

四元数关系旋转的知识图谱补全模型

陈恒^{1,2} 王维美¹ 李冠宇¹ 史一民¹

1 大连海事大学信息科学技术学院 辽宁 大连 116026

2 大连外国语大学语言智能研究中心 辽宁 大连 116044

(chenheng@dlufl.edu.cn)

摘要 知识图谱是真实世界三元组的结构化表示,通常三元组被表示成头实体、关系、尾实体的形式。针对知识图谱中广泛存在的数据稀疏问题,提出了一种将四元数作为关系旋转的知识图谱补全方法。文中使用极具表现力的超复数表示对实体和关系进行建模,以进行链接预测。这种超复数嵌入用于表示实体,关系则被建模为四元数空间中的旋转。具体来说,将每个关系定义为超复数空间中头实体到尾实体的旋转,用于推理和建模各种关系模式,包括对称/反对称、反转和组合。在公开的数据集 WN18RR 和 FB15K-237 上进行相关的链接预测实验,实验结果表明,在 WN18RR 数据集中,其平均倒数排名(Mean Reciprocal Rank, MRR)比 RotatE 的提高了 4.6%,其 Hit@10 比 RotatE 的提高了 1.7%;在 FB15K-237 数据集中,其平均倒数排名比 RotatE 的提高了 5.6%,其 Hit@3 比 RotatE 的提高了 1.4%。该实验证明,使用四元数作为关系旋转的知识图谱补全方法可以有效提高三元组预测精度。

关键词: 知识图谱;四元数;知识图谱补全;超复数表示;链接预测

中图法分类号 TP181

Knowledge Graph Completion Model Using Quaternion as Relational Rotation

CHEN Heng^{1,2}, WANG Wei-mei¹, LI Guan-yu¹ and SHI Yi-ming¹

1 Faculty of Information Science & Technology, Dalian Maritime University, Dalian, Liaoning 116026, China

2 Research Center for Language Intelligence, Dalian University of Foreign Languages, Dalian, Liaoning 116044, China

Abstract Knowledge graph is a structured representation of real-world triples. Typically, triples are represented in the form of head entity, relationship entity and tail entity. Aiming at the data sparse problem widely existing in knowledge graph, this paper proposes a knowledge graph completion method using quaternions as relational rotation. In this paper, we model entities and relations in the expressive hyper-complex representations for link prediction. This hyper-complex embedding is used to represent entities, and relations are modelled as rotations in quaternion space. Specifically, we define each relation as a rotation from the head entity to the tail entity in the hyper-complex space, which could be used to infer and model diverse relation patterns, including symmetry/anti-symmetry, reversal and combination. In the experiment, the public datasets WN18RR and FB15K-237 are used for the related link prediction experiment. Experimental results show that on the WN18RR dataset, its mean reciprocal rank (MRR) is 4.6% higher than RotatE, and its Hit@10 is 1.7% higher than RotatE. On the FB15K-237 dataset, its MRR is 5.6% higher than RotatE, its Hit@3 is 1.4% higher than RotatE. Experiments show that the knowledge graph completion method using quaternions as relational rotation can effectively improve the prediction accuracy of triples.

Keywords Knowledge graph, Quaternion, Knowledge graph completion, Hyper-complex representation, Link prediction

1 引言

目前,知识图谱(Knowledge Graph, KG)通常表示为多关系图,它是事实三元组的集合。现实世界中大量的事实可以

简单地表示为实体和实体间的关系。KG 在各领域有着重要应用,很多大型知识图谱如 YAGO^[1], Freebase^[2] 和 DBpedia^[3] 都使用三元组形式来表示实体间的关系,即 (h, r, t) , 例如 (Melbourne, cityOf, Australia)。KG 以结构化的方式描述

到稿日期:2020-03-16 返修日期:2020-07-08 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(61976032,61806038,61602076,61702072);辽宁省高等学校基本科研课题(2017JYT09);大连外国语大学科研创新团队(2016CXTD06)

This work was supported by the National Natural Science Foundation of China(61976032,61806038,61602076,61702072), Basic Scientific Research Projects of Liaoning University (2017JYT09) and Research and Innovation Team of Dalian University of Foreign Languages (2016CXTD06).

通信作者:李冠宇(rabitlee@163.com)

现实世界中的概念、实体及其链接关系,在个性化推荐^[4]、智能问答^[5]、智能搜索^[6]和自然语言处理^[7]任务中有非常重要的应用价值。虽然知识图谱包含了数以千万计的实体以及数以亿计的三元组事实,但是 KG 中仍然缺失大量数据,即缺少许多有效的三元组^[8-9]。因此,许多研究工作致力于知识图谱补全或链接预测任务,即判断一个三元组 (h, r, t) 是否有效^[10-11]。许多嵌入模型学习实体和关系的矢量或矩阵表示,以获得较为先进的链接预测结果。TransE^[12]将关系看作头部实体到尾部实体的转移,实体和关系由 k 维嵌入向量表示。此外,TransH^[13], Trans-R^[14]等模型分别是 TransE 模型的扩展。上述方法已被证明是可扩展且有效的^[15-16]。

与能量模型 SE^[17] (Structured Embeddings) 和 SME^[18] (Semantic Matching Energy) 相比, TransE^[12] 简单有效,在 FB15k 和 WN18 数据集上均具有更好的预测性能。受文献^[19]启发, TransE 在低维空间学习实体和关系的嵌入表示。TransE 将关系视为从头部实体到尾部实体的平移,即当 (h, r, t) 成立时, $h+r \approx t$ 。其他的翻译模型(如 TransH^[13], Trans-R^[14] 和 TransD^[20]) 在 TransE 的基础上使用投影矢量或矩阵将头实体和尾实体投影到关系矢量空间中。上述模型通常根据观察到的事实对知识图谱中的关系模式进行建模和推理。例如,某些关系是对称的(friend),而另一些关系则是反对称的(filiation);一些关系是其他关系的反转(hypernym and hyponym);或者某些关系是其他关系的组合(my father's wife is my mother)。在这 3 种关系模式中,如何寻找一种可以建模和推断所有关系模式的方法将变得至关重要。TransE 将关系视为头实体到尾实体的平移的方式,只能建模反转关系和组合关系,但不能建模和推理对称模式。Trans-H, TransR 是 TransE 模型的变形,可以建模对称/反对称模式,但不能推断反转和组合模式。DisMult^[21] 可以对头实体、关系和尾实体这 3 方进行交互建模,因此只能推理对称模式,无法建模反对称和反转模式。ComplEx^[22] 引入复值嵌入来建模对称和反对称关系,但是无法推理组合关系。RotatE^[16] 使用复数空间对 3 种关系模式进行建模,即将每个关系定义为复矢量空间中头实体到尾实体的旋转。本文使用四元数来建模和推理上述 3 种类型的关系模式。传统的复数由 1 个实部和 1 个虚部组成,而四元数具有 1 个实部和 3 个虚部。这种超复数嵌入用于表示实体,关系则被建模为超复数向量空间中的旋转。此外,四元数可以在三维空间中表达旋转,并且其比复数平面中的旋转具有更大的自由度^[23]。

为了评估本文方法,本文在 4 个标准数据集(WN18-RR, WN18, FB15K 和 FB15K(237))上进行知识图谱补全任务。链接预测实验的结果表明,使用四元数作为关系旋转的知识图谱补全方法优于大部分嵌入模型,其预测准确度更高。

2 相关工作和动机

2.1 相关工作

TransE 模型将知识图谱的关系看作头部实体到尾部实体的转移,给定三元组(Alaska, cityOf, America),即 $h+r \approx t$ 。在 TransE 中,假设 $\|h+r-t\|$ 无限接近于零,即 t 无限接近于 $h+r$ ^[24]。TransE 将关系表示为头实体和尾实体之间的双

射函数,其打分函数如式(1)所示。

$$f_r(h, t) = \|h+r-t\|_2^2 \quad (1)$$

其中, h, r, t 分别表示头实体、关系和尾实体的嵌入向量, $\|\cdot\|_2^2$ 代表 L2 范数。

作为首个将实体与关系嵌入低维空间的典型代表模型, TransE 利用较少的参数便能建模和推理复杂的关系模式。但是, TransE 在建模一对多和多对一关系时性能降低。为了解决这个问题, TransH 把 h 和 t 投影到一个超平面,将关系视为两个投影向量 h 和 t 之间的平移^[25]。TransR 则认为不同的关系应该具有不同的语义空间。因此, TransR 将实体和关系投影到不同的向量空间中,然后建模头部实体到尾部实体的翻译关系。TransH 和 TransR 打分函数如式(2)和式(3)所示。此外, TransX 模型是 TransE 模型的变形,如 TransH 和 TransR 等模型。

$$f_r(h, t) = \|(h-w_r^T h w_r) + r - (t-w_r^T t w_r)\|_2^2 \quad (2)$$

$$f_r(h, t) = \|(hM_r + r - tM_r)\|_2^2 \quad (3)$$

其中, h, r, t 分别表示头实体、关系和尾实体的嵌入向量, w_r^T 表示关系 r 的法向量, M_r 表示关系 r 的投影矩阵。

DisMult^[21] 模型使用减法运算符,只能捕获实体间的线性关系,通过矩阵乘法来提取多关系语义,其打分函数如式(4)所示。ComplEx^[22] 通过引入复数嵌入来扩展 Dist-Mult,以便更好地建模对称和反对称模式,但是无法推理组合关系,因为它通过关系 r 无法建模 h 到 t 的双射函数。ComplEx 的打分函数如式(5)所示。

$$f_r(h, t) = \langle h, r, t \rangle \quad (4)$$

$$f_r(h, t) = \text{Re}(\langle h, r, \bar{t} \rangle) \quad (5)$$

其中, h, r, t 分别表示头实体、关系和尾实体的嵌入向量, $\langle \cdot \rangle$ 表示点积, $\bar{\cdot}$ 表示向量的共轭, Re 表示复数的实部。

与上述模型不同, RotatE 可以有效地建模 3 种关系模式。受欧拉恒等式的启发, RotatE 将实体表示为复数矢量,将关系表示为复数矢量空间中的旋转。因此, RotatE 的打分函数如式(6)所示。

$$f_r(h, t) = -\|h \circ r - t\|_2^2 \quad (6)$$

其中, h, r, t 分别表示头实体、关系和尾实体的嵌入向量, \circ 表示哈达玛乘积。

与传统的复数表示不同,四元数能够在三维空间中表达旋转,并且比复数平面中的旋转具有更大的自由度和灵活性^[23]。这些事实表明,有关四元数的模型在建模几个关键关系模式方面具有优势。因此,我们考虑使用四元数代替复数来建模 3 种关系模式。表 1 列出了几种最新的知识图谱补全模型的得分函数以及这几种模型推断 3 种关系模式的能力。

2.2 动机

本文使用四元数表示实体,关系被建模为四元数空间中的旋转。RotatE 模型仅有一个旋转平面(复数平面),如图 1(a)所示。本文方法有两个旋转平面,如图 1(b)所示。由图 1(b)可知,我们将关系建模为超复数矢量空间中的旋转。如图 1 所示,四元数空间中实体和关系的旋转灵活性高于复数平面。此外,与旋转矩阵相比,四元数可以更有效地表示实体和关系,并且其值更加稳定。我们的动机来源于欧拉四元数扩展公式,即:

$$e^{\theta/2(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})} = \cos \theta/2 + (u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k}) \sin \theta/2$$

其中, $\mathbf{u} = (u_x, u_y, u_z)$, 这表明四元数可以在任意一个超复数空间 H^k 中进行旋转操作。具体来说, 我们将实体和关系映射到超复数向量空间, 并将每个关系定义为头实体到尾实体的旋转。给定一个三元组 (h, r, t) , 我们期望:

$$t \approx h \circ r \quad (7)$$

其中, $\mathbf{h}, \mathbf{r}, \mathbf{t} \in H^k$, $|r_i| = 1$, \circ 表示哈达玛乘积。特别地, 对于超复数空间中的每个维度, 其计算式如式(8)所示。

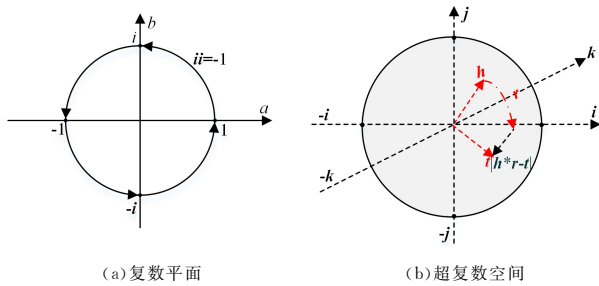
$$t_i = h_i r_i, \text{ where } h_i, r_i, t_i \in H \text{ and } |r_i| = 1 \quad (8)$$

表1 模型打分函数及推理关系模式的能力

Table 1 Scoring function of models and ability to infer all the three relationship patterns

Model	Score Function	Symmetry	Anti-symmetry	Reversal	Combination
TransE	$-\ \mathbf{h} + \mathbf{r} - \mathbf{t}\ $	×	√	√	√
TransX	—	√	√	×	×
DistMult	$\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle$	√	×	×	×
ComplEx	$Re(\langle \mathbf{h}, \mathbf{r}, \bar{\mathbf{t}} \rangle)$	√	√	√	×
QuatR	$\ \mathbf{h} \circ \mathbf{r} - \mathbf{t}\ $	√	√	√	√

注: $\langle \rangle$ 表示点积; \circ 表示 Hadamard 乘积; $\bar{\cdot}$ 表示复向量的共轭



注: a 代表实部轴, b 代表虚部轴

图1 复数平面和超复数空间图示

Fig. 1 Illustration of complex plane and super complex space

为了验证本文方法可以建模和推理 3 种关系模式表示, 我们做如下阐述。对称模式要求对称关系具有属性 $\mathbf{r} \circ \mathbf{r} = 1$, 解为 $r_i = \pm 1$ 。当我们在嵌入维度为 500 的 WN18 数据集上训练关系时, 关系的嵌入弧度为 π ($r_i = -1$) 或为 $0, 2\pi$ ($r_i = 1$)。反转模式则要求一对反转关系的嵌入以满足共轭属性。使用同样的操作, 我们在 WN18RR 训练集上训练本文方法。实验结果表明, 若两个反转关系的嵌入弧度为 0 或 2π , 则表示形式为 $\mathbf{r}_1 = \mathbf{r}_2^{-1}$ 。组合模式要求组合关系的嵌入弧度是另外两个关系弧度的相加。本文研究了本文方法在嵌入维度为 1000 的 FB15k-237 数据集上的训练结果。结果表明, 当 $\mathbf{r}_1 = \mathbf{r}_2 \circ \mathbf{r}_3$ 成立时, $\theta_1 = \theta_2 + \theta_3$ 。

事实证明, 本文方法可以有效地对 3 种关系模式进行建模: 对称/反对称、反转和组合。例如, 关系 \mathbf{r} 是对称的, 当且仅当 \mathbf{r} 的每个元素, 即 r_i , 满足 $r_i = e^{0/\sqrt{2}(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})} = \pm 1$; 关系 \mathbf{r}_1 和 \mathbf{r}_2 是反转的, 当且仅当它们的嵌入满足共轭属性 $\mathbf{r}_2 = \bar{\mathbf{r}}_1$; 关系 $\mathbf{r}_3 = e^{\theta_3/2(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})}$ 是 $\mathbf{r}_1 = e^{\theta_1/2(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})}$ 和 $\mathbf{r}_2 = e^{\theta_2/2(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})}$ 的组合, 当且仅当 $\mathbf{r}_3 = \mathbf{r}_1 \circ \mathbf{r}_2$, 即 $\theta_3 = \theta_1 + \theta_2$ 。

3 算法设计

本节将介绍本文提出的四元数方法。首先, 介绍四元数模

型可以处理的 3 个重要的关系模式。然后, 介绍哈密顿四元数以及四元数在知识图谱中的表示形式。最后, 介绍四元数模型及模型的优化, 即将关系定义为超复数向量空间中的旋转。

3.1 3 种关系模式

在本文中, 知识图谱补全旨在建模和推断 3 种关系模式: 对称、反转和组合。为了便于表示 3 种关系模式, 本文给出如下定义。

定义 1 对称(反对称)关系

若 $\forall e_i, e_j$, 有:

$$r(e_i, e_j) \Rightarrow r(e_j, e_i) \quad (9)$$

$$r(e_i, e_j) \Rightarrow \neg r(e_j, e_i) \quad (10)$$

则关系 r 是对称的(反对称的)。

定义 2 转关系

若 $\forall e_i, e_j$, 有:

$$r_1(e_i, e_j) \Rightarrow r_2(e_j, e_i) \quad (11)$$

$$r_2(e_i, e_j) \Rightarrow r_1(e_j, e_i) \quad (12)$$

则关系 r_1 是关系 r_2 的反转。

定义 3 组合关系

若 $\forall e_i, e_j, e_k$, 有:

$$r_2(e_i, e_j) \wedge r_3(e_j, e_k) \Rightarrow r_1(e_i, e_k) \quad (13)$$

则关系 r_1 是关系 r_2 和关系 r_3 的组合。

3.2 Hamilton's Quaternions (哈密顿四元数)

四元数是由实数加上 3 个虚数单位 $\mathbf{i}, \mathbf{j}, \mathbf{k}$ 组成的, 其表示形式为 $Q = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$, 其中, a, b, c, d 是实数, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ 是虚数单位, 此外, $\mathbf{ij} = \mathbf{k}, \mathbf{ji} = -\mathbf{k}, \mathbf{jk} = \mathbf{i}, \mathbf{ki} = \mathbf{j}, \mathbf{kj} = -\mathbf{i}, \mathbf{ik} = -\mathbf{j}$ 。

四元数重要运算规则如下。

共轭 $\bar{Q}: \bar{Q} = a - b\mathbf{i} - c\mathbf{j} - d\mathbf{k}$

四元数模: $|Q| = |\bar{Q}| = \sqrt{a^2 + b^2 + c^2 + d^2}$

四元数加法: $Q_1 = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ 和 $Q_2 = a' + b'\mathbf{i} + c'\mathbf{j} + d'\mathbf{k}$ 相加为:

$$Q_1 + Q_2 = (a + a') + (b + b')\mathbf{i} + (c + c')\mathbf{j} + (d + d')\mathbf{k}$$

四元数乘法: $Q_1 = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ 和 $Q_2 = a' + b'\mathbf{i} + c'\mathbf{j} + d'\mathbf{k}$ 相乘为:

$$Q_1 Q_2 = (aa' - bb' - cc' - dd') + (ab' + ba' + cd' - dc')\mathbf{i} + (ac' - bd' + ca' + db')\mathbf{j} + (ad' + bc' - cb' + da')\mathbf{k}$$

四元数除法: $Q_1 = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ 和 $Q_2 = a' + b'\mathbf{i} + c'\mathbf{j} + d'\mathbf{k}$ 相除为:

$$Q_1 / Q_2 = (aa' - bb' - cc' - dd') + (ab' + ba' - cd' - dc')\mathbf{i} + (ac' + ad' + ca' - db')\mathbf{j} + (ad' + bc' + cb' - da')\mathbf{k}$$

3.3 超复数空间中四元数表示及优化方法

本节介绍四元数旋转方法, 该方法可以对 3 种关系模式进行建模。受欧拉公式的启发, 我们将头实体、尾实体映射到超复数空间中, 即 $\mathbf{h}, \mathbf{t} \in H^k$, 然后将每个关系 r 定义为头实体 h 到尾实体 t 的旋转。给定一个三元组 (h, r, t) , 我们期望:

$$t \approx h \circ r, \text{ where } |r_i| = 1 \quad (14)$$

其中, \circ 是 Hadamard 乘积。对于向量中的每个元素, 我们都有 $t_i = h_i r_i$ 。另外, 算法 1 给出了四元数模型算法的伪代码。根据上述定义, 对于每个三元组 (h, r, t) , 我们将模型的目标函数定义为:

$$f_r(h, t) = \| \mathbf{h} \circ \mathbf{r} - \mathbf{t} \| \quad (15)$$

算法 1 四元数模型算法

Input: Training set $S = (h, r, t)$, E, R, γ and K represent entities, relations, margin, embeddings dim respectively

1. Initialize: $\mathbf{r} \leftarrow \text{torch.zeros}(|R|, K)$ for each $r \in R$
2. $\mathbf{r} \leftarrow \text{torch.nn.init.uniform}(r, (-\frac{\gamma}{K}, \frac{\gamma}{K}))$ for each $r \in R$
3. $\mathbf{e} \leftarrow \text{torch.zeros}(|E|, K)$ for each $e \in E$
4. loop:
 5. $\mathbf{e} \leftarrow \text{torch.nn.init.uniform}(r, (-\frac{\gamma}{K}, \frac{\gamma}{K}))$ for each $e \in E$
 6. $S_{\text{batch}} \leftarrow \text{sample}(S, b)$ // sample a minibatch of b
 7. $T_{\text{batch}} \leftarrow \emptyset$ // initialize the set of pairs of triples
 8. for $(h, r, t) \in S_{\text{batch}}$ do
 9. $(h', r, t') \leftarrow \text{sample}(S'_{(h, r, t)})$ // sample a corrupted triple
 10. $T_{\text{batch}} \leftarrow T_{\text{batch}} \cup \{((h, r, t), (h', r, t'))\}$
 11. end for
 12. Input $\leftarrow [h, r, t]$ // input matrix
 13. Calculate $f_r(h, t) = \| \mathbf{h} \circ \mathbf{r} - \mathbf{t} \|^2$
 14. Update loss function w. r. t:

$$-\frac{1}{K} \sum_{i=1}^n \log \sigma(f_r(\mathbf{h}_i', \mathbf{t}_i') - \gamma) - \log \sigma(\gamma - f_r(\mathbf{h}_i, \mathbf{t}_i))$$
 16. end loop

通过将每个关系定义为超复数空间中的旋转, 本文方法可以建模和推理所有 3 种类型的关系模式。最终, 我们得出以下定理。

定理 1 四元数方法可以推理和建模对称或反对称模式。

证明:

若 $r(e_i, e_j)$ 和 $r(e_j, e_i)$ 存在, 则有:

$$e_j = r \circ e_i \wedge e_i = r \circ e_j \Rightarrow r \circ r = 1 \quad (16)$$

否则, 若 $r(e_i, e_j)$ 和 $\neg r(e_j, e_i)$ 存在, 则有:

$$e_j = r \circ e_i \wedge e_i \neq r \circ e_j \Rightarrow r \circ r \neq 1 \quad (17)$$

定理 2 四元数方法可以推理和建模反转模式。

证明:

若 $r_1(e_i, e_j)$ 和 $r_2(e_j, e_i)$ 存在, 则有:

$$e_j = r_1 \circ e_i \wedge e_i = r_2 \circ e_j \Rightarrow r_1 = r_2^{-1} \quad (18)$$

定理 3 四元数方法可以推理和建模组合模式。

证明:

若 $r_1(e_i, e_k), r_2(e_i, e_j)$ 和 $r_3(e_j, e_k)$ 存在, 则有:

$$e_k = r_1 \circ e_i \wedge e_j = r_2 \circ e_i \wedge e_k = r_3 \circ e_j \Rightarrow r_1 = r_2 \circ r_3 \quad (19)$$

定理 4 四元数方法可以演化成 TransE (TransE 是本文模型的特例)。

证明:

令 $|h_i| = |t_i| = C$, 可以得到 $\mathbf{h}, \mathbf{r}, \mathbf{t}$, 其中:

$$\begin{aligned} \mathbf{h} &= C e^{\frac{\theta_h}{2}(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})} \\ &= C \cos \frac{\theta_h}{2} + (u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k}) C \sin \frac{\theta_h}{2} \end{aligned} \quad (20)$$

$$\begin{aligned} \mathbf{r} &= e^{\frac{\theta_r}{2}(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})} \\ &= \cos \frac{\theta_r}{2} + (u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k}) \sin \frac{\theta_r}{2} \end{aligned} \quad (21)$$

$$\mathbf{t} = C e^{\frac{\theta_t}{2}(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})}$$

$$= C \cos \frac{\theta_t}{2} + (u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k}) C \sin \frac{\theta_t}{2} \quad (22)$$

$$\begin{aligned} &\| \mathbf{h} \circ \mathbf{r} - \mathbf{t} \| \\ &= C \left\| e^{\frac{\theta_h + \theta_r}{2}(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})} - e^{\frac{\theta_t}{2}(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})} \right\| \end{aligned} \quad (23)$$

$$= C \left\| e^{\frac{\theta_h + \theta_r - \theta_t}{2}(u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k})} - 1 \right\| \quad (24)$$

$$\begin{aligned} &= C \left\| \cos \frac{\theta_h + \theta_r - \theta_t}{2} + (u_x \mathbf{i} + u_y \mathbf{j} + u_z \mathbf{k}) \right. \\ &\quad \left. \sin \frac{\theta_h + \theta_r - \theta_t}{2} - 1 \right\| \end{aligned} \quad (25)$$

$$= C \left\| \sqrt{\left(\cos \frac{\theta_h + \theta_r - \theta_t}{2} - 1 \right)^2 + \sin^2 \frac{\theta_h + \theta_r - \theta_t}{2}} \right\| \quad (26)$$

$$= C \left\| \sqrt{2 - 2 \cos \frac{\theta_h + \theta_r - \theta_t}{2}} \right\| \quad (27)$$

$$= 2C \left\| \sin \frac{\theta_h + \theta_r - \theta_t}{2} \right\| \quad (28)$$

若 (h, r, t) 在 TransE 中嵌入矢量分别是 $\mathbf{h}', \mathbf{r}', \mathbf{t}'$, 令 $\theta_h = c\mathbf{h}', \theta_r = c\mathbf{r}', \theta_t = c\mathbf{t}'$ 和 $C = 1/c$, 则有:

$$\lim_{c \rightarrow 0} \| \mathbf{h} \circ \mathbf{r} - \mathbf{t} \| = \| \mathbf{h}' + \mathbf{r}' - \mathbf{t}' \| \quad (29)$$

负采样技术在知识图谱嵌入^[22]和单词嵌入^[19]领域被证明效果显著^[16]。文中使用类似于负采样^[22]的损失函数来优化基于距离的四元数旋转模型:

$$L = -\log \sigma(\gamma - f_r(\mathbf{h}, \mathbf{t})) - \frac{1}{k} \sum_{i=1}^n \log \sigma(f_r(\mathbf{h}_i', \mathbf{t}_i') - \gamma) \quad (30)$$

其中, γ 是定值边缘超参数, σ 是 sigmoid 激活函数, (h_i', r, t_i') 是第 i 个负例三元组。

4 实验

4.1 数据集

本文使用从 Freebase 和 WordNet 中抽取的 4 个数据集, 即 FB15k^[12], WN18^[12], FB15K-237^[26] 和 WN18RR^[27], 来执行链接预测实验。参照文献^[26], WN18 包含许多可逆关系, 如 hypernym 和 hyponym。由于这些可逆关系会显著提升模型 QuatR 的预测能力, 我们将 WN18 中的所有可逆三元组过滤掉, 生成一个新数据集 WN18RR。同理, 过滤掉数据集 FB15k 中的所有可逆三元组, 得到 FB15K-237。

WN18 是 WordNet 的子集, 包含 40 943 个具有 18 种不同关系的实体以及 151 442 个事实三元组。此外, WN18 主要的关系模式是对称/反对称和反转模式。

FB15k 是 Freebase 的子集, 包含 14 951 个具有 1 345 种不同关系的实体以及 592 213 个事实三元组。此外, FB15k 主要的关系模式是对称/反对称和反转模式。

WN18RR 是 WN18 的新数据集, 包含 40 943 个具有 11 种不同关系的实体, 93 003 个事实三元组。由于 WN18RR 中的可逆关系被移除, 其主要的关系模式是对称/反对称和组合。

FB15K-237 是 FB15k 的新数据集, 包含 14 541 个具有 237 种不同关系的实体以及 310 116 个事实三元组。由于 FB15K-237 中可逆关系被移除, 其主要的关系模式是对称/反对称和组合。

数据集统计情况如表 2 所列。

表 2 数据集统计
Table 2 Data set statistics

Dataset	# entity	# relation	# validation	# training	# test
WN18RR	40943	11	3034	86835	3134
FB15k-237	14541	237	17535	272115	20466
WN18	40943	18	5000	141442	5000
FB15k	14951	1345	50000	483142	59071

4.2 参数设置

我们使用 Adam^[28] 作为优化器,并微调验证集上的超参数。本文算法使用 PyTorch 实现,并在 1755MHz 24GD6 GeForce 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 3000 次,其中嵌入维数 $k \in \{125, 250, 500, 1000\}$, 学习率 $\lambda \in \{0.00001, 0.00005, 0.0001, 0.0005\}$, 批处理大小 $b \in \{256, 512, 1024\}$, 定值边缘超参数 $\gamma \in \{3, 6, 9, 12, 18\}$ 。实体嵌入的实部和虚部统一初始化,且关系嵌入的弧度在 0 和 2π 之间。最优超参数设置如下:在 FB15K-237 上,将 Hit@10 最优设置为 $k=1000, b=1024, \gamma=9$, 初始化学学习率为 1×10^{-5} ;在 WN18RR 上,将 Hit@10 最优设置为 $k=500, b=512, \gamma=6$, 初始化学学习率为 5×10^{-5} 。

4.3 链接预测

4.3.1 实验设置

链接预测指预测知识图谱中丢失的事实三元组,即真实三元组丢失的头实体或尾实体。对于一个事实三元组

(Aliens, _ditrected_by, ?), 可以看出,其丢失了尾实体。在链接预测中,我们参照文献[12]和文献[14],首先将测试集中每个三元组的头实体、尾实体去掉,依次使用数据集中的所有实体来替代,然后利用打分函数逐个计算新三元组分数,最后得到每个新三元组的排名。

4.3.2 评估指标

依据文献[14],本文选用 MR, MRR 和 Hit@N 作为链接预测实验的评估指标。MR 代表正确实体的平均排名, MRR 代表正确实体的倒数平均排名, Hit@N 代表正确实体排第 N 所占的百分比。这 3 个指标中, MR 越低、MRR 越高或者 Hit@N 越高,模型的效果越好,学习能力就越强。实验中由于数据集中可能存在某些错误的三元组,我们采用文献[12]的标准,把可能错误的三元组从训练集、验证集、测试集中删除,最终得到真实三元组。Raw 代表原始数据集, Filter 代表删除之后的数据集,在链接预测实验中我们采用 Filter 标准。

4.3.3 实验分析

在链接预测实验中,我们选用 Window 7 操作系统, 64 位处理器, GPU 选取 1755MHz 24GD6 GeForce RTX 2080 Ti, 物理内存为 256GB。在链接预测实验中,我们利用超参数网格搜索算法训练模型 3000 次,以选择 MRR 和 Hit@10 最优超参数。数据集 WN18 和 FB15K 在不同模型下的链接预测结果如表 3 所列。表 4 列出了数据集 WN18RR 和 FB15K-237 的链接预测结果。此外,在表 3 中, [* *] 的实验结果来自文献[29], [*] 的实验结果来自文献[30]。在表 4 中, [*] 的实验结果来自文献[31], 其他实验结果来自文献[27]。

表 3 不同模型下在数据集 WN18 和 FB15K 上的链接预测结果

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

Model	FB15K					WN18				
	MR	MRR	H@3	H@10	H@10	MR	MRR	H@1	H@3	H@10
TransE [* *]	—	0.463	0.297	0.578	0.749	—	0.495	0.113	0.888	0.943
DistMult [*]	42	0.798	—	—	0.893	655	0.797	—	—	0.946
HoE ^[29]	—	0.524	0.402	0.613	0.739	—	0.938	0.930	0.945	0.949
ComplEx	—	0.692	0.599	0.759	0.840	—	0.941	0.936	0.945	0.947
ConvE ^[27]	51	0.657	0.558	0.723	0.831	374	0.943	0.935	0.946	0.956
RotatE	40	0.797	0.746	0.830	0.884	309	0.949	0.944	0.952	0.959
QuatR	36	0.820	0.750	0.878	0.905	265	0.960	0.956	0.967	0.952

表 4 不同模型下在数据集 WN18RR 和 FB15k-237 上的链接预测结果

Table 4 Link prediction results on datasets WN18RR and FB15k-237 under different models

Model	FB15K-237					WN18RR				
	MR	MRR	H@1	H@3	H@10	MR	MRR	H@1	H@3	H@10
TransE [*]	357	0.294	—	—	0.465	3384	0.226	—	—	0.501
DistMult	254	0.241	0.155	0.263	0.419	5110	0.430	0.390	0.440	0.490
ComplEx	339	0.247	0.158	0.275	0.428	5261	0.440	0.410	0.460	0.510
ConvE ^[27]	244	0.325	0.237	0.356	0.501	4187	0.430	0.400	0.440	0.520
RotatE	177	0.338	0.241	0.375	0.533	3340	0.476	0.428	0.492	0.571
QuatR	164	0.357	0.237	0.389	0.546	3024	0.498	0.421	0.504	0.588

从表 3 和表 4 可以看出,本文模型 QuatR 在 4 个基准数据集上获得了较为先进的实验结果,具体分析如下。1) 在 FB15k 和 WN18 上, QuatR 优于其他基准模型,但 Hit@10 除外。在 FB15k 上, QuatR 较 RotatE 在 MRR 上有 2.9% ($0.820 - 0.797 = 0.023$ 改善约 2.9%) 的提高,在 Hit@3 上有 4.8% 的提高;在 WN18 上, QuatR 较 RotatE 在 MRR 上有 1.2% 的提高,在 Hit@3 上有 1.5% 的提高。与 RotatE 相比,

QuatR 证实了四元数旋转比复数平面旋转拥有更大的灵活性。另外,实验结果也表明, QuatR 可以有效捕获对称性、反对称性和反转关系模式。2) 在 FB15K-237 和 WN18RR 上, QuatR 优于 RotatE 模型,但 Hit@1 除外。在 FB15K-237 上, QuatR 较 RotatE 在 MRR 上有 5.6% 的提高,在 Hit@3 上有 1.4% 的提高;在 WN18RR 上, QuatR 较 RotatE 在 MRR 上有 4.6% 的提高,在 Hit@10 上有 1.7% 的提高。此外,

TransE 的表现能力优于 ComplEx, 其原因在于 TransE 可以建模和推理组合模式, 而 ComplEx 无法推理该模式。在 FB15k-237 上, 主要关系是组合模式; 在 WN18RR 上, 主要关系是对称性和反对称性模式。因此, 本文方法 QuatR 可以有效地捕获对称性、反对称性和组合关系模式。

为证实四元数建模各种类型关系的能力, 我们在 WN18RR 上使用不同的关系进行表示学习。表 5 列出了 WN18RR 上每个关系的 MRR, 证实了四元数在建模和推理 3 种关系模式方面的优越的表示能力。另外, *similar_to* 的准确性为 100%, 这再次证明了四元数模型具有很强的表达力。

表 5 在数据集 WN18RR 上每个关系的 MRR 排名

Table 5 Filtered mean reciprocal rank for the models tested on each relation of WN18RR

Relation Name	RotatE	QuatR
hypernym	0.148	0.175
also_see	0.585	0.630
verb_group	0.943	0.951
similar_to	1.000	1.000
has_part	0.184	0.178
member_of_domain_usage	0.318	0.424
member_of_domain_region	0.200	0.178
instance_hypernym	0.318	0.385

4.4 模型分析

表 6 列出了本文方法和 RotatE 之间的参数量比较结果。在 WN18RR 和 FB15K-237 数据集上, 与基准模型 RotatE 相比, 本文模型参数的数量下降了 66.8% 和 51%。由表 6 可知, 本文模型可以极大减少复数模型 RotatE 中参数的数量, 在保持优越性能的同时, 最多可节省 67% 的参数。

表 6 在 WN18RR 和 FB15K-237 数据集上的参数量比较

Table 6 Number of parameters comparison on WN18RR and FB15K-237 datasets

Model	RotatE	Our
Space	C^k	H^k
FB15K-237	29.37M	9.75M (↓ 66.8%)
WN18RR	41.95M	20.73M (↓ 51%)

我们使用不同的评分函数重新进行实验。由表 7 可知, 在两个基准数据集 WN18RR 和 FB15K-237 上, MRR 和 Hit@10 的结果表明本文模型执行 Hadamard 乘积运算的性能更好。我们在头和尾四元数嵌入之间执行 Hadamard 乘积, 并将关系视为头实体到尾实体的旋转。因此, 有 $f_r(h, t) = \|h \circ r - t\|_2^2$ 。

表 7 不同打分函数下对 MRR 和 Hit@10 的分析

Table 7 Analysis of MRR and Hit@10 under different scoring functions

Analysis	FB15K-237		WN18RR	
	MRR	H@10	MRR	H@10
$\ h+r-t\ _2^2$	0.294	46.3	0.245	54.5
$\langle h, r, t \rangle$	0.241	41.7	0.420	49.1
$\ h \circ r - t\ _2^2$	0.357	54.6	0.498	58.8

结束语 本文针对知识图谱补全提出了一种四元数关系旋转模型, 贡献如下。

(1) 本文使用四元数代替传统的复数来建模和推理 3 种关系模式: 对称/反对称、反转和组合。四元数作为一种超复

数, 是复数在更高维度上的扩展。因此, 四元数可以在三维空间中表达旋转, 比复数平面中的旋转拥有更大的自由度和灵活性。

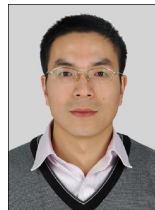
(2) 在 4 个基准数据集上, 我们进行了相关的链接预测实验。实验结果表明, 与其他基准模型相比, 本文方法在 MR, MRR, Hit@10 指标上有了显著提高。这表明四元数旋转模型可以有效解决知识图谱数据稀疏问题, 提高了知识图谱完备性。

为提高链接预测的准确率, 在今后的研究中, 我们将从以下两方面进行探索: 1) 考虑向四元数模型添加大量的实体描述文本; 2) 尝试将四元数与神经网络模型相结合来探索动态知识图谱补全方法。

参考文献

- [1] SUCHANEK F M, KASNECI G, WEIKUM G. Yago: a core of semantic knowledge[C]// Proceedings of the 16th International Conference on World Wide Web. 2007: 697-706.
- [2] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]// Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. 2008: 1247-1250.
- [3] LEHMANN J, ISELE R, JAKOB M, et al. DBpedia-a large-scale, multilingual knowledge base extracted from Wikipedia [J]. Semantic Web, 2015, 6(2): 167-195.
- [4] ZHANG F, YUAN N J, LIAN D, et al. Collaborative knowledge base embedding for recommender systems[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 353-362.
- [5] HAO Y, ZHANG Y, LIU K, et al. An end-to-end model for question answering over knowledge base with cross-attention combining global knowledge[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2017: 221-231.
- [6] XIONG C, POWER R, CALLAN J. Explicit semantic ranking for academic search via knowledge graph embedding[C]// Proceedings of the 26th International Conference on World Wide Web. 2017: 1271-1279.
- [7] YANG B, MITCHELL T. Leveraging knowledge bases in lstms for improving machine reading[J]. arXiv:1902.09091, 2019.
- [8] WANG W M, SHI Y M, LI G Y. Improved capsule network knowledge graph completion method[J/OL]. Computer Engineering: 1-7. [2020-03-13]. <https://doi.org/10.19678/j.issn.1000-3428.0055390>.
- [9] LI Z X, REN S Y, WANG H M, et al. Knowledge inference method based on unstructured text enhanced association rules [J]. Computer Science, 2019, 46(11): 209-215.
- [10] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]// Advances in Neural Information Processing Systems. 2013: 926-934.
- [11] WEST R, GABRILOVICH E, MURPHYK, et al. Knowledge base completion via search-based question answering[C]// Proceedings of the 23rd International Conference on World Wide

- Web. ACM, 2014; 515-526.
- [12] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]// Processings of the NIPS. Cambridge, MA: MIT Press, 2013; 2787-2795.
- [13] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes[C]// Processings of the AAAI. Menlo Park, CA: AAAI, 2014; 1112-1119.
- [14] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion[C]// Processings of the AAAI. Menlo Park, CA: AAAI, 2014; 2181-2187.
- [15] WANG Q, MAO Z, WANG B, et al. Knowledge graph embedding: A survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.
- [16] SUN Z, DENG Z H, NIE J Y, et al. Rotate: Knowledge graph embedding by relational rotation in complex space[J]. arXiv: 1902. 10197, 2019.
- [17] BORDES A, WESTON J, COLLOBERT R, et al. Learning structured embeddings of knowledge bases[C]// National Conference on Artificial Intelligence. 2011; 301-306.
- [18] MBORDES A, GLOTOT X, WESTON J, et al. A semantic matching energy function for learning with multi-relational data [J]. Neural Information Processing Systems, 2014, 94(2): 233-259.
- [19] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]// Advances in Neural Information Processing Systems. 2013; 3111-3119.
- [20] JI G, HE S, XU L, et al. Knowledge graph embedding via dynamic mapping matrix[C]// Processings of the ACL. Stroudsburg, PA: ACL, 2015; 687-696.
- [21] YANG B, YIH W, HE X, et al. Embedding entities and relations for learning and inference in knowledge bases[J]. arXiv: 1412. 6575, 2014.
- [22] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[C]// International Conference on Machine Learning. 2016; 2071-2080.
- [23] ZHANG S, TAY Y, YAO L, et al. Quaternion knowledge graph embeddings[C]// Advances in Neural Information Processing Systems. 2019; 2735-2745.
- [24] LIN H L, WANG Y Z, JIAY T, et al. Overview of knowledge fusion methods for network big data [J]. Journal of Computer, 2017, 40(1): 1-27.
- [25] GUAN S P, JIN X L, JIA Y T, et al. Research progress of knowledge reasoning for knowledge graph [J]. Journal of Software, 2018, 29(10): 2966-2994.
- [26] TOUTANOVA K, CHEN D. Observed Versus Latent Features for Knowledge Base and Text Inference[C]// Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality. 2015; 57-66.
- [27] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2D knowledge graph embeddings[C]// Processings of the AAAI. 2018; 1811-1818.
- [28] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv: 1412. 6980, 2014.
- [29] NICKEL M, ROSASCO L, POGGIO T A. Holographic embeddings of knowledge graphs[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2016; 1955-1961.
- [30] KADLEC R, BAJGAR O, KLEINDIENST J. Knowledge base completion: Baselines strike back[J]. arXiv: 1705. 10744, 2017.
- [31] NGUYEN D Q, NGUYEN T D, NGUYEN D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network[J]. arXiv: 1712. 02121, 2017.



CHEN Heng, born in 1982, Ph.D, associate professor. His main research interests include intelligent information processing and so on.



LI Guan-yu, born in 1963, Ph.D, professor. His main research interests include intelligent information processing and so on.