关系感知网络结合基于注意力的损失函数用于少样本知识图谱补全

Qiao Qiao, Yuepei Li, Kang Zhou, and Qi Li

Iowa State University, Ames, Iowa, USA
{qqiao1,liyp0095,kangzhou,qli}@iastate.edu

摘要 少样本知识图谱补全（FKGC）任务旨在使用少量参考实体对预测关系的未见事实。当前方法随机选择每个参考实体对的一个负样本以最小化基于边界的排序损失，如果负样本远离正样本并超出边界，则容易导致零损失问题。此外，实体在不同上下文中应具有不同的表示。为了解决这些问题，我们提出了一种新的基于注意力的损失（RANA ）关系感知网络框架。具体来说，为了更好地利用丰富的负样本和缓解零损失问题，我们战略性地选择相关负样本并设计了一个基于注意力的损失函数来进一步区分每个负样本的重要性。直观上，与正样本更相似的负样本将对模型有更大的贡献。此外，我们设计了一种动态关系感知实体编码器以学习上下文依赖的实体表示。实验表明，在两个基准数据集上 RANA 优于最先进的模型。

Keywords: 少量样本学习 · 知识图谱补全

1 介绍

知识图谱（KGs）包含丰富的三元组（事实），其中每个三元组 (h, r, t) 表示头实体 h 和尾实体 t 之间的关系 r。知识图谱如 Wikidata[16] 和 NELL[3] 已被应用于各种下游应用，例如关系抽取 [25]、命名实体识别 [24] 和节点分类 [10]。

知识图谱补全（KGC）旨在解决由于实体或关系缺失而导致的知识图谱不完整问题。KG 嵌入 [2, 15] 在 KGC 上取得了显著的性能。这些模型在有足够的训练三元组时表现良好，但知识图谱中很大一部分关系遵循长尾分布。例如，在 Wikidata [4] 中大约有 10%的关系不超过 10 个三元组。没有足够训练三元组的关系被称为少量样本关系。对于模型来说，预测具有有限训练三元组的关系至关重要且充满挑战。

少样本知识图谱补全（FKGC）方法被提出以解决少样本关系问题。给定关系 r 和少量参考实体对 (h, t)，FKGC 旨在为每个查询 $(h, ?)$ 排序候选

尾部实体 t。这些少样本参考实体对构成一个支持集，而查询则形成一个查询集。现有方法的一条研究路线集中在设计度量学习算法以计算实体对 [18, 23] 之间的相似性。另一条路线利用模型不可知元学习算法（MAML）[5] 来学习模型 [4, 9, 17] 的最优参数。

为了训练模型，当前的 FKGC 方法应用了一种基于边界的排序损失函数，该函数旨在通过一个边界将正样本三元组的得分与负样本三元组的得分分开。每个正样本三元组通过用随机选择的候选尾实体替换真实的尾实体来形成一个负样本三元组。这个损失函数不能有效地利用负样本。此外，由于候选者数量众多，可能会选择到不相关的负样本。这些不相关的负样本导致零损失，因为负样本三元组远离正样本三元组。因此，这些不相关的负样本不会对训练做出贡献，并减慢了收敛速度 [11]。例如，给定一个真实的三元组 *(科比 · 布莱恩特, 工作地点, 加利福尼亚)*，模型可以选择负样本尾实体，如纽约，泰国，伦敦等。因为泰国与真正的尾实体加利福尼亚不相关，所以加州和泰国之间的距离大于预定义的边界，相应的损失为零。因此泰国可能不会对训练有所贡献。

为了解决上述限制，我们提出一个称为瑞那纳 (**R** 关系-**A** 无关 **N** 网络与 **A** 注意力损失) 的框架。为了提高负样本的质量，我们建议首先过滤不相关的候选尾实体，然后随机采样多个负样本而不是一个。由于负样本的重要性不同且取决于它们与正样本的相似性，我们应用注意力机制为每个负样本分配权重，其中最相关的负样本的权重高于不太相关的负样本的权重。基于注意力的加权损失函数可以使模型有效地避免零损失问题，并因此学习更好的决策边界。

进一步，我们提出了一种基于上下文的动态关系感知实体编码器来学习不同关系中同一实体的不同表示。具体来说，给定目标关系及其支持集，实体编码器使用目标关系与相邻关系之间的相似性来区分相邻实体的影响，并动态编码实体的局部连接。最后，瑞纳 采用元学习使模型能够在少量梯度步骤和少数训练三元组的情况下，在新的关系上表现良好。

总之，我们的主要贡献是：

1. 我们提出了一种新的负采样策略和一种新颖的基于注意力的损失函数来解决零损失和收敛速度较慢的问题。
2. 我们提出了一种动态实体编码器来学习依赖上下文的实体表示并减少无关邻近实体的影响。
3. 在基准数据集上的实验结果表明，瑞纳 一致且显著地优于其他基线方法。

2 相关工作

2.1 基于嵌入的知识图谱补全

知识图谱嵌入旨在将实体和关系嵌入到一个低维连续向量空间中，同时保留它们的语义意义。现有方法可以分为以下几类：(1) 基于翻译的模型计算实体和关系之间的欧氏距离作为事实的可能性，如 TransE[2]，RotatE[13] 和 TransMS[20]；(2) 基于语义匹配的模型计算实体和关系之间的语义相似性作为事实的可能性，如 RESCAL[8]，DistMult[19] 和 PUDA[14]；以及 (3) 基于神经网络的模型将实体和关系输入深度神经网络以融合图网络结构和实体及关系的内容信息，如 SME[1]，CompLEx[15] 和 BertRL[22]。所有上述模型都需要足够的训练三元组，因此在少样本关系上的表现会受到影响。

2.2 少样本知识图谱补全

FKGC 要求模型根据少量训练事实预测新事实。现有方法分为两类：(1) 基于度量的方法旨在通过计算查询集和支持集之间的相似性来学习匹配度量。GMathching [18] 关注一次性 KGC，同时考虑了学习到的嵌入和局部图结构。FSRL [23] 和 FAAN [12] 将 GMatching 扩展到了少样本场景。(2) 基于优化的方法旨在学习一组良好的初始模型参数，使学习到的模型能够快速泛化到新关系。MetaR [4], GANA [9] 和 HiRe [17] 关注从支持集中的实体嵌入中提取特定关系的元信息并将其转移到查询集。

然而，所有这些方法都使用基于边界的排名损失，这不能有效避免低质量的负样本，导致零损失问题并影响收敛速度。负采样已被证明与正采样一样重要，对于确定优化目标 [21] 而言。特别是在少量样本设置下，鉴于有限的正样本，如何根据相应的正样本选择高质量的负样本至关重要。

3 初步的

3.1 问题定义

知识图谱 $\mathcal{G}$。知识图谱 $\mathcal{G}$ 是一组三元组 $\mathcal{T} = \{(h, r, t) \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}\}$，其中 $\mathcal{E}$ 和 $\mathcal{R}$ 分别代表实体集和关系集。关系集 $\mathcal{R}$ 包含少量样本关系和高频关系。背景知识图 $\mathcal{G}_{background}$ 是一组与所有高频关系相关的三元组。

知识图谱补全。KGK 任务是预测给定头实体 h 和查询关系 r：$(h, r, ?)$ 的尾部实体 t，或者是预测两个现有实体之间的未见关系 r：$(h, ?, t)$。在这项工作中，我们专注于尾实体预测。

少样本知识图谱补全。给定一个关系 $r \in \mathcal{R}$ 及其少量样本支持集 $\mathcal{S} = \{(h_i, t_i) \in \mathcal{T}\}$，FKGC 任务旨在为每个查询 $\mathcal{Q} = \{(h_i, ?) \in \mathcal{T}\}$ 预测尾实体 t。

少数样本关系的邻域。给定一个三元组 (h, r, t) 的少样本关系 r，r 的邻域定义为 $\{h, t, \mathcal{N}_h, \mathcal{N}_t\}$，其中 $\mathcal{N}_h$ 和 $\mathcal{N}_t$ 分别是 h, t 的一跳邻居的集合。所有 $\mathcal{N}_h$ 和 $\mathcal{N}_t$ 都来自背景知识图谱 $\mathcal{G}_{background}$。在 $\mathcal{N}_h$ 或 $\mathcal{N}_t$ 中的邻域由一个相邻关系 r_i 和一个相邻实体 c_i 组成。我们将每个实体（h 或 t）的邻居表示为 $\mathcal{N}_e = \{(r_i, c_i) | (e, r_i, c_i) \in \mathcal{G}_{background}\}$。

3.2 元学习设置

元学习旨在让模型在多个相关任务上进行训练，以便该模型可以使用少量训练数据快速学习新任务。我们利用了一种基于优化的元学习算法，称为 MAML[5]，其目标是通过使用预先良好初始化的元模型参数集 $\mathbf{\Theta}$ 来学习特定于任务的参数集 $\mathbf{\Theta}_i$。它可以分为两个阶段：元训练和元测试。在元训练期间，给定一个任务 $\mathcal{T}_i$，首先从 $\mathcal{T}_i$ 中采样出支持集 $\mathcal{S}_i$ 和查询集 $\mathcal{Q}_i$。然后，模型通过在支撑集 $\mathcal{S}_i$ 上进行一次梯度下降更新来学习一个任务特定的参数集 $\mathbf{\Theta}_i$：

$$\mathbf{\Theta}_i = \mathbf{\Theta} - \eta * \nabla \mathcal{L}_{\mathcal{S}_i}(\mathbf{\Theta}). \tag{1}$$

最后，通过对所有查询集任务进行元优化来学习元模型参数集 $\mathbf{\Theta}$，使用特定于任务的参数集 $\mathbf{\Theta}_i$。在元测试期间，该模型可以仅使用支持集 $\mathcal{S}$ 快速适应新任务。

在 FKGC 中，每个任务被定义为预测特定少样本关系的新三元组。元训练中的所有关系构成一个元训练集 $\mathcal{R}_{meta-training}$。由于目标是预测未见过的关系的事实，因此元验证 $\mathcal{R}_{meta-validation}$、元测试 $\mathcal{R}_{meta-testing}$ 和 $\mathcal{R}_{meta-training}$ 中的关系各不相同。

4 方法论

本节首先介绍三元组表示，旨在学习上下文相关的实体表示和一个好的初始化少样本关系表示。然后我们介绍一种新颖的负采样策略，目的是过滤无关的候选尾实体，并使用注意力机制区分每个负样本的重要性。最后，我

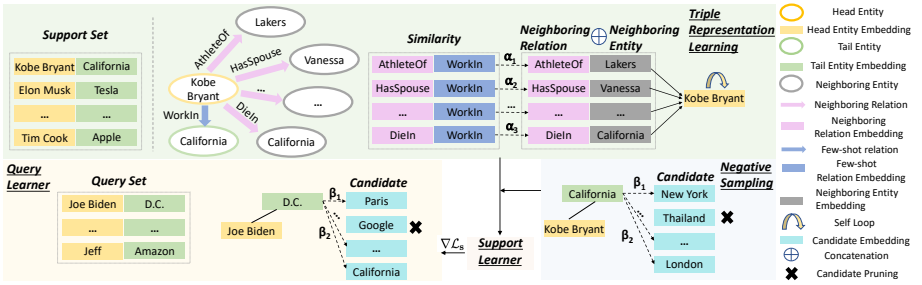


图 1: 瑞纳 的少量样本关系框架工作在

们介绍了元学习，旨在学习具有良好泛化能力的参数，从而使模型能够快速适应新的任务，只需少量参考三元组。图 1显示了少样本关系工作中的的 RANA 框架。

4.1 三重表示法

动态关系感知实体编码器。实体表示应依赖于上下文。例如，*(科比 · 布莱恩特，加利福尼亚)* 可以参与两种不同的关系，如工作内容和死亡，因此在这两种不同的关系上下文中，科比 · 布莱恩特应具有不同的嵌入。

此外，给定一个少量样本的关系，不同的邻居应对实体本身有不同的影响。例如，在图 1中，给定少样本关系工作内容和头实体科比 · 布莱恩特，其邻居 *(运动员 属于, 湖人队)* 应该获得更多关注，因为它揭示了关于科比 · 布莱恩特的工作信息，但邻居 *(有配偶, Vanessa)* 应该获得较少关注，因为这揭示了与少样本关系工作在无关的科比 · 布莱恩特的家庭信息。

为了应对这些问题，我们设计了一个动态关系感知实体编码器，该编码器通过结合邻近关系来学习实体在不同关系中的不同嵌入，并通过注意力机制区分每个邻居的重要性。

给定一个来自支持集 $\mathcal{S}$ 的实体对 (h, t)，少样本关系 r 的嵌入定义为:

$$\boldsymbol{r} = \boldsymbol{t} - \boldsymbol{h}, \tag{2}$$

其中 $\boldsymbol{h}$ 和 $\boldsymbol{t}$ 是由 TransE[2] 预训练的嵌入。

这里，我们以头实体 h 为例说明实体编码过程，此过程也适用于尾实体 t。

为了区分每个邻居的影响，我们使用多层感知器（MLP）网络来计算少量示例关系 r 与每个相邻关系 r_i 的相关性得分。

相关性得分定义如下：

$$m(r, r_i) = \boldsymbol{W_2}[\boldsymbol{tanh}(\boldsymbol{W_1}[\boldsymbol{r} \oplus \boldsymbol{r_i}])], \tag{3}$$

其中 $\oplus$ 表示连接操作，$\boldsymbol{r_i}$ 表示邻近关系的嵌入，而 $\boldsymbol{W_1}$ 和 $\boldsymbol{W_2}$ 是可训练参数。较高的相邻关系与少量样本关系之间的相关性分数意味着这种相邻关系对少量样本关系更为重要。

为了学习实体在不同关系中的不同表示，我们设计了头实体 h 的动态邻近嵌入 $\boldsymbol{A_{r_i,c_i}}$ 如下：

$$\boldsymbol{A_{r_i,c_i}} = \sum_{(r_i,c_i)\in\mathcal{N}_h} \alpha_i \boldsymbol{W_3}[\boldsymbol{r_i} \oplus \boldsymbol{c_i}], \tag{4}$$

其中 $\boldsymbol{W_3}$ 是可训练参数，α_i 是每个邻居的注意力分数：

$$\alpha_i = \frac{exp(m(r, r_i))}{\sum_{r_i\in\mathcal{N}_h} exp(m(r, r_i))}. \tag{5}$$

当相邻关系与少量示例关系更相关时，相应的邻居将获得更高的注意力 α_i。然后这个邻居将在邻近嵌入中扮演更重要的角色。

由于实体 h 自身的信息仍然有价值，我们将实体 h 的嵌入与 $\boldsymbol{A_{r_i,c_i}}$ 结合以得到最终表示 $\boldsymbol{h'}$ 如下：

$$\boldsymbol{h'} = \sigma(\boldsymbol{W_4}(\boldsymbol{h} + \boldsymbol{A_{r_i,c_i}})), \tag{6}$$

其中 $\boldsymbol{W_4}$ 是可训练参数而 $\sigma(\cdot)$ 是一个 sigmoid 函数。**少样本关系表示**。同一实体对可能涉及不同的关系，因此学习关系表示是必要的，并且它还能进一步帮助三元组表示的学习。

支持集中特定实体对的关系表示为：$\mathcal{S}$

$$\boldsymbol{R_{(h_i,t_i)}} = FC^{\sigma}_{\boldsymbol{W_5}}[\boldsymbol{h'_i} \oplus \boldsymbol{t'_i}], \tag{7}$$

其中，全连接层 $FC^{\sigma}_{\boldsymbol{W_5}}$ 的参数由 $\boldsymbol{W_5}$ 给出，并由 LeakyReLU 函数 $\sigma(\cdot)$ 激活。

支持集 $\boldsymbol{R^s}$ 中的关系表示是实体对在 $\mathcal{S}$ 中所有表示的平均值，

$$\boldsymbol{R^s} = \frac{\sum_{i=1}^{I} \boldsymbol{R_{(h_i,t_i)}}}{I}, \tag{8}$$

其中 I 是支持集中实体对的数量 $\mathcal{S}$。

4.2 负采样

由于在少样本设置下正样本有限，如何利用负样本变得更加关键。之前的 FKGC 方法使用基于边界的排序损失，并且为每个正样本随机选择一个负样本 [18, 4, 23, 12, 9]。但是随机选择可能会选到不相关的负样本并导致零损失问题。进一步来说，无论与正样本的相关性如何，所有负样本都会对模型训练产生相同的影响。为了解决这些问题，RANA 过滤掉无关的负样本，并使用注意力机制来区分每个负样本的重要性。

候选剪枝。由 [18] 构建的负样本候选集将候选实体限制为与支持集中真实尾部实体类型相同的那些，但这个广泛的候选集包括了许多无关的候选作为负样本。例如，给定一个事实 *(科比 · 布莱恩特, 工作地点, 加利福尼亚)*，之前的候选集被限制为地点和公司类型的实体，因为支持集中尾部实体的类型是公司或地点。然而，像泰国这样的候选人与加州无关，因此对模型训练没有帮助。

为了减少无关候选者的数量，并使模型在训练阶段能够选择高质量的负样本，瑞纳 通过真实尾实体 t 和候选尾实体 t^- 之间的相似性来过滤无关候选者。相似度由以下公式计算：

$$f(\boldsymbol{t}, \boldsymbol{t}^-) = \boldsymbol{t}^{-\boldsymbol{T}}\boldsymbol{t}, \tag{9}$$

其中 $\boldsymbol{t}$ 是真实尾实体的嵌入，$\boldsymbol{t}^-$ 是候选尾实体的嵌入。如果 $f(\boldsymbol{t}, \boldsymbol{t}^-) < \tau$，其中 τ 是一个阈值，那么应该过滤 t^-。

负样本的关注。为了充分利用负样本，RANA 选择多个负样本而不是一个，并通过注意力机制区分每个负三元组的贡献。

直观上，如果一个负样本与正样本更相关，这个负样本应在模型训练中扮演更重要的角色。因此，应给予这样的负样本更高的注意力。如图 1所示，给定一个正样本 *(科比 · 布莱恩特，加利福尼亚)*，负样本 *(科比 · 布莱恩特，纽约)* 比负样本 *(科比 · 布莱恩特，伦敦)* 与正样本更相关，因此模型应更多关注前者。

我们定义了一个缩放点积函数 $f(\boldsymbol{p_i}, \boldsymbol{n_{ij}})$ 来计算正样本 (h_i, t_i) 和其每一个负样本 (h_i, t_{ij}^-) 之间的相似度：

$$\boldsymbol{p_i} = \boldsymbol{h_i} \oplus \boldsymbol{t_i}, \qquad \boldsymbol{n_{ij}} = \boldsymbol{h_i} \oplus \boldsymbol{t_{ij}^-}, \qquad f(\boldsymbol{p_i}, \boldsymbol{n_{ij}}) = \frac{\boldsymbol{n_{ij}^T}\boldsymbol{p_i}}{\sqrt{|p|}}, \tag{10}$$

其中 $|p|$ 是 $\boldsymbol{p_i}$ 的维度。每个负三元组的注意力由以下定义：

$$\beta_{ij} = \frac{exp\, f(\boldsymbol{p_i}, \boldsymbol{n_{ij}})}{\sum_{j=1}^{J} exp\, f(\boldsymbol{p_i}, \boldsymbol{n_{ij}})}, \tag{11}$$

其中，J 是负样本的数量。

瑞纳 的损失。负采样在确定优化目标方面与正采样同样重要，但在基于边界的排名损失 [21] 中却被忽视了。为了缓解零损失和收敛速度较慢的问题，我们采样多个负三元组而不是一个，以增加生成相关负三元组的概率。

受 TransE [2] 的启发，我们首先按如下方式计算每个实体对 (h_i, t_i) 的距离：

$$d_{(h_i,t_i)} = ||\boldsymbol{h_i} + \boldsymbol{R} - \boldsymbol{t_i}||_{L2}, \tag{12}$$

由于较小的距离表示该三元组更有可能为真，因此该三元组应导致更高的分数。每个三元组的得分函数设计如下：

$$s_{(h_i,t_i)} = \gamma - d_{(h_i,t_i)}, \tag{13}$$

其中 γ 是一个超参数。

我们的基于日志的损失函数为：

$$\mathcal{L} = -\sum_{i=1}^{I} log\, \sigma(s_{(h_i,t_i)}) - \sum_{i=1}^{I}\sum_{j=1}^{J} \beta_{ij} log\, \sigma(-s_{(h_i,t_{ij}^-)}), \tag{14}$$

其中，$\sigma(\cdot)$ 是一个 sigmoid 函数，而 β_{ij} 则是通过方程 (11) 计算得到的每个负三元组的注意力。由于更相关的负三元组具有更高的注意力（β_{ij}），这个损失函数将使那些相关的负三元组在模型训练中产生更大的影响。

4.3 元学习

为了快速通过支持集学习新的关系，瑞纳 使用 MAML[5] 来优化可以适应少样本关系的模型参数。

支持学习者。支持学习者旨在学习少量样本关系的表示 $\boldsymbol{R^s}$，而 $\boldsymbol{R^s}$ 可以通过公式 (2)- 公式 (8) 计算。

查询学习者。遵循 MetaR 的 [4] 假设，关系是支持集和查询集之间的关键共同信息。因此我们旨在通过最小化损失函数来将支持集的关系 R^s 转移到查询集的关系 R^q 上。

Algorithm 1 训练框架

输入：训练任务 $\mathcal{R}_{meta-training}$，初始模型参数 $\boldsymbol{\Theta}$
预训练 KG 嵌入（不包括 $\mathcal{R}_{meta-training}$ 中的关系）

while not done **do**
  从 $\mathcal{R}_{meta-training}$ 中采样一个任务 $\mathcal{T}_i = \{\mathcal{S}_i, \mathcal{Q}_i\}$
  从 $\mathcal{S}_i$ 中通过方程 (2)- 方程 (8) 得到 $\boldsymbol{R^s}$
  通过方程 (9)- 方程 (11) 获取 $\mathcal{S}_i$ 的负样本
  通过方程 (12)- 方程 (14) 计算 $\mathcal{S}_i$ 的损失
  使用梯度下降和方程 (15) 更新特定任务关系 $\boldsymbol{R^q}$ 的嵌入
  通过方程 (9)- 方程 (11) 获取 $\mathcal{Q}_i$ 的负样本
  通过方程 (12)- 方程 (14) 计算 $\mathcal{Q}_i$ 的损失
  更新整个模型参数 $\boldsymbol{\Theta} \leftarrow \boldsymbol{\Theta} - \mu\nabla\mathcal{L}$
end while

在瑞娜 中，关系嵌入 $\boldsymbol{R^q}$ 可以通过梯度下降进行更新，

$$\boldsymbol{R^q} = \boldsymbol{R^s} - \eta * \boldsymbol{\nabla \mathcal{L}_s}, \tag{15}$$

其中超参数 η 表示步长，而 $\mathcal{L}_s$ 表示相应的支持集的损失，该损失由方程 (12)- 方程 (14) 计算。

为了更新所有参数瑞纳，我们使用更新的关系嵌入 $\boldsymbol{R^q}$ 来通过方程 (12) 到方程 (14) 计算相应查询集 $\mathcal{L}_q$ 的损失。

目标和训练过程。在元训练阶段，RANA 的目标是最小化所有任务的查询损失之和，总体损失为：

$$\mathcal{L} = arg \min_{\boldsymbol{\Theta}} \sum \mathcal{L}_q, \tag{16}$$

其中 $\boldsymbol{\Theta}$ 表示所有可训练的参数。

4.4 RANA 的算法

我们总结整体训练过程在算法 1中。

4.5 与 RotatE 的区别

RotatE [13] 是一种基于嵌入的 KGC 方法，它使用自对抗负采样技术来有效优化模型。我们的方法与 RotatE 在主要方面有所不同：我们将正样本三元组和负样本三元组之间的相似性视为每个负样本三元组的权重，而 RotatE

表 1: 数据集的统计。第 2 至 7 列分别表示实体数量、关系数量、三元组数量、在 $\mathcal{R}_{meta-training}$ 中的关系数量、在 $\mathcal{R}_{meta-validation}$ 中的关系数量以及在 $\mathcal{R}_{meta-testing}$ 中的关系数量。

Dataset	#Ent	#Rel	#Triples	#Train Rel	#Valid Rel	#Test Rel
NELL-One	$68,545$	358	$181,109$	51	5	11
Wiki-One	$4,838,244$	822	$5,859,240$	133	16	34

则考虑了负样本三元组的分布，并将概率作为每个负样本三元组的权重。因此，在 RotatE 中，负样本的权重与正样本无关。正如我们将在实验部分（第 5.5 节）所示，瑞娜 在少量示例设置下可以实现比 RotatE 自对抗负采样更好的性能。

5 实验

5.1 数据集和评估指标

我们在由 [18] 构建的 NELL-One 和 Wiki-One 数据集上进行了实验。在两个数据集中，选择三元组数量多于 50 但少于 500 的关系作为小样本关系，其余的关系被视为背景关系。我们分别使用 51/5/11 和 133/16/34 小样本关系用于训练/验证/测试 NELL-One 和 Wiki-One。两个数据集的统计信息显示在表 1中。为了评估 RANA 和所有基线的表现，我们使用了两个指标：平均倒数排名（MRR）和 Hits@K。MRR 是正确实体的平均倒数排名，Hits@K 是被排在前 k 位的正确实体的比例。

5.2 基线

传统的嵌入式方法力求通过建模知识图谱中的关系结构来学习实体和关系嵌入。我们将以下广泛使用的方法作为基线：TransE[2]，DistMult[19]，ComplEx[15]，SimplE[6] 和 RotatE[13]。所有这些方法都需要每个关系的充足训练三元组，并且不使用局部图结构来更新实体嵌入。

FKGC 方法 旨在通过利用深度神经网络来探索支持集和查询集之间的联系，从而学习长尾和未见的关系。我们将以下模型作为基线：GMatching[18]，MetaR[4]，FSRL[23]，FAAN[12]，GANA[9]，和 HiRe[17]。我们运行印度虎蛙 5 次并报告平均结果。

表 2: 5-shot KGC 的结果。**粗体**数字代表最佳结果，下划线 数字表示亚军结果。† 引用了来自 [12] 的结果，∗ 引用了他们原始论文中的结果。

	NELL-One				维基一号			
Model	平均倒数排名	命中率@10	命中率@5	命中率@1	平均倒数排名	击中率@10	命中率@5	命中率@1
TransE†	0.174	0.313	0.231	0.101	0.133	0.187	0.157	0.100
DistMult†	0.200	0.311	0.251	0.137	0.071	0.151	0.099	0.024
ComplEx†	0.184	0.297	0.229	0.118	0.080	0.181	0.122	0.032
SimplE†	0.158	0.285	0.226	0.097	0.093	0.180	0.128	0.043
RotatE†	0.176	0.329	0.247	0.101	0.049	0.090	0.064	0.026
GMatching†	0.176	0.294	0.233	0.113	0.263	0.387	0.337	0.197
MetaR†	0.209	0.355	0.280	0.141	0.323	0.418	0.385	0.270
FSRL†	0.153	0.319	0.212	0.073	0.158	0.287	0.206	0.097
FAAN†	0.279	0.428	0.364	0.200	0.341	0.463	0.395	0.281
GANA∗	0.344	0.517	0.437	0.246	0.351	0.446	0.407	0.299
HiRe∗	0.306	0.520	0.439	0.207	0.371	0.469	0.419	0.319
RANA	**0.361±0.011**	**0.573±0.009**	**0.475±0.010**	**0.253±0.013**	**0.379±0.008**	**0.480±0.012**	**0.437±0.008**	**0.329±0.011**

5.3 瑞纳 设置

预训练的实体和关系嵌入是从 TransE 获得的。嵌入维度分别设置为 NELL-One 的 50 和 Wiki-One 的 100。我们使用初始学习率为 0.01 的 Adam[7] 来更新参数。负样本的数量是 5, 边界 γ 是 12.0, 步长 η 是 1, 并且两个数据集上的邻居数量都是 25。在验证集上 MRR 最高的模型被用作最终模型。最优超参数通过网格搜索在验证集上调优。我们在配备特斯拉 V100 GPU（32G）的服务器上进行 RANA 。

5.4 总体评估结果与分析

所有模型在 NELL-One 和 Wiki-One 上的表现如表 2所示。与传统的基于嵌入的方法相比，结合图邻居对于学习实体嵌入是有效的。瑞纳 在这两个数据集上都优于其他 FKGC 模型。与第二名的结果相比，在 MRR、Hits@10、Hits@5 和 Hits@1 方面，瑞纳 分别在 NELL-One 上的提升为 4.9%、10.2%、8.2%、2.8%，在 Wiki-One 上的提升分别为 2.2%、2.3%、4.3%、3.1%。

5.5 消融研究

非洲爪蟾 由两个模块组成，包括一个动态关系感知实体编码器和负采样。为了研究每个组件的贡献，我们在不同的设置下进行了 5-shot KGC 实验。结果总结在表 3中。

表 3: 消融研究

	奈尔-一				Wiki-壹			
Model	平均倒数排名	命中率@10	命中率@5	命中率@1	均召回率	命中率@10	命中率@5	命中率@1
whole model	**0.372**	**0.580**	**0.477**	**0.257**	**0.387**	**0.486**	**0.443**	**0.339**
equation(4) w/o r_i	0.339	0.535	0.427	0.222	0.362	0.468	0.410	0.299
equation(5) with c_i	0.358	0.573	0.471	0.256	0.367	0.477	0.424	0.302
equation(4) w/o α_i	0.326	0.526	0.407	0.235	0.377	0.483	0.433	0.315
equation(14) with one negative sample	0.294	0.520	0.428	0.210	0.349	0.451	0.417	0.311
equation(14) w/o negative attention	0.293	0.494	0.416	0.213	0.298	0.387	0.371	0.257
w/o candidate pruning	0.298	0.507	0.425	0.217	0.311	0.445	0.360	0.243
w/o candidate pruning and negative attention	0.257	0.447	0.396	0.192	0.286	0.363	0.321	0.242
RotatE self-adversarial negative sampling	0.268	0.479	0.365	0.165	0.310	0.389	0.401	0.255

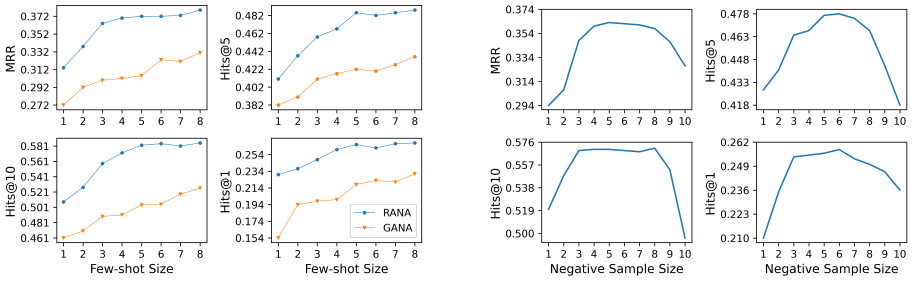
实体编码变体：我们分析了方程中邻近关系的影响。(4) 和 (5) 通过从方程 (4) 中移除 r_i 或在方程 (5) 中添加 c_i 而得。此外，我们从方程 (4) 中去除了注意力机制。结果表明邻近关系和注意力机制可以提升模型性能。这说明关系的语义信息能够改善实体表示，并且不同的关系应对实体本身有不同的影响。由于注意力机制的效果依赖于邻居，Wiki-One 的邻居比 NELL-One[9] 更稀疏，因此注意力机制在 Wiki-One 中的作用较小。

负采样变体：为了检验负采样和基于注意力的损失函数的有效性，我们进行了五项不同的实验。(A) 我们仅在方程 (14) 中使用一个负样本。(B) 我们从方程 (14) 中移除了负注意力机制。(C) 我们移除了候选剪枝阶段。(D) 我们同时移除了候选剪枝阶段和负注意力机制。(E) 我们用 RotatE[13] 自对抗负采样损失替换了方程 (14)。实验结果表明，负采样策略在瑞纳 的成功中起着关键作用。

5.6 少样本支持集大小和负样本的影响

为了分析支持集大小的影响，我们将瑞娜 与 GANA 在 NELL-one 上进行比较。图 2a显示了支持集大小从 1 到 8 的性能。RANA 在不同大小的支持集下优于 GANA，显示出 RANA 的有效性。经过 5 次射击后，RANA 的改进不显著。我们从关系团队教练中随机选择 20 个事实来分析 5 次射击设置中的错误。青蛙 在前 10 个预测中正确预测了 20 个尾实体中的 12 个。在其余的 8 个事实中，有 4 个的事实具有不正确的地面真实尾实体，另外 3 个的事实邻居少于 10 个。对于这些情况，增加支持集的大小不太可能改变结果。

我们进行了一项实验来分析负样本数量的影响。图 2b显示了瑞娜 在 NELL-One 上的性能，负样本数量从 1 到 10。当增加负样本数量时，性能最



(a) (b)

图 2: (a) 少样本支持集大小的影响,(b) 负样本大小的影响

初有所提高。在超过大小 6 后，由于类别不平衡问题，性能开始下降。经验上，我们推荐使用负样本数量为 3 到 5。

6 结论

在本文中，我们提出了一种带有基于注意力损失的关联感知网络用于 FKGC 任务。我们战略性地选择多个负样本而不是一个，并提出了基于注意力的损失来区分每个负样本的重要性。设计了一个动态关系感知实体编码器以学习依赖上下文的实体表示。实验结果表明，瑞纳 在两个基准数据集上优于其他 SOTA 基线。

参考文献

1. Bordes, A., Glorot, X., Weston, J., Bengio, Y.: A semantic matching energy function for learning with multi-relational data. Machine Learning **94**(2), 233–259 (2014)
2. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., Yakhnenko, O.: Translating embeddings for modeling multi-relational data. In: NeurIPS **26**, 2787–2795 (2013)
3. Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka, E., Mitchell, T.: Toward an architecture for never-ending language learning. In: AAAI. vol. 24, pp. 1306–1313 (2010)
4. Chen, M., Zhang, W., Zhang, W., Chen, Q., Chen, H.: Meta relational learning for few-shot link prediction in knowledge graphs. In: EMNLP-IJCNLP. pp. 4217–4226 (2019)
5. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: ICML. pp. 1126–1135. PMLR (2017)

6. Kazemi, S.M., Poole, D.: Simple embedding for link prediction in knowledge graphs. In: NeurIPS **31** (2018)
7. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2015)
8. Nickel, M., Tresp, V., Kriegel, H.P.: A three-way model for collective learning on multi-relational data. In: ICML (2011)
9. Niu, G., Li, Y., Tang, C., Geng, R., Dai, J., Liu, Q., Wang, H., Sun, J., Huang, F., Si, L.: Relational learning with gated and attentive neighbor aggregator for few-shot knowledge graph completion. In: SIGIR. pp. 213–222 (2021)
10. Rong, Y., Huang, W., Xu, T., Huang, J.: Dropedge: Towards deep graph convolutional networks on node classification. In: ICLR (2019)
11. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: CVPR. pp. 815–823 (2015)
12. Sheng, J., Guo, S., Chen, Z., Yue, J., Wang, L., Liu, T., Xu, H.: Adaptive attentional network for few-shot knowledge graph completion. In: EMNLP. pp. 1681–1691 (2020)
13. Sun, Z., Deng, Z.H., Nie, J.Y., Tang, J.: Rotate: Knowledge graph embedding by relational rotation in complex space. In: In ICLR (2018)
14. Tang, Z., Pei, S., Zhang, Z., Zhu, Y., Zhuang, F., Hoehndorf, R., Zhang, X.: Positive-unlabeled learning with adversarial data augmentation for knowledge graph completion. In: IJCAI. pp. 1935–1942 (2022)
15. Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., Bouchard, G.: Complex embeddings for simple link prediction. In: ICML. pp. 2071–2080. PMLR (2016)
16. Vrandečić, D., Krötzsch, M.: Wikidata: a free collaborative knowledgebase. CACM (2014)
17. Wu, H., Yin, J., Rajaratnam, B., Guo, J.: Hierarchical relational learning for few-shot knowledge graph completion. arXiv preprint arXiv:2209.01205 (2022)
18. Xiong, W., Yu, M., Chang, S., Guo, X., Wang, W.Y.: One-shot relational learning for knowledge graphs. In: EMNLP. pp. 1980–1990 (2018)
19. Yang, B., Yih, S.W.t., He, X., Gao, J., Deng, L.: Embedding entities and relations for learning and inference in knowledge bases. In: ICLR (2015)
20. Yang, S., Tian, J., Zhang, H., Yan, J., He, H., Jin, Y.: Transms: Knowledge graph embedding for complex relations by multidirectional semantics. In: IJCAI. pp. 1935–1942 (2019)
21. Yang, Z., Ding, M., Zhou, C., Yang, H., Zhou, J., Tang, J.: Understanding negative sampling in graph representation learning. In: SIGKDD. pp. 1666–1676 (2020)
22. Zha, H., Chen, Z., Yan, X.: Inductive relation prediction by bert. In: AAAI (2022)
23. Zhang, C., Yao, H., Huang, C., Jiang, M., Li, Z., Chawla, N.V.: Few-shot knowledge graph completion. In: AAAI. vol. 34, pp. 3041–3048 (2020)

24. Zhou, K., Li, Y., Li, Q.: Distantly supervised named entity recognition via confidence-based multi-class positive and unlabeled learning. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 7198–7211 (2022)

25. Zhou, K., Qiao, Q., Li, Y., Li, Q.: Improving distantly supervised relation extraction by natural language inference. In: Proceedings of the AAAI conference on artificial intelligence. vol. 37, pp. 14047–14055 (2023)