

语义区分和子图整合驱动的知识图谱表示学习方法

张杰勇¹, 尚斌¹, 杨易春², 刘彬¹, 刘东洋¹, 王鹏¹, 刘均³

(1. 空军工程大学信息与导航学院, 710077, 西安; 2. 西安市环境监测站, 710018, 西安;
3. 西安交通大学计算机科学与技术学院, 710049, 西安)

摘要: 针对现有基于图神经网络的知识图谱表示学习方法难以区分实体的多样化语义, 以及如何在嵌入学习过程中对所有关系进行均质化处理的问题, 提出了一种新颖的基于语义区分和子图整合知识图谱表示学习方法, 通过层次化建模, 分阶段挖掘图谱结构信息以学习实体表示。语义区分模块通过构建多个关系感知子图并整合其语义信息, 从而全面捕获实体属性。子图整合模块则通过评估不同子图的重要性, 实施选择性邻域聚合以保留具有区分度的语义特征。此外, 采用编码器对知识图谱三元组进行建模并执行链接预测任务, 并引入交叉熵损失函数训练模型, 以充分整合语义与结构信息。实验结果表明: 所提方法在 3 个标准数据集上相较于现有最优模型实现了全面性能提升。在 FB15K-237 数据集上, 平均倒数排名(M_{RR})、命中率 Hits@3 和 Hits@10 (正确答案出现在预测排名前 3、前 10 位的比例) 分别提升了 2.42%、3.90% 和 1.44%; 在 YAGO3-10 数据集上, M_{RR} 、Hits@3 和 Hits@10 分别提升了 2.17%、4.81% 和 2.01%; 在 WN18RR 数据集上, M_{RR} 、Hits@3 和 Hits@10 分别提升了 0.61%、0.59% 和 0.86%, 有效验证了所提方法的优越性。

关键词: 知识图谱; 表示学习; 图神经网络; 链接预测

中图分类号: TP391 **文献标志码:** A

DOI: 10.7652/xjtuxb202607018 **文章编号:** 0253-987X(2026)07-0196-11

Knowledge Graph Representation Learning Driven by Semantic Discrimination and Subgraph Integration

ZHANG Jieyong¹, SHANG Bin¹, YANG Yichun², LIU Bin¹,
LIU Dongyang¹, WANG Peng¹, LIU Jun³

(1. Information and Navigation School, Air Force Engineering University, Xi'an 710077, China;
2. Xi'an Environmental Monitoring Station, Xi'an 710018, China;
3. School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: To address the limitations of existing graph neural network-based knowledge graph representation learning methods in distinguishing diverse entity semantics and handling relations homogeneously during embedding learning, a novel knowledge graph representation learning method based on semantic discrimination and subgraph integration is proposed. Hierarchical modeling is employed to learn entity representations by progressively mining structural information from the graph. In the semantic discrimination module, multiple relation-aware subgraphs are constructed and

their semantic information is integrated to capture comprehensive entity attributes. In the subgraph integration module, the importance of different subgraphs is evaluated, and selective neighborhood aggregation is performed to preserve discriminative semantic features. Additionally, knowledge graph triples are modeled via an encoder for link prediction tasks, and a cross-entropy loss function is introduced during model training to effectively integrate semantic and structural information. Experimental results on three standard datasets demonstrate that the proposed method yields comprehensive performance improvements over current state-of-the-art models. Specifically, on the FB15K-237 dataset, improvements of 2.42%, 3.90%, and 1.44% were achieved in mean reciprocal rank (M_{RR}), Hits@3, and Hits@10 (denoting the proportion of queries with the correct answer in the topk predictions ($k=3, 10$)) respectively. On the YAGO3-10 dataset, the M_{RR} , Hits@3, and Hits@10 were improved by 2.17%, 4.81%, and 2.01%, respectively. On the WN18RR dataset, increases of 0.61%, 0.59%, and 0.86% were observed in the corresponding metrics. The superiority of the proposed method is thus validated.

Keywords: knowledge graph; representation learning; graph neural networks; link prediction

知识图谱(KG)^[1]是一种有向图结构数据库,以关系为边、实体为节点^[2],其通过三元组(头实体、关系、尾实体)表示现实世界中的事实^[3-4]。知识图谱能够存储人类的结构化知识,已在多个下游任务场景中得到应用,如问答系统^[5]、信息检索^[6]及推荐系统^[7]等。尽管 Yago3^[8]、Freebase^[9]等众多公开知识图谱的规模已十分庞大,但仍存在不完备性问题,核心原因是知识图谱中存在大量缺失的三元组^[10-12]。例如,在 Freebase 数据库中,94%的人物实体缺少已知父母关系,75%的人物实体无明确国籍信息。这些缺失的信息会对下游任务产生显著影响^[13],因此自动挖掘缺失三元组(又称链接预测)^[14-19]近年来已成为研究热点。

知识图谱表示学习(亦称知识图谱嵌入方法)是一种针对知识图谱链接预测任务的有效解决方案。该类方法通过学习实体与关系的连续向量表示,并基于所学特征与评分函数预测图谱中缺失的三元组。近年来,研究者提出了多种基于图神经网络的知识图谱表示学习模型:SEPA 模型引入球面几何框架并采用注意力机制聚焦查询表示^[20];MR-MKG 模型采用关系图注意力网络对多模态知识图谱进行编码,并设计了一个跨模态对齐模块来优化图像-文本对齐^[21];APKGC 模型通过专门设计的注意力惩罚模块调整语言模型中的注意力得分,并提出了一种自适应噪声采样模块来补充实体的多模态信息^[22];HyperFM 模型通过定制的超图转换器,以事实为中心的方式有效捕捉了不同数据模态之间的复杂交互,同时适应知识图谱的超关系结构^[23];MGTCa 模型提出了一种混合几何消息函数以增强邻域信息传递能力,并采用可

训练的卷积注意力网络来避免对知识图谱局部结构进行预验证的需求^[24];Knowformer 模型通过将关系组合信息显式注入实体表示,有效融合语义特征并区分实体在不同语境中的语义角色^[25];UniGE 模型提出基于最优传输理论与 Wasserstein 重心方法的嵌入对齐机制,并结合特征融合策略提升表示学习效果^[26];SG-GCN 模型通过引入枢轴节点强化高阶实体间的消息传递,并采用自适应消息组合机制实现邻域表示的高效聚合^[27];TGformer 模型利用上下文级子图关联共现实体形成的三元组,通过三元组级别与图级别双重架构学习嵌入表示^[14]。

尽管现有基于图神经网络的知识图谱表示学习方法已取得显著成效,但仍存在若干明显局限。①语义冲突问题。由于关系类型具有多样性,实体在不同三元组中往往承载不同的语义属性,这种多重语义可能引发语义冲突,即同一实体通过与不同关系相连会产生多种语义角色。若忽略此类语义冲突,模型将难以有效捕捉局部邻域中的潜在语义信息。②关系异质性问题。同质化处理不同类型的关系会导致特定语义特征在聚合过程中被弱化。这是由于不同关系具有异质特性,实体的邻域可根据关系类型划分为多个子图。不同子图对中心实体具有不同重要性,如“从事”关系子图的重要性显著高于“前往”关系。因此,深入探究不同子图的贡献度,将有助于实现邻域信息的选择性聚合。

基于上述分析,本文提出一种新颖的知识图谱表示学习模型 SDSI。该模型通过语义区分(SD)与子图整合(SI)两个核心模块,生成蕴含丰富局部结构信息的实体特征表示。其中,语义区分模块首先

将邻域实体按关系类型划分为多个子图,进而基于不同关系类型进行特征聚合;子图整合模块则通过评估各关系对中心实体的重要程度,深入挖掘不同关系类型的贡献差异,最终融合邻域信息与自身特征生成实体嵌入表示。通过上述模块的协同作用,模型能够有效学习具有区分度的实体特征,并最终将其应用于链接预测任务。

本研究提出的 SDSI 模型与现有基于注意力的知识图谱表示学习范式存在明确的延续性与发展关系。SDSI 模型是一个两阶段注意力驱动架构:第一阶段(SD 模块)在每个关系子图内使用注意力进行语义一致的特征聚合;第二阶段(SI 模块)用于评估并整合不同关系子图对中心实体的整体贡献。这种双层架构是对现有层级注意力思想的一种特定且深入的实例化与结构化,其创新点在于将“关系异质性”问题明确转化为一个可学习的子图重要性选择问题。因此,SDSI 模型通过“先按关系划分语义子图并内部聚合,再动态整合异质子图”的模块化设计,扩展了注意力机制在知识图谱中的应用粒度与深度。该架构创新使模型能够更精细地模拟知识图谱中复杂的局部结构,实现对现有注意力范式的自然延续与实质性推进。

1 问题定义

知识图谱(记为 $\mathcal{G}$)是一种包含多种关系的有向图,其形式化定义为 $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{T})$ 。其中,

$\mathcal{E} = \{h_1, h_2, \dots, h_A\}$ 表示全体实体构成的集合, A 为实体总数; $\mathcal{R} = \{r_1, r_2, \dots, r_B\}$ 表示所有关系的集合, B 为关系类型总数; $\mathcal{T} = \{(h, r, t) \mid (h, t) \in \mathcal{E}, r \in \mathcal{R}\}$ 代表 $\mathcal{G}$ 中所有三元组的集合,每个三元组 (h, r, t) 由头实体 h 、关系 r 和尾实体 t 组成。本研究沿袭已有工作的实验设置,采用链接预测任务进行评估:当给定三元组 (h, r, t) 中任一成分缺失时,模型需补全该三元组,并通过预设的评分函数 $\Phi(h, r, t)$ 进行评估。本文中该评分函数通过解码器实现。在模型结构方面,编码器的输入为实体特征矩阵 $E_1 \in \mathbf{R}^{A \times d}$,输出为蕴含丰富结构信息的更新后实体特征矩阵 $E_0 \in \mathbf{R}^{A \times d}$,其中 d 表示特征维度。最终,更新后的特征矩阵将被送入解码器以完成链接预测任务。

2 模型设计

本节将对模型进行理论阐述,其整体架构如图1所示。SDSI 采用编码器-解码器框架完成知识图谱表示学习任务。其中,编码器用于生成蕴含结构信息的实体特征,本文所提出的编码器由语义区分模块和子图整合模块两部分组成。

解码器以编码器学习得到的嵌入表示为输入,完成链接预测任务。该部分可采用多种现有知识图谱嵌入模型实现,如 ConvE^[10]、TransE^[13] 及 DistMult^[28] 等。

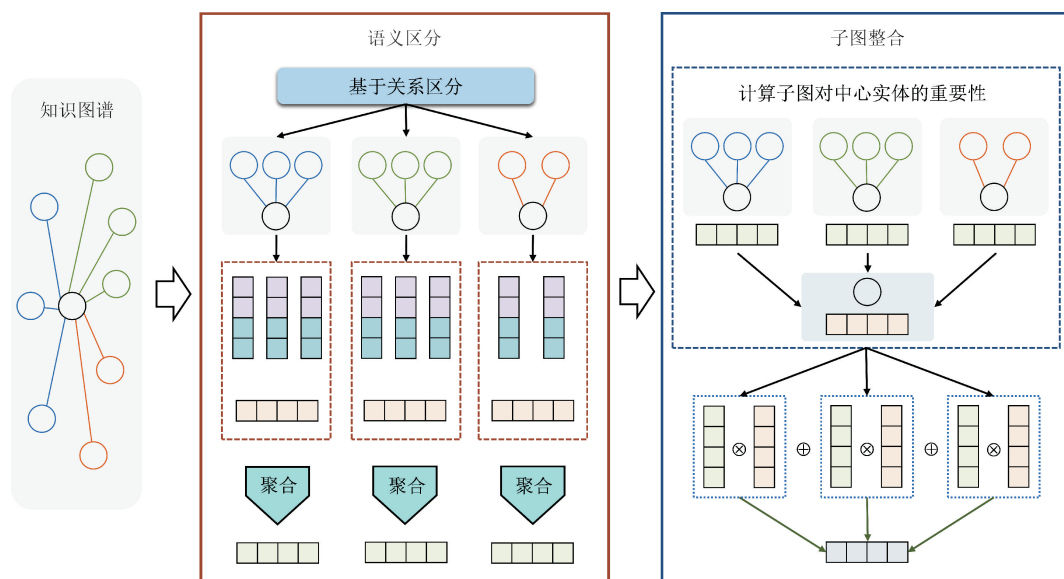


图1 语义区分和子图整合驱动的知识图谱表示学习方法框架

Fig. 1 Framework of the knowledge graph representation learning method with semantic discrimination and subgraph integration

语义区分模块首先根据实体语义的不同划分子图,进而计算各子图内部相邻实体间的重要性权重,并依此进行邻域信息聚合。为挖掘不同关系类型对中心实体的贡献程度,子图整合模块首先评估各关系子图的重要性,在此基础上聚合来自不同子图的邻域信息,最终生成实体嵌入表示。下面详细介绍所提的两个模块。

2.1 语义区分模块

SD 模块首先依据不同关系类型构建子图,随后对每个实体对应的子图进行特征聚合。通常,一个中心实体关联多个相邻实体(即链接),且这些链接包含不同的关系类型。由于每种关系都具有其特有的语义属性,可根据关系类型将中心实体的邻域实体划分为不同的集合,因此这些集合即构成关系子图。本研究的目标在于利用此类结构信息学习得到有效的实体特征表示。首先,初始化实体特征矩阵

$$E = [h_1, h_2, \dots, h_i, \dots, h_A] \quad (1)$$

式中: $h_i \in \mathbf{R}^d$ 为第 i 个实体的特征表示, $i = 1, 2, \dots, A$ 。根据邻居实体的不同语义, h_i 的邻居实体被划分为子图集合 $\mathcal{S}_i = \{\mathcal{S}_i^1, \dots, \mathcal{S}_i^m, \dots, \mathcal{S}_i^N\}$, 其中 N 表示子图的个数, $\mathcal{S}_i^N = \{t \in \mathcal{E} \mid (h, r, t) \in \mathcal{T}\}$ 是划分得到的第 i 个子图。

语义区分模块旨在提取各子图中的结构信息。对于特定子图 $\mathcal{S}_i^m$, 首先计算其中各邻居节点的重要性(即注意力得分), 继而基于得分进行链接信息聚合。对于实体 h_i 在子图中的某个邻居 $c_j^m = (r, t_j) \in \mathcal{S}_i^m$, 其注意力得分计算公式如下

$$\alpha_{i,j} = \frac{\exp(h_i ([r \parallel t_j] W_c)^T)}{\sum_{k \in \mathcal{S}_i^m} \exp(h_i ([r \parallel t_k] W_c)^T)} \quad (2)$$

式中: $r \in \mathbf{R}^d$; $t_j \in \mathbf{R}^d$; t_k 分别表示关系 r 和尾实体 t_j, t_k 的特征向量; $\parallel$ 表示向量拼接操作; $W_c \in \mathbf{R}^{2d \times d}$ 为可训练的转换矩阵; $\alpha_{i,j}$ 表示邻居 c_j^m 对实体 h_i 的注意力得分, 反映该邻居对实体 h_i 的重要程度。

需要说明的是, $\alpha_{i,j}$ 的计算限定在子图 $\mathcal{S}_i^m$ 范围内, 用于指导该子图内部的信息聚合。基于此, 对于第 n 个子图 $\mathcal{S}_i^n$, 其邻居信息聚合函数定义为

$$e_i^n = \sum_{j \in \mathcal{S}_i^n} \alpha_{i,j} c_j \quad (3)$$

多头注意力机制使模型能够从不同子空间协同捕捉信息, 从而提取更丰富的特征表征, 并有助于参数收敛过程的稳定。受图注意力网络(GAT)的启发, 本文采用多头注意力机制以获得多个子图表征

e_i^n , 并基于此提升特征的表示质量。具体而言, 使用 P 个相互独立的注意力头分别生成邻居特征, 并按如下方式进行聚合

$$a_i^n = [e_i^{1,n} \parallel e_i^{2,n} \parallel \dots \parallel e_i^{p,n} \parallel \dots \parallel e_i^{P,n}] W_p \quad (4)$$

式中: $e_i^{p,n}$ 表示由第 p 个注意力头生成的特征表示; $W_p \in \mathbf{R}^{Pd \times d}$ 是一个可训练的转换矩阵; P 为注意力头数; a_i^n 为根据关系 r 从子图 $\mathcal{S}_i^n$ 中聚合得到的特征。

2.2 子图整合模块

上述语义区分模块得到了各子图 $\mathcal{S}_i^n$ 对应的特征 a_i^n 。在此基础上, 子图整合(SI) 模块将对所有子图进行跨结构特征聚合。考虑到不同子图对中心实体 h_i 具有不同的贡献程度, 需要根据其重要性权重对信息进行聚合。各子图对应的注意力得分为

$$\beta_i^n = \frac{\exp(h_i W_x (a_i^n W_y)^T)}{\sum_{m \in \mathcal{S}_i} \exp(h_i W_x (a_i^m W_y)^T)} \quad (5)$$

式中: $W_x \in \mathbf{R}^{d \times d_n}$ 和 $W_y \in \mathbf{R}^{d \times d_n}$ 都为可训练的转换矩阵; m 为尾实体的序号; d_n 表示隐藏特征维度; β_i^n 表示子图 $\mathcal{S}_i^n$ 对实体 h_i 的注意力分数。该分数反映了子图 $\mathcal{S}_i^n$ 对中心实体的重要程度, 并用于生成实体 h_i 的最终特征表示。

为挖掘更深层次的潜在信息, 本模块引入了多头注意力机制, 使用 Q 个独立的注意力头并行计算 Q 组合注意力得分, 再通过求平均操作进行融合, 其计算方式如下

$$\beta_i^n = \frac{1}{Q} \sum_{q=1}^Q \beta_i^{q,n} \quad (6)$$

式中: $\beta_i^{q,n}$ 表示第 q 个注意力头生成的注意力得分。

对于中心实体 h_i , 语义区分模块获得子图集合 $\mathcal{S}_i = \{\mathcal{S}_i^1, \dots, \mathcal{S}_i^m, \dots, \mathcal{S}_i^N\}$ 并生成了 N 个特征表示。子图整合模块则对应计算得出 N 个重要性分数, 在此基础上, 可将所有子图信息按如下方式进行聚合

$$z_i = \frac{1}{N} \sum_{n=1}^N \beta_i^n a_i^n \quad (7)$$

此外, 为保留实体 h_i 的原始特征信息, 在聚合过程中进一步引入自连接机制, 具体操作为

$$h'_i = \sigma(h_i + z_i) \quad (8)$$

式中: $\sigma(\cdot)$ 表示激活函数; h'_i 为学习到的实体 h_i 的最终表示, 其包含了丰富的结构信息。

2.3 模型训练与优化

链接预测的核心目标是针对不完整的三元组 $(?, r, t)$ 推测头实体, 或对三元组 $(h, r, ?)$ 预测尾实体, 该任务通常通过解码器完成。本文选用

ConvE 模型作为解码器,需说明的是,该解码器可替换为其他知识图谱嵌入方法。对于给定三元组 (h, r, t) , 其评分函数定义如下

$$\Phi(\mathbf{h}, \mathbf{r}, \mathbf{t}) = \sigma(\text{vec}(\sigma([\bar{\mathbf{h}}'_i \parallel \bar{\mathbf{r}}] * \boldsymbol{\omega}))\mathbf{W}')\mathbf{t}' \quad (9)$$

式中: $\mathbf{r}$ 为关系 r 的特征表示; $\mathbf{h}'_i$ 与 $\mathbf{t}'$ 分别对应式(8)中头实体 h_i 和尾实体 t 更新之后的特征; $\sigma(\cdot)$ 表示激活函数; $\bar{\mathbf{h}}'_i$ 和 $\bar{\mathbf{r}}$ 分别是 $\mathbf{h}'_i$ 与 $\mathbf{t}'$ 经二维重塑后的结果; $*$ 表示卷积运算; $\boldsymbol{\omega}$ 为卷积核; $\mathbf{W}'$ 是可训练的转换矩阵。

接着,通过 $\text{sigmoid}(\cdot)$ 函数对评分函数进行处理,以得到相对分数

$$p(h, r, t) = \text{sigmoid}(\Phi(\mathbf{h}, \mathbf{r}, \mathbf{t})) \quad (10)$$

基于上式的得分,模型训练采用交叉熵损失函数,具体形式如下

$$\mathcal{L} = \sum_{(h, r, t) \in \mathcal{T}} -\frac{1}{|\epsilon|} \sum_{k=1}^{|\epsilon|} \left(y(h, r, t_k) \lg(p(h, r, t_k)) + (1 - y(h, r, t_k)) \lg(1 - p(h, r, t_k)) \right) \quad (11)$$

式中: $y(h, r, t_k) \in \{0, 1\}$ 表示三元组 (h, r, t_k) 的标签; $|\epsilon|$ 代表所有候选尾实体的数量。该损失函数的设计目标为:若三元组 (h, r, t_k) 为真,则 $y(h, r, t_k) = 1$;若三元组无效,则 $y(h, r, t_k) = 0$ 。

此外,本文采用标签平滑^[29]技术以缓解过拟合问题,并选用 Adam^[30]算法作为优化器。在特征以及每个全连接层之后的隐藏单元上均采用 Dropout^[31]技术;在各层之后采用批量归一化方法^[32],以实现模型训练的稳定性提升、正则化效果增强及收敛速率加快。

3 模型性能评估

3.1 数据集

为了验证所提出方法的有效性,在知识图谱表示学习领域广泛使用的 3 个基准数据集 FB15K-237、YAGO3-10 与 WN18RR 上开展了全面实验评估。表 1 展示了 3 个数据集的详细信息。

表 1 3 个数据集统计信息

Table 1 Statistics of the three datasets

数据集属性	FB15K-237	YAGO3-10	WN18RR
实体数	14 541	123 182	40 943
关系数	237	37	11
训练集数	272 115	1 079 040	86 835
验证集数	17 535	5 000	3 034
测试集数	20 466	5 000	3 134

FB15K-237 数据集是 FB15K 的子集,它源自大规模的 Freebase 知识库,通过精选和重构,保留了最具代表性的实体和关系。FB15K-237 通过移除 FB15K 中的逆关系,确保了评估结果更能真实反映模型的泛化与推理能力。

YAGO3-10 数据集是 YAGO3 的子集,特别关注其核心的 10 种关系。该数据集从多语言的 Wikipedia 中提取信息,并与英文词汇数据库 WordNet 以及地理数据库 GeoNames 进行融合,构建了一个统一且协调的知识库。

WN18RR 数据集是 WN18 的子集,它源自英文词汇数据库 WordNet 3.0,其实体对应的是 WordNet 中的同义词集,而关系则体现了这些同义词集之间的词汇语义关系。WN18RR 同样移除了 WN18 中的逆关系,以防止测试集泄漏问题。

3.2 评估指标

为了评估预测准确性,本研究采用 3 个通用评估指标:平均倒数排名(M_{RR})以及命中率 Hits@ k (k 取 1, 3, 10)。这 3 个评估指标最为常用并且能够准确体现模型的性能。其中, M_{RR} 表示从测试集中得到的排名 L 的倒数的平均值,取值范围为 $[0, 1]$, 计算公式为

$$M_{RR} = \frac{1}{L} \sum_{l \in L} \frac{1}{l} \quad (12)$$

Hits@ k 表示排名小于或者等于阈值 k 的预测的比例, Hits@ k 的取值范围为 $[0, 1]$, 其数值与模型性能呈正相关关系。该指标通常设置 $k = 1, 3, 5, 10$ 等进行多维度评估。当 $k = 1$ 时, Hits@1 特指模型预测结果中排名首位三元组所占的比例。这一特性使得 Hits@1 与 M_{RR} 指标之间存在紧密的数学关联。

3.3 实验设置

实体与关系的特征向量随机初始化,并在模型训练过程中同步更新。设置特征维度 $d = 200$, 隐藏特征维度 $d_n = 200$, 模型中的转换矩阵 $\mathbf{W}_e \in \mathbf{R}^{2d \times d}$ 、 $\mathbf{W}_p \in \mathbf{R}^{pd \times d}$ 、 $\mathbf{W}_q \in \mathbf{R}^{qd \times d}$ 、 $\mathbf{W}_x \in \mathbf{R}^{d \times d_n}$ 和 $\mathbf{W}_y \in \mathbf{R}^{d \times d_n}$ 均随机初始化,并通过训练过程进行优化。针对每个数据集,通过在验证集上进行网格搜索来确定最优超参数组合,其候选范围包括:注意力头个数 P 和 Q 均取值为 1, 2, 3, 4, 批处理大小设定为 128、256、512, 学习率从 0.001、0.002、0.003 中选取, 丢弃率在 0.1、0.2、0.3 范围内调整。所有实验均基于 PyTorch 框架实现,并在 NVIDIA GeForce

RTX2080Ti GPU 上完成。表 2 展示了各知识图谱数据集对应的最优超参数配置。

表 2 不同数据集上的超参数最优组合

Table 2 Optimal combination of hyper-parameters on different datasets

指标	FB15K-237	YAGO3-10	WN18RR
P	3	3	3
Q	3	2	2
学习率	0.001	0.001	0.002
批处理	256	128	512
丢弃率	0.2	0.2	0.3

3.4 基线方法

为了验证所提模型的性能,将其与多种前沿知识图谱表示学习方法进行了系统对比,包括知识图谱嵌入方法 TransE^[13]、ComplEx^[4]、ConvE^[10]、RotatE^[11]、Rethinking^[16]、CompoundE^[33]和基于图神经网络的知识图谱表示学习方法 CompGCN^[15]、SE-GNN^[17]、MRGAT^[18]、KRACL^[19]、SEPA^[20]、UniGE^[26]和 Tgformer^[14]。

3.5 链接预测实验结果

表 3 展示了 SDSI 及基线方法在 FB15K-237、YAGO3-10 和 WN18RR 这 3 个数据集上的链接预测结果。实验严格遵循已有研究的实验设置与数据划分标准。通过深入分析,可得到以下结论。

首先,与仅采用 ConvE 进行知识图谱表示学习的方法相比,SDSI 在所有数据集的各项评估指标上均取得稳定提升。这表明通过丰富的结构信息与局部领域特征学习实体表征,能够有效促进模型性能的改善,同时验证了本文所提出的语义区分与子图整合模块的有效性。

其次,相较于现有知识图谱表示学习方法,SDSI 在 3 个标准数据集上均取得了最优性能。具体而言,在 FB15K-237 数据集上,SDSI 模型相比当前最优结果在 3 项核心指标上提升了 1%~2%,其中 M_{RR} 从 0.372 提升至 0.381, Hits@10 从 0.557 提升至 0.565。同时,在 YAGO3-10 大规模数据集上,SDSI 模型在 M_{RR} 与 Hits@3 相较基线方法也实现了显著提升。此外,在稀疏数据集 WN18RR 上,SDSI 同样展现出优异的性能。这些结果充分证明了 SDSI 模型的有效性与普适性。

最后,与各基于图神经网络的基线模型相比,SDSI 模型在 3 个数据集上均表现出显著优势。这主要源于现有模型对子图信息的挖掘不够充分。区别于这些方法,SDSI 模型能够从基于不同关系类型构建的子图中提取潜在信息,并通过评估不同关系类型的贡献度来实现邻域信息表征的优化聚合。总体而言,SDSI 模型在各类知识图谱场景中均展现出卓越的适用性与有效性。

表 3 不同模型在 3 个数据集上的链接预测实验结果对比

Table 3 Link prediction results of different models on the three datasets

方法	模型	FB15K-237			YAGO3-10			WN18RR		
		M_{RR}	Hits@3	Hits@10	M_{RR}	Hits@3	Hits@10	M_{RR}	Hits@3	Hits@10
知识图谱 嵌入方法	TransE	0.294		0.465				0.226		0.501
	ComplEx	0.247	0.275	0.428	0.360	0.400	0.550	0.440	0.460	0.510
	ConvE	0.325	0.356	0.501	0.440	0.490	0.620	0.430	0.440	0.520
	RotatE	0.338	0.375	0.533	0.495	0.550	0.670	0.476	0.492	0.571
	Rethinking	0.355	0.389	0.535				0.472	0.485	0.544
	CompoundE	0.357	0.393	0.545				0.481	0.497	0.576
基于图神经 网络方法	CompGCN	0.355	0.390	0.535	0.489	0.500	0.582	0.479	0.494	0.546
	SE-GNN	0.362	0.396	0.546				0.481	0.506	0.569
	MRGAT	0.358	0.386	0.542	<u>0.552</u>	<u>0.561</u>	<u>0.698</u>	0.481	0.501	0.568
	KRACL	0.360	0.395	0.548				0.486	0.501	<u>0.580</u>
	SEPA	0.354	0.390	0.545				0.480	0.498	0.570
	UniGE	0.343	0.375	0.523	0.412	0.451	0.579	0.491	<u>0.512</u>	0.563
	TGformer	<u>0.372</u>	<u>0.410</u>	<u>0.557</u>				<u>0.493</u>	0.509	0.566
	SDSI	0.381	0.426	0.565	0.564	0.588	0.712	0.496	0.515	0.585

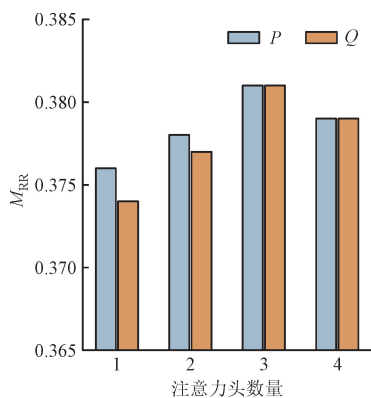
注:加粗表示最优结果,下划线表示次优结果。

3.6 参数敏感度分析

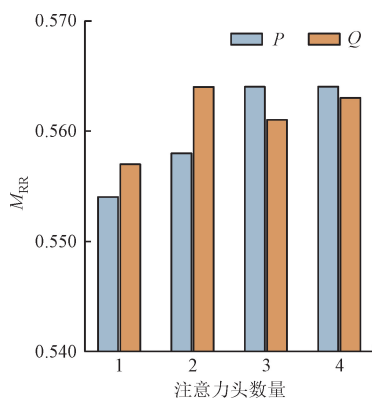
3.6.1 注意力头数量变化分析

在3个数据集上, M_{RR} 随注意力头数量 P 与 Q 变化的实验结果见图2。多头注意力机制的核心优势在于每个注意力单元能够聚焦于不同的特征子空间,从而提取具有区分度的结构与语义信息。本研究通过设置 P 、 Q 均为1、2、3、4进行链接预测实验,系统评估了注意力头数量对模型效能的影响。由图2可知,注意力头数量确实对SDSI模型的性能产生显著影响,这一现象与理论预期相符。在YAGO3-10和WN18RR数据集上, P 对模型性能的影响程度大于 Q 。而在FB15K-237数据集上, Q 的取值表现出更明显的影响效应,这主要源于该数据集包含的关系类型数量显著多于其他基准数据集,使得关系重要性评估在这一特定语境中具有更高权重。

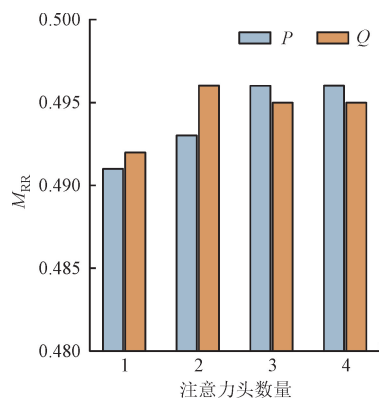
此外,从实验结果还观察到,模型性能并非随注意力头数量的增加而持续提升。当注意力头数量达到一定阈值后,性能增长趋于平缓甚至出现回落,这表明适中的注意力头配置已能满足大部分语义特征提取的需求,过度增加参数数量反而可能导致模型过拟合。



(a) FB15K-237



(b) YAGO3-10



(c) WN18RR

图2 注意力头数量对模型性能影响分析

Fig. 2 Analysis of the impact of attention head number on model performance

3.6.2 关系注意力分布分析

为探究多头注意力机制的内在作用机理,本节对实体与关系的注意力权重分布展开可视化分析。图3为可视化结果,其中色块颜色深度与注意力得分呈正相关。观察发现,各注意力头的颜色分布呈现显著差异,这与注意力机制自适应分配权重的特性相符,其本质是通过差异化加权来区分不同邻域信息的重要性。

图3呈现了与中心实体相关联的关系注意力分布图谱。纵轴表示不同的中心实体,横轴对应其连接的关系类型。分析表明:首先,针对同一实体,不同关系被赋予的注意力权重存在明显差异;其次,同一关系作用于不同实体时,所获得的注意力值也各不相同。以实体“06lvlf”为例,关系“education”和“people”获得较高关注度,而“nationality”和“base”两类关系对其表征学习的贡献度相对较低。这种分布特性印证了关系类型本身的复杂性会导致注意力矩阵呈现非均匀特性,因而通过差异化处理异构关系,能够有效捕捉特定的语义特征。

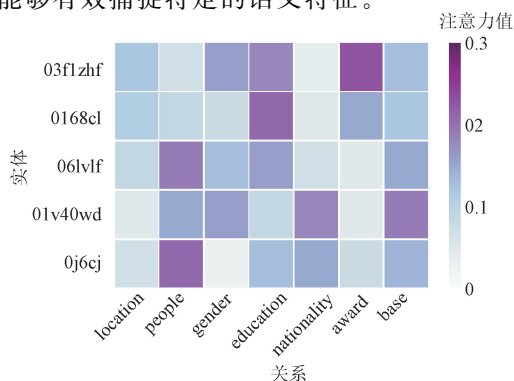


图3 关系注意力分布

Fig. 3 Attention distribution of relations

此外,在预测实体“0j6cj”的“gender”时,模型将较高的注意力权重分配给了“gender”子图,此异常主要源于结构密度偏差。SI 模块中的注意力机制在计算时会无差别地接收来自 SD 模块的各子图特征向量。如果子图结构复杂、实体众多,SD 模块可能学习到更具区分性和高维特征的表示,这种“丰富的”表示在注意力机制中更容易获得较高的相似度得分,导致模型可能在一定程度上混淆了结构的丰富性与语义的强相关性。

3.7 消融实验

为验证 SDSI 中各创新模块的有效性,通过删减对应组件构建了以下变体模型进行消融实验:①SDSI^{-SD} 变体移除了 SD 模块,改为直接根据关系重要性进行实体邻居信息聚合;②SDSI^{-SI} 变体则去除 SI 模块,转而依据邻居实体重要性进行信息聚合;③SDSI^{-IN} 变体切断了 SD 到 SI 模块的直接语义流,将 SD 生成的各子图表示映射为一个固定向量。

表 4 为 SDSI 模型及变体模型在 3 个数据集中的消融实验结果。可以看出,移除任一模块都会导致模型性能下降,这充分证明了 SD 与 SI 两个模块的有效性。具体而言,SD 模块负责从按关系类型划分的子图中整合信息,若去除该模块将导致模型无法捕捉子图中的邻域信息,从而影响性能表现;而 SI 模块的核心作用在于探究不同子图对中心实体的贡献程度,若缺少该模块,模型将难以区分关系的特定属性,进而降低预测精度。此外,在完整模型中,SI 模块的元注意力权重是基于 SD 模块输出的、具有明确关系语义标识的子图表示来计算的。这种设计使得注意力权重能够精准地关联到特定的关系语义上。而在 SDSI^{-IN} 变体中,SI 模块在进行整合时,难以将重要性与特定的关系语义直接关联起来,从而损失了部分建模精度。通过上述实验可以得出结论:所有创新模块均能有效提升模型性能,对本研究提出的整体模型具有重要贡献。

表 4 3 个数据集中的消融实验结果

Table 4 Ablation studies results on the three datasets

模型	FB15K-237			YAGO3-10			WN18RR		
	M_{RR}	Hits@3	Hits@10	M_{RR}	Hits@3	Hits@10	M_{RR}	Hits@3	Hits@10
SDSI ^{-SD}	0.375	0.422	0.559	0.557	0.580	0.703	0.492	0.510	0.580
SDSI ^{-SI}	0.373	0.420	0.557	0.560	0.583	0.706	0.494	0.512	0.582
SDSI ^{-IN}	0.377	0.424	0.561	0.561	0.585	0.707	0.494	0.513	0.582
SDSI	0.381	0.426	0.565	0.564	0.588	0.712	0.496	0.515	0.585

注:加粗表示最优结果。

3.8 计算复杂度分析

为分析 SDSI 模型的时间复杂度,本文采用 $O(\cdot)$ 记法对包含 n' 个实体、 m' 种关系和 t' 个三元组的知识图谱进行量化分析。具体而言,知识图谱中的 t' 个三元组对应整个图的 t' 条边,因此每个实体平均拥有 t'/n' 个邻居实体。对于每种关系,每个子图中的平均邻居实体数为 $t'/n'm'$ 。

在特定子图中,对关系和实体特征进行变换需要 $O(2t'/n'm')$ 次运算,计算注意力系数需要 $O(t'/n'm')$ 次运算;softmax 函数与特征聚合各需要 $O(t'/n'm')$ 次运算。此外,在每个子图中仅需 $O(1)$ 次运算。因此,单个子图的总运算量约为 $O(5t'/n'm')$ 。

对于具有 m' 种关系的实体,需要 $m'O(5t'/n'm') = O(5t'/n')$ 次运算。另外,聚合实体所有邻居信息需要 $O(t'/n')$ 次运算。最终,模型中每个实体的总运算量为 $O(6t'/n')$,整个知识图谱的总运算量为 $n'O(6t'/n') = O(6t')$ 。综上所述,SDSI 的时间复杂度为 $O(6t')$,其与知识图谱三元组数量 t' 相关。

3.9 可扩展性分析

SDSI 模型的核心创新在于通过 SD 与 SI 模块实现对知识图谱细粒度语义结构的高保真建模。然而,该设计在带来性能提升的同时,也引入了额外的计算复杂度,主要源于按关系类型划分邻域并执行子图内聚合,以及计算跨子图的元注意力权重。因此,提出 3 种策略以增强模型在大规模知识图谱下的可扩展性。

(1) 高效邻域采样策略。在 SD 模块中,不为每个关系类型聚合全部邻接点,而是为每个关系子图设定一个固定的采样预算。根据历史训练中学习到的子图注意力权重(SI 模块输出),对重要性高的关系子图进行过采样,对低重要性子图进行欠采样,在保持性能的同时提升效率。

(2) 模块化与并行化设计。SD 模块中对不同关系子图的处理是天然可并行的,可利用 GPU 进行并行计算,大幅压缩实际运行时间。可将 SI 模块的注意力计算设计为稀疏化的,仅计算与中心实体最相关的 Top-K 个子图间的交互。

(3) 分层图训练技术。采用子图划分算法将大规模图谱划分为多个紧凑的子图分区,在分区内应用完整的 SDSI 算法,在分区边界通过参数服务器交换并更新共享实体的嵌入。采用历史嵌入缓存技术,对低频更新的邻居实体使用缓存的历史嵌入进行计算,减少实时前向传播的计算量。

3.10 鲁棒性测试

为了全面评估本文提出的 SDSI 模型,设计了对关系标签存在噪声、模糊或定义不一致等数据缺陷的鲁棒性测试,以明确模型在非理想数据环境下的性能边界。为了模拟标注错误或数据存储混乱的情况,随机选取训练集中一定比例的三元组,将其关系标签替换为从关系集中随机抽取的其他关系。为了模拟语义模糊情况,将一定比例的三元组的关系标签替换为语义上最相近的错误关系。为了模拟定义不一致情况,将数据集中语义高度相似、常被混用的一对关系随机合并为其中一个。

表 5 为 SDSI 模型在 3 个数据集上的鲁棒性测试结果,其设置了不同的噪声并统计了 M_{RR} 。分析可得,3 类噪声对于模型性能均有一定的影响,其中 30% 随机噪声的破坏性最强。对于语义模糊噪声, M_{RR} 并没有出现明显下降,这直接得益于其 SD 模块。该模块通过关系类型划分邻域并进行子图内聚合,使得模型在学习时更关注于同一关系下的局部语义一致性。在定义不一致噪声场景下,SDSI 模型也表现出较高的性能保持,这主要归功于 SI 模块。当两种本应区分的关系被强制合并后,对应于原两种关系的子图信息将被混合。综上,SDSI 模型不仅在理想数据条件下具有先进的链接预测性能,在现实数据缺陷的非理想环境下也表现出更强的稳定性和可靠性,这为其在实际场景中的应用提供了重要依据。

表 5 鲁棒性测试
Table 5 Robustness test

数据类型	M_{RR}		
	FB15K-237	YAGO3-10	WN18RR
原始	0.381	0.564	0.496
10% 随机噪声	0.342	0.501	0.436
20% 随机噪声	0.312	0.457	0.401
30% 随机噪声	0.277	0.417	0.357
20% 语义模糊	0.301	0.451	0.391
20% 定义不一	0.327	0.490	0.426

4 结 语

本文提出了一种语义区分和子图整合模块驱动的知识图谱表示学习模型 SDSI。该模型通过语义区分模块,依据不同关系类型构建子图并聚合其中的语义信息,从而捕捉实体丰富的属性特征。为进一步挖掘子图对中心实体的差异化贡献,SDSI 采用子图整合模块,该模块通过计算各子图对实体重要性权重,融合邻域表征生成最终实体特征。在 3 个标准知识图谱数据集上进行的链接预测实验结果表明,SDSI 具有显著优越性。

未来工作中,将探索引入实体文本描述等多模态信息,以克服当前模型对纯结构化三元组依赖的局限。计划设计一个跨模态交互模块,通过模态知识对齐并融合图谱结构特征与文本语义特征,增强对实体细粒度语义的理解与区分能力,改善稀疏实体的表示学习,并提升模型在复杂语义场景下的推理鲁棒性。

参考文献:

- [1] Liang Ke, Meng Lingyuan, Liu Meng, et al. A survey of knowledge graph reasoning on graph types: static, dynamic, and multi-modal [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 46(12): 9456-9478.
- [2] 朱可, 秦涛, 朱奎宇, 等. 采用消息聚合的演化感知时序知识图谱推理方法 [J]. 西安交通大学学报, 2025, 59(11): 187-197.
Zhu Ke, Qin Tao, Zhu Kuiyu, et al. Evolution-aware temporal knowledge graph reasoning method based on message aggregation [J]. Journal of Xi'an Jiaotong University, 2025, 59(11): 187-197.
- [3] Shang Bin, Zhao Yinliang, Liu Jun, et al. LAFA: multimodal knowledge graph completion with link aware fusion and aggregation [C]//Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2024: 8957-8965.
- [4] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction [C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2016:2071-2080.

- [5] 吴信东, 黄满宗, 卜晨阳. BEKO: 大语言模型与知识图谱的双向增强 [J]. 计算机学报, 2025, 48(7): 1572-1588.
- Wu Xindong, Huang Manzong, Bu Chenyang. BEKO: bidirectional enhancement with a knowledge ocean for LLMs and KGs [J]. Chinese Journal of Computers, 2025, 48(7): 1572-1588.
- [6] Shang Bin, Zhao Yinliang, Wang Di, et al. Relation-aware multi-positive contrastive knowledge graph completion with embedding dimension scaling [C]//Proceedings of the 46th international ACM SIGIR conference on research and development in information retrieval. New York, USA: Association for Computing Machinery, 2023: 878-888.
- [7] 郑修林, 周鹏, 李培培, 等. 知识图谱补全技术及应用 [J]. 软件学报, 2025, 36(12): 5599-5628.
- Zheng Xiulin, Zhou Peng, Li Peipei, et al. Knowledge graph completion: techniques and applications [J]. Journal of Software, 2025, 36(12): 5599-5628.
- [8] Mahdisoltani F, Biega J, Suchanek F M. YAGO3: a knowledge base from multilingual wikipedias [EB/OL]. (2015-01-04) [2025-12-05]. <https://www.semantic-scholar.org/paper/YAGO3:-A-Knowledge-Base-from-Multilingual-Mahdisoltani-Biega/6c5b5adc3830ac45bf1d764603b1b71e5f729616>.
- [9] 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. New York, USA: ACM, 2008: 1247-1250.
- [10] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2D knowledge graph embeddings [C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2018: 1811-1818.
- [11] Sun Zhiqing, Deng Zhihong, Nie Jianyun, et al. RotatE: knowledge graph embedding by relational rotation in complex space [C]//ICLR 2019 Conference. New York, USA: ICLR, 2019: 1-18.
- [12] Shang Bin, Zhao Yinliang, Liu Jun. Learnable convolutional attention network for knowledge graph completion [J]. Knowledge-Based Systems, 2024, 285: 111360.
- [13] Bordes A, Usunier N, Garcia-duran A, et al. Translating embeddings for modeling multi-relational data [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2013: 2787-2795.
- [14] Shi Fobo, Li Duantengchuan, Wang Xiaoguang, et al. TGformer: a graph transformer framework for knowledge graph embedding [J]. IEEE Transactions on Knowledge and Data Engineering, 2025, 37(1): 526-541.
- [15] Vashishth S, Sanyal S, Nitin V, et al. Composition-based multi-relational graph convolutional networks [C]//ICLR2020 Conference. New York, USA: ICLR, 2020: 1-16.
- [16] Zhang Zhanqiu, Wang Jie, Ye Jieping, et al. Rethinking graph convolutional networks in knowledge graph completion [C]//Proceedings of the ACM Web Conference 2022. New York, USA: ACM, 2022: 798-807.
- [17] Li Ren, Cao Yanan, Zhu Qiannan, et al. How does knowledge graph embedding extrapolate to unseen data: a semantic evidence view [C]//Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence and Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence and the Twelfth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2022: 5781-5791.
- [18] Dai Guoquan, Wang Xizhao, Zou Xiaoying, et al. MRGAT: multi-relational graph attention network for knowledge graph completion [J]. Neural Networks, 2022, 154: 234-245.
- [19] Tan Zhaoxuan, Chen Zilong, Feng Shangbin, et al. KRACL: contrastive learning with graph context modeling for sparse knowledge graph completion [C]//Proceedings of the ACM Web Conference 2023. New York, USA: ACM, 2023: 2548-2559.
- [20] Gregucci C, Nayyeri M, Hernández D, et al. link prediction with attention applied on multiple knowledge graph embedding models [C]//Proceedings of the ACM Web Conference 2023. New York, USA: ACM, 2023: 2600-2610.
- [21] Lee J, Wang Yequan, LI Jing, et al. Multimodal reasoning with multimodal knowledge graph [C]//Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2024: 10767-10782.
- [22] Jian Yue, Luo Xiangyu, Li Zhifei, et al. APKGC: noise-enhanced multi-modal knowledge graph completion with attention penalty [C]//Proceedings of the Thirty-

- Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2025: 15005-15013.
- [23] Lu Yuhuan, Yu Weijian, Jing Xin, et al. HyperFM: fact-centric multimodal fusion for link prediction over hyper-relational knowledge graphs [C]//Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2025: 2818-2830.
- [24] Shang Bin, Zhao Yinliang, Liu Jun, et al. Mixed geometry message and trainable convolutional attention network for knowledge graph completion [C]//Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2024: 8966-8974.
- [25] Li Guangyao, Sun Zequn, Hu Wei, et al. Position-aware relational transformer for knowledge graph embedding [J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(8): 11580-11594.
- [26] Liu Yuhuan, Cao Zelin, Gao Xing, et al. Bridging the space gap: unifying geometry knowledge graph embedding with optimal transport [C]//Proceedings of the ACM Web Conference 2024. New York, USA: ACM, 2024: 2128-2137.
- [27] Yin Hong, Zhong Jiang, Li Rongzhen, et al. High-order neighbors aware representation learning for knowledge graph completion [J]. IEEE Transactions on Neural Networks and Learning Systems, 2025, 36(3): 5273-5287.
- [28] Yang B, Yih S W, He X, et al. Embedding entities and relations for learning and inference in knowledge bases [EB/OL]. (2014-12-20) [2025-12-05]. <https://arxiv.org/abs/1412.6575>.
- [29] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 2818-2826.
- [30] Kingma D P, Adam; A method for stochastic optimization [EB/OL]. (2014-12-22) [2025-12-05]. <https://arxiv.org/abs/1412.6980>.
- [31] Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting [J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [32] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2015: 448-456.
- [33] Ge Xiou, Wang Yuncheng, Wang Bin, et al. Compounding geometric operations for knowledge graph completion [C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2023: 6947-6965.

(编辑 亢列梅)