空间中的特征向量 $\mathbf{v}^H$ , 首先利用对数映射将其映射到该点所在位置的正切平面中, 然后就可以利用欧式空间中的向量和矩阵相乘进行计算, 然后再利用指数映射重新映射回双曲空间, 双曲矩阵向量的乘法计算方法如式(7)所示,

$$\mathbf{W} \otimes_K \mathbf{v}^H := \exp_{\mathbf{o}}^K \left( \mathbf{W} \log_{\mathbf{o}}^K(\mathbf{v}^H) \right) \quad (7)$$

与常数偏置项 $\mathbf{c}$ 双曲相加计算方法如式(8)所示,

$$\mathbf{v}^H \odot_K \mathbf{c} := \exp_{\mathbf{v}^H}^K(P_{\mathbf{o} \rightarrow \mathbf{v}^H}^K(\mathbf{c})) \quad (8)$$

其中,  $P(\cdot)$ 是平行转移运算。一般地, 对于双曲空间中的两点 $\mathbf{v}_i^H, \mathbf{v}_j^H \in \mathbb{B}^{d,K}$  ( $\mathbf{v}_i^H \neq \mathbf{v}_j^H$ ),  $P_{\mathbf{v}_i^H \rightarrow \mathbf{v}_j^H}^K(\mathbf{c})$ 将正切向量 $\mathbf{c} \in T_{\mathbf{v}_i^H}\mathbb{B}^{d,K}$ 平行转移到正切空间 $T_{\mathbf{v}_j^H}\mathbb{B}^{d,K}$ :

$$P_{\mathbf{v}_i^H \rightarrow \mathbf{v}_j^H}^K(\mathbf{c}) = \mathbf{c} - \frac{\langle \log_{\mathbf{v}_i^H}(\mathbf{v}_j^H), \mathbf{c} \rangle}{d_{\mathbb{B}}^K(\mathbf{v}_i^H, \mathbf{v}_j^H)^2} \cdot \left( \log_{\mathbf{v}_i^H}(\mathbf{v}_j^H) + \log_{\mathbf{v}_j^H}(\mathbf{v}_i^H) \right) \quad (9)$$

### 4.3 双曲空间中的邻居信息聚合

在图注意力网络GAT中, 注意力机制被看作一种更具表达能力的信息融合手段。区别于图卷积神经网络(Graph Convolutional Networks, GCN), GAT通过衡量中心节点与邻居节点的相关度区别对待不同的邻居节点, 并将其作为分配权重应用到聚合邻居结构信息和属性信息的操作中, 从而更加准确地刻画了中心节点, 提升了模型的表达能力。在知识图谱网络中, 除了邻居节点类型的多种多样之外, 不同节点之间的关系类型也并非单一的。同一实体在不同的关系中也扮演了不同的角色, 起到了不同的作用, 并且不同关系在描述对实体时, 其对实体的限定信息程度也是不同的, 例如在描述美国篮球职业联赛布鲁克林“篮网队”实体时, 关系“has\_players”比关系“based\_in\_city”更具有指向性, 因为一个球队的运动员可以唯一确定一支球队, 但是同一个城市却有可能会有两支球队, 比如篮网队所在的城市纽约还有尼克斯队。因此HyGAT-LP分别考虑两种层面的注意力机制, 即关系注意力和节点注意力。

对于知识图谱中的实体嵌入式矩阵用 $\mathbf{H} \in \mathbb{R}^{N_v \times d}$ 来表示, 其中 $N_v = |\mathbf{V}|$ 是总的实体个数,  $d$ 是每个实体嵌入式向量的特征维度, 矩阵的第 $i$ 行是实体 $v_i \in \mathbf{V}$ 的嵌入式向量 $\mathbf{v}_i = \mathbf{H}_i$ 。关系嵌入式

矩阵用  $\mathbf{R} \in \mathbb{R}^{N_r \times d}$  表示, 其中  $N_r = |\mathbf{R}|$  是总的关系类型个数,  $d$  是每种关系嵌入式向量的特征维度, 矩阵的第  $i$  行是关系  $r \in \mathbf{R}$  的嵌入式向量  $\mathbf{r}_i = \mathbf{R}_i$ 。对于实体  $v_i$ , 其周围的邻居实体和关系就是一定数量的知识三元组  $(v_h, r_{h,t}, v_t)$ 。为了学习实体  $v_i$  在双曲空间中的嵌入式向量  $\mathbf{v}_i^H$ , 跟  $v_i$  相关的每一个三元组都需要被学习。首先考虑关系层面的注意力, 它是每种关系在表示实体  $v_i$  时的权重, 即

$$\mathbf{a}_{v_i, r} = \mathbf{W}_1 \left[ \log_{\mathbf{o}}^K (\mathbf{v}_i^H) \parallel \log_{\mathbf{o}}^K (\mathbf{r}^H) \right] \quad (10)$$

$$\alpha_{v_i, r} = \text{softmax}(\mathbf{a}_{v_i, r}) = \frac{\exp(\sigma(\mathbf{q} \cdot \mathbf{a}_{v_i, r}))}{\sum_{r' \in \Gamma_{\mathbf{R}}(v_i)} \exp(\sigma(\mathbf{q} \cdot \mathbf{a}_{v_i, r'}))} \quad (11)$$

其中,  $\parallel$  表示拼接操作,  $\mathbf{W}_1 \in \mathbb{R}^{d \times 2d}$  和  $\mathbf{q} \in \mathbb{R}^d$  是训练参数矩阵和向量,  $\Gamma_{\mathbf{R}}(v_i)$  表示实体  $v_i$  的邻居关系,  $\sigma(\cdot)$  是欧式空间中的 LeakReLU 非线性激活函数,  $\exp(\cdot)$  是欧式空间中的指数函数。通过以上的计算, 就可以得到关系层面的注意力值  $\alpha_{v_i, r}$ , 它表示关系  $r$  对实体  $v_i$  嵌入式表示的权重系数。接着再考虑同一关系条件下不同邻居实体对中心实体嵌入表示的影响, 即实体层面的注意力。以实体“篮网队”的关系“has\_players”所连接的不同球员为例, 球星与普通球员对实体的嵌入式表示的影响也是不同的。模型先把邻居实体以关系类型聚类, 然后再计算实体层面的注意力, 即

$$\mathbf{b}_{v_i, r, v_j} = \mathbf{W}_2 \left[ \mathbf{a}_{v_i, r} \parallel \log_{\mathbf{o}}^K (\mathbf{v}_j^H) \right] \quad (12)$$

$$\beta_{r, v_j} = \text{softmax}(\mathbf{b}_{v_i, r, v_j}) = \frac{\exp(\sigma(\mathbf{z} \cdot \mathbf{b}_{v_i, r, v_j}))}{\sum_{v_j' \in \Gamma_{\mathbf{V}}(v_i, r)} \exp(\sigma(\mathbf{z} \cdot \mathbf{b}_{v_i, r, v_j'}))} \quad (13)$$

其中,  $\mathbf{v}_j^H$  是实体  $v_j$  的双曲嵌入式向量, 且是关系  $r$  下的尾实体,  $\mathbf{W}_2 \in \mathbb{R}^{d \times 2d}$  和  $\mathbf{z} \in \mathbb{R}^d$  是训练参数矩阵和向量,  $\Gamma_{\mathbf{V}}(v_i, r)$  表示在关系  $r$  下的尾实体集合,  $\mathbf{b}_{v_i, r, v_j}$  可以看作邻居三元组  $(v_i, r, v_j)$  的欧式空间中嵌入式表示,  $\beta_{r, v_j}$  是实体层面的注意力, 表示在关系  $r$  下尾实体  $v_j$  在所有邻居尾实体中对头实体  $v_i$  嵌入式表示的权重系数。通过以上计算, 结合关系层面和实体层面的注意力值进一步生成三元组层面的注意力, 即

$$\varepsilon_{v_i, r, v_j} = \alpha_{v_i, r} \cdot \beta_{r, v_j} \quad (14)$$

它表示邻居三元组对实体  $v_i$  嵌入式表示的权重系数。

GAT模型通过堆叠图卷积层实现迭代式地聚合邻居的特征, 从而更新当前节点的特征。为了保持信息在不同卷积层之间的传递性, 还需要加入自

己本身的特征信息。实体  $v_i$  融合邻居三元组结构与属性信息以及保持自身特征信息的方式如式(15)所示,

$$\text{AGG}^K(\mathbf{v}_i^H) = \exp_{\mathbf{v}_i^H}^K \left( \mathbf{W}_3 \left( \mathbf{v}_i + \sum_{v_j \in \mathbf{V}_{v_i, r}} \varepsilon_{v_i, r, v_j} \cdot \mathbf{b}_{v_i, r, v_j} \right) \right) \quad (15)$$

其中,  $\mathbf{W}_3 \in \mathbb{R}^{d \times d}$  是训练参数矩阵。

非线性激活函数可以增加模型的非线性表达能力, 进一步提高模型的性能。类似地, 当第  $l-1$  层和第  $l$  层中双曲曲率分别为  $-1/K_{l-1}$  和  $-1/K_l$  时, HyGAT-LP模型中双曲空间的非线性激活函数表达式如式(16)所示,

$$\sigma^{\otimes K_{l-1}, K_l}(\mathbf{v}_i^H) := \exp_{\mathbf{o}}^{K_l} \left( \sigma \left( \log_{\mathbf{o}}^{K_{l-1}} (\mathbf{v}_i^H) \right) \right) \quad (16)$$

首先利用对数映射  $\log_{\mathbf{o}}^{K_{l-1}}(\cdot)$ , 将输入的双曲嵌入式向量映射到原点  $\mathbf{o}$  所在的欧式切平面空间  $T_{\mathbf{o}}\mathbb{B}^{d, K}$ , 然后是非线性激活函数, 然后再指数映射  $\exp_{\mathbf{o}}^{K_l}(\cdot)$  回双曲空间  $\mathbb{B}^{d, K}$ 。双曲非线性激活函数可以在每一层平滑地变化曲率, 并保证最后可以映射回原点  $\mathbf{o}$  所在的双曲空间中, 与式(6)保持空间一致, 从而保证迭代更新的正确性。

#### 4.4 模型整体框架

基于以上章节对HyGAT-LP各个模块的介绍, HyGAT-LP整体框架如图2所示。

在知识图谱多关系图  $G = (V, R)$  中, 假设实体节点  $v_i$  为中心节点, 首先输入  $v_i$  的共  $T$  个邻居三元组的欧式嵌入式向量, 然后通过双曲空间和欧式空间的映射关系转换成双曲嵌入式向量, 并通过分开考虑关系注意力和实体注意力两种层面的注意力机制, 生成邻居三元组的双曲嵌入式向量和注意力, 用于融合邻居信息, HyGAT-LP中的信息传递过程如下:

$$\mathbf{h}_i^{l, H} = (\mathbf{W}^l \otimes_{K_{l-1}} \mathbf{v}_i^{l-1, H}) \odot_{K_{l-1}} \mathbf{b}_l \quad (17)$$

$$\mathbf{x}_i^{l, H} = \text{AGG}^{K_{l-1}}(\mathbf{h}_i^{l, H}) \quad (18)$$

$$\mathbf{v}_i^{l, H} = \sigma^{\otimes K_{l-1}, K_l}(\mathbf{x}_i^{l, H}) := \exp_{\mathbf{o}}^{K_l} \left( \sigma \left( \log_{\mathbf{o}}^{K_{l-1}} (\mathbf{v}_i^H) \right) \right) \quad (19)$$

主要包括双曲特征变换(式(17)), 基于注意力的邻居三元组信息融合(式(18))以及不同曲率的双曲非线性变换(式(19))。经过  $L$  层迭代更新后, 可以生成最终学习到的实体  $v_i$  的双曲嵌入式向量  $\mathbf{v}_i^{L, H}$ 。为了进一步提高注意力层的表达能力, 使用多头注意力机制, 即对以上注意力求解过程调用  $K$  组相互独立的注意力机制并进行求平均操作。

为了进行链路预测，利用式(20)计算三元组的得分作为衡量其成立可能性的依据，

$$f(v_h, r, v_t) = -d_{\mathbb{B}}\left(\mathbf{R}_d \otimes_K \mathbf{v}_h^{L,H}, \mathbf{v}_t^{L,H} \oplus_K \mathbf{r}^H\right)^2 + b_h + b_t \quad (20)$$

其中， $\mathbf{v}_h^{L,H}, \mathbf{v}_t^{L,H} \in \mathbb{B}^d$ 分别是头实体和尾实体的双曲嵌入式表示， $\mathbf{r}^H \in \mathbb{B}^d$ 是关系参数向量， $\mathbf{R}_d \in \mathbb{R}^{d \times d}$ 为对角关系矩阵，偏置项 $b_h$ 和 $b_t$ 是双曲面决策边界，确定了以 $v_h$ 为中心的超球的半径，当 $v_t$ 落到半径为 $\sqrt{b_h + b_t}$ 的超球面内时，三元组 $(v_h, r, v_t)$ 成立。由于头、尾实体的偏置并不一样，所以头尾实体决定不同的决策边界，关系特定的 $\mathbf{R}_d$ 和 $\mathbf{r}^H$ 决定了经关系调整后的实体嵌入的位置，而特定实体的决策边界的半径独立于关系。得分函数类似于现有的翻译模型的得分函数，只是加了偏置，这改变了模型的几何结构。偏置为每个实体定义了一个以实体的嵌入向量为中心的影响球体，这样实体不再被看作一个点。球体间的重合度量了两个实体间的相关性。关系可以看作在空间中去移动球体，只有由三元组中的关系连接的两个实体的球体才会重叠。

#### 4.5 模型训练与复杂度分析

为了训练模型，将知识图谱中每个真实有效三元组 $(v_h, r, v_t)$ 中的尾实体 $v_t$ 替换成实体集合 $\mathbf{V}$ 中随机选择的其他实体 $v_t' \in \mathbf{V} \setminus v_t$ 来生成 $(|\mathbf{V}| - 1)$ 个无效三元组 $(v_h, r, v_t')$ ，损失函数是伯努利负对数似然函数：

$$L(y, p) = -\frac{1}{|E_{\text{Train}}|} \sum_{i=1}^{|E_{\text{Train}}|} \cdot \left( y^{(i)} \lg(p^{(i)}) + (1 - y) \lg(1 - p^{(i)}) \right) \quad (21)$$

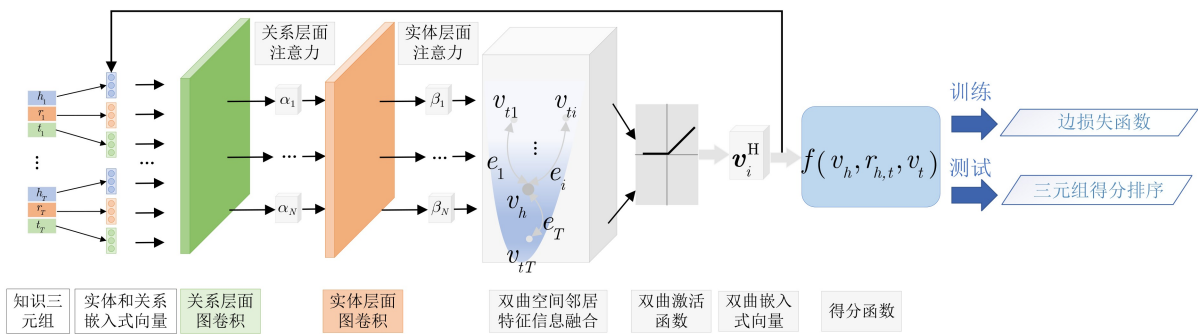


图2 HyGAT-LP模型框架图

表1 FB15k-237数据集和WN18RR数据集统计量信息

数据集	V	R	三元组数			ρ	
			训练集 $E_{\text{Train}}$	验证集 $E_{\text{Valid}}$	测试集 $E_{\text{Test}}$		总三元组数 $E$
FB15k-237	14541	237	272115	17535	20466	310116	1308.5
WN18RR	40943	11	86835	3034	3134	93003	8454.8

其中， $p^{(i)}$ 为衡量三元组成立的概率， $y^{(i)}$ 为标记三元组为正例或者负例的二值标签。

用黎曼随机梯度<sup>[22,23]</sup>优化双曲模型。相比较于欧式梯度 $\nabla_E L$ ，与欧式梯度更新 $\theta \leftarrow \theta - \eta \nabla_E L$ 不同，黎曼梯度 $\nabla_R L$ 在欧式梯度的基础上乘以庞加莱度量张量的逆，即 $\nabla_R L = 1/(\lambda_v^K)^2 \nabla_E L$ ，黎曼更新的第一步用 $\exp_{\theta}^K$ 将梯度 $\nabla_R L \in T_{\theta} \mathbb{B}^{d,K}$ 映射到庞加莱球的对应的测地线上，然后计算黎曼更新 $\theta \leftarrow \exp_{\theta}^K(-\eta \nabla_R L)$ 。

复杂度方面，计算注意力机制的总时间复杂度为 $O(N_v d^2)$ ，其中 $N_v$ 为实体个数。将邻居三元组通过注意力融合函数的图卷积操作的总时间复杂度为 $O(|E| d^2)$ ，其中 $|E|$ 为多关系图 $G = (V, R)$ 所含的三元组个数。模型的总体时间复杂度为 $O\left(\sum_{k=1}^K \sum_{l=1}^L (N_v d^2 + |E| d^2)\right)$ ，其中 $L$ 表示卷积层数， $K$ 表示多头注意力的头数。

## 5 实验结果及分析

### 5.1 数据集

为了评估所提算法的有效性，本文在FB15k-237和WN18RR这2个常用的基准数据集上开展实验。数据集的统计信息如表1所示。其中 $\rho$ 表示平均每一类关系的三元组数，体现了数据的稠密程度。FB15k-237数据集是知识图谱 Freebase的一个子集，包含真实世界中的事实，少部分关系具有层级特性，如part-of等，数据集稀疏(每一类关系对应的实体数量较小)，网络分层特性不明显。WN18RR数据集是知识图谱WordNet的子集，包含了单词之间的层级关系，如hypernym, has\_part等，数据

集稠密(每一类关系对应的实体数量较大), 具备天然的层次结构, 网络分层特性明显。

## 5.2 评价指标和参数设置

实验中使用倒数平均排序(Mean Reciprocal Rank, MRR)和Hits@1, Hits@3, Hits@10(排在前1, 3, 10名的有效实体的比例)作为评价指标。MRR是将排名取倒数使结果落在(0,1]之间, 值越大则模型效果越好。Hits@1, Hits@3, Hits@10表示在所有的候选集中正确答案排在第1, 3, 10名的比例, 值越大模型效果越好。将验证集上MRR最好的模型在测试集上运行来获取最终结果。

在训练阶段, 共使用 $L = 2$ 即两层卷积层来训练学习实体和关系的双曲嵌入式向量, 多头注意力机制对应头的数量为8, 卷积层中的非线性激活函数采用LeakyReLU函数, 模型采用Glorot方式进行模型参数初始化, Adam优化器最小化代价函数, 学习率为0.01。实体和关系的嵌入式向量维度为200, 批处理大小为128, 负采样数为50, 模型迭代次数为500。

## 5.3 实验结果和分析

### 5.3.1 模型总体性能对比

为评估所提算法对知识图谱链路预测任务的有效性, 实验选取转移距离模型TransE<sup>[7]</sup>, RotatE<sup>[8]</sup>, MuRE<sup>[20]</sup>, 语义匹配模型DistMult<sup>[10]</sup>, ComplEx<sup>[11]</sup>, TuckER<sup>[12]</sup>, 基于卷积神经网络的模型ConvE<sup>[13]</sup>、ConvKB<sup>[14]</sup>, 基于图神经网络的模型RGHAT<sup>[16]</sup>和基于双曲空间的转移距离模型MuRP<sup>[20]</sup>、ATTH<sup>[21]</sup>共11种模型作为基准算法与本文所提HyGAT-LP模型进行对比, 基准算法的实验数据参照原文献中的实验结果进行选取。

实验结果如表2所示, 其中加粗的为最优结果, 带下划线的为次优结果。所有对比算法所属的流形总共可以分为3类: 欧式空间 $\mathbb{R}$ 、复数域空间 $\mathbb{C}$ 和双曲空间 $\mathbb{B}$ 。实验结果显示, 欧式空间模型中, TransE模型的向量平移在以组合关系模式占主体的稀疏数据集FB15k-237上能有效捕捉三元组的全局特征; 但在对称关系模式占主体的WN18RR稠密数据集上, 特别在处理复杂关系类型时, 实体的嵌入式表示向量会趋向于相近, 导致MRR得分较低。与转移距离模型相反, DistMult和ComplEx所使用的双线性乘法运算擅长提取实体相似性特征, 在稠密数据集WN18RR上表现较好, 但在稀疏数据集FB15k-237上, 则难以提取足够的信息优化实体表示, MRR和Hits@10都有明显下滑。ConvE在两个数据集的几乎所有指标上都表现出色, 说明对实体和关系向量的拼接及2维转化有助于CNN提取局部模式与关系特征。ConvKB结合了CNN和TransE的平移特性, 但却没有解决实体嵌入式表示向量会趋向于相近的问题, 因此在WN18RR的MRR指标上同样表现出性能下滑。RotatE模型实现了推导对称/反对称、反向以及组合模式, 性能相较于以上几种模型有较大提升。TuckER模型是对以上模型的统一, 性能十分优秀。而RGHAT证明了基于图神经网络的模型相比较于以上几乎所有的非图神经网络模型性能更优, 则体现了图卷积网络和注意力机制提取的特征比传统算法提取的特征更加丰富和准确, 对知识图谱这类图结构数据的学习能力更强。

整体而言, 在分层特性不明显的FB15k-237数据集上, 基于双曲空间的模型和基于欧式空间的模

表2 FB15k-237数据集和WN18RR数据集上知识图谱链路预测结果

流形	模型	FB15k-237				WN18RR			
		MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
$\mathbb{R}$	TransE	0.294	0.186	0.321	0.465	0.226	0.427	0.441	0.532
	DistMult	0.241	0.155	0.263	0.419	0.444	0.412	0.470	0.504
	ConvE	0.325	0.237	0.356	0.501	0.456	0.419	0.470	0.531
	ConvKB	0.289	0.198	0.324	0.471	0.265	0.058	0.445	0.558
	MuRE	0.336	0.245	0.370	0.521	0.475	0.436	0.487	0.554
	TuckER	0.358	0.266	0.394	0.544	0.471	<b>0.443</b>	0.482	0.526
	RGHAT	<u>0.522</u>	<u>0.462</u>	<u>0.546</u>	<u>0.631</u>	0.483	0.425	<u>0.499</u>	<u>0.588</u>
$\mathbb{C}$	ComplEx	0.247	0.158	0.275	0.428	0.449	0.409	0.469	0.530
	RotatE	0.338	0.241	0.375	0.533	0.476	0.428	0.492	0.571
$\mathbb{B}$	MuRP	0.335	0.243	0.367	0.518	0.481	0.440	0.495	0.566
	ATTH	0.384	0.252	0.384	0.540	<u>0.486</u>	<b>0.443</b>	0.498	0.573
	HyGAT-LP	<b>0.529</b>	<b>0.467</b>	<b>0.566</b>	<b>0.650</b>	<b>0.521</b>	<u>0.441</u>	<b>0.511</b>	<b>0.619</b>

型以及基于复数域空间的模型性能相当, MuRP和 ATTH两种双曲空间模型与9种非双曲空间模型相比, MRR, Hits@1, Hits@3, Hits@10分别平均提升9.7%, 3.7%, 4.8%和5.5%, 但是在分层特性明显的WN18RR数据集上, MRR, Hits@1, Hits@3, Hits@10分别平均提升16.2%, 14.9%, 5%和4.7%, 说明双曲空间比欧式空间更适合建模知识图谱类型的树状层级多关系图数据。

本文所提的HyGAT-LP方法, 综合了双曲空间嵌入式表示学习和图注意力网络两者的优势, 且分层级地考虑了邻居三元组的影响, 在绝大多数的评价指标上都取得了最好的结果。HyGAT-LP模型与单独基于双曲空间的模型MuRP和ATTH相比, 在FB15k-237数据上, MRR, Hits@1, Hits@3, Hits@10 4种指标分别平均提升了47.1%, 88.7%, 50.7%和22.9%; 在WN18RR数据上, 4种指标分别平均提升了7.8%, 0%, 2.9%和8.7%。相对而言, 在FB15k-237数据集上提升幅度比在WN18RR数据集上提升幅度大, 验证了在同样的双曲空间条件下, 挖掘和利用邻居三元组信息能有效地学习不同逻辑模式的关系类型, 进一步丰富双曲空间嵌入式向量包含的特征信息, 从而提高链路预测的性能。

### 5.3.2 不同类型关系的性能度量分析

由于并非所有类型的关系都可以在实体上诱导出层级结构, 因此本文研究由每类关系形成的知识图谱的图曲率 $\xi_G$ <sup>[24]</sup>以及层次得分(Krackhardt hierarchy score, Khs)<sup>[25]</sup>, 进而分析每类关系的性能度量。图曲率是几何群论中的一个概念, 用于衡量图的树形结构化程度。层次得分仅针对有向网络定义, 并且测量其中存在有向路径( $x \rightarrow y$ )而不存在( $y \rightarrow x$ )的节点对( $x, y$ )的比重。对于所有有向非循环图, 该分数取值为1, 对于圈和自环, 该分数取值为0。图曲率 $\xi_G$ 越低, 层次得分Khs越高, 则该图的分层特性越明显, 图数据结构越趋向于树形结构。以WN18RR数据集为例, 此数据集中总共有11种类型的关系, 其中可以诱导出层级结构的关系有8种, 另外3种语义关系不具备层次结构, 每类数据的基本统计量信息如表3所示。

为了分析不同类型关系的性能度量受双曲空间和邻域三元组注意力机制的影响情况, 实验比较了MuRE模型、MuRP模型和HyGAT-LP模型的每类关系的MRR指标。对比结果如图3所示。从图3可以看出, 双曲空间嵌入式表示方式提升了层级关系的预测性能, 而在非层级关系的预测方面, 基于双曲空间和基于欧式空间的预测效果相似。HyGAT-LP模型相比较于MuRP模型而言, 对非层级关系的预

表3 在WN18RR数据集中每类关系的统计量信息

关系名	简记	总三元组数	Khs	$\xi_G$
hypernym	HY	37221	1.00	-2.46
has_part	HP	5142	1.00	-1.43
member_meronym	MM	7928	1.00	-2.90
synset_domain_topic_of	SDTO	3335	0.99	-0.69
instance_hypernym	IH	3150	1.00	-0.82
member_of_domain_region	MODR	983	1.00	-0.78
member_of_domain_usage	MODU	675	1.00	-0.74
also_see	AS	1396	0.36	-2.09
derivationally_related_form	DRF	31867	0.04	-3.84
similar_to	ST	86	0	-1.00
verb_group	VG	1220	0	-0.50

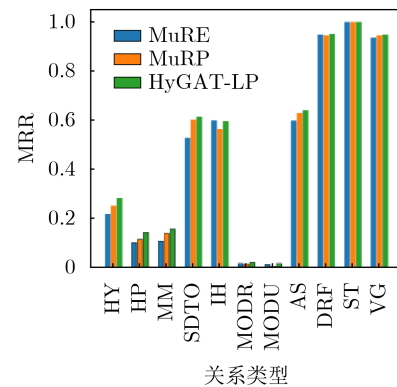


图3 WN18RR数据集上各类关系MRR指标对比图

测效果有进一步的提升, 且同时优于MuRE模型, 说明如果层级信息没有在数据集中起到决定性作用, 则模型的结构和深度对预测结果的影响会更加占据主导地位。

### 5.3.3 知识三元组实例的性能度量分析

在WN18RR数据集中, 当测试三元组分别为具备层级关系的(european\_union, member\_meronym, denmark)和非层级关系的(geology, derivationally\_related\_form, geologist)时, 在已知头实体和关系的条件下预测尾实体, 实验比较了MuRE模型、MuRP模型和HyGAT-LP模型对尾实体的排序情况如图4所示, 其中模型每训练迭代5次, 测试输出1次排序结果。从图4可以看出, 随着训练迭代次数的增加, 3个模型对目标尾实体预测的排名都呈逐步上涨的趋势, 其中双曲空间模型对层级关系三元组的预测优于欧式空间模型, 同等迭代次数条件下, 双曲空间模型对尾实体的排名基本上都比欧式空间模型对尾实体的排名更加靠前, 且双曲空间模型的最佳排名优于欧式空间的最佳排名。

HyGAT-LP模型在3个对比模型中收敛最快，且最佳排名最靠前。对于非层级关系三元组的预测，3个模型的最佳排名都是第1名，但是HyGAT-LP模型最快收敛到最佳排名，优于其余两个对比模型。

3个模型在最后一轮训练后的预测排名前10尾实体列表如表4所示。底色加深的实体为目标实体。从表4可以看出，MuRE模型和MuRP模型都

会选出与目标实体语义差别较大的实体，如层级关系中的united\_states和非层级关系中的give, film和chronologize，而HyGAT-LP模型不会出现这种情况，且选出的排名靠前的尾实体与目标尾实体更加相似，与头实体和关系形成的三元组更加符合现实意义。以上实验结果也从具体个体实例层面进一步验证了HyGAT-LP模型的有效性。

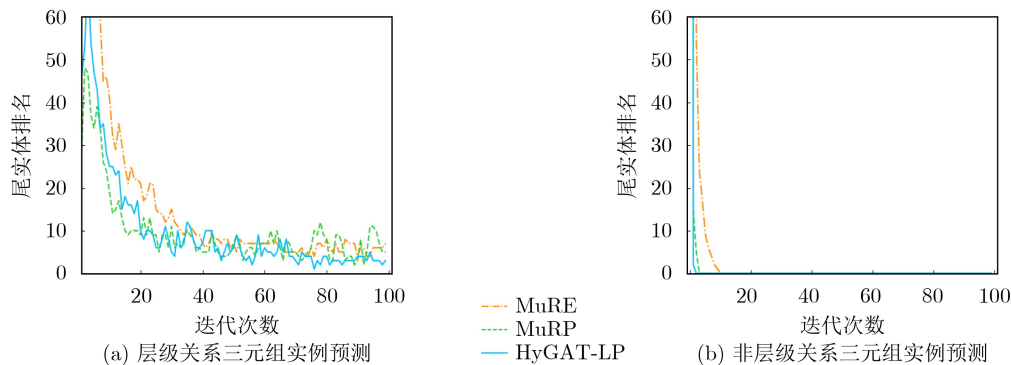


图4 实例三元组尾实体预测排名对比图

表4 预测排名前10尾实体情况

排序	(european_union, member_meronym, ?)			(geology, derivationally_related_form, ?)		
	MuRE	MuRP	HyGAT-LP	MuRE	MuRP	HyGAT-LP
1	england	noway	bulgaria	geologist	geologist	geologist
2	poland	bulgaria	denmark	give	scientist	paleontologist
3	scotland	denmark	europa	psychologist	anatomist	doctor
4	switzerland	switzerland	noway	explorer	meteorologist	scientist
5	denmark	poland	iceland	religionist	film	biologist
6	noway	iceland	belarus	theologian	biologist	geophysicist
7	bulgaria	scotland	switzerland	ophthalmologist	geomorphologic	geomorphologic
8	united_states	croatia	hungary	philosopher	paleontologist	philosopher
9	europa	belarus	ireland	diagnostician	chronologize	physicist
10	turkey	united_states	czechoslovakia	specialist	geophysicist	musician

## 6 结束语

现有基于欧式空间表示学习的知识图谱链路预测算法，仍然存在表示向量高度失真和三元组局部邻居信息利用不充分的问题。因此本文提出了HyGAT-LP算法，旨在学习知识图谱中实体和关系的双曲嵌入式表示，提升知识图谱链路预测的准确度，主要创新点在于设计并实现了基于双曲注意力网络模型，将多关系知识图谱嵌入到符合其所需几何特性的双曲空间，联合图注意力网络对局部结构化信息的挖掘能力，共同应对其复杂的网络结构和实体间复杂的关系。多个数据集上的实验结果表明，算法的性能稳定，与主流的基准算法相比有一定程度上的提高，特别是在树状分层特性更加明显

的WN18RR数据集上优势更为明显。未来的工作包括探索将本文所提算法扩展到对话系统、自动问答系统等知识图谱领域中。此外还将探索融合球形、欧式以及双曲等不同曲率空间的注意力网络模型，从而契合各式各样结构的图数据，使得算法的适用场景更加广阔。

## 参考文献

- [1] 王昊奋, 漆桂林, 陈华钧. 知识图谱: 方法、实践与应用[M]. 北京: 电子工业出版社, 2019: 1-3.  
WANG Haofen, QI Guilin, and CHEN Huajun. Knowledge Graph[M]. Beijing: Publishing House of Electronics History, 2019: 1-3.

- [2] REINANDA R, MELJ E, and DE RIJKE M. Knowledge graphs: An information retrieval perspective[J]. *Foundations and Trends® in Information Retrieval*, 2020, 14(4): 289–444. doi: [10.1561/1500000063](https://doi.org/10.1561/1500000063).
- [3] 张崇宇. 基于知识图谱的自动问答系统的应用研究与实现[D]. [硕士论文], 北京邮电大学, 2019.
- ZHANG Chongyu. Application research and implementation of automatic question answering system based on knowledge graph[D]. [Master dissertation], Beijing University of Posts and Telecommunications, 2019.
- [4] 秦川, 祝恒书, 庄福振, 等. 基于知识图谱的推荐系统研究综述[J]. *中国科学:信息科学*, 2020, 50(7): 937–956. doi: [10.1360/SSI-2019-0274](https://doi.org/10.1360/SSI-2019-0274).
- QIN Chuan, ZHU Hengshu, ZHUANG Fuzhen, *et al.* A survey on knowledge graph-based recommender systems[J]. *Scientia Sinica Informationis*, 2020, 50(7): 937–956. doi: [10.1360/SSI-2019-0274](https://doi.org/10.1360/SSI-2019-0274).
- [5] BRONSTEIN M M, BRUNA J, LECUN Y, *et al.* Geometric deep learning: Going beyond Euclidean data[J]. *IEEE Signal Processing Magazine*, 2017, 34(4): 18–42. doi: [10.1109/MSP.2017.2693418](https://doi.org/10.1109/MSP.2017.2693418).
- [6] WANG Quan, MAO Zhendong, WANG Bin, *et al.* Knowledge graph embedding: A survey of approaches and applications[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(12): 2724–2743. doi: [10.1109/TKDE.2017.2754499](https://doi.org/10.1109/TKDE.2017.2754499).
- [7] BORDES A, USUNIER N, GARCIA-DURÁN A, *et al.* Translating Embeddings for modeling multi-relational data[C]. Proceedings of the 26th International Conference on Neural Information Processing Systems (NIPS), Lake Tahoe, USA, 2013: 2787–2795. doi: [10.5555/2999792.2999923](https://doi.org/10.5555/2999792.2999923).
- [8] SUN Zhiqing, DENG Zhihong, NIE Jianyun, *et al.* RotatE: Knowledge graph embedding by relational rotation in complex space[EB/OL]. <https://arxiv.org/abs/1902.10197>, 2019.
- [9] NICKEL M, TRESP V, and KRIEGEL H. A three-way model for collective learning on multi-relational data[C]. Proceedings of the 28th International Conference on Machine Learning (ICML), Bellevue, USA, 2011: 809–816.
- [10] YANG Bishan, YIH W T, HE Xiaodong, *et al.* Embedding entities and relations for learning and inference in knowledge bases[EB/OL]. <https://arxiv.org/abs/1412.6575>, 2015.
- [11] TROUILLON T, WELBL J, RIEDEL S, *et al.* Complex embeddings for simple link prediction[C]. Proceedings of the 33rd International Conference on International Conference on Machine Learning (ICML), New York, USA, 2016: 2071–2080. doi: [10.5555/3045390.3045609](https://doi.org/10.5555/3045390.3045609).
- [12] BALAZEVIC I, ALLEN C, and HOSPEDALES T. TUCKER: Tensor factorization for knowledge graph completion[C]. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 2019: 5184–5194. doi: [10.18653/v1/D19-1522](https://doi.org/10.18653/v1/D19-1522).
- [13] DETTMERS T, MINERVINI P, STENETORP P, *et al.* Convolutional 2D knowledge graph embeddings[EB/OL]. <https://arxiv.org/abs/1707.01476>, 2018.
- [14] NGUYEN D Q, NGUYEN D Q, NGUYEN T D, *et al.* A convolutional neural network-based model for knowledge base completion and its application to search personalization[J]. *Semantic Web*, 2019, 10(5): 947–960. doi: [10.3233/SW-180318](https://doi.org/10.3233/SW-180318).
- [15] SCHLICHTKRULL M, KIPF T N, BLOEM P, *et al.* Modeling relational data with graph convolutional networks[C]. 15th International Conference on The Semantic Web, Heraklion, Greece, 2018: 593–607. doi: [10.1007/978-3-319-93417-4\\_38](https://doi.org/10.1007/978-3-319-93417-4_38).
- [16] ZHANG Zhao, ZHUANG Fuzhen, ZHU Hengshu, *et al.* Relational graph neural network with hierarchical attention for knowledge graph completion[C]. The AAAI Conference on Artificial Intelligence (AAAI), Palo Alto, USA, 2020: 9612–9619. doi: [10.1609/aaai.v34i05.6508](https://doi.org/10.1609/aaai.v34i05.6508).
- [17] WU Zonghan, PAN Shirui, CHEN Fengwen, *et al.* A comprehensive survey on graph neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(1): 4–24. doi: [10.1109/TNNLS.2020.2978386](https://doi.org/10.1109/TNNLS.2020.2978386).
- [18] 王强, 江昊, 羿舒文, 等. 复杂网络的双曲空间表征学习方法[J]. *软件学报*, 2021, 32(1): 93–117. doi: [10.13328/j.cnki.jos.006092](https://doi.org/10.13328/j.cnki.jos.006092).
- WANG Qiang, JIANG Hao, YI Shuwen, *et al.* Hyperbolic representation learning for complex networks[J]. *Journal of Software*, 2021, 32(1): 93–117. doi: [10.13328/j.cnki.jos.006092](https://doi.org/10.13328/j.cnki.jos.006092).
- [19] ZHANG Chengkun and GAO Junbin. Hype-HAN:

- Hyperbolic hierarchical attention network for semantic embedding[C]. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI), Yokohama, Japan, 2020: 3990–3996.
- [20] BALAŽEVIC I, ALLEN C, and HOSPEDALES T. Multi-relational poincaré graph embeddings[EB/OL]. <https://arxiv.org/abs/1905.09791>, 2019.
- [21] CHAMI I, WOLF A, JUAN Dacheng, *et al.* Low-dimensional hyperbolic knowledge graph embeddings[C]. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL), Seattle, USA, 2020: 6901–6914. doi: [10.18653/v1/2020.acl-main.617](https://doi.org/10.18653/v1/2020.acl-main.617).
- [22] BONNABEL S. Stochastic gradient descent on Riemannian manifolds[J]. *IEEE Transactions on Automatic Control*, 2013, 58(9): 2217–2229. doi: [10.1109/TAC.2013.2254619](https://doi.org/10.1109/TAC.2013.2254619).
- [23] SAKAI H and IIDUKA H. Riemannian adaptive optimization algorithm and its application to natural language processing[J]. *IEEE Transactions on Cybernetics*, 2021, 51(1): 1–12. doi: [10.1109/TCYB.2021.3049845](https://doi.org/10.1109/TCYB.2021.3049845).
- [24] ADCOCK A B, SULLIVAN B D, and MAHONEY M W. Tree-like structure in large social and information networks[C]. IEEE 13th International Conference on Data Mining, Dallas, USA, 2013: 1–10. doi: [10.1109/ICDM.2013.77](https://doi.org/10.1109/ICDM.2013.77).
- [25] KRACKHARDT D. *Computational Organization Theory*[M]. London: Psychology Press, 1994: 89–111.
- 吴 铮: 男, 1992年生, 博士生, 研究方向为图神经网络、网络大数据分析.
- 陈鸿昶: 男, 1964年生, 教授, 研究方向为智能信息处理、人工智能.
- 张建朋: 男, 1988年生, 助理研究员, 研究方向为数据挖掘、社会网络分析.

责任编辑: 陈 倩