

分 类 号 TP391

密 级

U D C

编 号 10486

武汉大学

硕 士 专 业 学 位 论 文

跨设备下的知识图谱  
表示学习研究

研 究 生 姓 名：杨世杰

学 号：2021282110124

指导教师姓名、职称：彭敏 教授

专 业 类 别（领 域）：计算机科学与技术

二〇二三年四月

# Cross-device knowledge graph represents learning research

Candidate: Yang Shijie

StudentNumber: 2021282110124

Supervisor: Prof.Peng Min

Major: Computer Science and Technology

Speciality: Natural Language Processing



School of Computer Science  
Wuhan University

April, 2023

# 论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本声明的法律结果由本人承担。

学位论文作者（签名）：

年 月 日

## 摘 要

知识图谱表示学习将图谱的实体和关系嵌入到连续的低维空间，同时能够保持其语义信息，有助于众多下游任务的应用，如知识图谱补全、推荐系统及问答系统等。

随着知识图谱规模的不断扩大，这些单独的图谱往往分散在不同的设备上，同时由于个人移动终端的快速发展，每个人的移动设备上也都部署有不同规模的图谱应用，出于合并成本的考量以及用户数据隐私的要求，这些跨设备的图谱往往不能进行有效的合并和知识融合。在跨设备的场景下，知识图谱表示学习面临如下的问题：在现有训练知识图谱训练出的模型如何更好处理各跨设备中存在未见实体和关系的新兴知识图谱。

为了解决这个问题，本文采用元学习训练方法来模拟训练知识图谱的未见实体和关系，以帮助模型学习处理未见组件的能力；同时引入本体信息来对未见组件的表示学习提供语义支持。主要解决了当前模型中的两个问题：(1) 当前处理未见组件的模型多数只能处理单独的未见实体或者未见关系，无法兼顾。而采用结构信息对未见组件表示的模型缺乏了知识图谱丰富的语义信息。(2) 跨设备下的分散知识图谱存储方式如何在模型训练中进行模拟，从而使模型获取更强大的泛化性能。

基于上述问题的分析本文设计了一个基于 GNN（图神经网络）并融合本体信息的表示学习框架，既能够表示未知组件的特征，也能从模型训练中采用元学习方法对任务进行模拟，从而获得对新兴知识图谱未知组件的泛化学习能力。在对未知关系进行表示时，通过对知识图谱高层语义体现的本体信息进行嵌入，结合关系图上的 GCN 进行拓扑信息和语义信息的学习，获得对未见关系的表达。对于未知实体，文章采用实体的邻接关系特征聚合进行初始化的表示，并使用 CompGCN 对每个实体和关系再次进行邻接信息的学习和更新，获得最终各组件的嵌入。

本文在两个基准数据集上根据场景任务要求抽取出特定的测试数据集，测试了模型的链接预测任务并与多个基准模型进行对比，从实验效果上可以发现本文模型均有较好的改善。同时，本文通过多个模块的消融实验以及多个维度的案例分析证明了本文模型的有效性。

**关键词：**表示学习；本体嵌入；元学习；图神经网络；归纳推理

## ABSTRACT

The knowledge graph representation learning embeds the entities and relationships of the graph into a continuous low-dimensional space while maintaining its semantic information. This has been helpful for many downstream applications, such as knowledge graph completion, recommendation systems, and question-and-answer systems.

As the size of knowledge graphs continues to expand, these separate graphs are often scattered across different devices. Moreover, due to the rapid development of personal mobile terminals, each person's mobile device also has various scaled graph applications. For cost consideration and user data privacy requirements, these cross-device graphs often cannot be effectively merged and fused. In the cross-device scenario, the knowledge graph representation learning faces the following issues: how to better deal with emerging knowledge graphs in which there are unseen entities and relationships that exist across various devices with the current trained knowledge graph model.

To solve this problem, this paper uses meta-learning training methods to simulate the unseen entities and relationships for training knowledge graphs to help the model learn the ability to deal with the unseen components. In addition, ontology information is introduced to provide semantic support for the representation learning of unseen components. This mainly solves two problems in the current models: (1) The current models that deal with unseen components can only handle individual unseen entities or unseen relationships and cannot consider both as a whole. The model that represents the structural information of the unseen components lacks semantic information of the rich knowledge graph. (2) How to simulate the decentralized storage method of cross-device knowledge graphs in model training, so that the model can obtain stronger generalizability.

Based on the analysis of the above issues, this paper designs a representation learning framework based on Graph Neural Networks (GNN) and ontology information. It can both represent the features of unknown components and simulate tasks using meta-learning methods in model training to obtain generalization learning ability for emerging knowledge graph unknown components. When representing unknown relationships, ontology information reflecting the high-level semantics of the knowledge graph is embedded, and topology information and semantic information learning are combined by using GCN on the relationship graph to obtain the expression of unseen relationships. For unknown entities, the paper uses the feature aggregation of adjacent relationship characteristics for initialization and representation, and uses CompGCN

to learn and update the adjacent relationship of each entity and relationship, and obtain the final embedding of each component.

The paper extracted specific test data sets based on the scene task requirements from two benchmark datasets and tested the model's link prediction task compared with many benchmark models. From the experimental results, the model had significantly improved performance. Furthermore, through multiple module ablation experiments and multi-dimensional case analysis, the paper proves the effectiveness of the model.

**Key words:** presentation learning; ontology embedding; meta-learning; graph neural networks; inductive reasoning

# 目 录

摘 要	I
ABSTRACT	II
1 绪论	1
1.1 研究背景和意义	1
1.2 国内外研究现状及趋势	3
1.2.1 基于翻译的方法	3
1.2.2 基于语义相似度的方法	5
1.2.3 融合辅助信息及归纳推理的方法	6
1.3 本文主要研究内容	7
1.4 本文组织结构	8
2 相关技术研究	10
2.1 知识图谱本体嵌入	10
2.2 归纳知识图谱表示学习	13
2.3 元学习训练方法	16
3 基于元学习本体增强的跨设备知识图谱表示模型	19
3.1 问题定义	19
3.2 模型整体架构	20
3.3 关系嵌入模块	21
3.3.1 本体信息嵌入	21
3.3.2 关系图构建	22
3.3.3 关系相关性系数聚合编码	24
3.4 实体嵌入模块	24
3.5 实体关系联合嵌入模块	25
3.6 元学习训练设定	26
3.7 链接预测任务及训练流程	26
3.8 本章小结	27
4 实验结果及分析	28
4.1 数据集	28
4.1.1 源数据集介绍	28

4.1.2	任务数据集构建 .....	29
4.1.3	数据集本体三元组获取 .....	29
4.2	实验设计及评价指标 .....	30
4.3	模型参数设置 .....	31
4.4	实验结果及分析 .....	32
4.5	模型消融实验 .....	34
4.6	未见实体案例分析 .....	36
4.7	未见关系案例分析 .....	36
4.8	本章小结 .....	37
5	总结与展望 .....	38
5.1	总结 .....	38
5.2	未来工作 .....	39
	参考文献 .....	40
	致谢 .....	44
	攻硕期间取得的学术成果和参与的项目 .....	45



# 1 绪论

## 1.1 研究背景和意义

近几年,知识图谱 (Knowledge Graph) 作为一种结构化的知识存储形式被众多信息检索系统作为不可或缺的基础组件而广泛使用<sup>[1]</sup>。语义网络的概念最早可追溯到 2001 年 Berners-Lee 的相关研究<sup>[2]</sup>,他主张推动和发展知识表示相关的如资源描述框架 (RDF) 等技术标准。早期类似基于图的知识结构广泛采用 RDF 标准进行表示,而随着互联网的出现及发展,语义网络开始转向语义互联网发展,并在 2012 年由谷歌提出名为知识图谱的新兴技术并应用在搜索引擎上。

知识图谱是使用图结构来描述知识的知识库。知识图谱由节点和边组成,其组成可以涵盖实体、概念、关系、属性等,作为一种结构化知识表示的形式,他将现实中的知识通过节点及节点间关系组成图来进行结构化的表示。一方面,知识图谱以图的形式将抽象的知识形象化的表示更直观地展示出知识的相互联系和内容,同时利用高效的图遍历算法可以完成对知识更全面的检索和查询;另一方面,知识图谱可以从数据中识别、推理、发现实体间及实体与概念间的复杂关系,将存储的知识转化为可计算的模型,凭此辅助进行知识问答、推荐系统等丰富的下游任务,如在电商系统中,通过对用户偏好图谱的构建可以更好的预测用户的购买意向进行相关商品推荐从而收获更好的经济效益。虽然知识图谱广泛使用的三元组存储已经可以将知识进行高效的结构化表示,但是这种传统的符号形式的表示仍旧会导致知识图谱在使用上备受局限。

为了解决这个问题,知识图谱表示的相关研究很快起步并获得了广泛的关注。知识图谱表示学习 (Knowledge Representation Learning, 也称知识图谱嵌入) 旨在将符号化的三元组  $(h, r, t)$  映射到低维稠密的向量空间,便于实体和关系之间的计算<sup>[3]</sup>,降维后获得的知识图谱嵌入可以作为其他网络模型的输入来辅助下游任务的效果提升,包括知识图补全 (KGC)、三元组分类、实体识别、关系抽取以及图谱外任务如推荐系统、问答系统等,例如谷歌构建的规模庞大的知识图谱也已经展示出该方法的强大能力。但当下表示学习的方法仍无法顾及到所有的应用场景,具有一定的使用局限性。

在传统的知识图谱表示学习应用场景下,学习的目标是获得知识图谱中的实体和关系最优的降维编码,而后可以利用这些编码进行下游的任务;其中任务中涉及的实体及关系往往是模型训练阶段都可以接触到的 (称为可见实体、可见关系)。经典的 KGE 模型如基于翻译的 TransE<sup>[4]</sup> 模型在同一空间向量中嵌入实体,通过头实体和尾实体的平移操作来进行关系嵌入,时空复杂度低,但在建模复杂关系的场景下弊端也比较明显,为此 TransE 相关的扩展模型<sup>[5]</sup> 也陆续被研究应用。其他的 KGE 模型如基于分解的模型 DistMult<sup>[6]</sup>、ComplEx<sup>[7]</sup> 以及旋转模型 RotatE<sup>[8]</sup> 等也都获得了广泛的关注和应用。

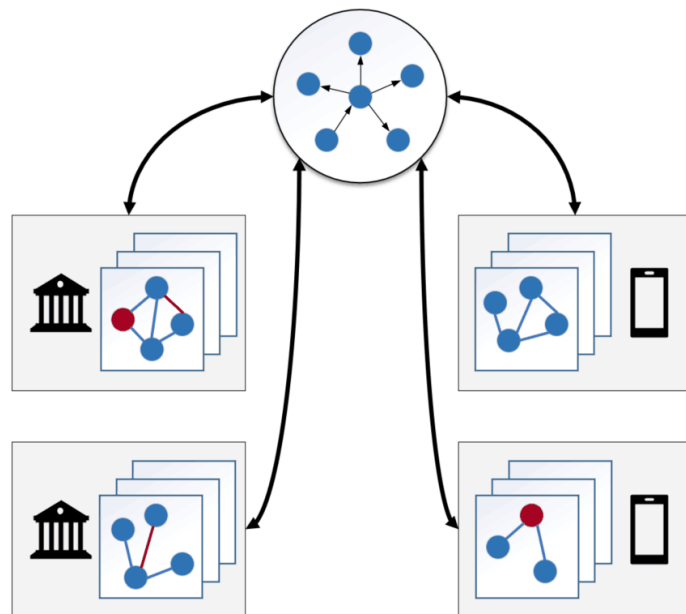


图 1.1 跨设备下的知识图谱分布

随着知识图谱的规模越来越庞大，单机的模型训练已经满足不了更加复杂的模型需要，因此知识图谱被逐渐部署在多台设备上并行处理来提高计算的效率，同时这些分散的图谱数据因为规模、成本等原因无法进行很有效的统一知识合并，而这些分散的跨设备的知识图谱会随着用户使用不断的进行更新和扩充，新的实体和关系会使基于传统的表示学习模型的效果下降。现如今个人移动设备的性能越发强大，为每个客户提供最符合个人爱好的个性化服务也是服务提供商的竞争优势之一，这种在目标图谱中同时包含已知实体、已知关系、未见实体、未见关系，本文称之为跨设备场景下的知识图谱。如何在跨设备的场景下的知识图谱上设计出一套高效的图谱嵌入模型，使得该模型能够尽可能满足新兴知识图谱的需要就成为了未来发展及现实需要的很重要的待解决命题。

为了处理知识表示遇到的未见关系和未见实体的问题，相关的归纳知识图谱表示方法研究也取得了一定的效果。**Ren**<sup>[9]</sup> 团队通过对语义证据的预测及实验验证研究了 KGE 的外推问题，通过建模对三种语义证据的加强，在知识图谱补全任务上预测不可见数据时都取得了更好的表现效果。此外 **Teru**<sup>[10]</sup> 等人通过切分实体子图分析子图的关系信息来获得实体表示，可以对包含未见实体的关系进行预测，但目前仅针对包含未见实体的任务，可以同时处理未见实体和未见关系的知识图谱表示模型缺少且效果不佳。因此本文提出了一种融合知识图谱结构信息和本体语义信息的知识图谱表示学习模型，可以同时数据中的未见关系和未见实体进行表示学习并取得不错的效果。

## 1.2 国内外研究现状及趋势

知识图谱表示学习一般包含三个步骤：1、初始化实体和关系的低维表示，其中实体通常表示为目标空间中的向量点，关系通常表示为空间中的操作如向量、张量、矩阵等。2、定义一个统一的得分函数，将图谱中的实体和关系组合形成多个三元组，对任意一个三元组可根据该得分函数计算出得分，一般来说在知识图谱中已知的事实三元组的得分应该最高。3、学习实体和关系的表示，通常采用梯度下降的方式求解。其中根据评判标准的不同，得分函数大体上分为三类：基于翻译的方法、基于语义相似度的方法以及融合辅助信息及归纳推理的方法。本节将分为三小节简单介绍相关的研究现状及趋势。

### 1.2.1 基于翻译的方法

知识图谱表示学习的关键是学习出实体和关系特征的低纬度表示，这些表示空间主要是指逐点空间，具体的像向量、矩阵和张量等表现形式；而其他类型的表示空间，如复向量空间、高斯空间也被研究者所使用<sup>[11]</sup>。基于翻译的方法通过测量实体向量间的距离来计算出三元组事实的可信度。这些基于翻译的方法通常会将关系的嵌入视为从头结点到尾结点的向量平移，例如 Bordes 等研究者在借鉴 Word2Vec<sup>[12]</sup> 中的语义平移后提出的 TransE 模型认为在同一表示空间中头结点向量和关系向量的向量和在距离上应该更靠近事实上的尾结点，并以此设置评分函数进行图谱嵌入，在 WordNet 和 FreeBase 两个数据集上的链接预测任务的效果相比于早期的 RESCAL<sup>[13]</sup> 模型会更好。同时由于 TransE 模型结构并不复杂，实现简单，在学术界和工业界都获得了广泛的研究及使用，使其成为经典的知识表示学习方法之一。

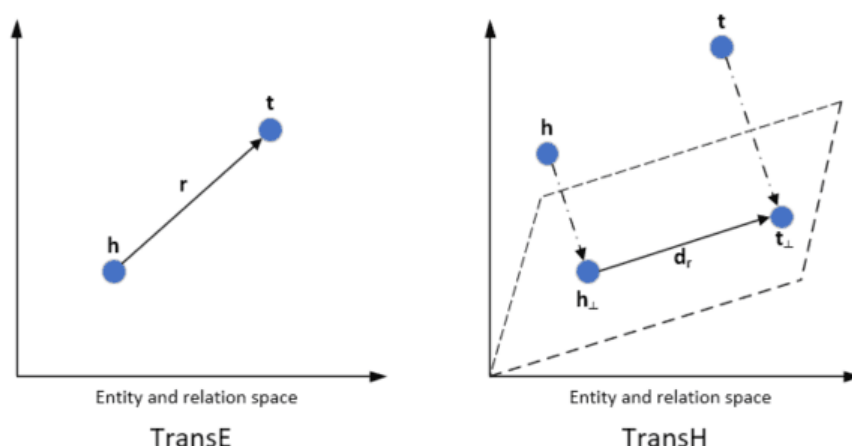


图 1.2 TransE 和 TransH 的实体关系映射

Bordes 提出了关系的四种类型：1 对 1、1 对多、多对 1、多对多。TransE 通过比较头结点与关系的向量和与尾结点的距离能够有效的处理 1 对 1 的简单关系类别，但同时缺

点也显而易见,在处理复杂关系时由于实体或者关系嵌入的相似性会导致实验效果的下降。假设建模 1 对多的复杂关系,有(海贼王,人物,梦奇·D·路飞),(海贼王,人物,娜美)两个三元组,在同一个表示空间中,由于实体  $h$  和关系  $r$  相同,TransE 会认为对应的所有尾实体都应该具有极为相似的向量表示,但很明显这个结论与事实相悖。为了弥补 TransE 模型的弊端,在该模型的基础上,相关的模型研究也有更新的发展。TransH 通过将头尾实体都映射到关系的超平面上,即使头实体和关系一致也不会导致尾实体的嵌入一致,规避了尾结点重叠的问题,可以在一定程度上处理多对多的关系。

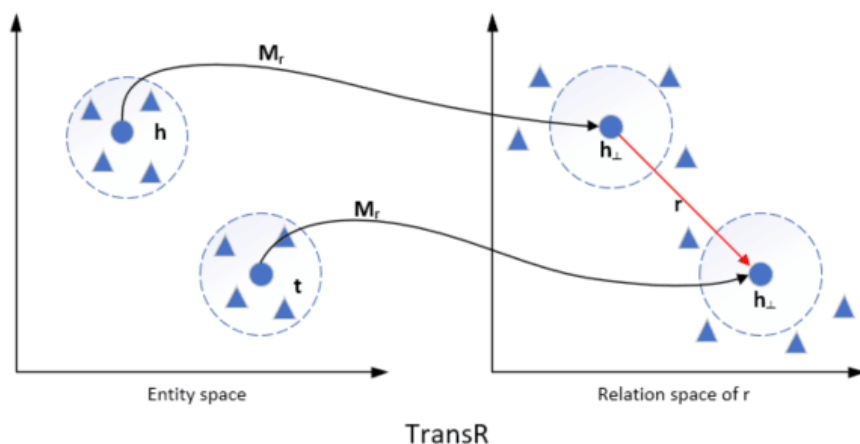


图 1.3 TransR 的实体关系映射

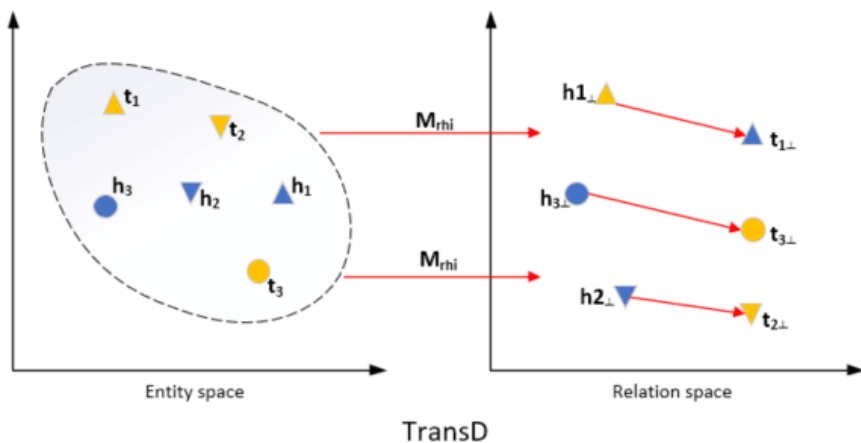


图 1.4 TransD 的实体关系映射

TransR<sup>[14]</sup> 舍弃了实体和关系映射在统统一语义空间中的假设,把 TransH 涉及到的实体在关系上特有的映射的概念延展成关系特有的语义空间映射,在 TransR 的设定中三元组的实体映射到一个语义空间,而关系则表示为关系语义空间中的转换向量,TransR 定义了关系关联的映射矩阵,在执行转移操作之前,分别通过实体向量在关系映射矩阵上的映射获得关系空间内的头实体与尾实体表示,最后通过计算头实体映射向量

加关系向量和尾实体映射向量的距离得到优化目标。但 TransR 模型将实体语义空间与关系语义空间的交互仅仅与关系矩阵相关联明显不合理，且由于空间投影使模型参数急剧增加，基于此 Ji 等人提出了 TransD<sup>[15]</sup> 模型，TransD 模型设置了两个投影矩阵，分别将头实体和尾实体投影到关系空间，同时只利用两个投影向量构建投影矩阵来减少参数量。

总体上，基于翻译的 Trans 模型都将关系作为实体向量间的翻译，但受限于结构的弊端往往不能很好处理复杂关系的场景，无法得到蕴含更多特征的知识图谱表示。

### 1.2.2 基于语义相似度的方法

基于语义相似度的方法通过标记实体和关系在向量空间表示所隐藏的语义特征的相似程度来作为评判三元组合理性的标准。传统的语义匹配的模型一般采用张量分解的方式来完成，其中的张量由一个多维的数组来表示<sup>[16]</sup>。

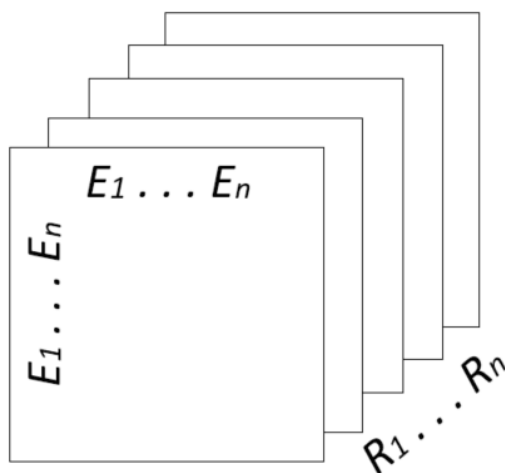


图 1.5 知识图谱的张量模型

RESCAL 模型将每个实体与一个向量联系起来，以捕捉其潜在的语义。每一种关系都用一个矩阵来表示，该矩阵模拟了潜在因素之间的双向相互作用。该模型通过潜在特征的成对交互解释三元组，首先将知识图谱中的三元组转换为三维张量  $X$ 。给定一个张量  $X \in R^{n \times n \times m}$ ，其中  $n$  和  $m$  分别表示实体和关系的数量，该张量的两个模保存连接的头尾实体，另一个模则保存关系。如果一个张量实体  $X_{ijk} = 1$  即表明知识图谱中存在三元组（第  $i$  个实体，第  $k$  个关系，第  $j$  个实体）；否则， $X_{ijk} = 0$  表示不存在或未见的三元组。然后对张量的每一个切片（关系数据）执行分解来获得实体的潜在语义表示，然而该模型的主要缺点在于对频繁出现的关系表现的很好，但对少见的关系表现的很差，并且容易导致过拟合，不适用存在大量稀有关系的数据集<sup>[17]</sup>。鉴于此，研究者提出了一个潜在因子模型 TATCE<sup>[18]</sup>，它能够将高容量三向模型与更易于控制的双向交互模型相结合，

并利用它们两者的优势。由于双向和三向模型不使用相同的数据模式，也不在嵌入中编码相同的信息，因此在 TATEC 中，他们在第一阶段使用了两种不同的嵌入，然后在后期将它们组合并微调。相比于三向模型 RESCAL、传统双向模型 TransE 以及同样采取双向三向模型结合的 NTN<sup>[19]</sup> 模型等，在链接预测任务上 TATEC 表现更优。为了简化模型计算的复杂度，DistMult 模型在 RESCAL 的基础上将关系矩阵简化为对角矩阵，从而减少双线性模型的参数量，达到和 TransE 基本相同。ComplEx 模型是对 DistMult 在关系模式上的改进，不仅具有学习对称关系的能力，还能够提供一些非对称性的表示。

近几年随着神经网络的兴起，由于其具有强大表示和泛化能力，也可以表达复杂的非线性投影。结合神经网络将知识图嵌入到连续的特征空间成为一个热门方向。其中代表性的如 ConvE<sup>[20]</sup>、ConvKB<sup>[21]</sup> 模型等，这些早期的结合神经网络模型的 KGE 方法采用卷积神经网络进行特性提取及嵌入，通过卷积神经网络捕捉知识库中的全局关系和翻译特性。将三元组输入到卷积层中进行卷积操作，输出特征并拼接为特征向量，然后经由评分函数进行打分。随着图卷积模型<sup>[22]</sup> 的提出及效果的不断提升，以图卷积网络为基础的 KGE 模型开始占据主流。R-GCN<sup>[23]</sup> 首次提出采用图卷积网络做图谱嵌入任务，使用 GCN 的信息传播的思想表示图谱，并用上层的卷积输出作为下层的卷积输入。虽然 R-GCN 的效果提升都比较微弱，但是因为 GCN 的出现，用它来做知识图谱表示学习也成为是一种必然。此后，基于 GCN 的 KGE 模型如 CompGCN<sup>[24]</sup> 也已经取代传统的神经网络 KGE 模型，不断取得更好的知识图谱嵌入效果。

### 1.2.3 融合辅助信息及归纳推理的方法

基于翻译的表示学习方法和基于语义相似度的表示学习方法通过三元组信息学习实体及关系的表示。这些方法的目的都是学习知识图谱存储的事实信息并映射到嵌入向量中，使嵌入向量能够更好地贴近这些事实，从而完成各种下游任务。然而，这些仅依赖于知识图谱三元组进行表示学习的方法存在明显的弊端。这些学到的嵌入表示只能处理知识图谱中可见的实体和关系，对于未见组件可靠性会大幅下降。为了能够对这些未见的组件进行有效的特征学习，研究者往往会引入除三元组之外的辅助信息。因为知识图谱不仅通过三元组进行数据存储和检索，还蕴含着丰富的其他信息可以辅助进行知识图谱表示学习，比如文本描述信息、关系路径信息及图结构信息等。

NTN 模型最早在表示学习中引入实体的描述信息。虽然该模型是利用图谱存储的实体的文本信息来对实体嵌入进行初始化操作。此后描述扩展的知识图谱表示模型 DKRL<sup>[25]</sup> 试图通过改进 TransE 模型使其能够进一步处理实体描述，DKRL 将实体嵌入由实体结构相关的变量及实体描述相关的变量联合表示，其中实体描述变量由描述文本通过连续词袋<sup>[26]</sup> 编码器或卷积神经网络编码器获得的词嵌入构成。另一种常见的辅助信息通过借助实体的关系信息来进行表示学习，关系信息揭示了实体之间的一个或多个语义关系。PtransE<sup>[27]</sup> 首次将图谱的多跳关系信息作为知识图谱嵌入的辅助信息，提出

了一个以关系路径为基础表示学习模型。该模型在 TransE 基础上改进，模型打分函数由两个部分组成，一方面是头尾实体直接关系的打分，另一方面计算头尾实体间其他多跳关系路径的打分。从而将 transE 的单步推理扩展为多步，获得了不错的效果提升。冯俊等人则将知识图谱视为一个大的有向图，提出了基于图感知的表示模型 GAKE<sup>[28]</sup>。该模型利用知识图的结构信息生成实体和关系的表示。

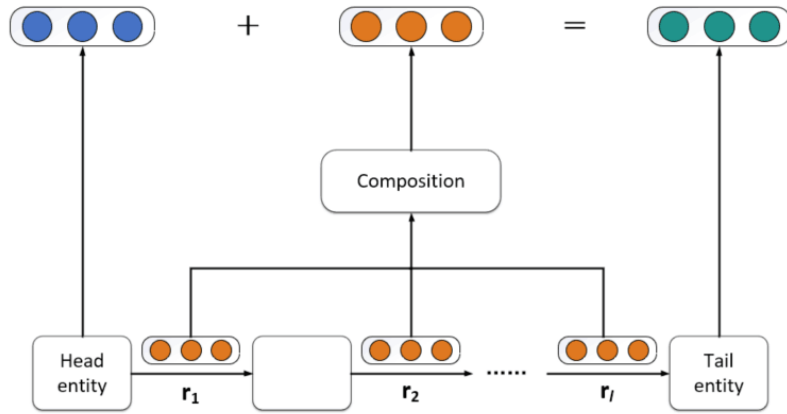


图 1.6 PtransE 的模型架构

除了知识图谱自身的额外信息作为辅助外，许多研究者也会引入图谱外的其他知识用于表示学习。比如上述的实体描述信息，也可以从新闻稿或者维基百科中获取。图谱外其他模态的信息也同样被用于辅助表示学习，比如实体的图像信息，代表性的 IKRL<sup>[29]</sup>模型实现了包含基于跨模态结构和基于图像表示，将图像编码到实体空间同时遵循平移原则。该模型提出的跨模态表示可以确保基于结构的表示和基于图像的表示映射在同一个表示空间中。

以往的知识图谱表示学习聚焦于如何从现有的图谱知识中学习更贴近知识的嵌入表示来完成相关的预测、补全、问答等任务，但这些学习往往学习到的不是更全局的表示，仅在现有的可见的图谱任务上才能获得较好的效果。而跨领域跨设备的场景要求更加苛刻的表示学习效果，需要更高层次、更适用性的知识提取，因此现有的研究开始向通用性知识以及辅助知识提取方向发展，该方向研究正在起步但无疑是未来的主要方向，本文的研究目标即在高层本体语义信息的辅助下进行更普适性的知识图谱表示学习。

### 1.3 本文主要研究内容

针对跨设备场景下知识图谱表示学习的调研不难发现传统的知识图谱表示学习方法无法很好处理未见的组件，而通过融合其他辅助信息来对未见组件表示的方法往往只能聚焦于某一类未见组件，而不能处理同时存在未见实体和未见关系的情况；同时在跨设备的场景要求下大大加强了对模型泛化性的要求，因此本文总结出以下两个主要面临



的难题:

- 1) 传统的知识图谱嵌入方法无法有效处理未知的实体和关系。当采用使用归纳推理的模型时,往往需要依赖规则和结构信息来处理单一未知组件,这些方法通常在子图的结构信息中学习未知组件的表示,但忽略了知识图谱中丰富的语义信息。如何在归纳推理模型中嵌入有效的语义信息是当前研究的一个难点。
- 2) 对于跨设备下的知识图谱表示学习,模型需要在训练数据集上训练获得参数,使其在存在未见组件的新型知识图谱上也可以正确地发挥作用。然而,在采用分散知识图谱存储方式时,需要思考如何进行模型训练,并在其训练中设置未见组件以帮助模型获取更强大的泛化性能,以更好地满足现实场景下的需求。

本文通过对近几年相关技术的研究认为,本体作为知识图谱最高层次的语义抽象可以作为所有设备的全局信息,并可用于对未知关系和实体进行一定程度的语义特征补充。此外,通过确定关系的相对位置关系,可以构建关系位置图,并嵌入本体的语义信息,从而对关系节点进行编码,以学习未见关系的特征表示。而后,基于实体及连接关系的结构信息,可以聚合连接未知实体的所有关系的特征,以表示未知实体。最后,为了模拟出具有未知关系和实体的训练场景,本文借用元学习的思想设计了多个训练任务,以学习最佳模型参数,并通过实验验证了模型的有效性。

综上,本文的研究内容主要有以下几个方面的贡献:

- 1) 分析了跨设备知识图谱表示学习当下遇到的一些问题及相关研究方法的一些进展,提出了一个融合了本体语义信息同时对未见关系和未见实体进行表示学习的嵌入学习模型框架。
- 2) 引入元学习的模型训练方法,设置多个单任务的训练任务,在各个任务中引入未见关系和未见实体用于模型训练,使得模型能够在单个训练集上训练并且能够快速作用在测试图谱上,获得更强的模型泛化能力。
- 3) 在多个数据集上进行了丰富的模型测试并与其他基准模型结果进行对比分析,验证了模型的有效性;同时通过一系列消融实验验证模型各部分的重要性,为之后的模型改进提供思路和支撑。

## 1.4 本文组织结构

本文的内容分为五章,以下主要概括各章的内容及安排:

第一章,绪论部分:主要介绍跨设备下的知识图谱表示学习相关的研究背景和现实意义,结合国内外研究者的相关研究分别介绍了传统的知识图谱表示学习方法中的基于翻译和基于语义相似度的方法以及相关一些通过辅助信息和归纳推理解决未见组件的一些研究进展,最后总结了研究面临的问题及本文的解决思路。

第二章,相关技术研究:分三个小结介绍了本文模型主要涉及到的三种技术的发展状况及原理分析。其中本体信息作为本文模型主要的语义信息来对未见组件进行语义



层次的补充；元学习训练方法已经被广泛运用在各个领域，本文介绍了涉及到的设计方法；最后本文模型作为归纳推理模型的延伸，介绍了归纳推理模型相关的研究进展。

第三章，基于元学习本体增强的跨领域知识图谱表示模型介绍：详细介绍了本文提出的知识图谱归纳表示学习模型的各个组成部分，主要包含了本体信息嵌入模块、关系结构图构建及基于 GCN 的关系特征学习、基于 CompGCN 的实体特征学习模块。学习到实体和关系的相关表示后采用多种 KGE 得分函数来进行打分和更新。最后说明了模型在进行元学习训练的任务设定及训练流程。

第四章，实验结果及分析：介绍了本文采用的可用于归纳推理知识表示学习任务评测的三个数据集的构建方法以及相关用于本体嵌入的本体三元组抽取的方法。本章将本文模型及相关基准模型在目标数据集上进行了充分的实验并比较实验结果，验证了模型的有效性。最后对本文模型的各组成模块进行消融实验，验证了各模块的重要性。

第五章，总结部分：综合全文的研究内容及实验结果，总结了模型的可取部分及仍存在的问题与可改进的方法，对未来的相关工作进行规划。

## 2 相关技术研究

经典的知识图谱嵌入模型通过提取三元组中的事实特征来对实体和关系编码，并使用如 TransE、RESCAL 等模型设计的评分函数来对嵌入后的三元组进行打分，以改进图谱嵌入的效果。但在跨设备知识图谱的应用场景下，图谱会引入训练数据中不存在的未见的实体和未见的关系，导致传统的嵌入模型无法学习到这些组件的特征表示。为了解决这一问题，研究者往往会借助除三元组之外的辅助信息来加强对未见组件的特征提取，比如实体的位置信息、邻接关系及图谱包含的其他文本信息等。由于这些信息与训练数据的没有很强的依赖关系，因此可以迁移到未见的组件上，从而学习出通用的特征来进行嵌入表示。然而，现存的相关模型为了摆脱数据强依赖的特性，在研究图谱中对未见组件编码进行特征表示时，忽略了知识图谱作为知识网络中包含的丰富的语义信息，没有很好进行多维度知识的融合。

为了解决这一问题，本文探索了将本体知识融合到学习图谱中可迁移的特征表示的过程中，并采用元学习的训练方法对场景进行模拟和快速学习。本节将简要介绍涉及到的元学习训练方法和归纳、本体嵌入知识图谱嵌入模型的相关内容。

### 2.1 知识图谱本体嵌入

本体作为表示和交换通用或主要知识的载体，通常以层次概念作为描述语义关系的主干和属性，并且可以选择性地定义一些逻辑约束，例如类的不相交性、属性域和可取值的范围。随着知识图谱在知识问答、推荐系统等下游任务上的广泛应用，像 NELL-995<sup>[30]</sup> 和 DBpedia<sup>[31]</sup> 这样的几个大型的知识图谱在研究和实际应用中被广泛使用。这些经典知识图谱通常会被从两个方面进行使用：一方面从实例出发，这些知识图谱存储了丰富的事实三元组，每个三元组中的实体由关系进行连接，如（奥巴马，政治家，美国）；而从本体的视角，知识图谱从广泛的事实三元组中进行提取构建抽象概念的语义元关系，如（政治家，领导，国家）。同时，知识图谱还能提供实体到本体的相关映射关系。这些存储在知识图谱中对知识的高高层抽象是对知识图谱语义结构的集中表示。

在知识图谱表示学习中，基于翻译的模型和基于语义匹配的模型已经被广泛应用，并且在一定程度上取得了不错的效果。然而，这两种模型都有一些局限性，它们忽略了知识图谱中包含的额外知识。知识图谱中的实体往往有着不同的类型信息，这些信息可以帮助模型更好地理解实体之间的关系。基于翻译的模型和基于语义匹配的模型通常只考虑了实体本身的信息，而没有利用实体类型的信息。另外，知识图谱中的实体之间通常存在层次关系，例如子类和父类之间的关系，这些层次关系可以为模型提供更多的语义信息，但是这些信息没有被基于翻译的模型和基于语义匹配的模型都没有很好地利

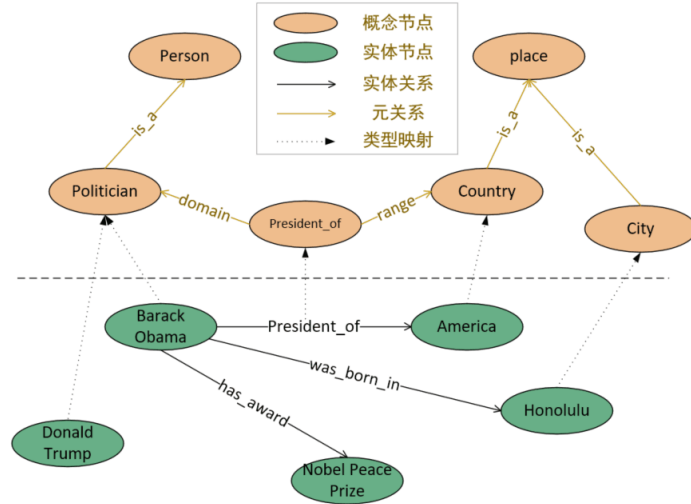


图 2.1 知识图谱的本体视角和实例视角

用。相反，利用这些全局信息进行表示的模型可以通过将语义或者结构信息整合到知识嵌入过程中来增强知识表示的学习效果。

一些利用本体进行知识图谱表示学习的方法将本体信息作为实体的额外类型信息，实体借用本体定义的更高层次的类型来表示，关系也使用语义类型进行表示。例如，将“马斯克”用本体定义的“人类”概念节点进行表示，“特斯拉”则被表示成所属的“公司”节点。这些简单的引入类型信息的方法是基于传统的基于三元组的模型的扩展，它集成了额外的文本信息来改进其性能。SSE<sup>[32]</sup> 模型结合实体的语义类别，将属于同一类别的实体平滑地嵌入到语义空间中。为了能够更有效的捕捉到这些高层本体类型信息带来的限制效果，该模型提出了两种在语义层面对嵌入变量的限制。首先该模型通过拉普拉斯特征映射的算法思想认为同一个语义类别下的实体的嵌入应该更加相近，即“法国”实体和“意大利”的实体因为同属于国家的语义所以他们对应的嵌入表示应该更为相似。因此，模型为每一个实体设置了一个语义分类矩阵  $W_{ij}^{(1)}$ ，如果任意两个实体在类型上相同，那么该矩阵中对应的值即为 1，否则为 0，基于该矩阵计算两个实体之间的语义平滑值：

$$\mathcal{R}_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|e_i - e_j\|_2^2 W_{ij}^{(1)} \quad (2.1)$$

其中  $e_i$  和  $e_j$  指代两个实体  $i$  和实体  $j$  对应的嵌入表示，通过使得两个同类实体的语义平滑值最小来实现拉普拉斯特征映射的语义限制。其次该模型还提出基于局部线性嵌入的语义约束思想，不同于拉普拉斯特征映射的对数据对局部不变性的设定，基于局部线性嵌入的思想认为一个实体可以通过其最接近的邻居节点通过一个线性的组合器来近似的表征，举例如“法国”的实体嵌入可以通过其语义相邻的节点如“中国”、“意大利”等国家节点通过线性组合的方式来近似的表示。同样，在该语义约束下 SSE 设置了一个类别邻接矩阵  $W_{ij}^{(2)}$ ，该矩阵指明实体的连接节点是否是该节点语义范畴内的最邻

近节点。然后在约束下计算实体的语义平滑值：

$$\mathcal{R}_2 = \sum_{i=1}^n \left\| e_i - \sum_{e_j \in \mathcal{N}(e_i)} w_{ij}^{(2)} e_j \right\|_2^2 \quad (2.2)$$

其中  $e_i$  和  $e_j$  指代两个实体  $i$  和实体  $j$  对应的嵌入表示，通过使得实体  $i$  和所有最邻近节点的线性组合距离最短来实现该猜想的约束。最后该模型分别通过在基于距离和基于语义相似度的嵌入模型上加入两种语义的约束来提高嵌入表示的效果。实验证明在高层语义类别的约束下的表示学习取得了很好的嵌入效果。虽然通过这两种类别信息的约束能够有效的引入了图谱除结构外额外的知识，但是该模型认为一个实体只属于一个类别，忽略了多个类别可能存在的联系，没有完全充分利用好高层的语义知识。

同样借助于本体的类型信息，JOIE<sup>[33]</sup> 模型从本体和实例三元组两个不同的层次分别进行嵌入学习，设计了跨视图和视图内建模模型，从知识库的多个方面进行学习。一方面学习了跨视图关联模型，实现本体概念与相应实例视图实体之间的连接。另一方面对视图内模型进行训练，以在单独的嵌入空间中捕获实例视图和本体视图的结构化知识，并为具有层次结构的本体启用了层次感知编码技术，虽然这两个模型都运用到了本体的类型信息，但是他们也使用了单一的类型，没有充分考虑到本体丰富的其他语义关系。

TKRL<sup>[34]</sup> 模型基于传统的翻译模型并结合本体的实体类型信息提出了一个 type-embedding 的嵌入方法。与 SSE 模型实体单一类别的设定不同，TKRL 模型认为一个实体可能有多个不同层次的类别，并且这些不同层次的类别信息应该被转化到与类别绑定的不同的特征空间中，因此三元组中的每个实体都应该包含多个实体类型相关的特征信息联合表示。而为了获取到实体的多类别的语义信息，TKRL 模型采用所有实体类型转换矩阵的加权和来计算出最后的转换矩阵：

$$M = \frac{\sum_{i=1}^n \alpha_i M_{c_i}}{\sum_{i=1}^n \alpha_i}, \alpha_i = \begin{cases} 1, & c_i \in C \\ 0, & c_i \notin C \end{cases} \quad (2.3)$$

其中  $n$  是一个实体所有可能的实体类型数量，指代第  $i$  个实体类型，是该实体类型对应的转化矩阵，是矩阵对应的权重值，最后的指代三元组给定三元组  $(h, r, t)$  头实体所有与关系  $r$  相连接的可能的实体类型的集合。TKRL 引入实体类型并对所有可能的类型信息进行聚合来增强嵌入模型的学习效果。

为了提高知识表示学习的性能，现有的大多数本体模型在知识嵌入过程中都包含了单一的本体信息，无法实现对所有可用本体信息的无缝嵌入。在现实世界中，知识图谱通常是不完整的，单一的本体信息对补充图谱的作用是有限的。本文能够在知识嵌入过程中无缝地整合所有可用的本体信息，充分补充图谱，提高复杂场景下的决策能力。同时本文使用了一种简单的本体形式，即 RDF Schema(RDFS) 中，而那些更复杂的 OWL 本体可以按照一定的标准转换为 RDFS 本体。本体可以用作知识图谱的模式，定义实体

类型、关系等。本文将本体集合表示为  $\mathcal{O} = \{\mathcal{C}, \mathcal{P}, \mathcal{T}_o\}$ ，其中  $\mathcal{C}$  是概念节点的集合 (即实体类型和实体关系)， $\mathcal{P}$  是属性的集合， $\mathcal{T}_o$  是本体三元组的集合。

## 2.2 归纳知识图谱表示学习

随着知识图谱日益增多的落地应用，图谱在分布的终端上独立的更新和训练，越来越多的出现了类似训练集中不存在但现实中又遇到的实体或关系等出现未见组件的情况，为了能够更好的将模型适用于这种现实的场景需求，一些通过归纳推理的模型也取得了一定的效果。

为了弥补仅从实体三元组学习嵌入导致不适用于训练集中不可见的实体的弊端，一些研究者首先尝试了从用额外的知识对嵌入表示学习进行补充使得对未见组件的嵌入可以从额外的知识中学习出一定的有效特征。例如 IKRL 模型在首次在原有的实体三元组的基础上补充了视觉的信息，将头尾实体的图像信息作为图像特征进行编码并和三元组学习的特征进行融合，在知识图谱补全任务和三元组分裂任务上都取得了很好的效果。不同于传统的知识图谱都是固定的且实体基本上不会随意新增的“closed-world”的知识图谱补全任务，ConMask<sup>[35]</sup> 为了解决知识图谱不断发展中很有可能出现的未见的实体提出了一种“open-world”下的表示学习方法。该模型为了能够实现包含未见实体的知识图谱补全任务，将学习的重点转移到了实体的描述文本上，模型包含三个部分：首先通过目标关系的名称圈出和任务相关联的单词并抹去不想干的单词；然后从实体的名字和实体的描述文本中学习到该实体的嵌入表示；最后通过计算各候选实体与目标实体在任务上的相似度来判断是否能组成事实三元组。虽然上述的方法在响应的知识图谱补全任务上都取得了一定的效果，但是没有足够的证据表明这些方法能够判断出文本和图像中的隐含事实。同时在一些情况下，假设新实体并没有提供描述文本或者图像信息时可能这些方法的效果就会大打折扣。

除了引入实体或者关系的某一特定类型的信息作为对未见组件特征知识来源的方法外，一些研究者通过对现有信息学习出逻辑上的规则来进行归纳表示学习，这些方法通过在训练集上学习出的独立于实体节点的规则来覆盖到未见的实体上进行表示。传统的基于规则的表示学习方法如 AMIE<sup>[36]</sup> 模型往往采用统计度量或者固定的人工设定的模式进行规则的学习，但是无疑这些方法有存在无法扩展的问题并且缺乏足够的推理表征能力。而最近的出现的一些可微规则学习方法如 Neural LP<sup>[37]</sup> 模型和 DURM<sup>[38]</sup> 模型通过端到端的方法学习规则逻辑，通过设计可微的逻辑规则学习模型并通过基于梯度的方法进行优化求解，但是这些方法不能解决知识图谱中缺边的问题同时处理不合理的规则候选方面存在不足。而不同于直接从事三元组中学习逻辑规则的方法，一些方法利用子图隐式地表示逻辑规则。GraIL 和 TACT<sup>[39]</sup> 通过对未见实体周围的封闭子图的学习逻辑规则，但是很显然随着实体邻居数目的指数增长，这些子图的规模可能会很庞大从而可能导致效率问题。

不同于引入额外知识及学习全局逻辑规则的方法，未见实体的已知邻居节点的特征也被视为归纳推理模型的一种新型的知识输入。为了能够处理在训练阶段未观察到的测试实体的问题，GNN 模型通过聚合所有邻接节点中已知部分的特征来作为未见组件的特征，该模型提供了一种在图上很好的通过聚合邻接特征来进行学习的思路。此后基于图神经网络的归纳知识图谱表示方法开始成为另外一类主要的研究路线。GNN 模型专门用于处理图数据，可以直接输入一个图网络并能够很容易的进行节点级别、边级别及图级别的预测任务，相比于 CNN 模型只能作用在具有相同结构的图像或者特定序列的语音和文字上，图数据没有固定的形式且邻居节点的也都是无序的，因此无法作用在复杂的知识图谱上，而 GNN 通过聚合和更新操作能够有效学习到图谱结构和节点特征的有效信息。

为了有效的学习到图的结构信息和节点的特征并对特征进行嵌入，GNN 主要包含了两个部分：聚合函数和更新函数，聚合函数可以对排名的邻接节点的特征进行聚合，常用的操作包含 sum、mean 以及 max 等。一次聚合操作可以将节点邻近一跳的邻接点的信息进行提取，而 GNN 一般包含多层，每一层都会接受上一层的信息在此进行聚合和传递，因此 n 层聚合后传递的信息会包含 n 层邻接节点的结构信息以及节点本身的特征信息。每层的特征聚合输出公式基本如下：

$$h_v^k = \sigma(W_k \sum \frac{h_u^{k-1}}{|N(v)|} + B_k h_v^{k-1}) \quad \text{where } k = 1, \dots, k-1 \quad (2.4)$$

从公式中可以看出每层的信息包含了两个部分，首先  $W_k \sum \frac{h_u^{k-1}}{|N(v)|}$  代表了对邻接点的特征聚合操作，而后一部分则是上一层聚合特征与一个可学习的权重参数的乘积，最后该两部分的特征通过一个激活函数来进行更新从而完成该层的节点特征输出。但是在 GNN 模型中聚合操作并没有区分邻接节点的重要性而是简单的进行了池化操作。为了增加对邻接点的关注度的区分，受到注意力机制在基于序列的任务上的优秀表现，GAT(Graph Attention Networks) 将该机制引入到图卷积网络中用于对图数据中的节点进行分类。注意力机制的好处之一是它可以处理可变大小的输入，并且聚焦于输入中最相关的部分来进行决策，而且在其他如 RNN 等模型在加入自注意力机制后也取得了不小的模型提升，可以证明该机制足以构建强大的模型。因此 GAT 使用注意力机制来判别邻居节点，计算图中的每个节点的隐藏表示，通过在节点-邻居对之间进行并行高效的计算而且可以指定邻居的任意权重来应用于具有不同度数的图节点，且相关论文的实验结果表明这种机制的引入同样可以高效地作用于归纳学习问题用于处理存在未知的组件等任务。

基于 GNN 的另外一种改进的方向聚焦于对图结构关系表示的补充上，图卷积神经网络集中于对节点特征的聚合和更新表示，在信息传递的过程中图的关系结构仅仅用于指明邻接点，关系的特征并没有参与到节点的更新过程中。为了增加关系信息对节点的

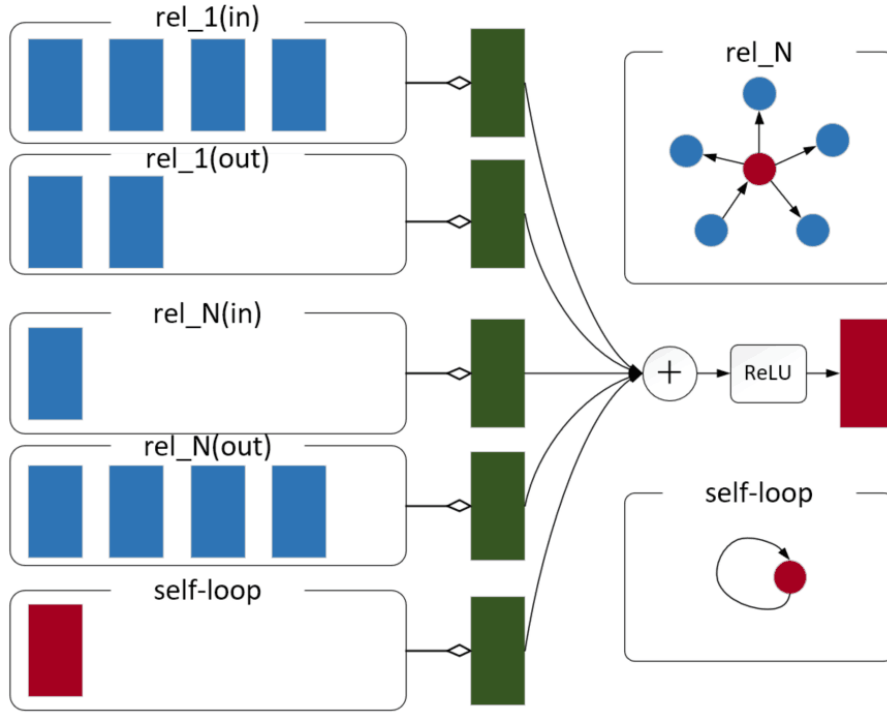


图 2.2 R-GCN 的特征传递

影响，R-GCN 在每层节点特征计算的过程中引入了邻接点及它们的关系：

$$h_i^{l+1} = \sigma \left( \sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{c_{i,r}} W_r^l h_j^l + W_0^l h_i^l \right) \quad (2.5)$$

与 GNN 对邻接点特征聚合的操作不同，R-GCN 引入了关系特定的转换  $W_r^l h_j^l$ ，这种转换取决于边的类型和方向， $c_{i,r}$  是问题相对应的归一化参数。为了确保第 1 层的节点表示可以受到相应层次的表示的影响，R-GCN 在基础的图关系上为每个节点添加了一个自连接的特殊关系。如图 2.2 所示，在 R-GCN 中每层对一个实体节点（红色块表示）进行特征生成的过程中，首先从邻接点获取特征（蓝色块表示）并根据该节点与邻接节点的关系类型进行特征转换得到该种关系对应的表示（绿色块表示），其中关系类型分别由入关系、出关系以及自循环关系组成；然后将所有关系转换后的邻接节点信息进行累加求和并通过一个如 ReLU 的激活函数即可获得该节点本层的输出表示。相比于 R-GCN，近期提出的 CompGCN 在 R-GCN 模型同时考虑关系特征和节点特征对图数据进行嵌入表示的基础上进一步引入了注意力机制，CompGCN 针对每种边类型和方向分别进行了注意力计算以加强对重要信息的关注，而且在计算效率方面，它减少了每个节点的嵌入大小并减少了依赖于固定卷积核的计算量，因此更适用于大规模图数据。

自图卷积网络广泛推广到图数据处理上后，其本身对于图数据隐藏知识的挖掘在一定程度上已经可以作用于归纳知识图谱表示学习上。近几年的归纳知识图谱表示学习的

相关方法也大多利用图卷积神经网络进一步提升归纳表示学习的效果,例如 INDIGO<sup>[40]</sup>模型在 GNN 的基础上进行改进,使用实体三元组与 GNN 的最内层和最外层的特征向量元素之间的一一对应关系对知识图谱进行编码。因此,预测的三元组可以直接从 GNN 的最后一层读出,而不需要额外打分函数,从而充分利用 GNN 的特征聚合能力。其他的相关的工作也有如 xiong(2018) 等人依赖邻域结构对未见关系进行编码学习的方法,但这些邻域的方法要么是专注于传统的知识图谱表示学习领域,要么是只能处理未见的关系而忽略未见的实体。本文通过本体全局信息的嵌入表示,可以同时对待见关系和位置实体进行辅助的嵌入表示,同时借助关系的位置结构信息加强知识的引入,从而训练处一个同时处理未见组件的归纳推理知识图谱表示学习模型。

### 2.3 元学习训练方法

学习最普适性的算法思想可以理解为 learning to learn, 它通过多个学习任务的训练来改进学习算法,而传统的机器学习算法则是在多个数据实例上进行模型的学习。

传统的机器学习方法会设置一个训练集  $\mathcal{D} = \{(x_1, y_1), \dots, (x_N, y_N)\}$ , 例如样本对 (输入的图片, 图片的标签), 本文的目的是训练出一个模型函数  $y = f_{\theta}(x)$ , 通过求解下述公式来获得其中的参数  $\theta$ :

$$\theta^* = \arg \min_{\theta} \mathcal{L}(\mathcal{D}; \theta, w) \quad (2.6)$$

其中的  $\mathcal{L}$  是一个损失函数来计算真实标签与模型预测标签之间的误差, 指代了模型如何学习的假设, 例如如何为参数  $\theta$  选择合适的优化器或者为选择函数类型等, 传统的机器学习方法实现过程中该部分由研究者手动设置; 而模型的泛化性能则通过评估该模型在已知标签上的测试集测试来衡量。传统的机器学习假设是模型的优化是对每个训练集  $\mathcal{D}$  从头开始执行, 并且模型如何学习的相关设定  $w$  是预先指定好的。然而, 模型如何学习的设定  $w$  会极大地影响到模型最后的准确性或数据效率等性能指标。元学习试图通过学习学习算法本身来改进这些指标, 而不是假设学习算法是预先指定或者固定的。此外元学习从任务的分布中学习, 而不是从头开始。

元学习 learning to learn 的思想精髓可以看做一个包含内外两层的双层优化问题, 双层优化<sup>[41]</sup>指代的是层次优化问题, 其中一个优化包含另一个优化作为约束<sup>[42][43]</sup>。经典的内外双层模型的算法如 MAML, 其算法流程如算法1所示。

在该视角下, 元学习任务可以通过如下公式来规范化:

$$\omega^* = \arg \min_{\omega} \sum_{i=1}^M \mathcal{L}^{\text{meta}}(\mathcal{D}_{\text{source}}^{\text{val}(i)}; \theta^{*(i)}, \omega) \quad (2.7)$$

$$s.t. \quad \theta^{*(i)}(\omega) = \arg \min_{\theta} \mathcal{L}^{\text{task}}(\mathcal{D}_{\text{source}}^{\text{train}(i)}; \theta, \omega) \quad (2.8)$$

其中  $\mathcal{L}^{\text{meta}}$  和  $\mathcal{L}^{\text{task}}$  分别指代外层的优化目标和内层的优化目标, 例如在分类任务下的交叉熵, 但是这两层的优化级别并不对称, 内层优化在基于外层参数  $\omega$  的优化过



**Data:**  $p(\mathcal{T})$ :distribution over tasks  
**Data:**  $\alpha, \beta$  step size hyperparameters  
 randomly initialize  $\theta$   
**while** not done **do**  
     Sample batch of tasks  $\mathcal{T}_i \sim p(\mathcal{T})$  **for**  $\mathcal{T}_i$  **do**  
         Evaluate  $\nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$  with respect to K examples  
         Compute adapted parameters with gradient descent:  $\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}_i}(f_{\theta})$   
     **end**  
     Update  $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'})$   
**end**

算法 1 Model-Agnostic Meta-Learning

程中不能对  $\omega$  进行修改。而公式中  $\omega$  可以指代如非凸优化<sup>[44]</sup> 的内层模型的初始化参数或者内层模型的其他可学习的超参数。因此，元学习的整个训练流程分为了两层的优化：内层模型接收外层模型的参数  $\omega$ ，然后根据自己的任务对该任务的训练集进行训练并在任务的测试集上计算出损失函数；外层模型接收内层模型计算出的损失函数对参数  $\omega$  进行更新来使得内层函数的损失函数达到最优。元学习的思想即通过外层模型的训练学习到内层模型一个更好的设定，可以让内层模型更好的完成各种任务。

上面提到了内层模型在训练的时候需要针对面向的问题提供相应的训练集和测试集，这里的以任务为训练单位的设定也是元学习方法区别于传统机器学习方法的一大特点。从训练任务的角度来说，元学习的目标就是学习一种通用的能够作用在跨任务上的学习算法，这些学习到的算法能够在新的任务上获得更好的表现效果。内层模型可以视为带有外层模型参数  $\omega$  的经典的机器学习算法，其数据集  $\mathcal{D} = (\mathcal{D}^{train}, \mathcal{D}^{val})$ ，针对单个任务的损失函数即为  $\mathcal{L}(\mathcal{D}; \omega) = \mathcal{L}(\mathcal{D}^{val}; \omega^*(\mathcal{D}^{train}, \omega), \omega)$ ，而在实际应用中通常只有一个训练集和测试集，所以一般会从源训练集中抽样出一组用于训练的任务，在这些任务中的训练集和测试集为了避免和最终模型训练后进行评估的测试集混淆，一般称之为 support 集和 query 集。

元学习发展至今已经在许多领域展示出了它对推动当代深度学习行业前沿的巨大潜力，尤其是在深度学习领域饱受诟病的如数据效率问题、知识迁移问题以及无监督学习问题<sup>[45]</sup> 等方面展现出了更优秀的解决方案。在多任务场景，元学习已经证明可以从一系列任务中提取出来单个任务不可知的知识从而改进相关新任务的学习效果<sup>[46]</sup>；而在单任务场景下，元学习可以从该任务的重复处理中进行多个阶段的改进<sup>[47][48][49]</sup>，也已成功应用在小样本图像识别<sup>[50]</sup>、数据效率提高<sup>[51]</sup> 和神经架构搜索 (NAS)<sup>[52]</sup> 等领域。而加利福尼亚大学的 xiong 等人 (2018)<sup>[53]</sup> 提出的工作是第一个针对知识图的少数镜头学习的研究。它是一种基于度量的模型，由相邻编码器和匹配处理器组成。邻居编码器通过实体的一跳邻居节点来增强实体的嵌入，匹配处理器通过 LSTM 块来实现多步匹

配。

本文讨论跨领域跨设备的知识图谱上知识表示的相关研究并尝试解决在有未见组件的下游链接预测任务，经典的知识图谱表示学习的链接预测任务采用的数据集会设置一个训练集和测试集，测试集中不包含未见组件。因此，本文从现有数据集上采样出符合问题条件的数据，并借助于元学习 *learning to learn* 的思想，将这些数据打包为一个个任务，用于学习和训练，以期能够从这些数据中学习到能够处理未见组件的知识图谱表示。

### 3 基于元学习本体增强的跨设备知识图谱表示模型

本章将详细介绍提出的基于元学习本体增强的跨领域知识图谱表示学习模型各个部分。本章首节对模型要解决的问题和测试的任务作出符号上的定义和解释；然后将介绍本文设计模型的整体框架以及对框架组成进行大概的描述；然后详细说明该知识表示模型的三个主要组成模块：关系编码模块、实体编码模块以及模型整体元学习训练设定与训练流程。此外对于训练集中可见的关系和实体，本文采用了传统的知识图谱表示学习的学习流程，即分别设置关系特征矩阵和实体特征矩阵，在初始化阶段对这两个矩阵按照各自目标维度随机赋值，在模型训练过程中通过 TransE 等打分函数对特征进行打分后进行更新，可以得到可见组件的特征嵌入。本章主要讲解本未见的组件进行编码的过程。该模型的主要创新点如下：（1）关系和实体的表示能够同时学习到实体三元组结构上的拓扑信息和本体层面上的语义信息。（2）能够通过元学习的训练流程模拟出目标问题的学习任务对模型参数进行更新。（3）能够同时对测试集中未见的关系和未见的实体嵌入编码。

#### 3.1 问题定义

一个单独的知识图谱通常定义为  $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$ ，其中  $\mathcal{E}$  指代所有图谱实体的集合， $\mathcal{R}$  指代图谱所有关系的集合而  $\mathcal{T}$  指代所有的实体三元组，而且对于三元组集合  $\mathcal{T} = \{(h, r, t) \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}\}$ ，即所有的三元组的头尾实体和关系均来自于实体集  $\mathcal{E}$  和关系集  $\mathcal{R}$  中。对于知识图谱上的链接预测任务，则是给定一个三元组的头结点和关系  $(h, r, ?)$  或者尾结点和关系  $(?, r, t)$  来预测缺失的实体节点  $e \in \mathcal{E}$ ，使得该缺失的实体节点能构成一个事实三元组  $(h, r, t)$  来完成对原始知识图谱的知识补全。为了评估表示学习模型在链接预测任务上的效果，通常会设置两组三元组数据，一组训练三元组  $\mathcal{T}_{support}$  用于对表示学习模型的参数进行学习，另一组测试三元组  $\mathcal{T}_{query}$  包含了训练集中不存在的事实三元组用于对模型的学习效果进行测试。例如对一个尾结点的预测任务，给定在测试三元组中的一个事实  $(h, r, t) \in \mathcal{T}_{query}$ ，通过模型计算所有可能预测三元组  $\{(h, r, e) | e \in \mathcal{E}, (h, r, e) \notin \mathcal{T}_{support} \cup \mathcal{T}_{query}\}$ ，如果在所有预测三元组中  $(h, r, t)$  的得分越高则说明该表示模型的效果越好。

而在跨设备的训练场景下，本文对传统的知识图谱链接预测任务的定义进行了场景适配。现给定一个基础的训练知识图谱  $\mathcal{G}^{train} = (\mathcal{E}^{train}, \mathcal{R}^{train}, \mathcal{T}^{train})$ ，训练的目标是在训练知识图谱上进行模型参数的学习，从而能够将该模型应用在包含未见组件的新兴测试知识图谱上，即  $\mathcal{G}^{test} = (\mathcal{E}^{test}, \mathcal{R}^{test}, \mathcal{T}_{support}^{test}, \mathcal{T}_{query}^{test})$ ，其中的实体集和关系集遵循  $(\mathcal{E}^{train} \neq \mathcal{E}^{test}, \mathcal{E}^{train} \cap \mathcal{E}^{test} \neq \emptyset)$  及  $(\mathcal{R}^{train} \neq \mathcal{R}^{test}, \mathcal{R}^{train} \cap \mathcal{R}^{test} \neq \emptyset)$ 。其中只用于标记

测试集中实体与关系的结构，不用于对模型的训练。而且在跨设备的设定下，测试数据集和训练数据集无法进行数据合并。

### 3.2 模型整体架构

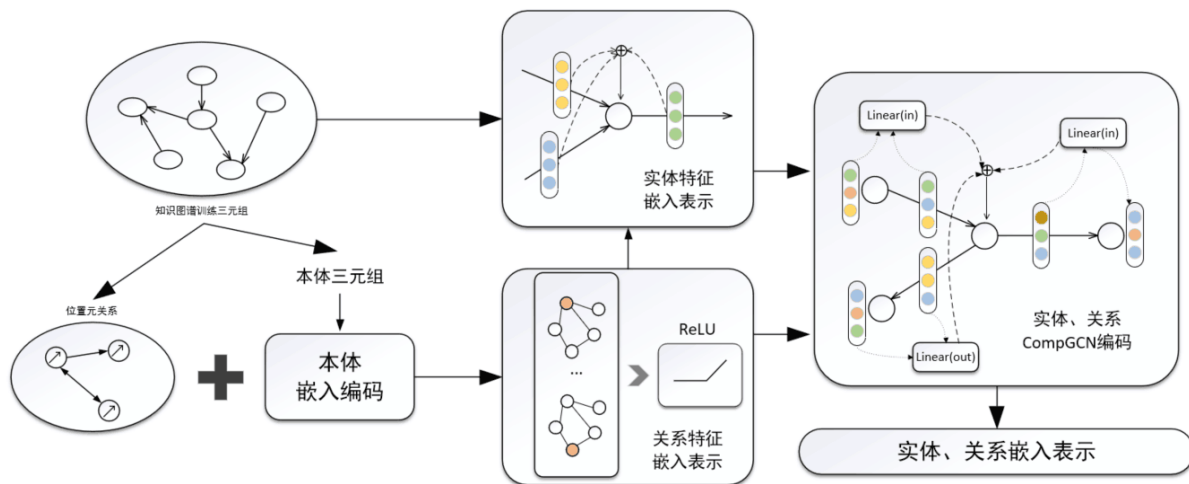


图 3.1 模型架构图

本文提出的模型整体结构图如图3.1所示，总体包含三个大的模块：关系编码模块、实体编码模块、实体关系联合嵌入模块。

为了对关系的特征进行提取，本文模型首先根据实例图谱中的关系的相对位置结构构建出关系位置图，该图中的节点为实例图谱中的关系。然后基于图谱的本体三元组信息对本体的关系及概念节点进行嵌入，为了补充本体的信息，同时采用概念节点的描述文本对本体嵌入进行加强；然后将本体信息和关系位置图进行结合，通过两层 GCN 模型对关系的语义信息和结构信息进行更新和学习，获得关系特征的嵌入表示。对实体的表示上，通过聚合实体邻接关系，通过学习关系方向调整矩阵的参数来对所有关系的特征进行提取获得实体特征的嵌入表示。最后为了加强实体与关系的联系，本文模型在实例图谱上进行两层 CompGCN 的特征提取获得了最终的实体和关系的嵌入，同时修改了 CompGCN 的输出层使得关系和实体的维度不要求一致，从而可以通过多种 KGE 模型作为打分函数计算损失函数对模型调优。

其中对本体嵌入的过程中，本文首先从基础的三元组数据中学习到结构化的嵌入表示，然后从三元组的概念节点的描述信息中使用词嵌入学习到概念节点的描述文本嵌入，为了将描述文本的嵌入补充到本体信息的结构化中去，本文使用一个共享参数的线性层将结构化嵌入表示和描述文本嵌入表示映射到同一个表示空间中。在映射后的线性层中，参照 TransE 的评分思想，本文采用三个距离打分函数将映射后的两种嵌入表示联合更新，学习到兼顾结构信息和描述文本信息的嵌入表示。最后将两种表示拼接作为本体信息进行后续操作。本体嵌入整体结构图如下：

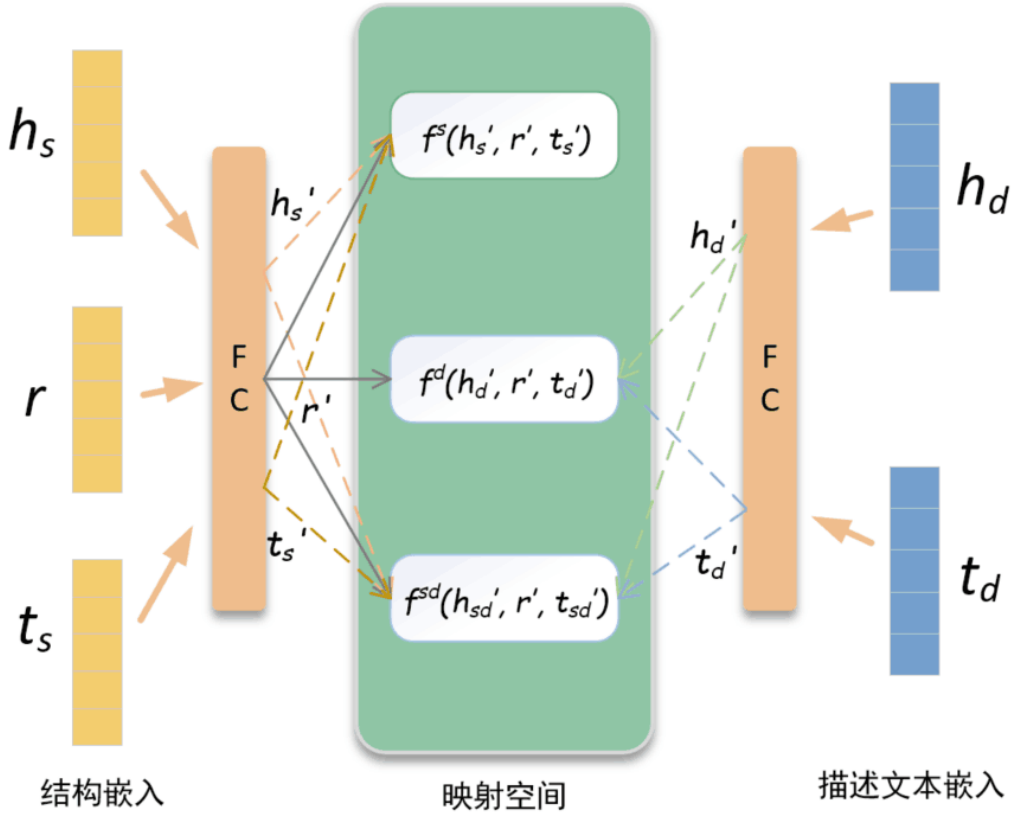


图 3.2 本体嵌入架构

### 3.3 关系嵌入模块

#### 3.3.1 本体信息嵌入

为了能从本体三元组的结构数据中捕捉到能够辅助实体三元组编码，特别是包含未见组件的本体概念信息，本文首先需要将本体三元组转换为响应的语义信息编码。与传统的知识图谱嵌入表示类似，对于一个本体三元组  $(c_i, p, c_j)$ ，本体语义信息编码的目的就是设计出一个打分函数  $f(c_i, p, c_j)$  作为编码模型的激活函数，本文分别采用了传统的知识图谱嵌入方法如 TransE 对本体三元组进行编码，然后将编码后的本体信息用于辅助关系编码。例如按照 TransE 模型的设定，本体三元组的元关系属性是头尾两个实体节点的转移量，将打分函数设置为：

$$f_{TransE}(c_i, p, c_j) = -\|c_i + p + c_j\| \quad (3.1)$$

其中  $c_i, p, c_j$ ，是一个本体三元组响应的概念编码和元关系编码。同时为了提高所有本体三元组的嵌入效果，本文采用自对抗负抽样损失函数来对计算损失更新模型：

$$\mathcal{L}_{\mathcal{O}} = \frac{1}{|\mathcal{T}_{\mathcal{O}}|} \sum_{(c_i, p, c_j) \in \mathcal{T}_{\mathcal{O}}} [\gamma_o + f(c'_i, p, c'_j) + f(c_i, p, c_j)] \quad (3.2)$$

公式中  $\gamma_o$  是控制正负样本得分的参数, 同时  $c'_i, c'_j$  是不存在与本体三元组中的负样本, 为了生成这些负样本, 本文在所有的本体概念中分别遮盖住已存在的本体三元组的头节点和尾节点, 然后从所有本体节点中随机筛选出其他节点组成负样本。本体信息除了结构化的本体三元组之外, 还有许多对本体的详细描述文本, 如本体节点 “companyceo” 的描述文本 “specifies that a particular CEO is the CEO of a particular company”。这些描述文本可以在对本体信息提取的时候提供额外的语义信息, 因此本文通过文本描述进一步加强了对本体三元组的语义嵌入。但是描述文本的建模和一般的三元组建模因为模型的差异一定无法直接进行融合, 因此对于一个特定的本体三元组  $(c_i, p, c_j)$ , 本文首先获得了三元组结构的嵌入  $h_s/r/t_s \in \mathbb{R}^{d_1}$  以及对每一个本体节点描述文本的向量表示  $h_d/t_d \in \mathbb{R}^{d_2}$  来指代文本描述信息, 为了能够让这两个不同层面的嵌入进行融合, 本文引入一个全连接层将两个不同的嵌入同时映射到同一个表示维度上, 映射后的结构嵌入和文本嵌入分别表示为  $h'_s$  和  $h'_d$ , 在统一后的表示空间中使用 TransE 对三元组结构的嵌入进行打分:

$$f^s = - \| h'_s + r' - t'_s \| \quad (3.3)$$

对描述文本的嵌入进行打分:

$$f^d = - \| h'_d + r' - t'_d \| \quad (3.4)$$

同时为了为了使这两种类型的表示相互兼容和互补, 本文遵循 DKRL 模型的设定来定义交叉和相加得分函数:

$$f^{sd} = - \| h'_s + r' - t'_d \| - \| h'_d + r' - t'_s \| \quad (3.5)$$

所有的四个得分函数综合可以保证两个层面的嵌入表示可以在相同空间里进行学习和更新, 最后本体嵌入的得分函数表示为:

$$f^{sd} = - \| h'_s + r' - t'_d \| - \| h'_d + r' - t'_s \| \quad (3.6)$$

因此在本体嵌入的损失函数也响应的转化为:

$$\mathcal{L}_{\mathcal{O}}^{ont} = \frac{1}{|\mathcal{T}_{\mathcal{O}}|} \sum_{(c_i, p, c_j) \in \mathcal{T}_{\mathcal{O}}} [\gamma_o + f'(c'_i, p, c'_j) + f'(c_i, p, c_j)] \quad (3.7)$$

经过训练后每个本体节点都有两个层面的嵌入: 三元组结构嵌入和描述文本嵌入, 然后本文将这两种映射后的嵌入进行拼接作为本体节点的最终的向量表示。上述的描述文本的向量表示, 本文采用了 CNN 网络进行生成。

### 3.3.2 关系图构建

对未见的关系, 由于在训练集中没有相关的三元组可以提供特征学习的支持, 所以必须要从除三元组以外的其他层面获取到可以对未见关系编码的特征表示。而关系作为

构成知识图谱的一种结构信息，关系受相连节点的影响会与其他关系存在联系，这些联系可以作为关系的一种拓扑特征信息，类似的研究也有使用节点的度作为节点特征进行学习的尝试。这些结构性的信息会减少对训练集数据的依赖，在对关系特征进行学习的时候可以通过聚合相连其他关系的信息来对关系进行特征表示，能够较好作用在未见的关系上。

为了对关系的结构特征进行学习，本文继承了 TACT 模型对于关系的拓扑设定，将相邻的两个关系按照关系的方向抽取出四种元关系类型，并将关系视为节点，关系与关系之间由元关系进行连接。

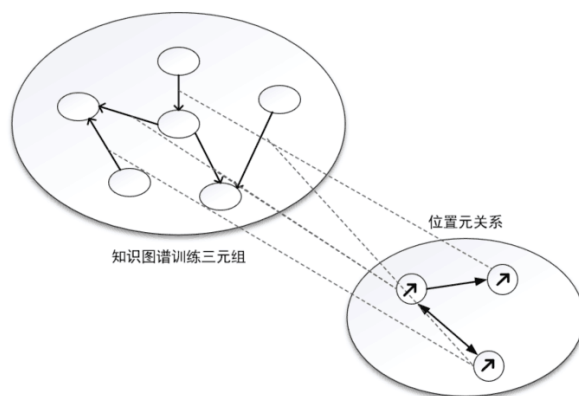


图 3.3 从实体关系构建关系图

这些元关系根据关系的相对关系设置为四类，如图3.4所示，分别为 tail-head、head-tail、tail-tail、head-head。元关系的 head 和 tail 都代表了两个相邻关系的指向，比如 (relation1, tail-head, relation2) 代表同一个实体连接的两个相邻关系 1 和关系 2，且关系 1 指向该实体而关系 2 则从该实体指向其他实体。对于在训练三元组中的两个关系，如果它们符合其中一种的相对未见关系，则在关系位置图中创建一个关系结点，并通过与关系进行连接。

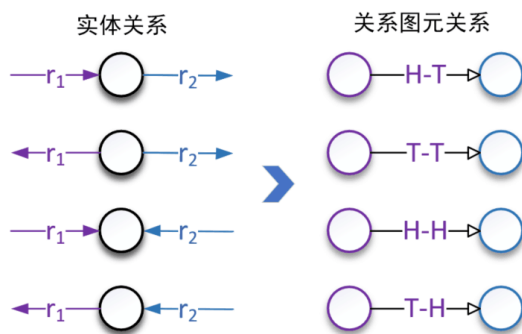


图 3.4 实体关系与关系图元关系的映射

因此根据输入的实体三元组，可以将原始的图结构转变为关系相关图，同时在构建

图的过程中没有使用到任何额外的实体属性或关系属性，可以作用到任何未见或已知的关系上。在关系相关图中的节点代表关系，边表示在原始三元组任意两个关系对应的元关系。

### 3.3.3 关系相关性系数聚合编码

在位置本体图中结点为原三元组中的关系，其位置结构部分的特征可以通过聚合关系结点的其他相邻关系结点的特征获取该关系的特征。为了能够将本体的嵌入作为对关系嵌入的语义补充，在构建好的关系位置图上本文通过对关系节点到本体概念的映射获取到各关系节点的初始化特征表示，然后通过多层 GCN 网络聚合邻接节点的特征来加强对关系节点的特征学习，对于节点的更新公式如下：

$$h_v = f \left( \sum_{(u,r) \in \mathcal{N}(v)} W h_u \right) \quad (3.8)$$

其中  $h_u$  是相邻关系节点的特征，每层 GCN 会有两步的操作，首先左乘归一化后的邻接矩阵对邻居节点的特征进行聚合，然后右乘一个可学习的线性转换矩阵  $W_r$  将输入的特征映射到目标特征空间中，最后使用一个非线性的激活函数来获取本层的特征输出。

为了区分在链接预测任务中不同元关系的的重要性，本文在对不同元关系连接的关系特征进行聚合时设置一个可学习的权重参数，根据任务的表现来学习不同元关系对应的重要程度，计算公式转化为：

$$h_v = f \left( \sum_{(u,r) \in \mathcal{N}(v)} W_{\lambda(r)} h_u \right) \quad (3.9)$$

其中  $w_r$  是两个关系节点相连的元关系类型相对应的参数，根据本文设定的四种不同的元关系，该系数由四个不同的参数控制：

$$W_{dir(r)} = \begin{cases} W_{t-h}, & r \in R_{tail-head} \\ W_{h-t}, & r \in R_{head-tail} \\ W_{t-t}, & r \in R_{tail-tail} \\ W_{h-h}, & r \in R_{head-head} \end{cases} \quad (3.10)$$

## 3.4 实体嵌入模块

对于未见的实体特征提取，本文认为对一个未见的实体可通过它连接的关系来进行预测。如图3.5所示，右侧未见实体 X 的关系结构和 Tom 的关系相似，可以推断出 3 是一个类似 Tom 的一个学生节点，所以本文通过聚合实体所有关系来对实体进行表示。

同时考虑到实体关系的方向性，本文通过以下公式来进行实体特征的聚合：

$$h_e = \frac{1}{|\mathcal{N}(e)|} \sum_{r \in \mathcal{N}(e)} W_{dir(r)}^{ent} h_r \quad (3.11)$$



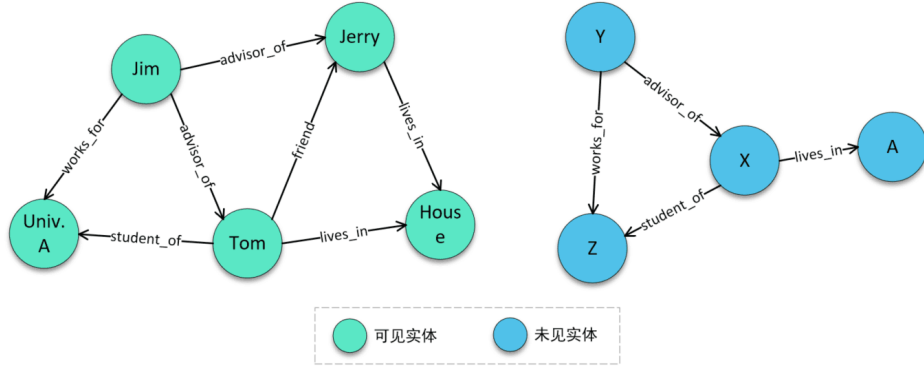


图 3.5 未见实体的关系特征

$\mathcal{N}(e)$  是实体  $e$  所有连接关系的集合； $W_{dir(r)}^{ent}$  是用于将关系特征转化为实体特征且区分方向的一个转换参数， $W_{dir(r)}^{ent}$  在关系  $r$  为指向实体  $e$  时表示为  $W_{in}^{ent}$ ，当关系  $r$  由实体  $e$  指向其他实体时记作  $W_{out}^{ent}$ 。

### 3.5 实体关系联合嵌入模块

为了将关系特征聚合获得的实体特征进行抽取，本文在将输出的特征作为输入放到一个 GNN 网络中进行更新和编码得到学习后的嵌入，该 GNN 网络也以 CompGCN 模型为基础，但是将模型原有的实体关系聚合器修改为了一个线性转化层可以更好适应以多种传统 KGE 方法如 RotatE 来作为解码器进行下游任务，GNN 层的实体嵌入更新如下：

$$m_e^{l+1} = \sum_{(r,t) \in \mathcal{O}(e)} W_{out}^l [h_r^l; h_t^l] + \sum_{(r,h) \in \mathcal{I}(e)} W_{in}^l [h_r^l; h_h^l] \quad (3.12)$$

$\mathcal{O}(e)$  是实体  $e$  的所有指向其他节点的关系和实体的集合， $\mathcal{I}(e)$  是所有指向其他节点  $e$  的关系和实体的集合， $W_{in}^l$  和  $W_{out}^l$  分别是在 GNN 网络的第  $l$  层对实体的入关系和出关系的可学习的参数， $[\cdot]$  而指代两个向量的连接操作。经由一层的 GNN 网络对特征的聚合后，实体的特征转换公式如下：

$$h_e^{l+1} = \sigma \left( \frac{m_e^{l+1}}{|\mathcal{O}(e)| + |\mathcal{I}(e)|} + W_{self}^l h_e^l \right) \quad (3.13)$$

$W_{self}^l$  是针对实体  $e$  本身特征自循环更新的模型学习参数， $\sigma$  指代模型 GNN 模型的激活函数。同时，关系也在该层 GCN 中也进行了更新操作：

$$h_r^{l+1} = \sigma(W_{self}^l h_r^l) \quad (3.14)$$

经过两层 GCN 对关系嵌入和实体嵌入的更新，该 GNN 模块即可输出该训练任务下所有可见组件以及未见组件的知识提取表示，用于本次任务的 loss 的计算及模型参数更新。

### 3.6 元学习训练设定

为了能够对知识图谱中的未见关系和未见实体进行表示学习，借鉴于元学习 *learning to learn* 的思想，本文从训练集中抽取了一系列包含未见实体和关系的训练任务来模拟测试环境，并且在该训练环境中对模型进行训练。

每个训练任务  $S^i = (\mathcal{E}^i, \mathcal{R}^i, \mathcal{T}_{sup}^i, \mathcal{T}_{que}^i)$  包含训练实体集、关系集、support 训练三元组集以及 query 测试三元组集。为了模拟未见的组件，将部分实体和关系标记为未见，每个训练任务被重新定义为如下：

$$S^i = (\mathcal{E}^i = (\hat{\mathcal{E}}^i, \tilde{\mathcal{E}}^i), \mathcal{R}^i = (\hat{\mathcal{R}}^i, \tilde{\mathcal{R}}^i), \mathcal{T}_{sup}^i, \mathcal{T}_{que}^i) \quad (3.15)$$

$\hat{\mathcal{E}}^i \in \mathcal{E}^{train}$  指代实体集中的可见实体而  $\tilde{\mathcal{E}}^i \notin \mathcal{E}^{train}$  指代未见的实体； $\hat{\mathcal{R}}^i \in \mathcal{R}^{train}$  指代关系集中的可见关系  $\tilde{\mathcal{R}}^i \notin \mathcal{R}^{train}$  指代未见的关系。因此所有训练的总目标是基于每个任务的 support 集上进行对未见组件的训练使得在 query 集上的评估得分最高：

$$\max_{\theta} \mathbb{E}_{S^i \sim p(S)} \left[ \sum_{(h,r,t) \in \mathcal{T}_{que}^i} \frac{1}{|\mathcal{T}_{que}^i|} \mathcal{M}_{\theta}(h, r, t | \mathcal{T}_{sup}^i) \right] \quad (3.16)$$

其中  $\mathcal{M}$  可以通过在 support 集上学习到响应的模型参数然后在 query 集上对三语组进行可行性打分，结合前几节模型中学习到的两个分散的 embedding 对三元组进行打分。本文采用自对抗负抽样损失函数来对计算损失更新模型：

$$\begin{aligned} \mathcal{L}(S^i) = & \frac{1}{|\mathcal{T}_{que}^i|} \sum_{(h,r,t) \in \mathcal{T}_{que}^i} -\log \sigma(\gamma + s(h, r, t)) \\ & - \sum_{i=1}^n p(h'_i, r, t'_i) \log \sigma(-\gamma - s(h, r, t)) \end{aligned} \quad (3.17)$$

$n$  指代负采样的数量， $p(h'_i, r, t'_i)$  指代负采样的权重系数， $s(h, r, t)$  是用于计算 embedding 在 query 三元组上的得分函数。

### 3.7 链接预测任务及训练流程

链接预测任务主要是基于知识图谱已有事实来推断出知识图谱中当前不存在的事实三元组，根据待预测节点的不同可区分为头结点预测和尾结点预测，通过比较测试的事实三元组在模型预测的事实三元组排名中出现的概率来评估模型链接预测任务的好坏。本文在进行每次任务损失的过程中，通过正负样本的联合计算出损失，并以此进行梯度下降求解。负样本损失计算公式如下：

$$\mathcal{L}_{neg} = \sum_{i=1}^n (\text{softmax}(W[s_{tail}; s_{head}]) \log \sigma(-[s_{tail}; s_{head}])) \quad (3.18)$$

其中  $s_{tail}$  和  $s_{head}$  是对一个测试三元组对应的尾结点负采样得分和头结点负采样得分,  $n$  代表负采样的数量,  $[\cdot]$  操作将正负样本的得分矩阵进行简单的连接操作,  $W$  是对得分进行微调的权重参数, 本文默认设置为 1。之后和正样本的得分求得算术平均即为该测试三元组的单个损失。

对于模型的训练流程, 本文按照元学习单任务训练方法进行, 每个任务的目标都是使该任务下的连接预测效果最优, 具体流程如下:

- 1) 初始化模型参数, 主要包含了训练数据集中可见实体和关系的嵌入初始化、元关系的权重参数以及两个不同阶段的 GCN 层和 CompGCN 层的传递参数。
- 2) 获得任务的支持集和查询集, 主要从原训练集中抽取单个任务的子图, 并划分为支持集和查询集, 同时对关系和节点的标签进行设置以模拟出未见组件。
- 3) 单次任务训练获得实体关系的嵌入, 通过 KGE 模型计算查询集损失, 通过损失函数进行梯度下降求解。
- 4) 更新模型参数, 重复下一个任务的训练直至效果不再更优。
- 5) 对测试集进行模型测试, 输出任务得分。

### 3.8 本章小结

本章从本文模型的总体架构开始逐步详细介绍了模型的各个组成部分, 包括主要的关系编码模块、实体编码模块以及元学习训练任务的定义和训练流程的说明。在关系编码模块说明了本文模型如何对关系的拓扑结构进行提取以及如何通过 GCN 层对关系的本体信息和拓扑信息进行融合学习, 从而学习到未见关系的有效表示; 在实体编码模块, 通过未见实体的邻接关系的特征聚合进行表示, 并且在实例知识图谱之上再次通过 GNN 来聚合其他关系和实体信息完成最后的特征表示。之后的章节中将对本文模型在测试数据集上进行充分的实验, 检验本文模型的有效性。

## 4 实验结果及分析

本章主要对本文提出的基于元学习本体增强的跨领域知识图谱表示模型(MOKER)进行实验的验证和效果的苹果。主要验证了该模型在面向跨设备情景下的测试数据集上的实验效果,并和目前已经提出的可以处理类似问题的归纳知识图谱表示模型的效果进行对比,结果显示该模型在引入多层次的特征信息后在实验数据上明显优于现有其他模型。最后通过对模型的多个消融实验也证明了模型各个部分的重要性。

### 4.1 数据集

由于传统的知识图谱数据集基本面向“closed-world”下的设定,即测试三元组中的关系和实体都是在训练三元组数据中已知的,不存在跨设备场景下的未见的组件,因此本文基于三个包含本体层面三元组的知识图谱 DBpedia 和 NELL-995 构建了包含未未知组件的新数据集 DB\_Ext 和 NELL\_Ext 对模型的效果进行测试,该三个数据集分别从经典的数据集 DBpedia 和 NELL-995 数据集中抽取部分子集得到。

#### 4.1.1 源数据集介绍

**DBpedia:** DBpedia 数据集是一个基于维基百科的语义知识库,该数据集由一个目前由一个开源社区维护并使用维基百科上的各种文章和各种在线网络资源进行扩展。DBpedia 项目从 wikipedia 中提取结构化的数据,然后通过提取器将这些结构化的属性如图像、标签、描述文本等信息转化为事实三元组并进行存储,最新的数据集快照已经包含了 8.5 亿条事实三元组数据。除了丰富的事实三元组外,DBpedia 强调按照本体论构建知识库,其本体类型数达到了 768 个类别,主要包含人物、地点、工作、组织等本体类型。这些实体、属性和关系通常以 RDF 图的形式进行存储,并以 SPARQL 作为标准查询语言。同时该开源社区提供了网页公布数据集最新的版本和统计数据,同时支持数据集各版本检索和下载,非常便于研究者使用。

**NELL-995:** NELL-995 数据集是从 NELL 数据集中抽取用于知识图谱补全任务的一个基准数据集,NELL(Never Ending Language Learning)是一个由卡内基梅隆大学主导的自动化学习系统,用于从网络中的非结构文本中自动的提取知识。NELL-995 数据集建立在 NELL 知识库上,包含 995 个实体和 129 个关系,涵盖了广泛的领域,如地理、医学、体育、音乐和文化等。NELL-995 数据集具有对实体和关系进行细粒度类型分类的属性。每个实体都被分配到一组多层次的关系类型中,并且与其他实体之间的关系也相应地被分类。因此,NELL-995 数据集是学习细粒度知识表示和关系推理模型的理想数据集。

#### 4.1.2 任务数据集构建

为了抽取出包含未知组件组件的数据子集，本文首先将源数据集所有的三元组构建一个大的实体知识图谱，然后随机选取 100 个根实体节点各延伸 10 个相邻接点构成测试子图并从子图中的关系和实体中抽取 1/10 的关系和实体标记为测试集专属，从训练集中剔除这些专属的三元组从而构建出未见的组件。除此之外，为了专门测试模型对未见关系及未见实体的单独表示性能，在测试集中的 query 集设置中，本文将测试三元组分为三类，分别只包含了一种单独的未见组件及同时包含两种未见组件的 query 三元组部分。最终每个数据集中都包含了两个基本的子知识图谱  $\mathcal{G}^{train}$  和  $\mathcal{G}^{test}$ ，训练知识图谱和测试图谱都单独从标准知识图谱中抽取，其中测试图谱中的三元组包含了训练图谱中不存在的实体和关系。同时将测试图谱的测试三元组分为了三类：1、所有测试三元组中只包含了未见的实体 (unseen\_ent)；2、所有测试三元组中只包含了未见的关系 (unseen\_rel)；3、所有测试三元组中同时包含了未见的实体和关系 (unseen\_both)。该测试数据集中的统计数据由下表所示。其中 DB\_Ext 数据集包含的测试三元组数量仅含未见的实体的三元组 243 个，仅含未见的关系的三元组 10 个，同时包含两个未见组件的三元组 243 个；NELL\_Ext 数据集包含的测试三元组数量仅含未见的实体的三元组 565 个，仅含未见的关系的三元组 12 个，同时包含两个未见组件的三元组 115 个。

#### 4.1.3 数据集本体三元组获取

为了获得知识图谱相关的本体三元组数据，在现存的可找到的本体三元组数据的基础上，本文扩展了类型相关的三元组数据，首先根据所有实体三元组的数据上映射到实体的类型信息，那么对于知识图谱中的所有关系来说，(type, relation, type) 类型三元组中的 type 头结点即为关系在本体层次上的 domain 关系，type 为结点即为关系在本体层次上的 range 关系，同时在进行构建的过程中本文根据 type 元关系出现的频率进行阈值的把控。除了根据 type 构建的本体三元组本文还引入了关系与关系的联系和 type 本身的本体三元组中的 isa、synonym 关系共同设置为 generalizations 关系加入到本体三元组数据中，最终构建出了由 domain、range、generalizations 三种元关系构成的本体三元组数据。其中三元组的统计数量如下表 4.1 所示。

表 4.1 数据集统计数据 (括号中为未见组件数量)

	训练图谱			测试图谱			
	实体数	关系数	三元组数	实体数	关系数	support 三元组数	query 三元组数
NELL_Ext	1583	153	5269	851(753)	140(30)	2160	692
DB_Ext	795	115	1508	913(884)	128(46)	1930	496

## 4.2 实验设计及评价指标

本实验包含本体嵌入表示学习和未见组件表示学习两个阶段，第一个阶段主要学习到融合描述文本信息的本体嵌入表示，第二阶段使用本体信息进行未见组件的表示学习。

第一阶段首先采用传统的表示学习方法对本体三元组数据进行初步嵌入表示；然后对本体节点的描述信息从预训练词嵌入 glove 中获得各描述文本单词的初始化词向量，使用 TF-IDF 统计方法判断单词的重要程度对各单词的词向量进行加权聚合计算获得本体节点描述文本的初始化向量嵌入；最后本体三元组的结构化表示嵌入和本体描述文本的表示嵌入映射到同一个空间中进行评分和更新，获得拼接后的最终的本体嵌入。

第二阶段对未见组件进行表示学习，为了在元学习中模拟出存在未见组件的场景，在每个元学习任务的设置中都人为抹除了一些实体和关系的标签，使得这些实体和关系必须通过的模型学习得到嵌入表示而不是从传统嵌入方法的嵌入矩阵中取得，具体的一个 task 实验流程如下：

- 1) 从训练数据集中随机挑选任意 10 个起始节点，然后根据这 10 个起始节点的关系延伸出邻居节点补充到子图中。
- 2) 每次选取 64 个子图组成一个 batch 进行一个 task 任务的训练。
- 3) 通过模型获得这个 batch 的关系表示和节点表示。
- 4) 计算 batch 单个子项的 loss 和整个 batch 的 loss。
- 5) 根据 loss 进行梯度下降求解最优参数。

对于模型在链接预测任务上的评价指标，本文选取了 MRR 和 Hit@10 作为评判的标准。其中 MRR 通过计算预测三元组排名的倒数来进行计算，即对测试三元组中的所有事实三元组，如果该三元组在预测排名中得分高那么排名就靠前，对应的倒数也会比较大，因此 MRR 作为评价指标越大则证明总体链接预测的排名较好。而 Hits@n 指的是在预测三元组中排名小于 n 的三元组的平均占比，计算公式如下：

$$HITS@n = \frac{1}{|S|} \sum_{i=1}^{|S|} \mathbb{I}(\text{rank}_i \leq n) \quad (4.1)$$

假如 n 设置为 10，那么统计事实三元组在预测三元组中前 n 名的个数，最后再除以总个数就得到了 Hits@10 的结果，其中  $\mathbb{I}(\cdot)$  为 indicator 函数（若条件真则函数值为 1，否则为 0）。

参与比较的模型如下：

**GraIL**：不直接学习实体节点对应的嵌入也没有使用任何节点的属性，而是在测试三元组候选关系的周围构建子图，通过子图的结构和结构化的节点特征来对三元组进行预测，使得模型能够很好的作用在未知的实体三元组预测任务上。

**CoMPILE**：对 GraIL 模型的子图归纳模型进行了改进，加强了对子图关系方向性的

限制，同时在对未知节点特征聚合的信息传递过程中增加了先前模型中忽略的关系特征。

**DURM:** 提出一个可同时学习规则逻辑及其置信度得分方法的可微模型，并且可以使用梯度优化的方法来优化归纳逻辑编程任务，可用于处理含有未知实体的链接预测任务。

**Neural-LP:** 基于知识库构建一个可学习逻辑规则的可微模型，因为逻辑规则是独立于实体和关系的，因此理论上可作用在任何未见的实体上，在归纳图谱补全任务上相比于传统的对实体进行结构信息表示学习的方法有很大提升。

**MaKEr:** 通过对关系结构的特征学习聚合关系邻接关系的特征对关系进行表示，聚合关系特征对实体编码，借用拓扑结构的信息一定程度上实现对未见实体和未见关系表示。

### 4.3 模型参数设置

对于用于对比的 **baseline** 模型的参数设置，本文遵循了相关论文给出的最优超参数设置，对于本文提出的模型的相关参数设置如下表4.2所示：

**表 4.2 模型超参数设置**

本体嵌入参数	设置值	元学习训练参数	设置值	嵌入参数	元学习训练参数
lr	0.00005	lr	0.001	dim	300
ent_str_dim	150	train_bs	64	num_gcn	2
ent_text_dim	300	eval_bs	16	num_comgcn	2
mapping_size	300	num_step	100000	gcn_dim	300
training_epochs	1000	early_stop_patience	20	hid_drop	0.3
batch_size	100	num_sample_for_estimate_size	10	-	-

其中主要包含一下三个方面的参数设置：

- 1) 在本体三元组进行嵌入时模型学习率 **lr**、本体三元组概念节点结构嵌入维度 **ent\_str\_dim**、本体概念节点文本嵌入维度 **ent\_text\_dim**、线性隐藏层维度 **mapping\_size** 以及训练的 **epoch** 数量和 **batch** 的大小。
- 2) 对知识图谱表示元学习训练相关的任务学习率 **lr**、单任务支持集的 **batch** 数量 **train\_bs**、单任务查询集的 **batch** 数量 **eval\_bs**、元训练总任务数 **num\_step**、元训练提前结束无效训练任务计数 **early\_stop\_patience** 以及单个 **batch** 采样的根节点数。
- 3) 对实体和关系嵌入基本维度的设置 **dim**、关系位置图对关系进行 **GCN** 的层数 **num\_gcn**、**GCN** 中间传递维度 **gcn\_dim** 以及 **GCN** 层的丢弃率 **hid\_drop**、对关系和实体进行联合学习的 **ComGCN** 的层数。

在对三元组进行链接预测的评定过程中，本文通过对 **ComGCN** 第二层输出的改

进可支持多种 KGE 模型作为评分函数，实际采用的 KGE 模型包含 TransE、DistMult、ComplEx 及 RotatE，实体和关系的维度根据采用的 KGE 模型在基础嵌入维度上进行调整，具体如下表4.3所示：

表 4.3 评分函数

模型名	实体维度	关系维度	评分函数
TranE	dim	dim	$F = -\ h + r - t\ $
DistMult	dim	dim	$F = h^T \text{diag}(r) t$
ComplEx	$2 * \text{dim}$	$2 * \text{dim}$	$F = \text{Re}(h^T \text{diag}(r) t)$
RotatE	$2 * \text{dim}$	dim	$F = -\ h \circ r - t\ $

其中评分函数中的  $h$ 、 $r$ 、 $t$  分别指代头实体、关系和尾实体的嵌入表示， $\text{Re}$  表示复向量的实部分量， $\circ$  操作表示哈达玛积。

#### 4.4 实验结果及分析

各模型在 NELL\_Ext 上的链接预测任务实验结果如下表4.4所示，按照测试数据集的三元组对未知组件的包含情况，分为了只包含未知实体的结果 ( $u_{\text{ent}}$ )、只包含未知关系的结果 ( $u_{\text{rel}}$ ) 以及同时包含未知实体和未见关系的结果 ( $u_{\text{both}}$ )。表格中加粗的为最优的实验效果，带下划线则是该类基本模型中表现最优的得分，模型括号后面的指代的是在评分阶段采用的 KGE 评分函数。各模型在 DB\_Ext 上的链接预测任务实验结果如下表4.5所示：上面两个表格展示了各模型在 NELL\_Ext 和 DB\_Ext 上的链接预测结果，其中 Hits@10 在计算的过程中本文选择了 50 个候选进行评估，并且根据不同类型的查询三元组 (即  $u_{\text{ent}}$ 、 $u_{\text{rel}}$  和  $u_{\text{both}}$ ) 显示了不同的模型得分。而且对于 GraIL、Neural-LP、DRUM 和 CoMPILE 在对应的论文中只在包含未知实体的数据集上做了实验，因此在对存在未见关系的测试集结果上进行留空处理，但是仍在未见实体的测试集中比较相对应的效果。上述结果均为响应模型运行 4 此后取平均值的结果。

从结果中表明本文提出的模型 MOKER 相比于其他 baseline 模型有所改进且在不同的 KGE 评分模型上基本上有不同程度的提升，而且最优的结果基本上由 MOKER 与 RotatE 模型组合得出，RotatE 模型相比于其他的 KGE 模型采用了更加复杂的关系和实体的映射关系因此能表现更多的没有重叠的特征信息，同时也证明了本文模型的有效性。

在处理未见的实体的测试集上，本文比较了基于规则学习处理未见实体的 Neural-LP 模型和 DRUM 模型以及基于子图推理的 GraIL 模型和 CoMPILE 模型。从结果上可以明显看出基于规则的模型依赖于针对数据集学习出的规则，模型效果受数据集影响较大，且需要大量的数据集或者样本平衡要求十分严格，总体效果没有基于子图的模型



表 4.4 NELL\_Ext 数据集结果

NELL_Ext						
	u_ent		u_rel		u_both	
	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10
Neural-LP	30.48	47.96	-	-	-	-
DRUM	31.82	48.32	-	-	-	-
GraIL	71.62	92.92	-	-	-	-
CoMPILE	<u>75.94</u>	<u>93.62</u>	-	-	-	-
MaKEr(TransE)	70.82	92.00	24.56	54.17	21.53	51.74
MaKEr(DistMult)	70.63	91.33	27.02	60.00	<b>41.39</b>	57.65
MaKEr(ComplEx)	72.24	91.91	18.27	34.17	29.39	59.65
MaKEr(RotatE)	<u>77.09</u>	<u>94.64</u>	<u>31.53</u>	<u>55.00</u>	31.45	<u>62.35</u>
MOKER(TransE)	78.28	94.86	20.72	53.34	27.11	55.85
MOKER(DistMult)	75.98	92.46	19.30	22.50	31.37	55.65
MOKER(ComplEx)	73.61	90.60	24.44	38.33	29.70	54.96
MOKER(RotatE)	<b>79.92</b>	<b>94.73</b>	<b>45.07</b>	<b>75.63</b>	<u>40.33</u>	<b>67.06</b>

效果好，在 NELL\_Ext 数据集上结果差距较大。CoMPILE 模型在 GraIL 模型的基础上强调了关系的重要性，总体效果上要比 GraIL 表现更好。但上述四个模型首先无法处理未见关系，而且基于子图的归纳推理模型强调了测试三元组中头尾实体间的局部子图信息，没有完全利用到实体周围的结构特征信息以及关系信息。MaKEr 模型虽然考虑到了关系对实体的重要性，但是在关系表示的特征学习上效果比本文模型要差，因此通过邻接关系聚合特征表示未知实体的效果在 MRR 的评价指标上比本文模型平均低了 4.42%，在 Hits@10 的表现上本文模型在对应 RotatE 评分函数上也领先了 1.85%，说明了本文模型在未见实体表示上的有效性。

在处理未见关系和同时处理两种未知组件的测试集上，本文比较了通过结构信息对关系进行编码并通过聚合关系信息对未见实体进行编码的 MaKEr 模型，对于未见的关系，本文通过引入额外的本体知识作为关系语义信息的补充，同时通过关系的位置结构利用 GCN 对关系的表示进行更新以学习到未见关系周围的结构拓扑信息。从两个数据集上的链接预测任务结果来看，相比于 MaKEr 模型仅学习结构信息对关系表示，本文引入的本体信息能够有效对关系的表示进行语义补充，在处理未见关系时 RotatE 模型中表现出明显的优势，其 MRR 得分比 MaKEr 的得分高 14.54%，而在 Hits@10 的得分方面，70.63 的得分比 MaKEr 的得分高出 20% 左右。表明本文模型能够更好地捕捉关系的语义和结构信息，在考虑局部和全局关系时都具有不错的性能。在同时含有两种未

表 4.5 DB\_Ext 数据集结果

	DB_Ext					
	u_ent		u_rel		u_both	
	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10
Neural-LP	57.15	73.46	-	-	-	-
DRUM	59.88	73.25	-	-	-	-
GraIL	59.44	<u>80.86</u>	-	-	-	-
CoMPILE	<u>60.66</u>	79.93	-	-	-	-
MaKEr(TransE)	54.4	83.7	31.13	54.00	38.66	66.5
MaKEr(DistMult)	46.24	81.07	16.43	11.00	32.16	56.71
MaKEr(ComplEx)	53.79	82.47	19.95	29.00	36.88	59.26
MaKEr(RotatE)	<u>59.55</u>	<u>86.09</u>	<u>32.93</u>	<u>55.00</u>	<b>41.27</b>	<u>66.54</u>
MOKER(TransE)	64.63	89.60	<b>44.77</b>	70.25	34.92	<b>69.42</b>
MOKER(DistMult)	56.73	80.25	13.68	11.00	33.94	61.74
MOKER(ComplEx)	52.22	77.34	14.40	15.00	31.72	59.49
MOKER(RotatE)	<b>66.43</b>	<b>89.67</b>	41.80	<b>74.00</b>	<u>35.11</u>	63.81

知组件的测试集上，相比于 MaKEr 模型，本文模型在 NELL\_Ext 数据集的 Hits@10 评分上有平均 1.82% 的提升，在 DB\_Ext 数据集的 MRR 和 Hits@10 分别有平均 1.18% 和 0.53% 的提升。

除此之外，本文实验发现，TransE 和 RotatE 的模型效果比 DistMult 和 ComplEx 模型更好，尤其在关系处理方面。在本体嵌入实验中，采用 RotatE 作为评分函数，对实体知识图谱进行表示。DistMult 和 ComplEx 模型因其复杂性不适合与 RotatE 的本体嵌入配合，而效果反而下降。相比之下，TransE 模型简洁易操作，在进行特征提取时更加适合。而 RotatE 作为近几年提出的 KGE 模型，一方面本身与本体嵌入方法相同，另一方面个更全面的对实体和关系进行低维嵌入表示，因此取得了本文模型中的最优效果。

## 4.5 模型消融实验

本节将介绍针对模型的几个重要模块的多项消融实验，以展示本文模型的不同部分的重要性，主要设置了 5 项不同的消融设置实验：(1) 去除元学习的设置；(2) 去除本体的设置；(3) 去除实体聚合表示的设置；(4) 去除关系 GCN 聚合的设置；(5) 同时去除本体和元学习的设置；获得的实验结果如表 4.6 所示：

其中个消融实验的具体设置如下：

- 1) 去除元学习的设置：在模型训练阶段，本文采用了元学习的设置在训练集上提取

表 4.6 在 NELL\_Ext 上的消融实验结果

NELL_Ext						
	u_ent		u_rel		u_both	
	MRR	Hits@10	MRR	Hits@10	MRR	Hits@10
MOKER(TransE)	<b>78.28</b>	<b>94.86</b>	20.72	53.34	<b>27.11</b>	<b>55.85</b>
no_meta_TransE	29.31	43.82	12.65	30.83	10.18	22.26
no_ont_TransE	76.89	94.07	14.96	27.50	14.06	28.78
no_ent_TransE	76.85	94.07	18.00	55.83	19.43	50.7
no_gcn_TransE	74.68	92.27	<b>23.88</b>	<b>61.67</b>	19.85	46.26
no_meta_ont_TransE	26.57	38.69	11.15	25.00	12.62	27.74

任务子图并通过随机标签来模拟出未见的组件从而训练出模型处理未见组件的能力。在该消融实验下，取消掉了对子图标签的模拟效果即按照传统 KGE 模型的训练方法，在训练集上对所有三元组进行嵌入和损失计算。

- 2) 去除本体的设置：本文模型的一个创新点即在对关系嵌入表示的时候通过关系位置图加入本体信息进行学习，在该消融实验设置下，将本体信息替换为随机表示对未见关系进行特征学习。
- 3) 去除实体聚合表示的设置：本文模型在处理未见实体时，采用该实体周围的关系信息聚合表示，该消融实验设置下对未见实体进行随机化嵌入设置，考察未见实体的表示模块。
- 4) 去除关系 GCN 聚合的设置：在构建完关系的位置图并引入本体嵌入后，本非模型通过 GCN 来加强关系对邻接结构的特征学习，该消融实验下去除两层 GCN 层，直接使用本体信息。
- 5) 同时去除本体和元学习设置：同时结合消融实验 (1) 和 (2) 的设置。

消融实验结果与 MOKER(TransE) 模型在 NELL\_Ext 上的实验结果进行比较，从消融实验结果可以看出，基本所有的消融设置都会导致性能下降，表明这些设计的重要性。但是在去除关系 GCN 聚合的设置上，在仅含有未见关系的测试集上效果反而提升，分析可知该层 GCN 的作用是在关系本体嵌入的基础上学习关系邻接关系的结构信息，这些结构信息的引入一定程度上会影响仅对关系的表示效果导致在仅包含未见关系的测试集上的效果下降。但是这些结构信息的引入在中未见实体的表示中，因为实体需要聚合邻接关系进行表示，因此引入的结构信息会对实体的表示效果进行提升，可以发现去掉 GCN 在包含未见实体和同时包含未见实体和未见关系的测试集上效果都有明显的下降，因此也侧面证明了 GCN 模型对实验效果提升的必要性。此外，本文观察到元学习设置对模型性能至关重要，表明在推广到测试知识图谱的任务上对模型进行元训练的有

效性。

## 4.6 未见实体案例分析

在图4.1中展示了本体提出的 MOKER 模型和去除元学习和本体嵌入的传统 TransE-KGE 模型产生的在 NELL-Ext 数据集的测试集上实体嵌入的可视化。在这张图中，不同颜色展示了不同类型的实体，圆点代表未见的实体，叉号则代表了该类型下的已知实体。可以看出 MOKER 产生的嵌入分布与对应类型更加一致，在嵌入映射的距离上更近紧凑，而 TransE-KGE 产生的嵌入则混合了不同实体类型。MOKER 将嵌入映射到了不同的聚类中，而 TransE-KGE 中不同实体类型的嵌入则混合在一起。此外，本文的模型可以将未见过的实体的嵌入与同一类型的已知实体聚类。不同类型实体的聚类表明，MOKER 能够用包含合理语义和信息性知识的嵌入来表示未见过的实体。

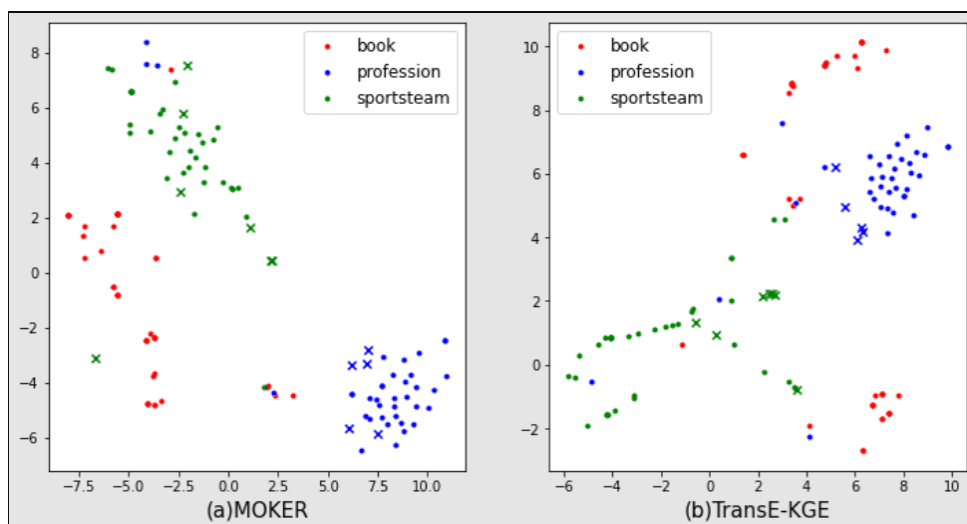


图 4.1 实体嵌入可视化分析

## 4.7 未见关系案例分析

对于未见关系，本文选取了部分关系同样映射到了二维空间上进行可视化分析，如图4.2所示，其中圆点代表通过训练集中可见的关系，叉号代表测试集中的未见关系。本文模型通过采用关系位置图根据关系的相对位置对关系局部的邻接关系进行建模学习到了关系的拓扑结构信息，同时引入本体信息作为语义补充，因此同一实体的具有相近语义的相邻关系在距离上应该表现为更加相近，从图上可以看出，对于未见关系 `has_office_in_city` 贴近于具有类似语义的关系 `has_office_in_coutry`，未见关系 `person_born_in_city` 更贴近于已知关系 `person_born_in_location`。而对于已知关系 `parent_of_person` 和 `father_of_person` 和未见关系 `mother_of_person`，本文可知一个实体如果存在 `parent_of_person` 的关系那么该实体节点的邻接关系中只能存在其中一个 `father` 或者 `mother` 的关系，因此在图中本文可以观察到 `mother_of_person` 在距离上更接近于 `par-`

ent\_of\_person 关系，而远离 father\_of\_person 关系。由此可见，本文模型通过关系的位置图上联合本体语义信息有效的学习到了对应关系的语义关系，且其中相似的关系在向量空间中靠近，显示了本文提出的 MOKER 在嵌入未知关系方面的有效性。

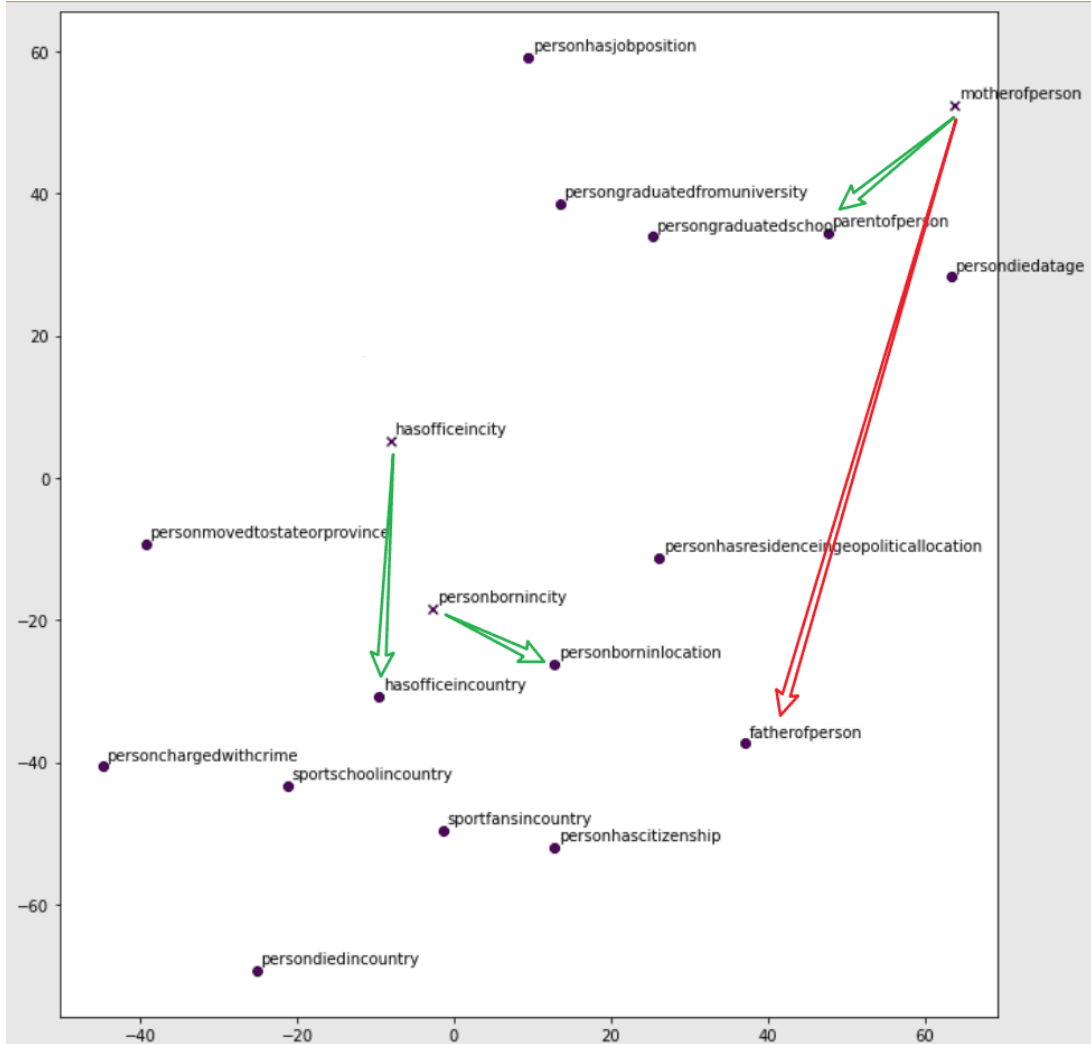


图 4.2 实体嵌入可视化分析

## 4.8 本章小结

本章将本文提出的模型在测试数据集上 NELL\_Ext 和 DB\_Ext 上进行了链接预测任务的相关实验，测试数据集从 NELL-995 和 DBpedia 上抽取子集并添加了测试集中的未见组件。与多个基准模型相比，本文提出的模型在任务得分上均有不同程度的提升，并通过对实验结果的分析验证了模型对于表示学习效果增强的有效性。同时通过对各个模型组件的消融实验发现明显的效果下降，证明了模型模块的重要性；最后对未见实体和未见关系的案例分析可知得到的嵌入表示符合模型原理设定，再次证明了该模型对未见组件表示上的突出能力。

## 5 总结与展望

### 5.1 总结

当下知识图谱技术的应用日益渗透到每个人的日常生活中，从搜索引擎到推荐系统，不断探索知识图谱知识的应用与补充是未来发展不可或缺的一个研究方向。但是出于数据隐私、成本等多方面的考量，大到公司多个图谱存储服务器，小到人们每日使用的移动终端，我们无法将所有分散知识图谱添加的实体和关系完全覆盖，如何在训练集训练能够快速应用于包含未见组件的新兴知识图谱的研究是无可避免的需求。

本文采用元学习的方法在训练集中模拟了新兴知识图谱中可能出现的未见关系和未见实体，使得模型能够在训练中获得处理未见组件的能力，同时元学习的方法能够一次训练快速应用，在算力珍贵的时代能够最大提高训练的效率，降低成本消耗。对未见关系的表示能力学习上，本文通过对关系相对位置关系的应用构建出以关系为结点的图，图上的关系之间通过预定义的元关系进行连接，体现出关系的拓扑信息；知识图谱作为丰富语义信息的载体，为了对拓扑信息进行补充，在关系图的基础上，本文通过对知识图谱高层语义体现的本体信息进行嵌入作为关系语义信息的体现，然后在关系图上使用 GCN 对关系节点的嵌入进行拓扑信息和语义信息的学习，获得对未见关系的表达。对未见实体的表示能力学习上，本文通过对图谱的结构分析认为，相似类型的实体在邻接结构上往往表现出高度的相似性，因此采用实体的邻接关系特征聚合作为未见实体的初始化表示，避免了对未见实体周围组件的苛刻性要求。但关系的表示仍然不足以支撑对实体特征的体现，模型最后在实例知识图谱上，采用 CompGCN 对每一个实体和关系再次进行邻接信息的学习和更新，获得最终的各组件的嵌入。本文最后通过两个数据集上的实验和多个基准模型的比较和链接预测任务得分以及案例分析证明了本文模型的有效性。本文的工作总结如下：

- 1) 在对知识图谱组件的表示学习中引入了本体信息作为知识图谱的语义信息补充，通过对拓扑信息和语义信息的结合能够有效对未见组件进行表示学习，同时采用 GNN 层对初步表示学习的效果进行加强，有效学习到了各组件的特征信息。
- 2) 采用元学习的训练方式对模型进行训练，在元学习的单任务设定中对未见组件进行模拟，使得模型能够学习到处理未见组件的能力同时提高训练的效率降低成本。
- 3) 在两个知识图谱链接预测基准数据集的基础上构建满足跨设备场景要求下的测试数据集，并与多个基准模型进行实验效果比较，通过对结果分析可知本文模型相比于其他基准模型均有不同程度的提升，证明了本文模型的有效性。

## 5.2 未来工作

虽然本文在测试数据集上的效果相比于其他基准模型已经获得了明显的提升，但是在本文的整个研究过程中仍旧发现了当下模型在应用方面值得继续探究的几个研究方向：

- 1) 本体嵌入的评分函数对模型效果的影响：在对本体的嵌入过程中，本文采用了目前传统 KGE 方法中效果最好的 RotatE 模型作为评分函数，但是从实验结果中我们可以发现对于 DistMult 和 ComplEx 评分函数的得分会比更简单的 TransE 更低，那么不同的本体嵌入方法可能对模型的效果会产生影响，是否本体嵌入评分函数和模型评分函数相对应会更获得更好的模型效果值得探究。
- 2) 数据集本体信息的提取：本文模型研究本体信息对知识图谱表示学习的可用性，本体三元组是必不可少的一环，但当下的一些基准知识图谱有些并没有相应的本体三元组，如被广泛使用的 FB15K-237 数据集，因为其源数据集已经停止维护本体类型信息也都比较杂乱无法使用，虽然本文在处理测试数据集时采用了对实例三元组类型信息补充了本体三元组中的关系信息，但面对一个数据集如何更充分获取到其本体信息仍旧是待解决的难题。
- 3) 规则信息的引入：从 DB\_Ext 的基准模型测试信息可以看出基于规则提取的模型在未见的实体上也能够不错的表现效果，而且规则理论上可以同时作用于实体和关系，是对未见组件表示学习进行约束的良好工具，如何将规则信息融入到本文模型更好的学习到未见组件的表示也是未来工作的一个值得期待的方向。

## 参考文献

- [1] ZOU X. A survey on application of knowledge graph[A]. Journal of Physics: Conference Series : Vol 1487[C], 2020 : 012016.
- [2] BERNERS-LEE T, HENDLER J, LASSILA O. The semantic web[J]. Scientific american, 2001, 284(5) : 34 – 43.
- [3] 钱育蓉张正航 □ □. 基于 TransE 的表示学习方法研究综述 [J]. 计算机应用研究, 2021, 38(656-663).
- [4] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[J]. Advances in neural information processing systems, 2013, 26.
- [5] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes[A]. Proceedings of the AAAI conference on artificial intelligence : Vol 28[C], 2014.
- [6] YANG B, YIH W-T, HE X, et al. Embedding entities and relations for learning and inference in knowledge bases[J]. arXiv preprint arXiv:1412.6575, 2014.
- [7] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[A]. International conference on machine learning[C], 2016 : 2071 – 2080.
- [8] SUN Z, DENG Z-H, NIE J-Y, et al. Rotate: Knowledge graph embedding by relational rotation in complex space[J]. arXiv preprint arXiv:1902.10197, 2019.
- [9] LI R, CAO Y, ZHU Q, et al. How does knowledge graph embedding extrapolate to unseen data: a semantic evidence view[A]. Proceedings of the AAAI Conference on Artificial Intelligence : Vol 36[C], 2022 : 5781 – 5791.
- [10] TERU K, DENIS E, HAMILTON W. Inductive relation prediction by subgraph reasoning[A]. International Conference on Machine Learning[C], 2020 : 9448 – 9457.
- [11] DAI Y, WANG S, XIONG N N, et al. A survey on knowledge graph embedding: Approaches, applications and benchmarks[J]. Electronics, 2020, 9(5) : 750.
- [12] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[J]. Advances in neural information processing systems, 2013, 26.
- [13] NICKEL M, TRESP V, KRIEGEL H-P, et al. A three-way model for collective learning on multi-relational data.[A]. Icml : Vol 11[C], 2011 : 3104482 – 3104584.
- [14] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph



- completion[A]. Proceedings of the AAAI conference on artificial intelligence : Vol 29[C], 2015.
- [15] JI G, HE S, XU L, et al. Knowledge graph embedding via dynamic mapping matrix[A]. 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)[C], 2015 : 687 – 696.
- [16] KOLDA T G, BADER B W. Tensor decompositions and applications[J]. SIAM review, 2009, 51(3) : 455 – 500.
- [17] CHOUDHARY S, LUTHRA T, MITTAL A, et al. A survey of knowledge graph embedding and their applications[J]. arXiv preprint arXiv:2107.07842, 2021.
- [18] GARCÍA-DURÁN A, BORDES A, USUNIER N. Effective blending of two and three-way interactions for modeling multi-relational data[A]. Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part I 14[C], 2014 : 434 – 449.
- [19] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[J]. Advances in neural information processing systems, 2013, 26.
- [20] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2d knowledge graph embeddings[A]. Proceedings of the AAAI conference on artificial intelligence : Vol 32[C], 2018.
- [21] 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 preprint arXiv:1712.02121, 2017.
- [22] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint arXiv:1609.02907, 2016.
- [23] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[A]. The Semantic Web: 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, June 3–7, 2018, Proceedings 15[C], 2018 : 593 – 607.
- [24] VASHISHTH S, SANYAL S, NITIN V, et al. Composition-based multi-relational graph convolutional networks[J]. arXiv preprint arXiv:1911.03082, 2019.
- [25] XIE R, LIU Z, JIA J, et al. Representation learning of knowledge graphs with entity descriptions[A]. Proceedings of the AAAI Conference on Artificial Intelligence : Vol 30[C], 2016.
- [26] VALVERDE-REBAZA J C, de ANDRADE LOPES A. Link prediction in complex networks based on cluster information[A]. Advances in Artificial Intelligence-SBIA 2012:

- 21th Brazilian Symposium on Artificial Intelligence, Curitiba, Brazil, October 20-25, 2012. Proceedings[C], 2012 : 92 – 101.
- [27] LIN Y, LIU Z, LUAN H, et al. Modeling relation paths for representation learning of knowledge bases[J]. arXiv preprint arXiv:1506.00379, 2015.
- [28] FENG J, HUANG M, YANG Y, et al. GAKE: Graph aware knowledge embedding[A]. Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers[C], 2016 : 641 – 651.
- [29] XIE R, LIU Z, LUAN H, et al. Image-embodied knowledge representation learning[J]. arXiv preprint arXiv:1609.07028, 2016.
- [30] XIONG W, HOANG T, WANG W Y. Deeppath: A reinforcement learning method for knowledge graph reasoning[J]. arXiv preprint arXiv:1707.06690, 2017.
- [31] AUER S, BIZER C, KOBILAROV G, et al. Dbpedia: A nucleus for a web of open data[A]. The Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007+ ASWC 2007, Busan, Korea, November 11-15, 2007. Proceedings[C], 2007 : 722 – 735.
- [32] GUO S, WANG Q, WANG B, et al. SSE: Semantically smooth embedding for knowledge graphs[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 29(4) : 884 – 897.
- [33] HAO J, CHEN M, YU W, et al. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts[A]. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining[C], 2019 : 1709 – 1719.
- [34] XIE R, LIU Z, SUN M, et al. Representation learning of knowledge graphs with hierarchical types.[A]. IJCAI : Vol 2016[C], 2016 : 2965 – 2971.
- [35] SHI B, WENINGER T. Open-world knowledge graph completion[A]. Proceedings of the AAAI conference on artificial intelligence : Vol 32[C], 2018.
- [36] GALÁRRAGA L A, TEFLIOUDI C, HOSE K, et al. AMIE: association rule mining under incomplete evidence in ontological knowledge bases[A]. Proceedings of the 22nd international conference on World Wide Web[C], 2013 : 413 – 422.
- [37] YANG F, YANG Z, COHEN W W. Differentiable learning of logical rules for knowledge base reasoning[J]. Advances in neural information processing systems, 2017, 30.
- [38] SADEGHIAN A, ARMANDPOUR M, DING P, et al. Drum: End-to-end differentiable rule mining on knowledge graphs[J]. Advances in Neural Information Processing Systems, 2019, 32.
- [39] CHEN J, HE H, WU F, et al. Topology-aware correlations between relations for inductive

- link prediction in knowledge graphs[A]. Proceedings of the AAAI Conference on Artificial Intelligence : Vol 35[C], 2021 : 6271 – 6278.
- [40] LIU S, GRAU B, HORROCKS I, et al. Indigo: Gnn-based inductive knowledge graph completion using pair-wise encoding[J]. Advances in Neural Information Processing Systems, 2021, 34 : 2034 – 2045.
- [41] STACKELBERG H V, OTHERS. Theory of the market economy[J], 1952.
- [42] FRANCESCHI L, FRASCONI P, SALZO S, et al. Bilevel programming for hyperparameter optimization and meta-learning[A]. International Conference on Machine Learning[C], 2018 : 1568 – 1577.
- [43] SINHA A, MALO P, DEB K. A review on bilevel optimization: From classical to evolutionary approaches and applications[J]. IEEE Transactions on Evolutionary Computation, 2017, 22(2) : 276 – 295.
- [44] FINN C, ABBEEL P, LEVINE S. Model-agnostic meta-learning for fast adaptation of deep networks[A]. International conference on machine learning[C], 2017 : 1126 – 1135.
- [45] MARCUS G. Deep learning: A critical appraisal[J]. arXiv preprint arXiv:1801.00631, 2018.
- [46] THRUN S, PRATT L. Learning to learn: Introduction and overview[J]. Learning to learn, 1998 : 3 – 17.
- [47] ANDRYCHOWICZ M, DENIL M, GOMEZ S, et al. Learning to learn by gradient descent by gradient descent[J]. Advances in neural information processing systems, 2016, 29.
- [48] LIU H, SIMONYAN K, YANG Y. Darts: Differentiable architecture search[J]. arXiv preprint arXiv:1806.09055, 2018.
- [49] ZHOU W, LI Y, YANG Y, et al. Online meta-critic learning for off-policy actor-critic methods[J]. Advances in neural information processing systems, 2020, 33 : 17662 – 17673.
- [50] SNELL J, SWERSKY K, ZEMEL R. Prototypical networks for few-shot learning[J]. Advances in neural information processing systems, 2017, 30.
- [51] HOUTHOOFT R, CHEN Y, ISOLA P, et al. Evolved policy gradients[J]. Advances in Neural Information Processing Systems, 2018, 31.
- [52] REAL E, AGGARWAL A, HUANG Y, et al. Regularized evolution for image classifier architecture search[A]. Proceedings of the aaai conference on artificial intelligence : Vol 33[C], 2019 : 4780 – 4789.
- [53] XIONG W, YU M, CHANG S, et al. One-shot relational learning for knowledge graphs[J]. arXiv preprint arXiv:1808.09040, 2018.

## 致谢

以简短的文字表达作者对完成论文和学业提供帮助的老师、同学、领导、同事及亲属的感激之情。

## 攻硕期间取得的学术成果和参与的项目

- [1] 在这里列出你攻博期间发表的科研成果
- [2] 用 \item 分隔就好

## 武汉大学学位论文使用授权协议书

本学位论文作者愿意遵守武汉大学关于保存、使用学位论文的管理办法及规定，即：学校有权保存学位论文的印刷本和电子版，并提供文献检索与阅览服务；学校可以采用影印、缩印、数字化或其它复制手段保存论文；在以教学与科研服务为目的前提下，学校可以在校园网内公布部分或全部内容。

- 1、 在本论文提交当年，同意在校园网内以及中国高等教育文献保障系统（CALIS）、高校学位论文系统提供查询及前十六页浏览服务。
- 2、 在本论文提交 ☐ 当年/ ☐ 一年/ ☐ 两年/ ☐ 三年以后，同意在校园网内允许读者在线浏览并下载全文，学校可以为存在馆际合作关系的兄弟高校用户提供文献传递服务和交换服务。（保密论文解密后遵守此规定）

论文作者（签名）： \_\_\_\_\_

学 号： \_\_\_\_\_

学 院： \_\_\_\_\_

日期：            年    月    日