

融合实体多层语义的四元数表示学习方法

冯艳¹, 蒲飞^{2,†}, 杨柏林³

浙江工商大学计算机科学与技术学院

杭州, 310016

2607220876@qq.com, puf2008@gmail.com, ybl@zjgsu.edu.cn

摘要

知识图谱作为人工智能领域中一个重要的研究分支, 通常采用三元组(头实体, 关系, 尾实体)的形式描述和存储客观世界中的信息。由于知识图谱普遍存在数据稀疏的问题, 即知识图谱不完备问题, 因此需要借助机器学习和表示学习来补全知识图谱。针对这个问题, 本文提出一种融合实体多层语义的四元数表示学习方法, 通过将实体和关系建模在超复数空间内, 用四元数和单位四元数分别表示头尾实体和关系, 利用关系单位四元数对头实体和尾实体分别做右等斜旋转和逆方向的右等斜旋转来融合实体的多层语义信息, 在能够推理和建模多种复杂关系模式(对称/反对称、反转和组合关系模式等)的同时, 也能捕获到实体和关系之间隐含的多层交互信息。本文对比几个主流的知识图谱嵌入模型, 在四个公开数据集WN18、WN18RR、FB15K和FB15K-237上进行链接预测实验, 实验结果显示, 本文提出的方法相对于已有的知识图谱嵌入模型的性能有了明显改进和提升。

关键词: 知识图谱; 表示学习; 四元数; 多层语义; 链接预测

Quaternion Representation Learning Method to Fusion Multilayer Semantics of Entities

Yan Feng¹, Fei Pu^{2,†}, Bailin Yang³

School of Computer Science and Technology, Zhejiang Gongshang University
Hangzhou, 310016

2607220876@qq.com, puf2008@gmail.com, ybl@zjgsu.edu.cn

Abstract

As an important research branch in the field of artificial intelligence, Knowledge graph usually uses the form of triplets (head entities, relationships, tail entities) to describe and store information in the objective world. Due to the widespread problem of sparse data in knowledge graphs, so needed use machine learning and representation learning to complement the knowledge graph. To solve this problem, this paper proposes a quaternion representation learning method to fusion multilayer semantics of entities, by modeling entities and relationships in the hypercomplex number space, quaternions and unit quaternions are used to represent the head and tail entities and relationships, and the relationship unit quaternions are used to make right isoclinal rotation and reverse right isoclinal rotation on the head and tail entities respectively to fuse the context Semantic information of entities, While reasoning and modeling a variety of complex relationship patterns (symmetric and anti-symmetric, reverse and composite relationship patterns and so on), it can also capture the implicit multilayer interaction information between entities and relationships. This article compares

several mainstream knowledge graph embedding models and conducts link prediction experiments on four public datasets WN18, WN18RR, FB15K, and FB15K-237. The experimental results show that the proposed method has significantly improved and promoted the performance of existing knowledge graph embedding models.

Keywords: Knowledge graph , Representation learning , Quaternion , Multilayer semantic , Link prediction

1 引言

随着大数据时代的发展，互联网中源源不断产生形式多样的数据信息，为了满足学术研究和生产生活的发展要求，需要找到一种能够有效挖掘、组织和管理海量信息的方式。学术界和工业界将注意力放在了具有极强文本处理和开放互通能力的知识图谱(Knowledge Graph,KG)上，知识图谱作为人工智能应用领域的关键技术，本质上是一种语义网络，旨在将客观世界中的事物描述成易于机器理解和处理的形式，以达到人与机器无障碍沟通的目的。知识图谱的构建过程能够模拟人的认知思维，极大的促进了知识挖掘与知识推理等相关领域的发展。目前互联网中已有一些大规模知识图谱供各行各业使用，并在问答系统、信息检索、推荐系统、自然语言处理等领域得到了广泛的应用，例如LinkedData(Bizer et al, 2008)，WordNet(Miller and George A, 1995)，Freebase(Bollacker et al, 2008)，DBpedia(Auer S et al, 2007)，YAGO(Suchanek et al, 2007)等。知识图谱通常采用三元组 (h, r, t) 的形式作为存储知识的基本格式， h 和 t 分别代表头实体和尾实体， r 代表头尾实体之间的关系。

知识图谱的构建与应用依赖于知识表示，但是基于三元组的存储形式存在着计算复杂度高、可重用率低和数据稀疏等问题，并且无法完全描述实体和关系间隐含的语义信息。随着深度学习技术的发展，知识图谱表示学习技术利用机器学习的手段将实体和关系表示为低维稠密的向量，不仅解决了数据稀疏等问题，也使其在知识图谱应用层面发挥着重要作用。虽然知识图谱中包含着大量现实世界中的概念、实体及其之间的关联关系，但由于客观世界知识的增长速度飞快，导致大部分知识图谱都存在数据稀疏的问题，即缺少许多有效的三元组。因此，许多研究工作致力于链接预测任务以完成知识图谱补全，即根据已知的三元组信息去预测缺失的部分。

知识图谱嵌入模型将实体和关系表示为低维稠密的向量，通过学习与优化获得最为准确的表示形式。经典的翻译模型TransE模型(Bordes et al, 2013)相对于能量模型SE(Bordes et al, 2011)和SME模型(Bordes et al, 2014)，通过简单有效的嵌入方法在FB15k和WN18数据集上获得了更好的预测性能。TransE模型将关系描述为实数空间中头实体到尾实体的平移，即当三元组 (h, r, t) 成立时，能够达到： $h+r \approx t$ 。TransE模型在建模复杂关系模式时(1-N, N-1, N-N)效果很差，因此，在TransE模型基础之上，TransH模型(Wang et al, 2014)、TransR模型(Lin et al, 2015)和TransD模型(Ji et al, 2015)等一系列翻译模型被提出。但以上这些模型都是对从知识图谱中观察到的关系模式进行建模和推理，而忽略了一些隐含的关系模式，例如对称关系(朋友)/反对称关系(父子)，反转关系(上位词和下位词)，组合关系(爸爸的妻子是妈妈)。如何实现对这三种复杂关系模式同时建模变得至关重要。TransE模型通过头尾实体之间的平移操作能够建模反转关系和组合关系，但无法建模对称关系，TransH模型、TransR模型和TransD模型作为对TransE模型的扩展，能够建模对称和反对称关系，但是在建模反转和组合关系时性能较差。DisMult模型(Yang et al, 2014)通过学习头尾实体和关系三者之间的交互来推理和建模对称关系，但对其他关系模式无能为力。ComplEx模型(Trouillon T et al, 2016)通过引入复数嵌入空间来建模对称和反对称关系模式，但是无法推理和建模组合关系。RotatE模型(Sun et al, 2019)通过将关系定义为复数空间中从头实体到尾实体的旋转操作，具有了对上

©2023 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

基金项目：浙江省重点研发计划项目(No.2021C02004, No.2019C01004)；浙江工商大学“数字+”学科建设管理项目(No.SZJ2022A009)

†通讯作者

述三种复杂关系模式同时建模的能力。传统的复数由一个实部和一个虚部组成，为了增强模型对复杂关系模式的建模能力，QuatE模型(Zhang et al, 2019)引入了超复数空间，用具有一个实部和三个虚部的四元数来描述头尾实体之间的旋转过程，四元数空间相对于传统的复数空间，能在具有更高自由度和灵活度的三维空间中表示旋转，拥有了对实体和关系更为准确的表示形式。

因此，本文选择采用四元数和单位四元数分别表示头尾实体和关系，能够同时建模多种复杂关系模式。此外，还利用单位关系四元数对头实体和尾实体做右等斜旋转和逆方向的右等斜旋转，以此来捕获实体和关系之间的交互信息。为了评估本文提出的方法，本文在WN18、WN18RR、FB15K、FB15K-237四个标准数据集上对比多个知识图谱嵌入模型进行了链接预测实验，实验结果显示本文提出的方法准确率更高，性能优于大部分嵌入模型。

2 相关工作和研究动机

2.1 相关工作

现有的知识图谱嵌入模型可大致分为翻译模型、语义匹配模型和神经网络模型三种，这些模型的评分函数不同，所能建模和推理的关系模式也各不相同。

TransE模型(Bordes et al, 2013)将实体和关系建模在同一实数空间，采用头实体和尾实体之间的平移距离来刻画三元组成立的可能性，并将这种平移操作解释为一种翻译过程，TransE模型的评分函数如公式(1)：

$$f_r(h, t) = \|h + r - t\| \quad (1)$$

其中 h ， t ， r 分别是头尾实体和关系的嵌入向量，在最理想情况下，TransE模型 $h + r$ 无限接近 t 。

作为首个翻译模型，TransE模型通过简单有效的操作和较少的参数实现了对逆关系和组合关系的推理与建模，但是不支持建模对称/反对称关系模式。为了解决这个问题，基于TransE模型的一系列扩展模型被提出。TransH模型(Wang et al, 2014)引入了超平面的概念，即为每个关系都设定了一个专属的投影超平面，区分了实体在不同关系中的表示形式。TransH模型的评分函数定义如下：

$$f_r(h, t) = \|(h - W_r^T h W_r) + r - (t - W_r^T t W_r)\| \quad (2)$$

其中， W_r 是关系 r 的法向量。

TransR模型(Lin et al, 2015)认为，不同的关系应具有不同的语义空间，并利用投影矩阵代替投影向量，为每个关系 r 都设置一个投影矩阵 M_r 来构造其专属的语义空间，并将评分函数定义为：

$$f_r(h, t) = \|M_r h - M_r t\| \quad (3)$$

DistMult模型(Yang et al, 2014)基于双线性规则，将关系矩阵设置为对角矩阵，通过矩阵乘法来捕获实体间的线性关系，通过对头实体、尾实体和关系三者之间的交互信息进行学习，能够推理和建模对称关系模式，其评分函数定义如下：

$$f_r(h, t) = h^T \text{diag}(r) t \quad (4)$$

ComplEx模型(Trouillon T et al, 2016)通过引入复数空间来完善DistMult模型，使同一实体在作为头实体和尾实体时的嵌入向量为复共轭，能够同时捕获对称和非对称关系，但无法建模和推理组合关系。其评分函数定义如下：

$$f_r(h, t) = \text{Re}(h^T \text{diag}(r) \bar{t}) \quad (5)$$

其中， h ， r ， t 是头实体、关系和尾实体的嵌入向量， \bar{t} 是尾实体向量的共轭， Re 表示复数的实部。

与上述模型不同，RotatE模型(Sun et al, 2019)能够同时有效建模3种复杂关系模式，受欧

拉恒等式的启发，RotatE模型将关系表示为复数空间中头实体到尾实体之间的旋转过程，并采用Hadamard乘积加强向量之间的交互，其评分函数定义如下：

$$f_r(h, t) = \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\| \quad (6)$$

为了进行更高空间自由度的建模，QuatE模型将复数空间扩展至四元数空间，将实体和关系向量表示为四元数形式，评分函数定义如下：

$$f_r(h, t) = \mathbf{Q}_h \otimes \mathbf{Q}_r \cdot \mathbf{Q}_t \quad (7)$$

其中， \otimes 表示向量之间的Hamilton乘积， \cdot 表示向量之间的点积操作。QuatE模型(Zhang et al, 2019)利用四元数的几何特性拥有了两个旋转平面，极大地丰富了对实体和关系的语义表示，而RotatE模型只有一个旋转平面，并且RotatE模型中的欧拉角旋转易引起万向节锁死现象，而四元数旋转可以有效避免这一问题(Hamilton W R, 1844)。

除了以上模型，HAKE模型(Zhang et al, 2020)从实体层次语义感知的角度出发，将实体和关系建模在极坐标空间中，由极坐标向量分别刻画不同层级的实体（模数部分）和同一层级的不同实体（相位部分），其评分函数定义如下：

$$f_r(\mathbf{h}, \mathbf{t}) = d_r(\mathbf{h}, \mathbf{t}) = -d_{r,m}(\mathbf{h}, \mathbf{t}) - \lambda d_{r,p}(\mathbf{h}, \mathbf{t}) \quad (8)$$

通过以上对主流模型描述可知，提高模型的建模能力与挖掘实体的多层含义在知识图谱补全过程中尤为重要。四元数不同于传统复数的表示，能够在超复数空间中进行更高自由度和灵活度的旋转，在建模复杂关系模式方面具有明显优势。因此，我们考虑用四元数代替复数来表示实体和关系的嵌入向量，并通过四元数旋转来融合实体的多层语义信息。表1列出了几种最新的知识图谱补全模型推理与建模3种复杂关系模式的能力。

Model	Symmerty	Anti-symmerty	Reversal	Combination
TransE	×	✓	✓	✓
TranaX	✓	✓	×	×
DistMult	✓	×	×	×
ComplEx	✓	✓	✓	×
RotatE	✓	✓	✓	✓
MLQuatE	✓	✓	✓	✓

表1：模型推理关系模式的能力

Table 1: The ability of models to infer relational patterns

2.2 研究动机

为了能够更好的建模各种复杂关系模式（对称/反对称、反转和组合等），以及学习到实体的多种语义信息，例如实体“小米”在三元组（粮食作物，种类，小米）中表示一种供人们食用的粮食，而在三元组（手机，品牌，小米）中则表示一种手机品牌。本文采用超复数空间作为嵌入空间，用四元数和单位四元数表示实体和关系，相较于仅有一个欧拉角旋转平面的复数空间，如图1(a)所示，本文的方法具有两个旋转平面，如图1(b)，可以避免由欧拉角引起的万向节锁死问题（失去一个自由度），并且对实体和关系建模的灵活度与自由度更高。此外，相较于矩阵表示旋转，在超复数空间中，任何四维旋转都可以分解为一对左/右等斜旋转，即从左/右乘以单位四元数计算每四维旋转，因此四元数表示旋转使实体和关系的嵌入向量更加有效和数值稳定(Hamilton W R, 1844)。因此，本文提出了MLQuatE模型（Quaternion Representation Learning Method to Fusion Multilayer Semantics of Entities），具体来说，我们将实体和关系映射到超复数空间中，利用关系向量对头实体向量做右等斜旋转，再取其和尾实体的内积值，然后再利用关系向量对尾实体做逆方向的右等斜旋转，并取其和头实体的内积值，最后将这两次旋转的值相加用来衡量三元组成立的可能性。相对于在超复数空间中用四元数进行旋转操作来提高建模能力的QuatE模型，本文提出的模型在将实体和关系建模在超复数

空间的基础之上，还利用关系四元数分别对头尾实体都做了右等斜旋转和逆向的右等斜旋转，一方面实现了在四元数空间的旋转操作，增强了建模能力，另一方面通过右等斜旋转的计算，实体和关系向量实现了各个维度的交互，融合了实体更全面的各层次含义。因此，本文提出的方法不仅利用四元数的特性拥有了更强的建模能力，并且通过融合两次旋转操作来捕获实体和关系之间更多潜在交互的多层语义信息。

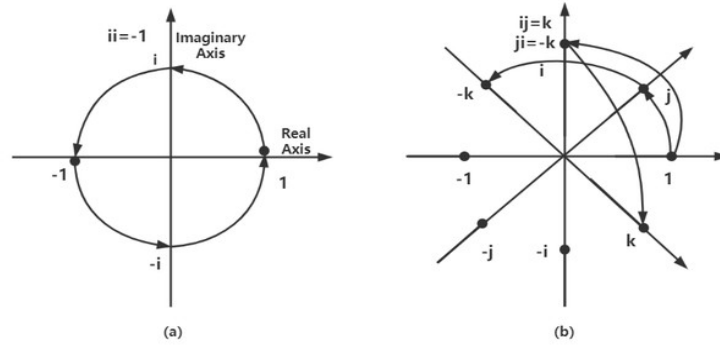


图1: 复数空间和超复数空间图示

Figure 1: Illustration of complex space and hypercomplex number space

3 算法设计

3.1 四元数基础知识

四元数是复数的不可交换延伸，由1个实部和3个虚部组成。四元数 Q 的表示形式定义为： $Q = a + bi + cj + dk$ ，其中 a, b, c, d 是实数， i, j, k 是虚数单位，并且满足条件： $ij = k, ji = -k, jk = i, kj = -i, ki = j, ik = -j$ 。

四元数的常用运算规则定义如下：

四元数的共轭：

$$\bar{Q} = a - bi - cj - dk \quad (9)$$

四元数的模：

$$|Q| = |\bar{Q}| = \sqrt{a^2 + b^2 + c^2 + d^2} \quad (10)$$

四元数的逆：

$$Q^{-1} = \frac{\bar{Q}}{|Q|^2} \quad (11)$$

四元数 $Q_1 = a_1 + b_1i + c_1j + d_1k$ 和 $Q_2 = a_2 + b_2i + c_2j + d_2k$ 的加法定义为：

$$Q_1 + Q_2 = (a_1 + a_2) + (b_1 + b_2)i + (c_1 + c_2)j + (d_1 + d_2)k \quad (12)$$

四元数 $Q_1 = a_1 + b_1i + c_1j + d_1k$ 和 $Q_2 = a_2 + b_2i + c_2j + d_2k$ 的减法定义为：

$$Q_1 - Q_2 = (a_1 - a_2) + (b_1 - b_2)i + (c_1 - c_2)j + (d_1 - d_2)k \quad (13)$$

四元数 $Q_1 = a_1 + b_1\mathbf{i} + c_1\mathbf{j} + d_1\mathbf{k}$ 和 $Q_2 = a_2 + b_2\mathbf{i} + c_2\mathbf{j} + d_2\mathbf{k}$ 的Hamilton乘积运算定义为:

$$\begin{aligned} Q_1 \otimes Q_2 = & (a_1a_2 - b_1b_2 - c_1c_2 - d_1d_2) \\ & + (b_1a_2 + a_1b_2 - d_1c_2 + c_1d_2)\mathbf{i} \\ & + (c_1a_2 + d_1b_2 + a_1c_2 - b_1d_2)\mathbf{j} \\ & + (d_1a_2 - c_1b_2 + b_1c_2 + a_1d_2)\mathbf{k} \end{aligned} \quad (14)$$

四元数 $Q_1 = a_1 + b_1\mathbf{i} + c_1\mathbf{j} + d_1\mathbf{k}$ 和 $Q_2 = a_2 + b_2\mathbf{i} + c_2\mathbf{j} + d_2\mathbf{k}$ 的点积运算定义为:

$$Q_1 \cdot Q_2 = a_1 \cdot a_2 + (b_1 \cdot b_2)\mathbf{i} + (c_1 \cdot c_2)\mathbf{j} + (d_1 \cdot d_2)\mathbf{k} \quad (15)$$

3.2 3种复杂关系模式定义

定义1 $\forall x, y \in E, r \in R$, 关系 r 是对称的 (反对称的), 如果

$$r(x, y) \Rightarrow r(y, x) (r(x, y) \Rightarrow \neg(y, x)) \quad (16)$$

定义2 $\forall x, y \in E, r_1, r_2 \in R$, 关系 r_1 和 r_2 是反转的, 如果

$$r_2(x, y) \Rightarrow r_1(y, x) \quad (17)$$

定义3 $\forall x, y, z \in E, r_1, r_2, r_3 \in R$, 关系 r_1 由关系 r_2 和关系 r_3 组成, 如果

$$r_2(x, y) \wedge r_3(y, z) \Rightarrow r_1(x, z) \quad (18)$$

3.3 四元数建模和推理3种复杂关系模式的证明

定理1 四元数可以建模和推理对称和反对称关系模式。

证明: 若 $r(x, y)$ 和 $r(y, x)$ 存在, 则有:

$$y = r \cdot x \wedge x = r \cdot y \Rightarrow r \cdot r = 1 \quad (19)$$

否则, 若 $r(x, y)$ 和 $\neg r(y, x)$ 存在, 则有:

$$y = r \cdot x \wedge x \neq r \cdot y \Rightarrow r \cdot r \neq 1 \quad (20)$$

定理2 四元数可以建模和推理反转关系模式。

证明: 若 $r_1 = (x, y)$ 和 $r_2 = (y, x)$ 存在, 则有:

$$y = r_1 \cdot x \wedge x = r_2 \cdot y \Rightarrow r_1 = r_2^{-1} \quad (21)$$

定理3 四元数可以建模和推理组合关系模式。

证明: 若 $r_1 = (x, z)$, $r_2 = (x, y)$ 和 $r_3 = (y, z)$ 存在, 则有:

$$z = r_1 \cdot x \wedge y = r_2 \cdot x \wedge z = r_3 \cdot y \Rightarrow r_1 = r_2 \cdot r_3 \quad (22)$$

3.4 模型算法

本文利用四元数最直观的几何和物理解释, 将头尾实体和关系投影在超复数空间中, 判断一个三元组是否合理分为以下三个步骤: (1)利用单位关系四元数对头部四元数做右等斜旋转, 取旋转后的头部四元数和尾部四元数之间的内积; (2)利用单位关系四元数的逆对尾部四元数做右等斜旋转, 取旋转后的尾部四元数和头部四元数之间的内积; (3)将步骤(1)和步骤(2)的值相加作为三元组的得分。若KG中存在三元组, 则模型会用关系 (关系的逆) 旋转头实体 (尾实体), 使旋转后的头实体 (尾实体) 和尾实体 (头实体) 之间的夹角尽可能小, 从而使三元组得分尽可能高。否则, 我们令两个实体呈正交状态, 则它们的内积为零。

更具体地说, 我们使用四元数 $Q \in H^{N \times k}$ 表示实体嵌入, $W \in H^{M \times k}$ 表示关系嵌入, 其中 k 是嵌入维度。给定一个三元组 (h, r, t) , 头实体 h 和尾实体 t 分别用 $Q_h = a_h + b_h\mathbf{i} + c_h\mathbf{j} + d_h\mathbf{k}$ 和 $Q_t = a_t + b_t\mathbf{i} + c_t\mathbf{j} + d_t\mathbf{k}$ 来表示, 关系 r 用 $W_r = a_r + b_r\mathbf{i} + c_r\mathbf{j} + d_r\mathbf{k}$ 来表示, 其

中, $a_h, b_h, c_h, d_h, a_t, b_t, c_t, d_t, a_r, b_r, c_r, d_r \in R^k$ 。我们首先将关系四元数 W_r 单位化为单位四元数, 以消除关系向量的缩放影响:

$$W'_r = \frac{W_r}{|W_r|} = \frac{a_r + b_r i + c_r j + d_r k}{\sqrt{a_r^2 + b_r^2 + c_r^2 + d_r^2}} = a'_r + b'_r i + c'_r j + d'_r k \quad (23)$$

然后, 利用关系 W_r 对头实体 Q_h 进行右等斜旋转, 即实体向量和关系向量做Hamilton乘积操作:

$$\begin{aligned} Q'_h(p, q, u, v) = Q_h \otimes W'_r = & (p \circ a'_r + q \circ b'_r + u \circ c'_r + v \circ d'_r) \\ & + (p \circ b'_r + q \circ a'_r + u \circ d'_r + v \circ c'_r) i \\ & + (p \circ c'_r + q \circ d'_r + u \circ a'_r + v \circ b'_r) j \\ & + (p \circ d'_r + q \circ c'_r + u \circ b'_r + v \circ a'_r) k \end{aligned} \quad (24)$$

其中, \circ 表示两个向量之间逐元素的乘法。单位四元数的右乘法是四元数上的右等斜旋转, 我们也可以交换 Q_h 和 W_r , 即对 Q_h 做左等斜旋转, 这不会从根本上改变几何意义。等斜旋转是双平面旋转的一种特殊情况, 其中每个平面的角度都相等。

最后, 我们计算经过旋转过的头实体和尾实体之间的内积值:

$$Q'_h \cdot Q_t = (p \cdot a_t) + (q \cdot b_t) + (u \cdot c_t) + (v \cdot d_t) \quad (25)$$

同样的, 也用关系的逆对尾实体做右等斜旋转, 再与头实体做内积操作。我们通过向量的Hamilton乘积和内积运算鼓励实体和关系向量各个维度的充分交互, 并将对实体向量的两次旋转操作进行叠加, 以此来刻画实体的多层语义信息。因此我们将评分函数定义为:

$$f_r(h, t) = Q_h \otimes \frac{W_r}{|W_r|} \cdot Q_t + Q_t \otimes \left(\frac{W_r}{|W_r|} \right)^{-1} \cdot Q_h \quad (26)$$

我们定义正则化逻辑损失函数来优化模型参数:

$$L(Q, W) = \sum_{r(h,t) \in \Omega \cup \Omega^-} \log(1 + \exp(-Y_{hrt} \theta(h, r, t))) + \lambda_1 \|Q\|_2^2 + \lambda_2 \|W\|_2^2 \quad (27)$$

在损失函数中, 采用正则化率为 λ_1 和 λ_2 的范数分别对 Q 和 W 做正则化处理, Ω^- 是使用例如均匀采样、伯努利采样(Wang et al, 2014)或对抗采样(Sun et al, 2019)等负采样策略采样得到的负样本。另外, 由于评分函数最后取得的是旋转后的头实体向量和尾实体向量之间的所有分量的总和(内积), 因此损失函数位于欧几里得空间。

4 实验设计与结果分析

4.1 数据集

本文的链接预测实验使用WN18、FB15K、WN18RR和FB15K-237四个数据集, 表2列出了这四个数据集中的实体和关系数量以及分别划分训练集、验证集和测试集所包含的三元组数量。

WN18数据集信息来源于英语词汇数据库WordNet, WordNet是由普林斯顿大学在1985年建立的包含大量实体和关系的在线词典数据库系统。WN18数据集共包含了151442个事实三元组, 主要的关系模式是对称/反对称和反转关系模式。

FB15K数据集信息来源于FreeBase知识库, FreeBase知识库是一个可以由用户添加内容的人机共同维护的数据库, 包含了音乐、电影和体育等多方面的数据信息, 更贴合现实生活中的客观事实。FB15K数据集共包含了592213个事实三元组, 主要的关系模式是对称/反对称和反转关系模式。

数据集	#实体	#关系	#训练集	#验证集	#测试集
WN18	40943	18	141442	5000	5000
FB15K	14951	1345	483142	50000	59071
WN18RR	40943	11	86835	3034	3134
FB15K-237	14951	237	272115	17535	20466

表2: 数据集统计

Table 2: Dataset statistics

WN18RR数据集是WN18数据集的子集, 参考文献(Toutanova K and Chen D, 2015)由于WN18数据集中包含许多可逆关系, 例如“上位词”和“下位词”, 这导致模型在训练时, 大量的测试三元组能在训练集中找到他们的逆三元组, 这种数据泄露增加了实验效果的不准确性, 因此将WN18数据集中的可逆关系剔除之后得到了WN18RR数据集。WN18RR数据集中包含了93003个事实三元组, 主要的关系模式是对称/反对称和组合关系模式。

FB15K-237数据集是FB15K数据集的子集, 同样是剔除了FB15K数据集中的可逆关系, FB15K-237数据集包含了310116个事实三元组, 主要的关系模式是对称/反对称和组合关系模式。

4.2 对比模型

我们选取了七个主流的知识图谱嵌入模型作为对比模型。对于翻译模型, 选取了TransE(Bordes et al, 2013)、RotatE(Sun et al, 2019)、QuatE模型(Zhang et al, 2019), 对于语义匹配模型, 选取了DistMult(Yang et al, 2014)、ComplEx模型(Trouillon T et al, 2016), 对于神经网络模型, 选取了ConvE模型(Dettmers T, 2018), 并对以上模型和MLQuatE模型在链接预测实验中的实验结果做出对比和分析。

4.3 实验环境与参数设置

本文基于Window10操作系统, 64位处理器, 算法使用PyTorch1.6实现, 在RTX 2080 Ti GPU和256GB物理内存的服务器上进行链接预测实验。

我们使用超参数网格搜索训练模型5000次, 其中嵌入维度 $k \in (50, 100, 200, 250, 300)$, 实体嵌入和关系嵌入的实部和虚部统一初始化, 学习率固定为0.1, 无需进一步调优, 正则化率 $\lambda_1, \lambda_2 \in (0, 0.01, 0.05, 0.1, 0.2)$, 每个正样本的负采样数量 $\in (1, 5, 10, 20)$ 。

4.4 链接预测实验

4.4.1 实验设置

链接预测是知识图谱补全的核心任务, 即根据已知的不完整的三元组信息推测出缺失的部分。链接预测可以分为关系预测和实体预测, 关系预测是指通过已知的头尾实体信息对未知的关系信息进行预测。而实体预测又分为头实体预测和尾实体预测, 即根据已知的头实体(尾实体)和关系预测出缺失的尾实体(头实体)。显而易见实体预测更具有训练价值和意义且更能衡量一个模型的性能。因此本文采用更具挑战性的实体预测对模型进行测试。在链接预测实验中, 本文首先用实体集中的实体替换测试集中事实三元组的头实体和尾实体, 然后利用评分函数逐一计算新三元组的得分, 最后得到这些新三元组合理性的排名。

4.4.2 评估指标

评估指标用来衡量模型的实验效果。本文选取较为常用的指标MR、MRR、Hits@N ($N=1, 3, 10$) 作为评估链接预测实验结果的指标。MR (Mean Rank) 是对所有正确三元组在整体预测结果序列中的位置序号取平均值, MR 值越小, 预测效果越好。MRR (Mean Reciprocal Rank) 是对所有正确三元组在整体预测结果序列中的倒数位置序号取平均值, MRR 值越大, 预测效果越好。Hits@N 是获取序号小于N 的正确三元组在整体三元组中的比例, Hits@N 值越大, 预测效果越好。由于数据集中可能存在一些错误的三元组, 因此在进行实验之前, 本文参考文献(Bordes et al, 2013), 剔除掉训练集、验证集和测试集中可能存在错误的三元组, 以避免得到不稳定的评估结果。

4.4.3 实验结果与分析

表3列出了不同模型在数据集WN18和FB15K上的链接预测结果，表4列出了不同模型在数据集WN18RR和FB15K-237上的链接预测结果。由于对比模型的实验和本文的实验使用了相同的数据集和评价指标，因此我们采用了部分已发表文献中的实验结果作为对比依据。在表3和表4中，MLQuatE模型的实验结果取在最优参数上5次实验结果的平均值，我们用粗体标明最优结果，下划线标明次优结果，“-”表示该模型论文中没有给出相关指标数据。

Model	WN18					FB15K				
	MR	MRR	Hit@1	Hit@3	Hit@10	MR	MRR	Hit@1	Hit@3	Hit@10
TransE	-	0.495	0.113	0.888	0.943	-	0.463	0.297	0.578	0.749
DistMult	665	0.797	-	-	0.946	42	0.798	-	-	0.893
ComplEx	-	0.941	0.936	0.945	0.947	-	0.692	0.599	0.759	0.840
ConvE	374	0.943	0.935	0.946	0.956	51	0.657	0.558	0.723	0.831
RotatE	309	0.949	0.944	0.952	0.959	40	0.797	<u>0.746</u>	0.829	0.884
QuatE	<u>162</u>	<u>0.950</u>	<u>0.945</u>	<u>0.954</u>	0.959	17	0.782	0.711	<u>0.835</u>	<u>0.900</u>
MLQuatE	160	0.957	0.951	0.956	0.959	<u>19</u>	<u>0.784</u>	0.751	0.858	0.905

表3: 不同模型在数据集WN18和FB15K上的链接预测结果

Table 3: Link prediction results of different models on datasets WN18 and FB15K

Model	WN18RR					FB15K-237				
	MR	MRR	Hit@1	Hit@3	Hit@10	MR	MRR	Hit@1	Hit@3	Hit@10
TransE	3384	0.226	-	-	0.501	357	0.294	-	-	0.465
DistMult	5110	0.43	0.39	0.44	0.49	254	0.241	0.155	0.263	0.419
ComplEx	5261	0.44	0.41	0.46	0.51	339	0.247	0.158	0.275	0.428
ConvE	4187	0.43	0.40	0.44	0.52	244	0.325	0.237	0.356	0.501
RotatE	3340	0.476	0.428	0.492	0.571	177	0.338	0.241	0.375	0.533
QuatE	2314	<u>0.488</u>	<u>0.438</u>	<u>0.508</u>	0.582	87	<u>0.348</u>	<u>0.248</u>	<u>0.382</u>	<u>0.550</u>
MLQuatE	<u>2318</u>	0.490	0.444	0.511	<u>0.581</u>	<u>88</u>	0.365	0.269	0.401	0.559

表4: 不同模型在数据集WN18RR和FB15K-237上的链接预测结果

Table 4: Link prediction results of different models on datasets WN18RR and FB15K-237

从表3和表4中的实验结果可以看出，本文提出的MLQuatE模型在四个基准数据集上取得了较为先进的结果，具体分析如下：(1)在数据集WN18上，MLQuatE模型在各项指标上均显著优于TransE、DistMult、ComplEx和ConvE等基础模型，与在复数空间表示旋转的RotatE模型相比，MLQuatE模型的各项指标均优于或持平于RotatE模型，这说明在超复数空间进行嵌入的有效性。与同样使用四元数表示嵌入向量的QuatE模型相比，MLQuatE模型除Hit@10指标与QuatE模型持平以外，其余指标均优于QuatE模型，这说明了本文通过融合正向右等斜旋转和逆向右等斜旋转来捕获实体和关系多层语义信息的有效性。在数据集FB15K上，MLQuatE模型的Hit@1、Hit@3和Hit@10指标均优于其他模型，其中，MLQuatE模型较QuatE模型在Hit@1指标上有5.6%(0.751-0.711=0.04,提升5.6%)的提升，较QuatE模型在Hit@3指标上有2.8%的提升，较RotatE模型在Hit@3指标上有3.5%的提升，较RotatE模型在Hit@10指标上有2.4%的提升。指标MR和MRR虽然不是最优结果，但也取得了次优结果。综上所述，MLQuatE模型能够很好的推理和建模对称/反对称和反转关系模式。(2)除数据集WN18RR上的MR和Hit@10指标以及数据集FB15K-237上的MR指标外，MLQuatE模型在其余指标上的结果均明显优于其他模型。其中在数据集FB15K-

237上, MLQuatE模型较QuatE模型在MRR指标上有4.9%的提升, 在Hit@1和Hit@3指标上分别有约8.5%和5%的提升。较RotatE模型在MRR指标上有约8%的提升, 在Hit@1和Hit@3指标上分别有约12%和6.9%的提升。这说明MLQuatE模型在建模对称/反对称和组合关系模式时也具有明显优势。

5 结论与未来工作

针对知识图谱存在的数据稀疏问题, 本文提出了一种融合实体多层语义的四元数表示学习方法进行知识图谱补全, 主要贡献如下。(1) 本文采用超复数空间代替传统的复数空间作为嵌入空间, 用四元数代替复数来表示实体和关系的嵌入向量, 不仅能够有效建模3种复杂关系模式: 对称/反对称、反转和组合, 也使旋转操作更加自由和灵活。除此之外, 本文利用单位关系四元数对头尾实体分别做右等斜旋转和逆方向的右等斜旋转, 以此来融合实体的多层语义信息。(2) 在4个基准数据集上展开了链接预测实验, 实验结果表明, 相较于其他主流的知识图谱嵌入模型, 本文提出的模型在多项指标上均取得了优异的结果。这表明融合实体多层语义的四元数表示学习方法能够有效完成知识图谱补全任务, 解决知识图谱数据稀疏性问题。

在今后的研究工作中, 我们将从以下两个方面进行改进和创新:(1) 考虑对四元数进行扩充, 例如对偶四元数(Cao Z et al, 2021)或者八元数等, 进一步提升对实体和关系向量表示的精确度;(2) 考虑将四元数与神经网络模型结合, 动态抽取和融入关系路径, 增强模型的多跳推理能力。

参考文献

- Auer S. Bizer C. Kobilarov G. Lehmann J. Cyganiak R. Ives Z. 2007. *Dbpedia: A nucleus for a web of open data*. Springer. 722–735.
- Bizer. Christian and Heath. Tom and Idehen. Kingsley and Berners-Lee. Tim. 2008. *Linked data on the web (LDOW2008)*. Proceedings of the 17th international conference on World Wide Web. 1265–1266.
- Bollacker. Kurt and Evans. Colin and Paritosh. Praveen and Sturge. Tim and Taylor. Jamie. 2008. *Freebase: a collaboratively created graph database for structuring human knowledge*. Proceedings of the 2008 ACM SIGMOD international conference on Management of data. 1247–1250
- Bordes, Antoine and Glorot, Xavier and Weston, Jason and Bengio, Yoshua. 2014. *A semantic matching energy function for learning with multi-relational data: Application to word-sense disambiguation*, volume 94. Machine Learning. 233–259.
- Bordes. Antoine and Usunier. Nicolas and Garcia-Duran. Alberto and Weston. Jason and Yakhnenko. Oksana. 2013. *Translating embeddings for modeling multi-relational data*, volume 26. Advances in neural information processing systems.
- Bordes. Antoine and Weston. Jason and Collobert. Ronan and Bengio. Yoshua. 2011. *Learning structured embeddings of knowledge bases*, 25(1):301–306. Proceedings of the AAAI conference on artificial intelligence.
- Cao Z. Xu Q. Yang Z. Cao X. Huang Q. 2021. *Dual quaternion knowledge graph embeddings*. Proceedings of the AAAI conference on artificial intelligence.
- Dettmers T. Minervini P. Stenetorp P. Riedel S. 2018. *Convolutional 2d knowledge graph embeddings*. Proceedings of the AAAI conference on artificial intelligence, volume 32.
- Hamilton W R. 1844. *LXXVIII. On quaternions; or on a new system of imaginaries in Algebra: To the editors of the Philosophical Magazine and Journal*. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 25(169): 489–495.
- Ji. Guoliang and He. Shizhu and Xu. Liheng and Liu. Kang and Zhao. Jun. 2015. *Knowledge graph embedding via dynamic mapping matrix*. Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers). 687–696.
- Lin. Yankai and Liu. Zhiyuan and Sun. Maosong and Liu. Yang and Zhu. Xuan. 2015. *Learning entity and relation embeddings for knowledge graph completion*, volume 29. Proceedings of the AAAI conference on artificial intelligence.

- Miller, George A. 1995. *WordNet: a lexical database for English*. Communications of the ACM. 39–41.
- Suchanek, Fabian M and Kasneci, Gjergji and Weikum, Gerhard. 2007. *Yago: a core of semantic knowledge*. Proceedings of the 16th international conference on World Wide Web. 697–706.
- Sun, Zhiqing and Deng, Zhi-Hong and Nie, Jian-Yun and Tang, Jian. 2019. *Rotate: Knowledge graph embedding by relational rotation in complex space*. arXiv preprint arXiv:1902.10197.
- Toutanova, K. and Chen, D. 2015. *Observed versus latent features for knowledge base and text inference*[C]. Proceedings of the 3rd workshop on continuous vector space models and their compositionality. 57–66.
- Trouillon, T. and Welbl, J. and Riedel, S. and Gaussier, É. and Bouchard, G. 2016. *Complex embeddings for simple link prediction*. International conference on machine learning. 2071–2080.
- Wang, Zhen and Zhang, Jianwen and Feng, Jianlin and Chen, Zheng. 2014. *Knowledge graph embedding by translating on hyperplanes*, volume 28. Proceedings of the AAAI conference on artificial intelligence.
- Yang, Bishan and Yih, Wen-tau and He, Xiaodong and Gao, Jianfeng and Deng, Li. 2014. *Embedding entities and relations for learning and inference in knowledge bases*. arXiv preprint arXiv:1412.6575.
- Zhang, Shuai and Tay, Yi and Yao, Lina and Liu, Qi. 2019. *Quaternion knowledge graph embeddings*. Advances in neural information processing systems, volume 32.
- Zhang, Z. and Cai, J. and Zhang, Y. and Wang, J. 2020. *Learning hierarchy-aware knowledge graph embeddings for link prediction*. Proceedings of the AAAI conference on artificial intelligence. 34(03): 3065–3072