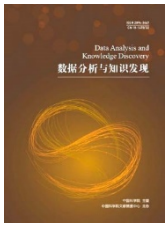

数据分析与知识发现
Data Analysis and Knowledge Discovery
ISSN 2096-3467,CN 10-1478/G2

《数据分析与知识发现》网络首发论文

题目：　结合类型感知注意力的少样本知识图谱补全
作者：　普祥和，王红斌，线岩团
网络首发日期：　2022-11-02
引用格式：　普祥和，王红斌，线岩团．结合类型感知注意力的少样本知识图谱补全[J/OL]．数据分析与知识发现.
https://kns.cnki.net/kcms/detail/10.1478.G2.20221101.1617.002.html

网络首发时间：2022-11-02 10:18:17
网络首发地址：https://kns.cnki.net/kcms/detail/10.1478.G2.20221101.1617.002.html

结合类型感知注意力的少样本知识图谱补全

普祥和[1,2,3]，王红斌[1,2,3]，线岩团[1,2,3]
[1](昆明理工大学 信息工程与自动化学院 昆明 650500)
[2](昆明理工大学 云南省人工智能重点实验室 昆明 650500)
[3](昆明理工大学 云南省计算机技术应用重点实验室 昆明 650500)

摘要：[目的]现有少样本知识图谱补全方法在处理复杂关系时不能很好地区分邻居重要性，导致预测性能不佳，考虑充分利用实体邻居信息，提高少样本知识图谱补全方法的性能。**[方法]**通过类型感知邻居编码器学习实体邻居中包含的隐含类型信息，得到类型感知注意力，增强实体表示；利用 Transformer 编码器捕获任务关系的不同含义；通过联合匹配原型网络聚合参考集得到参考集表示并进行实体预测。**[结果]**在 NELL 和 Wiki 两个公共数据集上通过实体预测任务进行了实验验证，实验结果表明，MRR 指标分别较 baseline 方法提高了 1.6 和 1.2 个百分点。**[局限]**未对与实体相关性较低的邻居进行筛选，使得类型感知注意力权重的分配受噪声影响。**[结论]**实验证明本文方法能够通过学习更丰富的实体邻居信息来有效提高少样本知识图谱补全的性能。

关键词：少样本知识图谱补全；实体预测；类型感知注意力
分类号：TP391
DOI: 10.11925/infotech.2096-3467.2022.0799

Few-Shot Knowledge Graph Completion Combined with Type-Aware Attention

Pu Xianghe[1,2,3], Wang Hongbin[1,2,3], Xian Yantuan[1,2,3]
[1](Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)
[2] (Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, China)
[2] (Yunnan Key Laboratory of Computer Technology Application, Kunming University of Science and Technology, Kunming 650500, China)

Abstract：[Objective]The existing few-shot knowledge graph completion methods cannot distinguish the importance of neighbors well when dealing with complex relations, resulting in low prediction performance. Consider making full use of entity neighbor information to improve the performance of the few-shot knowledge graph completion methods. **[Methods]**First, we use type-aware neighbor coder to learn the implicit type information contained in the entity neighbor to obtain type-aware attention and enhance entity representation. Then, use Transformer encoder to capture different meanings of task relation. Finally, the reference set representation is obtained by jointly matching prototype network aggregation reference set, then make entity prediction. **[Results]**The proposed method is tested on NELL and Wiki datasets through entity prediction

tasks. The experimental results show that the MRR is 1.6 and 1.2 percentage points higher than the baseline method, respectively. **[Limitations]** Neighbors with lower physical relevance were not filtered, causing the distribution of type-aware attention weights affected by noise. **[Conclusions]** Experiments show that this method can effectively improve the performance of few-shot knowledge graph completion by learning more abundant entity neighbor information.
Keywords：Few-Shot Knowledge Graph Completion; Entity Prediction; Type-Aware Attention

1 引言

随着知识图谱的不断发展，目前已被广泛应用于智能问答、推荐系统、搜索引擎等不同领域[1]。对于大型知识图谱或知识库，如FreeBase[2]、DBpedia[3]、NELL[4]和Wikidata[5]等，虽然它们包含了大量关于现实世界的事实三元组，但仍然存在不完整问题，如图1所示，三元组(sportsteam:raptors, teamcoach, ?)缺少了相应的尾实体，通过人工查找方式来逐个补全这些缺失的三元组代价高昂且不切实际。为了解决这一问题，知识图谱补全任务应运而生。在实际应用数据中，往往还存在着大量出现频次很低的关系，即少样本关系，这些关系只出现在很少的三元组中，越是这些少样本关系，越有可能需要进行补全。传统知识图谱补全方法在处理这些关系时，由于缺乏足够的训练数据导致方法性能不佳。

近年来，一些新的方法被提出[6-10]，这些方法通过注意力机制等方式显示建模实体的图结构特征来增强实体表示，以解决少样本知识图谱补全问题。然而，这些方法对实体的图结构特征进行建模时没有充分利用实体邻居信息，面对数据中的一对多、多对多等复杂情况，不能很好地区分不同邻居的重要性。

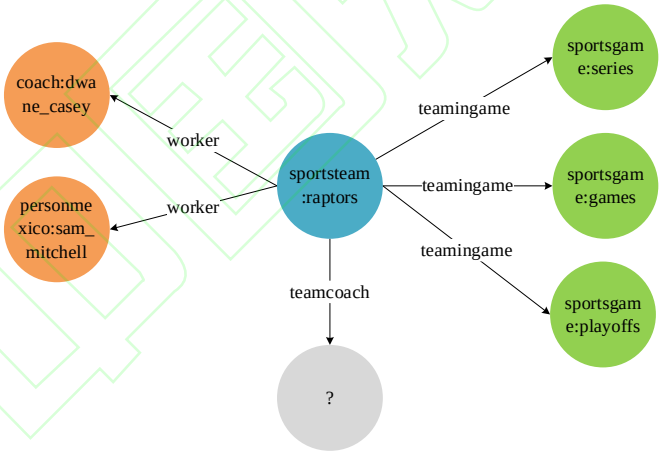

图 1 实体 raptors 及其邻居示意图

Fig. 1 Entity raptors and the neighbors

图 1 为实体 raptors 及其邻居示意图，邻居包括邻居关系和邻居实体。对于当前待补全的三元组(sportsteam:raptors, teamcoach, ?)，任务关系为 teamcoach，表示 raptors 的教练，左半邻居通过关系“worker”与 raptors 相连，表示 raptors 的员工，右半邻居则通过关系“teamingame”相连，表示 raptors 参与的比赛。FAAN 模型[9]认为，“教练”与“员工”描述的都是与 raptors 相关的人物属性，因此，将更关注左半邻居。但左半邻居均通过关系“worker”与 raptors 相连，计算得到邻居关系和任务关系的相似度相同，故分配的邻居权重也相等。通过观察两个邻居实体发现，dwane_casey 在与 raptors 相关的邻居中存在隐含类型“coach”，而 sam_mitchell 在与 raptors 相关的邻居中存在隐含类型“personmexico”，由于

“coach”与任务关系teamcoach更相似，故在进行补全时需更多关注dwane _casey所在的邻居。

为了说明邻居关系相同但邻居实体不同的复杂情况，本文统计了数据集NELL和Wiki中存在这一复杂情况的实体规模，如表1和表2所示。其中，用m表示一个实体具有一组m个邻居，它们的邻居关系相同但邻居实体不同。

表1 数据集NELL中的复杂实体规模

Table 1 Scale of complex entity in dataset NELL

m	2	3	4	10+
实体数	10173	3671	1886	2204

表2 数据集Wiki中的复杂实体规模

Table 2 Scale of complex entity in dataset Wiki

m	2	3	4	10+
实体数	111012	21233	5879	14047

由表1和表2可知，在数据集NELL中，有10173个实体至少具有一组两个邻居，它们的邻居关系相同但邻居实体不同；在数据集Wiki中有14047个实体至少具有一组十个以上的邻居，它们的邻居关系相同但邻居实体不同。由此可见，知识图谱中普遍存在着与图1类似的邻居结构复杂的情况，仅凭邻居关系不能很好地区分不同邻居的重要性。因此，本文认为，实体邻居中存在隐含的类型表示，在对邻居进行编码时关注实体邻居的类型表示，能够对不同邻居进行进一步区分，从而更好地聚合邻居信息来增强实体表示，进而改善少样本知识图谱补全的性能。基于此，本文提出一种结合类型感知的少样本知识图谱补全方法，用于学习邻居实体中隐含的类型信息，得到类型感知注意力，以解决先前方法在学习邻居特征过程中遇到一对多、多对多复杂情况时，不能很好地区分邻居重要性的问题。

2 相关工作

知识图谱补全(Knowledge Graph Completion，KGC)旨在通过机器学习方法来自动寻找补全缺失的三元组。通过知识图谱补全可以缓解知识图谱中数据稀疏的问题，使知识图谱更完整，能够更好的应用于相应的领域中，因此知识图谱补全成为了新的研究热点之一。常用的知识图谱补全方法为基于知识表示学习的方法，又称知识图谱嵌入[11]。基于知识表示学习的模型可以分为基于距离的模型，双线性模型和神经网络模型。

基于距离的模型将关系建模为头实体到尾实体的距离变换，并通过变换后的距离差来定义得分函数。典型的基于距离的模型为TransE模型[12]，其假设$\boldsymbol{h}+\boldsymbol{r}\approx\boldsymbol{t}$，即对于一个三元组，如果头实体向量和关系向量之和与尾实体向量越接近，那么该三元组越有可能是一个正确的三元组，反之则可能是一个错误的三元组。TransH模型[13]在TransE基础上，将实体投影到一个超平面，并将关系视为两个投影之间的距离变换，以解决TransE无法处理一对多和多对多情况的问题。TransR[14]模型将实体和关系投影到不同的向量空间中，并将关系建模为关系空间中两个实体投影的距离变换。TransD[15]模型则进一步建模了关系的多语义情况。为更好地建立对称和反对称关系模型，RotatE[16]模型将每个关系定义为复数量空间中头实体到尾实体的旋转。HAKE[17]模型则将实体和关系映射到极坐标系上以建模知识图谱中存在的语义分层现象。SimplE[18]模型通过张量分解的方法，独立学习每个实体的两个嵌入，使得到的嵌入具有可解释性。

双线性模型[19]采用乘积形式的得分函数来衡量实体和关系的语义相关性。典型的双线性模型为RESCAL[20]模型，RESCAL模型使用三阶张量来对实体和关系进行表示学习。为了降低计算量，DistMult[21]模型将矩阵 $\boldsymbol{M}_r$ 替换为对角矩阵，HolE[22]模型则使用循环相关运算来代替中间的关系矩阵乘积。ComplEx[23]模型则通过引入复数嵌入来扩展DistMult模型，能够更好地建模对称和反对称关系。

神经网络模型将 $\boldsymbol{h}$, $\boldsymbol{r}$ 和 $\boldsymbol{t}$ 同时输入神经网络，通过神经网络来判断三元组的得分。ConvE[24]模型使用卷积神经网络来定义得分函数，此外，也有一些工作根据知识图谱图结构的特点，通过图神经网络GNN来结合知识图谱进行知识图谱补全工作。例如TransGCN[25]模型利用关系嵌入对图结构中节点的现有实体嵌入进行变换，使其成为一个齐次图。通过定义两个变换运算符，可以更清楚地描述变换。

传统基于知识表示学习的方法通常需要对所有关系有足够多的训练三元组，因此在解决少样本问题时，模型的性能会受到限制。先前的少样本学习研究主要集中在计算机视觉、模仿学习和情感分析等领域，受此启发，最近的一些工作试图将少样本学习方法应用于知识图谱中的长尾关系，即少样本知识图谱补全。

Xiong等人提出了一种匹配网络GMatching[6]，利用邻居编码器编码实体的单跳邻居来增强实体嵌入，并使用一个LSTM匹配处理器通过LSTM块执行多步匹配，但其假设所有邻居对实体嵌入的贡献相等，导致分配的邻居权重完全相等。Wang等人提出B-GMatching方法[10]引入BERT预训练语言模型来增强GMatching中实体和关系的语义表示。Chen等人提出了一种新的元关系学习框架MetaR[7]，该框架通过在任务之间提取共享知识并将其从一些现有事实转移到不完整的事实中。为了改进参考集的语义表示，Zhang等人提出FSRL[8]模型将GMatching扩展到少样本情况下，通过注意力机制进一步捕获局部图结构。虽然FSRL模型在基准数据集Wiki和NELL上取得了很好的效果，但Sheng等人注意到，上述模型为邻居分配的权重在所有任务关系中都不会改变，这些工作都将静态权重分配给了邻居，在涉及不同任务关系时只能得到静态的实体表示，而实际上实体的邻居可能会对不同的任务关系产生不同的影响，为了解决这一问题，Sheng等人提出了少样本知识图谱补全中的动态属性概念，并设计了一个自适应注意力网络FAAN[9]来学习动态表征。FAAN通过一个自适应邻居编码器动态调整实体的表示以适应不同的任务关系，并通过Transformer编码器使参考集的表示适应不同的查询。然而，这一方法在面对一对多、多对多等复杂关系时，仍然不能很好地区分邻居重要性，影响了少样本知识图谱补全的性能。

针对该问题，本文提出用于少样本知识图谱补全的类型感知注意力网络，通过一个类型感知邻居编码器，在给定任务关系及其参考和查询三元组的情况下，学习邻居实体中隐含的类型信息，得到类型感知注意力，以进一步区分不同实体邻居的重要性，增强实体嵌入表示，并在参考集聚合过程中，通过实体级聚合和关系级聚合以共同帮助查询对选择与其更相似的参考集，进而解决先前方法在学习单跳邻居特征过程中遇到一对多、多对多复杂情况时，不能很好的区分邻居重要性的问题，具体方法细节见第 4 节。

3 任务定义

给定一个知识图谱 G ， G 中包含一系列三元组 $T=\{(h,r,t)\in E\times R\times E\}$ ，其中 E 和 R 分别表示实体集和关系集。知识图谱补全任务是在给定头实体 h 和查询关系 r $(h,r,?)$ 的情况下，预测尾实体 t ，或预测给定头实体 h 和尾实体 t 之间看不见的

关系$(h,?,t)$。本文工作中主要针对前一种情况，即预测给定关系下的未知事实。少样本知识图谱补全的目的是在给定与关系r相关的少量实体对$(h_k,t_k)\in R_r$的情况下，使得真实尾实体t_{true}的排名高于假候选实体t_{false}。形式上遵循先前工作对任务的定义：

给定一个关系$r\in R$，并给出它的参考集$S_r=\{(h_k,t_k)\,|\,(h_k,r,t_k)\in T\}$，当$|S_r|=K$且$K$很小的情况下，给定$(h,r,?)$并从候选实体集$C$中预测$t$，该任务称为$K$样本知识图谱补全。

该任务的目标是在给定少样本参考实体对S_r的情况下，使真实尾实体的排名高于假候选实体。为了模拟该任务，每个训练任务对应一个关系$r\in R$，r具有自己的参考和查询实体对，即$D_r=\{S_r,Q_r\}$，其中S_r仅由K个参考实体对$(h_k,t_k)\in R_r$组成。查询集$Q_r=\{h_m,t_m\,/\,C_{h_m,r}\}$包含真实尾实体$t_m$和相应的候选实体$C_{h_m,r}$。每个候选项都是基于实体类型约束进行选择得到的实体。因此，可以通过给出$(h_m,r,?)$和它的参考集S_r，对候选集$C_{h_m,r}$中的实体进行排序来训练模型。所有训练过程中的任务构成元训练集，记为$T_{\text{mtr}}=\{D_r\}$。

通过元训练集对模型进行训练后，得到的模型可以用来预测测试集中出现的新关系$r'\in R'$的相关事实。测试集中的关系在元训练集中不可见，即$R'\cup R=\varnothing$。每个测试关系r'也有自己的一些少样本参考对和查询对，即$D_{r'}=\{S_{r'},Q_{r'}\}$。所有测试过程中的任务构成元测试集，记为$T_{\text{mte}}=\{D_{r'}\}$。此外，模型还可以访问一个背景知识图谱$G'$，它是一个$G$的子集，但对所有$T_{\text{mtr}}$和$T_{\text{mte}}$都不可见。

4 本文方法

本文方法整体框架如图2所示，主要由三个部分组成：图2左半部分为类型感知邻居编码器，用于从邻居实体和邻居关系中学习邻居的隐含类型信息并增强实体的向量表示；图2的中间部分为Transformer编码器，用于学习从类型感知邻居编码器中得到的实体对特定于任务关系的向量表示；图2右半部分为联合匹配原型网络，用于聚合参考对得到参考集，并将查询对与参考集进行比较，从而学习一个用于预测的度量函数。

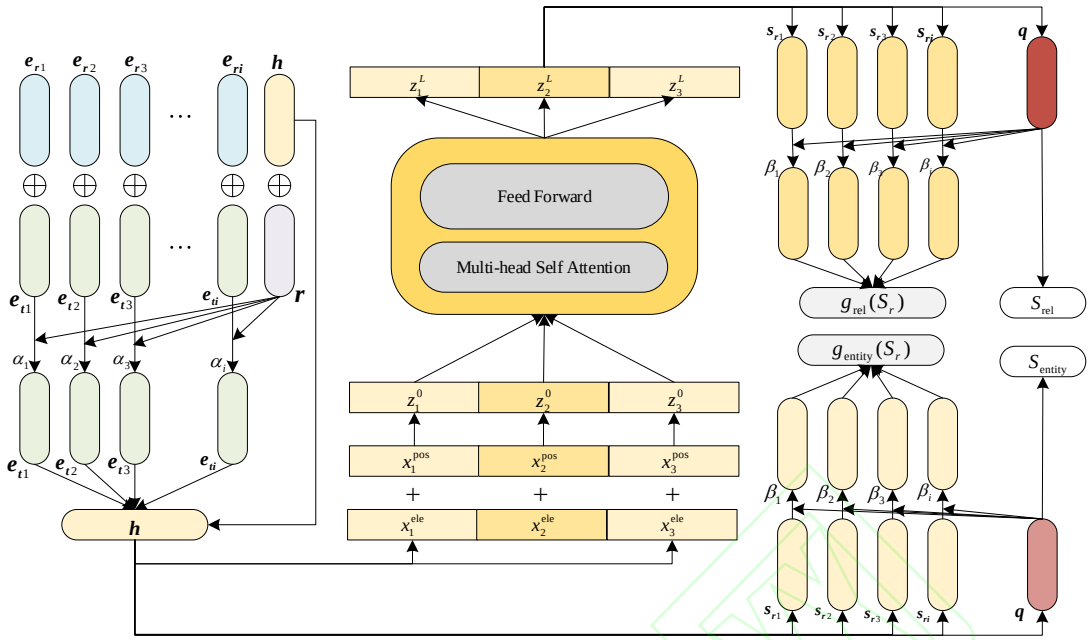

图 2 本文模型整体结构图

Fig. 2 Overall framework of the model

4.1 类型感知邻居编码器

在先前的少样本知识图谱补全工作中，首先需要对实体邻居进行编码，通过注意力机制改进图结构信息的学习过程，并进一步考虑实体的动态特性，增强实体的嵌入表示。现有的少样本知识图谱补全方法相比传统基于知识表示学习的方法，虽然能够从很少的样本中学到有用的信息，且取得了可观的效果，但这些方法仍没有考虑到邻居实体中包含的隐含类型信息，面对一对多、多对多等复杂情况时，不能很好的区分不同邻居的重要性。为了解决这一问题，本文设计了一个类型感知邻居编码器，用于联合邻居实体和邻居关系的信息来学习邻居类型表示，并将其扩展到多头，进而改进增强实体嵌入表示的过程。

给定与任务关系 r 相关的实体对 (h,t)，记 $N_h=\{(e_r,e_t)|(h,e_r,e_t)\in G'\}$ 为头实体 h 的单跳邻居，e_r 和 e_t 分别表示 h 的邻居关系和邻居实体。类型感知邻居编码器需要通过学习邻居关系 e_r 和邻居实体 e_t 的特征，得到类型表示，并根据类型表示来分配邻居权重对邻居进行编码，最终输出 h 的表示。如图1所示，对于邻居关系worker，dwane_casey特定于worker的类型为教练，而sam_mitchell特定于worker的类型为人事经理。因此，本文提出邻居实体特定于邻居关系的类型表示 $\boldsymbol{c}_i$ 如式(1)所示：

$$\boldsymbol{c}_i=\tanh(\boldsymbol{W}_1(\boldsymbol{e}_{ri}\oplus\boldsymbol{e}_{ti})+\boldsymbol{b}_1) \tag{1}$$

其中，$\oplus$ 为向量拼接操作，$\boldsymbol{W}_1\in\mathrm{R}^{2d\times d}$ 与 $\boldsymbol{b}_1$ 为可训练参数，则 $\boldsymbol{c}_i$ 为第 i 个邻居实体特定于与其对应邻居关系的类型表示。

对于给定的实体对 $(\boldsymbol{h},\boldsymbol{t})$，首先通过TransE模型将任务关系 $\boldsymbol{r}$ 进行表示，如式(2)所示[12]，并提出头实体 $\boldsymbol{h}$ 特定于任务关系 $\boldsymbol{r}$ 的类型表示 $\boldsymbol{c}$ 如式(3)所示：

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

$$\boldsymbol{c}=\tanh(\boldsymbol{W}_2(\boldsymbol{h}\oplus\boldsymbol{r})+\boldsymbol{b}_2) \tag{3}$$

其中，$\boldsymbol{W}_2\in\mathrm{R}^{2d\times d}$ 与 $\boldsymbol{b}_2$ 为可训练参数，则 $\boldsymbol{c}$ 为头实体特定于任务关系的类型表示。

对于一个事实三元组，其整体上下文信息应具有同一类型的语义环境，因此，得到头实体及其邻居实体的类型表示 $\boldsymbol{c}$ 和 $\boldsymbol{c}_i$ 后，本文提出第 i 个邻居的类型得分 $\boldsymbol{d}_i$ 如式(4)所示：

$$\boldsymbol{d}_i = \left\| \boldsymbol{c} + \boldsymbol{r} - \boldsymbol{c}_i \right\|_2 \tag{4}$$

对不同邻居的类型得分通过softmax函数获得类型注意力，用以区分不同邻居的重要性，并提出基于类型注意力的实体的聚合表示，如式(5)-式(7)所示：

$$\alpha_i = \frac{\exp(\boldsymbol{d}_i)}{\sum_{j \in N_h} \exp(\boldsymbol{d}_j)} \tag{5}$$

$$\boldsymbol{e}^{(k)} = \sum_{i \in N_h} \alpha_i \boldsymbol{e}_{ti} \tag{6}$$

$$\boldsymbol{e} = \Big\|_{k=1}^{K} \boldsymbol{e}^{(k)} \tag{7}$$

其中，$\boldsymbol{e}^{(k)}$ 为第 k 个头得到的实体聚合表示，$\boldsymbol{e}$ 为最终的实体聚合表示。此外，定义了一个参数 a 来调整预训练实体嵌入表示和编码得到的实体聚合表示之间的权重，并将其相加作为最终的实体增强表示，如式(8)-式(9)所示：

$$a = \sigma(\boldsymbol{W}_3^{\mathrm{T}} \boldsymbol{e}) \tag{8}$$

$$\boldsymbol{h} = \mathrm{ReLU}(a * \boldsymbol{e} + (1 - a) * \boldsymbol{h}) \tag{9}$$

其中，σ 为Sigmoid激活函数，$\boldsymbol{W}_3 \in \mathrm{R}^{\mathrm{d} \times \mathrm{d}}$ 为权重矩阵。通过以上方法得到实体增强表示的过程同样适用于尾实体 t 。

4.2 Transformer 编码器

Transformer[26]能够通过叠加注意力网络和全连接层，关注输入序列中每一部分与其它部分间的关联，得到优化的特征向量。通过类型感知邻居编码器得到实体增强的表示后，根据动态知识图谱嵌入方法[27]，需要进一步对实体对进行表示。

对由类型感知邻居编码器得到的与任务关系 r 相关的实体对 $(\boldsymbol{h},\boldsymbol{t}) \in R_r$ ，将其与任务关系 $\boldsymbol{r}$ 拼接成序列 $X = (\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3)$ 。对于每个 X 中的元素 $\boldsymbol{x}_i$ ，其对应的输入由式(10)得到[27]：

$$\boldsymbol{z}_i^0 = \boldsymbol{x}_i^{\mathrm{ele}} + \boldsymbol{x}_i^{\mathrm{pos}} \tag{10}$$

其中，$\boldsymbol{x}_i^{\mathrm{ele}}$ 由类型感知邻居编码器获得，$\boldsymbol{x}_i^{\mathrm{pos}}$ 为相应的位置嵌入。得到输入的表示后，将其输入到一个 L 层Transformer编码器中来对 X 进行编码以得到实体对的表示，如式(11)所示[27]：

$$\boldsymbol{z}_i^l = \mathrm{TransformerEncoder}(\boldsymbol{z}_i^{l-1}) \tag{11}$$

其中，$\boldsymbol{z}_i^l$ 是第 l 层的元素 $\boldsymbol{x}_i$ 对应的隐藏状态，Transformer编码器采用多头自注意力机制。

为了进行少样本知识图谱补全任务，将掩码作用于任务关系 $\boldsymbol{x}_2$ 上，以获得有意义的实体对嵌入表示。取最终的隐藏状态 $\boldsymbol{z}_2^L$ 作为当前实体对的表示。通过Transformer编码器编码实体对特定于任务关系的表示，学习三元组整体的上下文类型信息，使得任务关系在面对不同实体对时，能表现出不同的类型信息。

4.3 联合匹配原型网络

原型网络[28]能够将少量样本投影到同一空间，计算每个样本类别的中心作为原型，然后通过比较查询到不同原型的距离，达到分类的效果，因此常用于少样本学习任务中。为了比较查询对和参考集来进行实体预测，本文采用了一种联

合匹配原型网络，该网络在原型网络的基础上，分别将参考对的类型感知编码器输出和Transformer编码器输出进行聚合得到实体级原型和关系级原型，作为少样本参考集表示。具体而言，将FAAN模型的自适应匹配处理器[9]进行扩展，使其分别用于实体级聚合和关系级聚合，得到实体级原型 $g_{\text{entity}}(\boldsymbol{S}_r)$ 和关系级原型 $g_{\text{rel}}(\boldsymbol{S}_r)$ 如式(12)-式(15)所示：

$$\beta_{i,\text{rel}} = \frac{\exp\{\boldsymbol{q}_{\text{rel}} \cdot \boldsymbol{s}_{ri,\text{rel}}\}}{\sum_j \exp\{\boldsymbol{q}_{\text{rel}} \cdot \boldsymbol{s}_{rj,\text{rel}}\}} \tag{12}$$

$$g_{\text{rel}}(\boldsymbol{S}_r) = \sum_{i \in S_r} \beta_{i,\text{rel}} \boldsymbol{s}_{ri,\text{rel}} \tag{13}$$

$$\beta_{i,\text{entity}} = \frac{\exp\{\cos(\boldsymbol{q}_{\text{entity}}, \boldsymbol{s}_{ri,\text{entity}})\}}{\sum_j \exp\{\cos(\boldsymbol{q}_{\text{entity}}, \boldsymbol{s}_{rj,\text{entity}})\}} \tag{14}$$

$$g_{\text{entity}}(\boldsymbol{S}_r) = \sum_{i \in S_r} \beta_{i,\text{entity}} \boldsymbol{s}_{ri,\text{entity}} \tag{15}$$

其中，$\cdot$ 为逐元素点积，$\boldsymbol{s}_{ri,\text{entity}}$ 为参考集 S_r 中第 i 个参考对的实体表示，$\boldsymbol{q}_{\text{entity}}$ 为查询对的实体表示，两者由式(9)得到；$\boldsymbol{s}_{ri,\text{rel}}$ 为参考集 S_r 中第 i 个参考对的表示，$\boldsymbol{q}_{\text{rel}}$ 为查询对的表示，两者由式(11)得到；$g_{\text{rel}}(\boldsymbol{S}_r)$ 和 $g_{\text{entity}}(\boldsymbol{S}_r)$ 为关系级原型和实体级原型，作为参考集的聚合表示。

为了进行预测，对FAAN模型的自适应匹配处理器[9]进行改进，用于进一步计算查询对 q 和参考集表示 $g(S_r)$ 的相似性，如式(16)-式(17)所示：

$$S_{\text{entity}} = \cos(\boldsymbol{q}_{\text{entity}}, g_{\text{entity}}(\boldsymbol{S}_r)) \tag{16}$$

$$S_{\text{rel}} = \boldsymbol{q}_{\text{rel}} \cdot g_{\text{rel}}(\boldsymbol{S}_r) \tag{17}$$

由式(16)和式(17)可知，查询对与参考集的实体级表示和关系级表示更相似，该查询对更有可能真实存在。

4.4 目标函数

通过元训练集 T_{mtr} 在模型上进行训练可以获得一个用于预测的度量函数。对于每个少样本任务关系 r，从训练集中随机采样 K 个实体对作为参考集 S_r，训练集剩余的实体对作为正例查询对 $Q_r = \{(h_p, t_p)\}$。通过随机破坏 (h_p, t_p) 的尾实体，构造负例实体对 $Q_r^- = \{(h_p, t_f)\}$，其中 $t_f \neq t_p$。 S_r、Q_r 和 Q_r^- 构成元训练集 T_{mtr}。相应的损失为：

$$L_{\text{rel}} = \sum_r \sum_{q \in Q_r} \sum_{q^- \in Q_r^-} [\gamma_{\text{rel}} + S_{\text{rel}}^- - S_{\text{rel}}]_+ \tag{18}$$

$$L_{\text{entity}} = \sum_r \sum_{q \in Q_r} \sum_{q^- \in Q_r^-} [\gamma_{\text{entity}} + S_{\text{entity}}^- - S_{\text{entity}}]_+ \tag{19}$$

$$L = L_{\text{rel}} + \lambda_1 L_{\text{head}} + \lambda_2 L_{\text{tail}} + \frac{\theta}{2n} \sum_w w^2 \tag{20}$$

其中，对于 L_{entity} 将其细分为 L_{head} 和 L_{tail}，$[\cdot]_+$ 为最大值函数，γ_{rel} 和 γ_{entity} 为正负例间的边界，λ_1 和 λ_2 为相应的权重参数，并对模型参数进行 L_2 正则化避免过拟合。

5 实验与结果分析

5.1 数据集

本文在两个公共基准数据集 NELL[4]和 Wiki[5]上对所提方法进行实验，数据集规模如表 3 所示。在两个数据集中，选取三元组数目在 50 到 500 之间的关系作为少样本任务，在 NELL 和 Wiki 中分别有 67 和 183 个任务，进一步划分其中的 51/5/11 和 133/16/34 个任务关系作为训练集、验证集和测试集。

表 3 数据集

Table 3 Datasets

数据集	实体数	关系数	三元组数	任务数
NELL	68545	358	181109	67
Wiki	4838244	822	5859240	183

5.2 对比方法

为了验证方法的有效性，本文将所提方法与以下较为常用的知识图谱补全方法进行比较。

（1）基于知识表示学习的方法

TransE[12]：一种基于距离的模型，将尾实体向量建模为头实体向量与关系向量的和。

DistMult[21]：一种双线性模型，在RESCAL模型的基础上将关系矩阵简化为对角矩阵。

ComplEx[23]：通过引入复数值嵌入来扩展 DistMult模型，以更好地建立模型的非对称关系。

SimplE[18]：一种基于张量分解的模型，能独立学习每个实体的两个嵌入，并使得到的嵌入具有可解释性。

RotatE[16]：一种基于距离的模型，将关系定义成复数空间中头实体到尾实体的旋转变换。

这类方法通过对知识图谱建模来学习实体和关系的嵌入表示，并根据得分函数计算三元组得分以进行知识图谱补全，但这些方法需要每个关系有足够多的三元组来进行训练，当训练三元组较少时，模型的性能会受到较大的影响。

（2）少样本知识图谱补全方法

GMatching[6]：使用了邻居编码器和匹配网络，但模型为所有邻居分配的权重都相等。其中，MaxP、MeanP和Max分别表示在得到参考集的一般嵌入时，使用最大池化方法、平均池化方法和使用查询与所有参考对中的相似度得分最高的一个参考对作为参考集的嵌入。

FSRL[8]：引入了注意力机制以对邻居分配不同的权重。

MetaR[7]：通过新的优化策略将共享知识从参考集转移到查询上来进行预测。

FAAN[9]：通过一个自适应注意力网络使模型能够学习实体和关系的动态表征。

B-GMatching[10]：引入BERT预训练的语言表示模型来增强GMatching中实体和关系的语义表示。

这类方法在 NELL 和 Wiki 数据集上取得了很好的少样本知识图谱补全效果，但上述方法在对邻居编码时都没有考虑到邻居实体中包含的隐含类型信息。

5.3 评价指标

本文使用Hits@k命中率和平均倒数排名MRR以检验不同方法的性能，k分别取1、5和10，具体如下：

$$\mathrm{MRR}=\frac{1}{|S|}\sum_{i=1}^{|S|}\frac{1}{\mathrm{rank}_i} \tag{21}$$

$$\mathrm{Hits@}k=\frac{1}{|S|}\sum_{i=1}^{|S|}I(\mathrm{rank}_i\le k) \tag{22}$$

其中，$|S|$是三元组集合的大小，rank_i表示第i个三元组的排名，$I(\mathrm{rank}_i\le k)$表示当第i个三元组的排名小于k时值为1，否则值为0。

5.4 参数设置

本文通过PyTorch框架进行标准的5-shot的少样本知识图谱补全实验。参数设置大部分与FAAN模型相同，选择TransE模型的预训练结果作为模型的预训练嵌入，通过Adam优化器[29]更新模型参数，在前10000个训练步骤中进行预热，然后线性衰减，对每训练10000步的结果在验证集进行验证，并调整超参数。

对于NELL数据集，嵌入维度为100，邻居编码器Dropout概率为0.3，Transformer编码器层数为3，Transformer编码器多头数目为4，Dropout概率为0.1，初始学习率为1e-4，$\gamma_{\mathrm{head}}=\gamma_{\mathrm{rel}}=\gamma_{\mathrm{tail}}=5$，$\lambda_1=\lambda_2=0.1$。

对于Wiki数据集，嵌入维度为50，邻居编码器Dropout概率为0.3，Transformer编码器层数为4，Transformer编码器多头数目为8，Dropout概率为0.2，初始学习率为6e-5，$\gamma_{\mathrm{head}}=\gamma_{\mathrm{tail}}=0.1$，$\gamma_{\mathrm{rel}}=5$，$\lambda_1=\lambda_2=1$。

两个数据集的最大邻居数量为50，参数正则化的θ取值为0.00001。

5.5 实验结果分析

通过在NELL和Wiki数据集上进行实验，得到各模型的实验结果如表4和表5所示。

表4 $K=5$时不同模型在NELL数据集上的实验结果

Table 4 Experimental results of different models on NELL when $K=5$

	NELL			
	MRR	Hits@10	Hits@5	Hits@1
TransE[12]	.174	.313	.231	.101
DistMult[21]	.200	.311	.251	.137
ComplEx[23]	.184	.297	.229	.118
SimplE[18]	.158	.285	.226	.097
RotatE[16]	.176	.329	.247	.101
GMatching (MaxP)[6]	.176	.294	.233	.113
GMatching (MeanP)[6]	.141	.272	.201	.080
GMatching(Max)[6]	.147	.244	.197	.090
FSRL[8]	.153	.319	.212	.073
MetaR[7]	.209	.355	.280	.141
FAAN[9]	.279	**.428**	.364	.200
B-GMatching[10]	.198	.276	.280	.134
本文方法(Multi-head 1)	.295	**.430**	**.372**	.221
本文方法(Multi-head 2)	**.296**	.426	.362	**.227**
本文方法(Multi-head 3)	.291	.418	.362	.218

表5 $K=5$时不同模型在Wiki数据集上的实验结果

Table 5 Experimental results of different models on Wiki when $K=5$

	Wiki			
	MRR	Hits@10	Hits@5	Hits@1
TransE[11]	.133	.187	.157	.100
DistMult[20]	.071	.151	.099	.024
ComplEx[22]	.080	.181	.122	.032
SimplE[17]	.093	.180	.128	.043
RotatE[15]	.049	.090	.064	.026
GMatching (MaxP)[6]	.263	.387	.337	.197
GMatching (MeanP)[6]	.254	.374	.314	.193
GMatching(Max)[6]	.245	.372	.295	.185
FSRL[8]	.158	.287	.206	.097
MetaR[7]	.323	.418	.385	.270
FAAN[9]	.309/**.341**	.451/.463	.380/.395	.239/**.281**
B-GMatching[10]	.220	.336	.275	.172
本文方法(Multi-head 1)	.321	**.468**	.389	.251
本文方法(Multi-head 2)	**.322**	.463	**.397**	.251
本文方法(Multi-head 3)	.321	.451	.386	**.257**

*注：在验证FAAN模型在Wiki数据集上的实验效果时，其结果与原文中所给实验结果存在一定差异，斜线前为实际实验结果，斜线后为原文所给实验结果，本文以实际实验结果为准。

由表4和表5可知，与传统基于知识表示学习的方法相比，少样本知识图谱补全方法取得了较好的性能，表明少样本知识图谱补全方法更能针对并解决知识图谱中存在的长尾问题。与目前效果最好的少样本知识图谱补全方法FAAN相比，本文方法较在NELL数据集上，MRR指标提高了1.6%，Hits@10指标提高了0.2%，Hits@5指标提高了0.8%，Hits@1指标提高了2.1%；在Wiki数据集上，MRR指标提高了1.2%，Hits@10指标提高了1.7%，Hits@5指标提高了0.9%，Hits@1指标提高了1.2%。以上实现结果表明，同时利用邻居实体和邻居关系来挖掘实体隐含类型信息能够在复杂的邻居关系中学习到更好的表示，并提高少样本知识图谱补全的性能。

在表4和表5中，Multi-head表示类型感知邻居编码器的多头数目，通过增大多头数目，NELL数据集的Hits@1指标进一步提升0.6个百分点，但Hits@10和Hits@5指标分别下降了0.4和1.0个百分点；Wiki数据集的Hits@5指标和Hits@1指标分别提升0.8和0.6个百分点，但Hits@10指标下降了1.7个百分点，即随着多头数目的增加，模型将正确实体排进前10名的能力降低，但对于已排进前10名的实体，模型能够将其进一步排在更靠前的位置。这是由于更大的多头数目使得类型特征的学习过程更加复杂，降低了泛化能力，但复杂的学习过程能使邻居编码器学到更丰富的类型信息，并提高对实体进行排名的能力。

此外，本文方法中的类型感知注意力会受实体邻居类型的影响，当实体邻居与当前实体和任务关系相关性较低甚至无关时，实体邻居包含的类型信息也与实体相关性较低，这些相关性较低的邻居在一定程度上引入了噪声。在本文的工作中，实体的所有邻居都被用于进行聚合，与实体相关性较低的邻居同样需要被聚合到实体表示中，使得类型感知编码器并不总能有效地学习到类型信息，影响实

体邻居聚合的效果，进而限制着本文方法的性能。

综上所述，本文方法性能在一定程度上受到噪声的限制，但仍在合理的参数设置范围内较先前方法取得了一定的优越性，更能针对并解决少样本知识图谱补全问题。

5.6 少样本大小影响实验

参考集中样本个数 K 的大小对少样本学习有着重要的影响，参考集越小，越需要模型具有更强的泛化能力，从更少的参考样本中学习有效的信息。因此，本文通过在NELL数据集上进行一系列实验，并测试比较了在不同 K 大小情况下，本文方法与FAAN模型的性能，来分析少样本大小 K 的影响，如图3所示。

由图3可知，随着 K 的增加，两种模型的性能整体都呈上升趋势。这表明，适当增大参考集的大小可以有效提高少样本知识图谱补全的性能。对于不同 K 取值，本文方法相比FAAN模型也在各指标上取得了较优的效果，验证了本文方法的有效性。此外，当模型通过计算得分得到不同候选实体的排名时，如果正确实体的得分较高，但仍是一个错误的实体得分最高，则会导致知识图谱补全的可信度较低。而本文方法在 K=4,5 的情况下，Hits@1指标能显著高于FAAN模型，表明本文方法能够将更多的正确实体排在候选实体中的第1位，提高知识图谱补全的可信度。

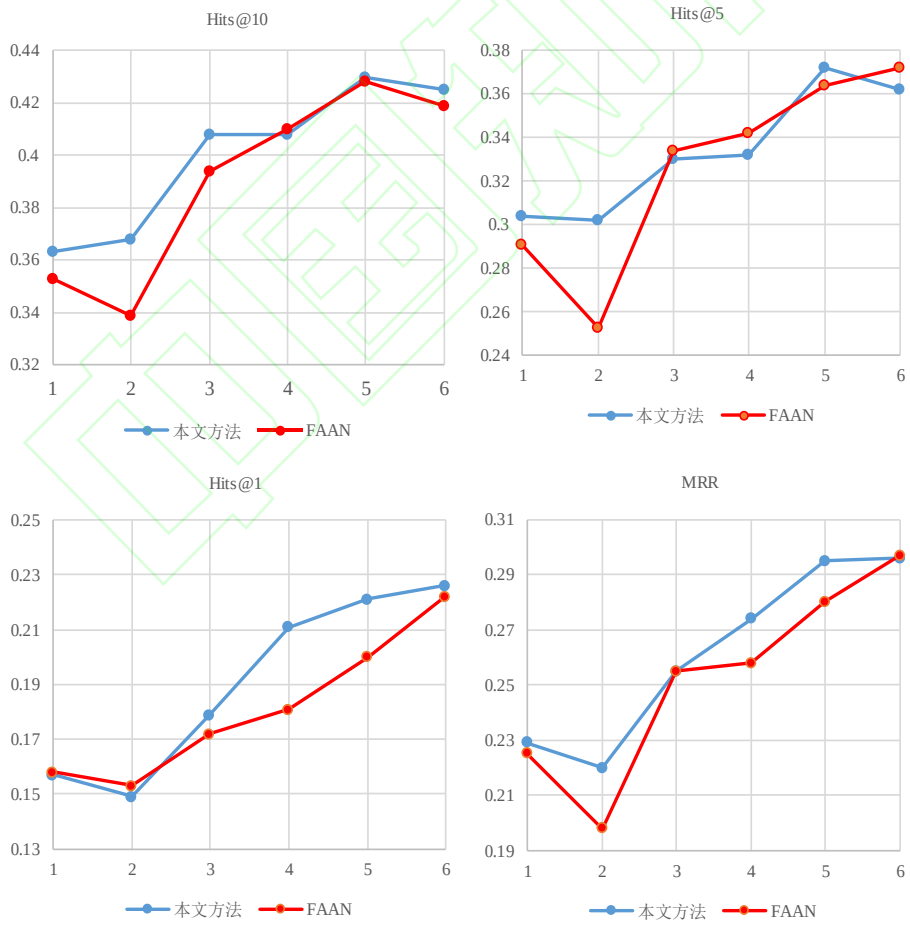

图 3 少样本大小影响实验结果

Fig3. Experimental results of the effects of few-shot size

5.7 不同关系实验

为了了解本文方法在不同任务关系上的性能，在NELL数据集的测试集上对不同关系进行少样本知识图谱补全实验，以观察不同方法在不同任务关系上的表现，并给出本文方法与FAAN模型对比的结果，如表6所示。

由表6可知，两种方法在不同关系上的结果具有很高的方差，这是因为不同关系的候选集大小各不相同，存在候选集较小或容易预测的关系，如表6中的关系1，使得结果较高，而其他关系候选集较多，难以预测。此外，在多数关系中(关系1,5,6,7,8,11)，本文方法相比FAAN模型具有更好的性能，表明本文方法更适用于少样本知识图谱补全任务。

表 6 不同关系比较实验结果

Table 6 Experimental results of different relations

Relation_id	Model	Hits@10	Hits@5	Hits@1	MRR
1	本文方法	**1.000**	**0.971**	**0.971**	**0.976**
	FAAN	**1.000**	**0.971**	**0.971**	0.974
2	本文方法	**0.179**	0.104	0.029	0.077
	FAAN	0.168	**0.133**	**0.058**	**0.101**
3	本文方法	**0.781**	**0.781**	0.234	0.467
	FAAN	0.766	0.719	**0.406**	**0.533**
4	本文方法	0.007	0.007	**0.007**	0.009
	FAAN	**0.014**	**0.014**	**0.007**	**0.012**
5	本文方法	0.341	**0.307**	**0.157**	**0.231**
	FAAN	**0.346**	0.286	0.152	0.220
6	本文方法	**0.617**	**0.527**	**0.346**	**0.442**
	FAAN	0.602	0.505	0.220	0.352
7	本文方法	**0.745**	**0.694**	**0.531**	**0.602**
	FAAN	0.724	0.643	0.480	0.558
8	本文方法	**0.296**	**0.163**	**0.059**	**0.135**
	FAAN	0.281	**0.163**	0.022	0.112
9	本文方法	**0.857**	**0.813**	0.407	0.572
	FAAN	0.835	0.758	**0.505**	**0.615**
10	本文方法	0.128	0.074	0.033	0.061
	FAAN	**0.134**	**0.092**	**0.042**	**0.073**
11	本文方法	**0.659**	**0.601**	**0.322**	**0.435**
	FAAN	0.649	0.596	0.303	0.432

5.8 消融实验

为了验证本文方法中各部分的效果，在NELL数据集上进行了相应的消融实验，实验结果如表7所示。其中，分别从整体方法中去掉调整预训练实体嵌入表示和编码得到的实体聚合表示的权重参数 a；去掉实体级原型(-EP)；去掉关系级原型(-RP)；去掉类型感知注意力，即将邻居实体特定于邻居关系的表示 c_i 求平均作为邻居编码器的输出(-TAA)；将整个类型感知编码器替换为FAAN的自适应邻居编码器，保留联合匹配原型网络(FAAN+)。

表 7 消融实验结果

Table 7 Ablation experiment results

	MRR	Hits@10	Hits@5	Hits@1
本文方法(- a)	. 289	. 423	. 371	. 215
本文方法(-EP)	. 289	. 427	**. 373**	. 214
本文方法(-RP)	. 224	. 330	. 270	. 165

本文方法(-TAA)	.278	.420	.351	.208
本文方法(FAAN+)	.289	.426	.368	.218
FAAN	.279	.428	.364	.200
本文方法	**.295**	**.430**	.372	**.221**

由表7可知，在去除实体级原型和关系级原型后，模型的性能有所下降，特别是去除关系级原型，即不经过Transformer编码器进行编码，而直接使用依据邻居编码器的输出得到的实体级原型进行预测，导致模型性能下降7.1%；在去除类型感知注意力后，模型性能下降了1.7%；去除权重参数 a 后，模型性能下降了0.6%；使用FAAN模型的自适应邻居编码器和实体级原型与关系级原型，相比FAAN模型提升了1%。综上所述，本文方法能够通过类型感知编码器学习邻居实体的特定类型表示，得到类型感知注意力，用来增强实体嵌入的表示，并通过Transformer编码器使任务关系能捕获不同三元组的上下文信息，最后通过联合匹配原型网络聚合参考集得到实体级原型和关系级原型进行预测，并取得了较好的效果。

5.9 可视化分析

为了观察类型感知邻居编码器中不同邻居的重要性，将NELL数据集导入Neo4j图形数据库上进行了相应的可视化工作，得到由1000个实体及这些实体间的关系构成的知识图谱示意图，如图4所示。

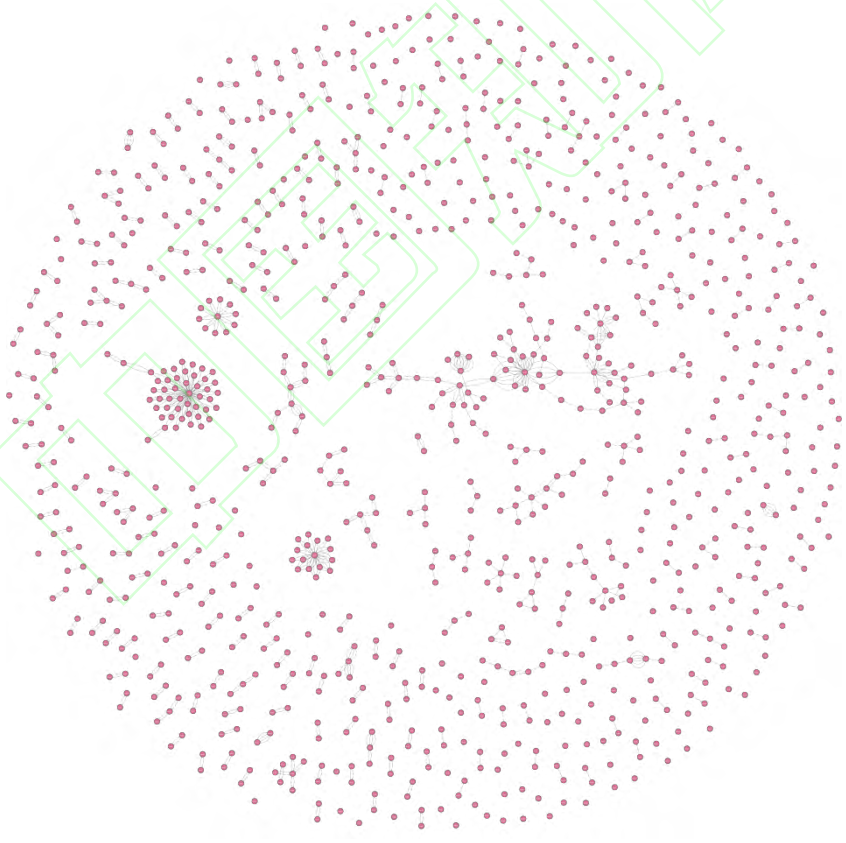

图 4 NELL 数据集可视化(节点个数取 1000)

Fig4. Visualization of NELL dataset(the number of nodes is 1000)

为了便于进一步分析，在图4上选取其中一个实体及其邻居，并标注由4.1节式(5)计算得到的注意力权重，如图5所示。图5给出一查询三元组(product:quicktime, produceby, company:apple002)在类型感知邻居编码器中的类

型注意力分布情况，其中，邻居关系中的“_inv”表示逆关系，mutual proxy for和mutual proxy for_inv表示共同代理，synonym for和synonym for_inv表示同义词。

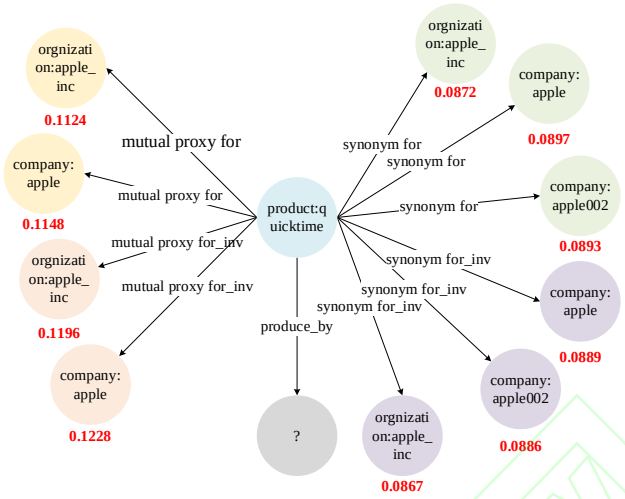

图 5 注意力分布

Fig5. Attention distribution

由图5可知，对于头实体quicktime，左半部分邻居的注意力分数显著高于右半部分，这是因为左半部分的邻居实体与头实体quicktime存在代理关系，与任务关系produce_by更加相似，而右半部分邻居表示的为quicktime的同义词，与任务关系相似性较低。进一步，对于邻居实体apple，其主要以类型“公司”出现在quicktime的邻居中，而邻居实体apple_inc主要以类型“组织”出现在quicktime的邻居中，因此类型感知邻居编码器会更加关注实体apple，并为与其相关的邻居分配更高的权重。综上所述，可视化实验结果展示了本文方法在对实体邻居进行注意力权重分配时的详细情况，并验证了本文方法的有效性。

6 结语

本文针对以往少样本知识图谱补全模型在面对数据中存在的一对多、多对多等复杂情况时，因分配出相同的邻居注意力权重而不能很好地区分不同邻居重要性的问题，在FAAN模型基础上提出了一个用于少样本知识图谱补全的类型感知注意力网络，该方法通过一个新的类型感知邻居编码器学习邻居实体中包含的隐含语义信息来优化实体编码阶段的过程，增强实体的表示以适应包含复杂关系的任务，并通过一个联合匹配原型网络学习实体级原型和关系级原型，以更准确的将查询对与参考集进行匹配。本文在两个基准数据集上进行了相应的实验，实验表明，本文的方法能够通过学习更丰富的邻居信息来提高模型在少样本情况下补全知识图谱中缺失三元组的性能，在大多数指标上优于baseline模型。本文工作在实体聚合过程中没有对实体邻居进行合适的筛选以去除噪声，未来将针对实体邻居的自适应聚合开展进一步研究。

参考文献：

[1] 李涓子, 侯磊. 知识图谱研究综述[J]. 山西大学学报, 2017, 40(3): 454-459.

(Li Juanzi, Hou Lei. Reviews on Knowledge Graph Research [J]. Journal of Shanxi University, 2017, 40(3): 454-459.)

[2] Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]. In: Proceedings of the 2008 ACM SIGMOD international conference on Management of data, Vancouver. New York: SIGMOD, 2008: 1247-1250.
[3] Lehmann J, Isele R, Jakob M, et al. DBpedia a large-scale, multilingual knowledge base extracted from wikipedia[J]. Semantic web, 2015, 6(2): 167-195.
[4] Carlson A, Betteridge J, Kisiel B, et al. Toward an architecture for never-ending language learning[C]. In: Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, Georgia. Menlo Park: AAAI, 2010: 1306-1313.
[5] Vrandečić D, Krötzsch M. Wikidata: a free collaborative knowledgebase[J]. Communications of the ACM, 2014, 57(10): 78-85.
[6] Xiong W, Yu M, Chang S, et al. One-shot relational learning for knowledge graphs[C]. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels. Stroudsburg: EMNLP, 2018: 1980–1990.
[7] Chen M, Zhang W, Zhang W, et al. Meta relational learning for few-shot link prediction in knowledge graphs[C]. In: Proceedings of the Empirical Methods in Natural Language Processing 2019 and 9th International Joint Conference on Natural Language Processing, Hong Kong. Stroudsburg: ACL, 2019: 4216–4225.
[8] Zhang C, Yao H, Huang C, et al. Few-shot knowledge graph completion[C]. In: Proceedings of The Thirty-Fourth AAAI Conference on Artificial Intelligence, New York. Menlo Park: AAAI, 2020: 3041–3048.
[9] Sheng J, Guo S, Chen Z, et al. Adaptive Attentional Network for Few-Shot Knowledge Graph Completion[C]. In: Proceedings of the Empirical Methods in Natural Language Processing, Online. Stroudsburg: ACL, 2020: 1681-1691.
[10] Wang F, Xie Y, Zhang K, et al. Bert-based Knowledge Graph Completion Algorithm for Few-Shot[C]. In: Proceedings of the 2nd International Conference on Big Data Economy and Information Management: 2nd International Conference on Big Data Economy and Information Management (BDEIM), Sanya. BDEIM, 2021: 217-224.
[11] 王桢. 基于嵌入模型的知识图谱补全[D]. 广州:中山大学, 2017:1-78.
(Wang Zhen. Embedding Model based Knowledge Graph Completion[D]. Guangzhou: Sun Yat-sen University, 2017:1-78.)
[12] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data[C]. In: Proceedings of the Neural Information Processing Systems, Nevada. La Jolla: NIPS, 2020: 2787–2795.
[13] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes[C]. In: Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence and the Twenty-Sixth Innovative Applications of Artificial Intelligence Conference, Quebec Ciry. Palo Alto: AAAI, 2014: 1112–1119.
[14] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion[C]. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence and the Twenty-Seventh Innovative Applications of Artificial Intelligence Conference, Austin. Palo Alto: AAAI, 2015: 2181–2187.
[15] Ji G, He S, Xu L, et al, Knowledge graph embedding via dynamic mapping matrix[C]. In: Proceedings of the 53rd annual meeting of the Association for Computational Linguistics and 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Languages processing, Beijing. Association for Computational Linguistics: Stroudsburg, ACL 2015: 687–696.
[16] Sun Z, Deng Z H, Nie J Y, RotatE: Knowledge graph embedding by relational rotation in complex space[J]. arXiv preprint arXiv:1902.10197, 2019.
[17] Zhang Z, Cai J, Zhang Y, et al. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction[C]. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, New York. Palo Alto:

AAAI, 2020: 3065-3072.
[18] Kazemi S, Poole D, Simple embedding for link prediction in knowledge graphs[C]. Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, Canada. NeurIPS, 2018: 4289-4300.
[19] Ji S, Pan S, Cambria E, et al. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2): 494-514
[20] Nickel M, Tresp V, and Kriegel H, A three-way model for collective learning on multi-relational data[C]. In: Proceeding of the Twenty-Eighth International Conference on Machine Learning, Bellevue. Madison: IMLS, 2011: 809–816.
[21] Yang B, Yih W, He X, et al. Embedding entities and relations for learning and inference in knowledge bases[J]. arXiv preprint arXiv:1412.6575, 2014.
[22] Nickel M, Rosasco L, Poggio T, Holographic embeddings of knowledge graphs[C]. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence and the Twenty-Eighth Innovative Applications of Artificial Intelligence Conference, Phoenix. Palo Alto: AAAI, 2016: 1955–1961.
[23] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction[C]. In: Proceedings of the 33rd International Conference on Machine Learning, New York. New York: International Machine Learning Society, 2016: 2071-2080.
[24] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2d knowledge graph embeddings[C]. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans. Palo Alto: AAAI, 2018:1811–1818.
[25] Cai L, Yan B, Mai G, et al. TransGCN: Coupling transformation assumptions with graph convolutional networks for link prediction[C]. In: Proceedings of the 10th International Conference on Knowledge Capture, New York. New York: ACM, 2019: 131–138.
[26] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]. Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, Long Beach. NeurIPS, 2017: 5998-6008.
[27] Wang Q, Huang P, Wang H, et al. Coke: Contextualized knowledge graph embedding[J]. arXiv preprint arXiv:1911.02168, 2019.
[28] Snell J, Swersky K, Zemel R. Prototypical networks for few-shot learning[J]. arXiv preprint arXiv:1703.05175, 2017.
[29] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

通讯作者（Corresponding author）：王红斌(Wang Hongbin)，ORCID：0000-0003-2176-2998，E-mail：whbin2007@126.com。

基金项目：本文系云南省重点研发计划项目（项目编号：202202AD080003）和国家自然科学基金项目（项目编号：61966020）的研究成果之一。

This work was supported by the National Key Research and Development Plans Project of Yunnan Province (Grant No. 202202AD080003) and the National Natural Science Foundation of China (Grant No. 61966020).

作者贡献声明：

普祥和：论文思想的具体实现与论文撰写；

王红斌：提出论文思想和论文修改；
线岩团：参与论文的修改完善。

利益冲突声明：
所有作者声明不存在利益冲突关系。