

文章编号: XXXX-XXXX (2019) 00-0000-00

## 聚合邻域信息的联合知识表示模型

彭敏 黄婷 田纲 张鼎 罗娟 银源

(武汉大学 计算机学院, 湖北 武汉 430072)

**摘要:** 知识表示学习在关系抽取、自动问答等自然语言处理任务中获得了广泛的关注, 该技术旨在将知识库中的实体与关系表示为稠密低维实值向量。然而, 已有的模型在建库知识库中三元组时, 或是忽略三元组的邻域信息, 导致无法处理关联知识较少的罕见实体, 或是在引入邻域信息时不能自适应地为每个实体抽取最相关的邻节点属性, 导致引入了冗余信息。基于以上问题, 本文在知识表示模型 TransE 的基础上提出了聚合邻域信息的联合知识表示模型 TransE-NA (Neighborhood Aggregation on TransE)。该模型首先根据实体的稀疏度确定其邻节点数量, 然后根据实体的邻边关系选取对应邻节点上最相关的属性作为实体的邻域信息。在链接预测和三元组分类任务上的实验结果表明, 本文的模型效果超越了基线模型, 验证了本文模型能有效聚合邻域信息, 缓解数据稀疏问题, 改善知识表示性能。

**关键词:** 知识表示; 邻域信息; 知识图谱

中图分类号: TP391

文献标识码: A

## Neighborhood Aggregation for Knowledge Graph Representation

PENG Min, HUANG Ting, TIAN Gang, ZHANG Ding, LUO Juan, YIN Yuan

(School of Computer Science, Wuhan University, Wuhan, Hubei 430072, China)

**Abstract:** Knowledge representation learning, which aims to encode entities and relations into a dense, real-valued and low-dimensional semantic space, has drawn massive attention in natural language processing tasks, such as relation extraction and question answering. However, when modeling triplets in the knowledge graph, some methods ignore the neighbor information and thus cannot handle rare entities with few facts, other methods incorporate neighbor information but cannot extract most relevant attributes of neighbors for each entity adaptively, resulting in the incorporation of redundant information. To tackle these issues, we propose a new model named TransE-NA (Neighborhood Aggregation on TransE) based on TransE, which firstly determines the number of neighbors according to sparse degrees of entities and then aggregates the neighbor information of entities from the most relevant attributes of neighbors according to the corresponding relations. Experimental results over link prediction and triplet classification show that our approach outperforms baselines and our method can aggregate neighbor information to alleviate data sparsity problem and improve the performance effectively.

**Key words:** knowledge representation; neighborhood aggregation; knowledge graph

### 1 引言

目前, 知识库在智能问答<sup>[1]</sup>、信息抽取<sup>[2-3]</sup>

和命名实体消歧<sup>[4]</sup>等自然语言处理任务中发挥着重要作用。流行的知识库有 Freebase<sup>[5]</sup>、WordNet<sup>[6]</sup>

收稿日期: 2019-XX-XX; 定稿日期: 2019-XX-XX

基金项目: 国家重点研发计划项目(2018YFC1604000, 2018YFC1604003); 国家自然科学基金(61772382)

和 YAGO<sup>[7]</sup>等。知识库以网络的形式将人类知识组织起来,使用三元组(头实体,关系,尾实体)表达知识。然而,这种网络表示形式往往面临计算效率低和数据稀疏等问题。近年来,知识表示学习技术获得了广泛的关注,该技术学习旨在将知识库中的实体和关系表示成稠密低维实值向量,能够显著提升知识获取、融合和推理的性能。其中,以 TransE<sup>[8]</sup>为代表的翻译模型,将关系建模为从头实体到尾实体的翻译向量。由于其简单有效,它已成为知识表示学习的代表模型。在 TransE 基础上扩展的其他模型如 TransH<sup>[9]</sup>、TransR<sup>[10]</sup>、CTransR<sup>[10]</sup>和 TransD<sup>[11]</sup>等也取得了很好的效果。然而,以上模型假设三元组之间是相互独立的,在建模时忽略了三元组周围的邻域信息,故而无法处理关联知识较少的罕见实体。

事实证明<sup>[12-14]</sup>,三元组周围的邻域信息能提供有效语义信息,提升知识推理性能,缓解数据稀疏问题。如图 1 所示,在预测 Adam Ant 的职业时,如果已知此人弹奏的乐器和音乐风格,则可推知此人的职业是音乐家。目前已有一些工作试图利用邻域信息辅助知识推理。例如 GAKE<sup>[12]</sup>将知识图谱建模为有向图,提出了融合邻域信息的通用框架。TransE-NMM<sup>[13]</sup>将实体编码为其邻节点向量的加权混合表示。TransN<sup>[14]</sup>为实体与关系增设一个上下文向量(context embedding),以此来编码邻域对象之间的语义关联。

然而,以上工作仍存在两个问题有待解决:

(1) 实体是多种属性的综合体,在其通过不同的关系充当不同实体的邻节点时,所提供的属性信息应该是不同的。模型应当根据实体的邻边关系选择邻节点最相关的属性作为邻域信息,而以上工作忽略了这一点;(2) 直观上讲,关联三元组多的实体会更依赖于三元组的结构信息,而关联三元组较少的实体会更依赖于其邻节点提供的邻域信息。模型应该灵活调整实体的邻节点数量来平衡上述两种信息,以获取更有价值的语义表示。

类似于以往模型<sup>[13-14]</sup>,本文在 TransE 的基础上引入邻域信息,提出了一个新的聚合邻域信息的联合表示模型 TransE-NA (Neighborhood Aggregation on TransE)。受到 TranSparse<sup>[15]</sup>的启发,本模型首先根据实体的稀疏度决定其邻节点数量。稀疏度越高的实体,其邻节点数量越多;稀疏度越低的实体,其邻节点数量越少。然后,模型为每个关系设置一个投影矩阵。通过邻边关系对应的投影矩阵的转换,模型能够为实体选取邻节点最相关的属性作为邻域信息。链接预测和三元组分类任务上的实验结果表明,本模型的性能超过了基准模型 TransE-NMM,验证了本模型在聚合

邻域信息上的有效性。同时与当前最先进的模型 TransN 相比,本模型有很强的竞争力。

本文的主要贡献如下:

(1) 本文提出了聚合邻域信息的联合知识表示模型 TransE-NA。该模型能够根据邻边关系为实体选择邻节点最相关的属性作为邻域信息。

(2) 本文提出了邻节点选择机制。根据实体稀疏度决定其邻节点数量,有效平衡了三元组结构信息与邻域信息。

(3) 实验结果表明,本文模型性能超过了基准模型,并且与当前最好的模型相比有很强的竞争力。

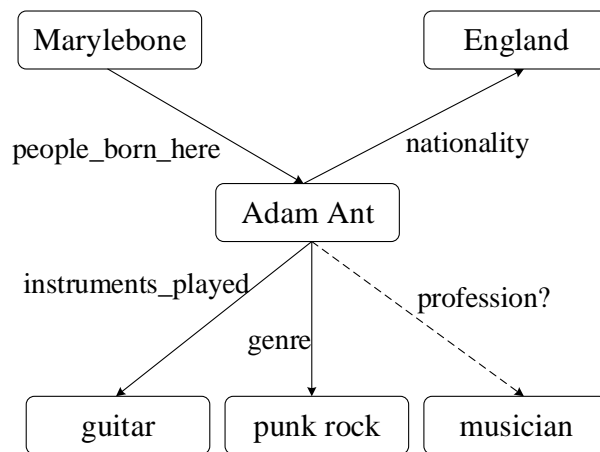


图 1 FB15k 中实体及其邻节点

## 2 相关工作

### 2.1 知识表示学习

近年来,知识表示学习逐渐成为研究热点,研究者提出了多种模型对知识库中的实体和关系建模。本节主要介绍仅利用知识库结构信息的知识表示代表模型。

非结构表示(Unstructured model, UM)<sup>[16]</sup>仅建模头、尾实体向量的语义关联,无法区分不同的关系。结构表示(Structured Embedding, SE)<sup>[17]</sup>为每个关系设置两个矩阵用于头、尾实体的投影操作,并在关系空间中计算投影后的向量距离。然而该模型不能准确捕获实体间的关系信息。单层神经网络模型(Single Layer Model, SLM)<sup>[18]</sup>试图通过采用非线性神经网络刻画三元组的得分函数来改进 SE 模型,但是引入了更高的复杂度。语义匹配能量模型(Semantic Matching Energy, SME)<sup>[19]</sup>采用神经网络编码实体与关系向量,得分函数有两种形式:线性形式(linear)与非线性形式(bilinear)。隐变量模型(Latent Factor Model, LFM)<sup>[20-21]</sup>为关系设置变换矩阵,试图捕获实体

与关系的二阶联系。张量神经网络模型 (Neural Tensor Network, NTN)<sup>[18]</sup>扩展了 SLM 模型, 利用双线性张量刻画头、尾实体间的二阶联系。但是在应用于大型知识图谱上时, 其复杂度过高。

近年来, 以 TransE<sup>[8]</sup>为代表的翻译模型因为其简单有效, 逐渐成为知识表示学习的代表模型。TransE 将关系向量建模为从头实体向量到尾实体向量的翻译向量。但是该模型过于简单, 无法区分复杂关系中的不同实体。为此 TransH<sup>[9]</sup>将实体投影到关系超平面上, 让每个实体在面对不同的关系时拥有不同的表示。TransR<sup>[10]</sup>将实体与关系建模在不同的语义空间上, 通过关系投影矩阵将实体投影到关系空间中。TransD<sup>[11]</sup>提出实体和关系投影向量, 为每个实体关系对动态构建投影矩阵。TranSparse<sup>[15]</sup>提出稀疏投影矩阵, 以解决知识图谱的异构性与不均衡性问题。

虽然上述模型在知识表示任务上取得了一定成就, 但是它们假设三元组相互独立并对其进行单独建模, 没有考虑三元组的邻域信息, 从而无法有效处理关联知识较少的罕见实体。本文提出的 TransE-NA 模型有效利用邻域信息辅助建模三元组, 为罕见实体提供更多语义信息, 缓解了数据稀疏问题, 在链接预测和三元组分类任务上表现优异。

## 2.2 引入邻域信息的知识表示学习

为有效利用三元组的邻域信息, 缓解数据稀疏问题, 提高知识推理和融合的能力, 越来越多的工作开始聚焦于引入邻域信息的知识表示学习。

GAKE<sup>[12]</sup>将知识图谱建模为有向图, 引入了三种邻域信息来辅助建模知识图谱, 分别是邻节点, 关系路径和邻边关系。然而 GAKE 将实体与关系都视为目标节点, 忽略了实体与关系之间的语义联系, 从而不能学到合适的语义表示<sup>[24]</sup>。TransE-NMM<sup>[13]</sup>在 TransE 的基础上引入邻域信息, 将实体建模为其邻节点向量的混合加权表示。然而这种简单的加权混合不能很好地提取有效的邻域信息。TransN<sup>[14]</sup>为了更好地建模邻域信息, 为每个实体与关系设置两种向量: 一种为对象向量 (object embedding), 用于表示对象的语义信息; 一种为上下文向量 (context embedding), 用于捕获对象之间的关联信息。但是这种表示方式使得每个实体与关系拥有两个表示向量, 造成了额外的计算开销。还有模型聚合了关系路径辅助知识推理。这些模型采用乘法、加法和递归神经网络 (RNN)<sup>[22]</sup>等组合策略, 将关系路径编码到与知识库中实体和关系相同的语义空间, 为知识图谱

建模提供有效信息。比如 PTransE<sup>[23]</sup>在关系路径建模中考虑了以上三种组合方式, 将关系路径建模为从头实体到尾实体的翻译向量, 取得了较好的实验效果。事实上, 本文的 TransE-NMM 模型也可以视为引入了长度为 1 的关系路径信息。

以上模型虽然通过引入邻域信息在性能上有所提升, 但是无法根据邻边关系选择邻节点最相关的属性信息。其次在邻节点选择上, 没有根据实体的稀疏情况综合考虑, 从而在缓解数据稀疏问题上能力受限。本文模型自适应地为每个实体引入有效邻域信息, 并根据实体的稀疏度决定其邻节点数量, 有效缓解了数据稀疏问题, 提高了知识表示性能。

## 3 模型

### 3.1 符号表示

本节定义常用符号。E 表示实体集, R 表示关系集。G 表示知识图谱, 包含正确三元组 (h, r, t), 其中 h, t ∈ E, r ∈ R。r<sup>-1</sup> 表示关系 r 的逆关系。R<sup>-1</sup> 表示逆关系集, 其定义为 R<sup>-1</sup> = {r<sup>-1</sup> | r ∈ R}。K 表示包含逆关系的知识图谱, 其定义为 K = {(h, r, t) | (r, h, e) ∈ h}。N<sub>e</sub> 表示实体 e 的邻节点集合, 定义为 N<sub>e</sub> = {(e', r') | r' ∈ R ∪ R<sup>-1</sup>, (e', r', e) ∈ K}。

### 3.2 结构表示模型 TransE

本节简要介绍知识表示模型的背景知识。知识表示模型旨在将关系数据 (实体和关系) 嵌入到稠密低维实值语义空间中。模型通常会定义一个得分函数 f<sub>r</sub>(h, t), 用于衡量三元组 (h, r, t) 的正确性。一个好的得分函数应该满足以下条件: 对于正确三元组, 其分数尽可能高; 对于错误三元组, 其得分越低越好。

TransE 是知识表示的代表模型, 模型简单、计算复杂度低, 并能在知识推理任务上取得很好的效果。TransE 受词向量空间平移不变现象的启发, 将关系表示为从头实体到尾实体的翻译。对于正确三元组 (h, r, t), 其对应的表示向量应满足 h + r ≈ t。故而对于三元组 (h, r, t), TransE 定义得分函数为:

$$f_r(\mathbf{h}, \mathbf{t}) = -\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{l_2}^2 \quad \#(1)$$

用于衡量向量 h + r 到 t 的 l<sub>1</sub> 或 l<sub>2</sub> 距离。

### 3.3 引入邻域信息的联合表示模型 TransE-NA

TransE-NA 在 TransE 的基础上引入邻域信息改善数据稀疏问题, 提高知识表示的区分度。本节分别介绍邻节点选择、基于邻节点的实体表示

以及融合邻域信息后的得分函数。

### 3.3.1 邻节点选择

在融合邻域信息之前，有必要先为每个实体设定邻节点数目。原因有以下三点：（1）知识图谱具有异质性。有些实体会关联非常多的三元组，而有些实体关联的三元组数目非常少。不同实体对邻域信息的依赖程度也有所不同。图 2 展示了子图 FB15k 的统计数据，横轴表示实体索引，纵轴表示每个实体关联的三元组数目。从图 2 可以看出，各个实体关联的三元组数目相差较大。而以往的方法在选定实体的邻节点数目时忽略了实体的异质性，这显然是不合理的；（2）并非所有的邻节点都能提供有用信息；（3）融合所有的邻节点造成的计算开销巨大且不适用于实际情况，灵活调整实体的邻节点数量不但能降低计算开销，还能提高模型训练速度。

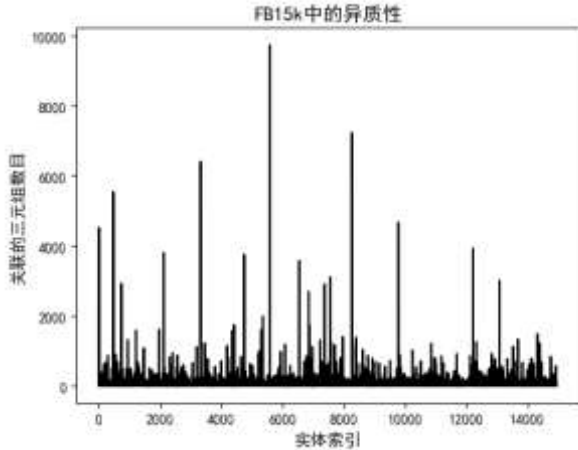


图 2 FB15k 中实体的异质性

直观上，关联三元组较少的实体往往需要更多的邻节点信息作为补充，关联三元组较多的实体会更依赖于训练好的结构向量。因此，模型根据实体的稀疏度确定其邻节点数量。稀疏度越高的实体所拥有的邻节点越多，反之亦然。实体的稀疏度定义为：

$$\theta_e = 1 - (1 - \theta_{\min}) C_e / C_e \#(2)$$

其中， $\theta_{\min}$  ( $0 \leq \theta_{\min} \leq 1$ ) 表示最低稀疏度， $C_e$  表示实体  $e$  在数据集中出现的次数， $C_e \#$  表示最频繁的实体出现的次数。考虑到知识库中  $C_e \#$  往往非常大，使得几乎所有实体通过式 (2) 计算出的稀疏度均为 1。本文在实验过程中将一些非常频繁的实体的稀疏度设置为 0。 $C_e \#$  则表示剩余实体中最频繁的实体出现的次数。实体  $e$  的邻节点数目定义如下：

$$\xi_e = \theta_e \times |N_e| \#(3)$$

其中， $|N_e|$  表示实体  $e$  原本所拥有的邻节点数目， $\xi_e$  表示根据实体稀疏度调整后的邻节点数目。

在确定邻节点数目之后，模型仍旧无法判断哪些邻节点能够提供有用信息。本文考虑两种方式解决该问题：逐步筛选和随机挑选。前者做法与 TransN 类似。在训练初期编码所有的邻节点，在训练过程中，根据邻节点的权重逐步删除权重较低的邻节点，直到邻节点数量达到指定值。这种做法虽然能选取更有价值的邻节点，但是存在以下弊端：（1）计算开销大，不利于并行计算。知识库中有些实体的邻节点数目高达数千个，编码所有的邻节点所造成的计算开销非常大。而且各个实体的邻节点数目参差不齐，不利于并行编码；（2）运算速度慢，不适用于大型知识图谱。逐步筛选会导致运算速度变慢，当模型被应用到大型知识图谱时，运算效率会很低。基于以上原因，在确定邻节点数目之后，模型采用随机挑选的方式选出邻节点。该方法具有以下优势：（1）本文的邻节点选择策略合理降低了邻节点数目，从而降低了运算开销；（2）由于没有筛选过程，本模型的运行速度加快，适用于大型知识图谱。最后从实验结果可以看出，随机挑选方式依旧能选取有效邻域信息用于辅助建模。

### 3.3.2 基于邻节点的实体表示

在选定实体的邻节点之后，本节将着手于邻域信息的表示。如 TransR 所述，实体有许多属性，不同的关系关注实体的不同属性。受此启发，本文认为，当实体通过不同的邻边关系作为不同实体的邻节点时，所提供的属性信息也是不同的。最相关的属性信息应当由实体相连的邻边关系所决定。

在模型中，三元组  $(h, r, t)$  的向量表示为  $\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^n$ 。每个关系  $r$  对应一个投影矩阵  $\mathbf{M}_r \in \mathbb{R}^{n \times n}$ ，用于选择邻节点相应的属性信息。所以，实体  $e$  在预测关系  $r$  时所对应的邻域信息  $\mathbf{e}_{e,r}$  表示为：

$$\mathbf{e}_{e,r} = \sum_{(e',r') \in N_e(\xi_e)} \alpha_{r,r'} \mathbf{e}' \mathbf{M}_{r'} \#(4)$$

其中， $N_e(\xi_e)$  是  $N_e$  的子集，包含  $\xi_e$  个邻节点。 $\alpha_{r,r'}$  是权重系数，用于权衡不同邻节点的重要程度。比如，在图 1 中，当预测 Adam Ant 的职业时，比起民族和出生地这两种关系信息，知道音乐风格和演奏乐器等关系信息会更有利。本文根据 Wang 等提出的注意力机制<sup>[14]</sup>对  $\alpha_{r,r'}$  定义如下：

$$\alpha_{r,r'} = \frac{\exp(\psi_{r,r'})}{\sum_i \exp(\psi_{r,r_i})} = \frac{\exp(\mathbf{r}\mathbf{W}\mathbf{r}')}{\sum_i \exp(\mathbf{r}\mathbf{W}\mathbf{r}_i)} \quad \#(5)$$

其中  $\mathbf{W} \in \mathbb{R}^{n \times n}$  表示权重矩阵。

在得到邻域信息的表示之后, 模型引入 Xu 等提出的门机制<sup>[25]</sup>融合实体  $e$  的结构表示向量  $\mathbf{e}$  和邻域表示向量  $\mathbf{e}_{e,r}$ 。实体  $e$  聚合邻节点的联合表示向量  $\mathbf{g}_{e,r}$  为:

$$\mathbf{g}_{e,r} = \mathbf{g}_e \odot \mathbf{e} + (1 - \mathbf{g}_e) \odot \mathbf{e}_{e,r} \quad \#(6)$$

其中  $\mathbf{g}_e \in \mathbb{R}^n$  是平衡两种信息源的门控向量,  $\odot$  表示元素乘法。为了约束  $\mathbf{g}_e$  的每一维在  $[0,1]$  之间, 本文采用 sigmoid 函数:

$$\mathbf{g}_e = \sigma(\tilde{\mathbf{g}}_e) \quad \#(7)$$

其中  $\tilde{\mathbf{g}}_e \in \mathbb{R}^n$  是存储在查找表中的实值向量, 服从均匀分布。 $\tilde{\mathbf{g}}_e$  被训练完后在测试时保持不变。

### 3.3.3 得分函数

在新模型 TransE-NA 中, 融合邻域信息的三元组  $(h,r,t)$  的得分函数定义为:

$$g_r(\mathbf{h}, \mathbf{t}) = -\|\mathbf{g}_{h,r} + \mathbf{r} - \mathbf{g}_{t,r}\|_{l_2}^2 \quad \#(8)$$

其中  $\mathbf{g}_{h,r}$  与  $\mathbf{g}_{t,r}$  的定义见式 (6)。参照以往的模型<sup>[8-10]</sup>, 在实验过程中, 本文对实体和关系向量进行了如下约束:  $\|\mathbf{h}\|_2 \leq 1$ ,  $\|\mathbf{r}\|_2 \leq 1$ ,  $\|\mathbf{t}\|_2 \leq 1$ 。

### 3.4 参数优化

类似于 TransE, 本文采用最大间隔法定义如下损失函数:

$$L = \sum_{(h,r,t) \in G} \sum_{(h',r',t') \in G'} [\gamma + g_r(\mathbf{h}', \mathbf{t}') - g_r(\mathbf{h}, \mathbf{t})]_+ \quad \#(9)$$

其中  $[x]_+ = \max(0, x)$ ,  $\gamma$  是分离正确三元组和错误三元组的间隔,  $G$  表示正确三元组集合,  $G'$  表示错误三元组集合。

由于知识图谱只包含正确的三元组, 本文需要人工构造错误三元组, 可通过替换正确三元组的头实体或尾实体来完成错误三元组的构造工作。正确三元组集合  $G$  可从给定的知识图谱数据集中获得。错误三元组集合  $G'$  的构造遵循以下规则:  $G' = \{(h', r, t) | h' \in E, (h, r, t) \in G\} \cup \{(h, r, t') | t' \in E, (h, r, t) \in G\}$ 。本文采用 Wang 等提出的“bern”策略<sup>[9]</sup>降低生成假的错误三元组的概率。“bern”策略根据关系类别为每个关系设置一个概率值, 根据该值决定替换关系对应的三元组的头实体或者尾实体。

模型采用随机梯度下降 (SGD) 优化目标函数。为了加速收敛和避免过拟合, 模型采用 TransE 的训练结果初始化实体与关系向量。

## 4 实验结果及分析

为了验证模型 TransE-NA 的有效性, 本文将在链接预测和三元组分类两个任务上评估模型性能, 并将实验结果与基准模型 TransE-NMM 以及其他最先进的模型的结果相比较。

### 4.1 实验数据

实验用到的数据集主要来自两个大型知识图谱: WordNet<sup>[6]</sup>和 Freebase<sup>[5]</sup>。WordNet 是覆盖范围广泛的英文语义知识库。WordNet 中的实体是具有不同概念的同义词, 关系表示同义实体之间的语义联系。本文采用 WordNet 的两个子集: WN11<sup>[18]</sup>和 WN18<sup>[16]</sup>。Freebase 是包含常见信息的世界知识库, 如: (robert\_noyce, profession, physicist) 和 (hans\_dreier, location, hollywood) 等。本文采用 Freebase 的两个子集: FB13<sup>[18]</sup>和 FB15k<sup>[16]</sup>。以上四个数据集的统计信息见表 1。

表 1 数据集的统计信息

数据集	关系	实体	训练集	验证集	测试集
FB15k	1,345	14,951	483,142	50,000	59,071
FB13	13	75,043	316,232	5,908	23,733
WN11	11	38,696	112,581	2,609	10,544
WN18	18	40,943	141,442	5,000	5,000

### 4.2 实验设置

#### 4.2.1 对比模型

本实验的对比模型分为三类: (1) 本文提出的 TransE-NA 模型; (2) 仅利用三元组结构信息的表示模型: Unstructured<sup>[16]</sup>、SE<sup>[17]</sup>、SME (linear)<sup>[19]</sup>、SME (bilinear)<sup>[19]</sup>、TransE<sup>[8]</sup>、TransH<sup>[9]</sup>、TransR<sup>[10]</sup>和 CTransR<sup>[10]</sup>; (3) 引入邻域信息的表示模型: GAKE<sup>[12]</sup>、TransE-NMM<sup>[13]</sup>和 TransN<sup>[14]</sup>。

#### 4.2.2 参数设置

在实验过程中, 本文设定以下参数的取值范围: batch\_size 大小  $B \in \{100, 200, 1000, 3000, 4800, 1000\}$ , 最大间隔  $\gamma \in \{0.5, 1, 2, 4, 5, 7, 10\}$ , 实体与关系向量的维度  $n \in \{50, 80, 100\}$ , 梯度下降学习率  $\lambda \in \{0.1, 0.01, 0.001\}$ , 相异性度量  $L \in \{L_1, L_2\}$ , 最低稀疏度  $\theta_{\min} \in \{0.0, 0.1, 0.2, 0.3, 0.5\}$ 。

在链接预测任务中, WN18 数据集上的最优参数为:  $B = 3000$ ,  $\gamma = 4$ ,  $n = 50$ ,  $\lambda = 0.001$ ,  $L = L_1$ ,  $\theta_{\min} = 0.2$ 。FB15k 数据集上的最优参数为:  $B = 5000$ ,  $\gamma = 2$ ,  $n = 100$ ,  $\lambda = 0.001$ ,  $L = L_1$ ,  $\theta_{\min} = 0.0$ 。在三元组分类任务中, WN11 数据集上的最优参数为:

$B = 5000$ ,  $\gamma = 7$ ,  $n = 50$ ,  $\lambda = 0.001$ ,  $L = L_1$ ,  $\theta_{\min} = 0.5$ 。

FB13 数据集上的最优参数为:  $B = 5000$ ,  $\gamma = 2$ ,

$n = 50$ ,  $\lambda = 0.01$ ,  $L = L_2$ ,  $\theta_{\min} = 0.0$ 。

由于 TransE-NMM 没有给出在 WN18 和

FB15k 数据集上链接预测任务的实验结果, 本文采用 Wang 等人<sup>[14]</sup>复现该算法得出的结果进行比较。由于其他对比模型所用数据集是一样的, 本文直接采用对比模型给出的实验结果。

表 2 链接预测的结果

数据集	WN18				FB15k			
	Mean Rank		Hits@10		Mean Rank		Hits@10	
	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt
Unstructured <sup>[16]</sup>	315	304	35.3	38.2	1,074	979	4.5	6.3
SE <sup>[17]</sup>	1,011	985	68.5	80.5	273	162	28.8	39.8
SME(linear) <sup>[19]</sup>	545	533	65.1	74.1	274	154	30.7	40.8
SME(bilinear) <sup>[19]</sup>	526	509	54.7	61.3	284	158	31.3	41.3
TransE <sup>[8]</sup>	263	251	75.4	89.2	243	125	34.9	47.1
TransH <sup>[9]</sup>	401	388	73.0	82.3	212	87	45.7	64.4
TransR <sup>[10]</sup>	238	225	79.8	92.0	198	77	48.2	68.7
CTransR <sup>[10]</sup>	231	218	79.4	92.3	199	<b>75</b>	48.4	70.2
GAKE <sup>[12]</sup>	-	-	-	-	228	119	44.5	64.8
TransN <sup>[14]</sup>	<b>142</b>	<b>134</b>	<b>86.2</b>	<b>97.1</b>	<b>181</b>	81	50.8	71.1
TransE-NMM <sup>[13]</sup>	259	249	84.2	95.0	190	101	49.5	65.7
TransE-NA	210	198	81.6	95.3	186	81	<b>53.7</b>	<b>78.6</b>

### 4.3 链接预测

#### 4.3.1 评估准则

链接预测任务旨在评估模型在知识图谱补全上的性能, 具体做法为预测三元组缺失的头实体或者尾实体。给定测试三元组  $(h, r, t)$ , 本文依次将实体集  $E$  中的实体取出, 替换给定三元组的头实体或者尾实体。然后根据式 (8) 定义的得分函数计算这些错误三元组和给定三元组的分数, 并将分数按降序排列。本任务的重点在于计算正确三元组的排名, 排名越高, 说明模型在该任务上的表现越好。类似于 TransE<sup>[8]</sup>, 本文采用 *Mean Rank* 和 *Hits@10* 作为评估准则。其中, *Mean Rank* 表示正确三元组的平均排名, *Hits@10* 表示正确三元组排在前 10 的比率。*Mean Rank* 越低, *Hits@10* 越高, 说明模型在该任务上的性能越好。评估设置也分为两种: “Raw” 和 “Filt”。“Raw” 设置如上述一样, 将正确三元组与替换产生的所有错误三元组进行排名。但是正如 TransE 所述, 替换产生的三元组不一定全是错误的, 有些也可能作为正确三元组存在于知识图谱中。所以这些假的错误三元组的排名可能会比给定正确三元组高, 反而降低总体 *Mean Rank* 指标。故而在排名之前剔除掉这些假的错误三元组, 这种设置也叫

“Filt”。本文将在这两种设置下评估模型的效果。

#### 4.3.2 实验结果分析

本模型在数据集 WN18 和 FB15k 上的实验结果见表 2。从实验结果可以看出: (1) TransE-NA 在两个数据集上的性能均超过了基准模型 TransE-NMM, 其中在 FB15k 数据集上 “Filt” 设置下的 *Hits@10* 指标提升了 12.9%。这验证了根据邻边关系矩阵抽取实体的邻节点相关属性的有效性, 同时说明了本文的邻节点选择策略有助于缓解数据稀疏问题; (2) 与仅利用三元组结构信息的模型中效果最好的 TransD 相比, 本文的模型在各个指标上均有提升。这验证了利用邻域信息辅助知识推理融合的有效性。而且本文模型仅基于 TransE 进行改进, 有理由认为, 当本文的邻域聚合器应用到比 TransE 更优秀的表示模型如 TransD 时, 效果会更好; (3) 虽然 GAKE 运用了邻节点、关系路径和邻边关系三种邻域信息, 其性能依旧远低于本文仅利用了邻节点的 TransE-NA 模型。这说明了在链接预测任务上, 基于翻译的表示模型性能优于基于图的表示模型; (4) TransE-NA 在 FB15k 数据集上的表现高于目前利用邻域信息的最好的模型 TransN, 但是在 WN18 数据集上表现略显逊色。本文分析原因如下: 实验过程中发现 WN18 数据集上存在许多

ORC (one-relation-circle) 结构<sup>[26]</sup>, 即由一些如反向、对称和传播关系等特殊关系组成的结构。Zhang<sup>[26]</sup>等人给出的数据表明, WN18 中由对称关系组成的 ORC 结构占据总数据的 19.84%。而以 TransE 为基础的翻译模型无法处理这种 ORC 结构。例如对于对称关系  $r$ , 存在  $(e_1, r, e_2) \in G$  和  $(e_2, r, e_1) \in G$ 。在 TransE 中, 对应的向量则满足:  $\mathbf{e}_1 + \mathbf{r} \approx \mathbf{e}_2$  和  $\mathbf{e}_2 + \mathbf{r} \approx \mathbf{e}_1$ 。从数学角度可推出:  $\mathbf{e}_1 \approx \mathbf{e}_2$  且  $\mathbf{r} \approx \mathbf{0}$ 。所以, 由对称关系连接的邻节点也无法辅助建模。TransN 通过给每个实体与关系单独设置用于编码邻域信息的上下文向量来弱化这种结构造成的影响, 但是造成了额外的计算开销, 不适用于大型知识图谱。FB15k 中也存在 ORC 结构, 但是由对称关系组成的 ORC 结构较少, 且 FB15k 上实体的平均邻节点个数大于 WN18, 选取到无用邻节点的可能性较小, 故而影响不大。所以在

该 FB15k 数据集上本模型的表现更好。有理由认为, 当本文的邻域聚合器应用到能处理 ORC 结构的模型上时, 实验效果会更好。

为了更加细致地评估和理解模型在此任务上的表现, 本文参照 TransE 提出的方式将 FB15k 数据集上的关系分为以下四类: 1-1, 1-N, N-1 和 N-N, 并比较在不同类别的关系下, 模型在 Hits@10 指标上的表现, 实验结果如表 3 所示。从实验结果可以看出: TransE-NA 在各关系类别下的表现均超过以往模型。相比于目前引入邻域信息表现最好的模型 TransN, 本模型在 N-1 和 N-N 关系下的头实体预测以及 1-N 和 N-N 关系下的尾实体预测上的性能有明显提高, 验证了本模型在面对复杂关系时有一定优势。而且每个实体与关系只有一个表示向量, 降低了模型的复杂度, 更适用于大型知识图谱建模。

表 3 FB15k 中各类别关系下的链接预测结果

任务 关系类型	头实体预测 (Hits@10)				尾实体预测 (Hits@10)			
	1-1	1-N	N-1	N-N	1-1	1-N	N-1	N-N
Unstructured <sup>[16]</sup>	34.5	2.5	6.1	6.6	34.3	4.2	1.9	6.6
SE <sup>[17]</sup>	35.6	62.6	17.2	37.5	34.9	14.6	68.3	41.3
SME(linear) <sup>[19]</sup>	35.1	53.7	19.0	40.3	32.7	14.9	61.6	43.3
SME(bilinear) <sup>[19]</sup>	30.9	69.6	19.9	38.6	28.2	13.1	76.0	41.8
TransE <sup>[8]</sup>	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50.0
TransH <sup>[9]</sup>	66.8	87.6	28.7	64.5	65.5	39.8	83.3	67.2
TransR <sup>[10]</sup>	78.8	89.2	34.1	69.2	79.2	37.4	90.4	72.1
CTransR <sup>[10]</sup>	81.5	89.0	34.7	71.2	80.8	38.6	90.1	73.8
TransN <sup>[14]</sup>	82.7	95.3	36.1	70.3	81.1	40.2	<b>94.8</b>	74.6
TransE-NMM <sup>[13]</sup>	59.9	87.1	25.0	48.5	59.1	36.3	93.2	53.7
TransE-NA	<b>85.8</b>	<b>95.9</b>	<b>42.6</b>	<b>79.3</b>	<b>84.9</b>	<b>53.8</b>	94.2	<b>82.4</b>

## 4.4 三元组分类

### 4.4.1 评估准则

三元组分类是一个判断未知三元组是否正确的二分类任务, 旨在评估模型在知识推理上的性能。在该任务上, 本文为每个关系  $r$  设置一个阈值  $\delta_r$ 。给定三元组  $(h, r, t)$ , 根据式 (8) 计算其得分。如果分数大于阈值  $\delta_r$ , 则判定为正确三元组; 反之, 则判定为错误三元组。本文通过在验证集上最大化分类准确率来确定每个关系  $r$  对应的阈值  $\delta_r$ , 然后在测试集上根据该阈值测试三元组的正确性。

### 4.4.2 实验结果分析

实验结果如表 4 所示, 由于 TransN 没有进行三元组分类实验, 本文只与基准模型 TransN-EMM 做对比。从表中可以看出: 模型在 WN11 数据集上的表现优于基准模型, 而在 FB13 数据集上的表现稍显逊色。本文分析原因如下: FB13 数据集涵盖的 13 种关系分别为 gender、nationality、place\_of\_death、place\_of\_birth、location、profession、institution、cause\_of\_death、religion、parents、children、ethnicity 和 spouse。比起其他数据集包含的关系, FB13 数据集的关系之间关联性不够强, 即很难通过已知关系去推测未知关系。所以, FB13 数据集上的有效邻域信息更难挖掘。为提升知识推理的性能, 提高正确三

元组和错误三元组的区分度，本文将进一步考虑融合如实体描述、实体类别等多源信息。

表 4 三元组分类结果

数据集	WN11	FB13
SE <sup>[17]</sup>	53.0	75.2
SME(bilinear) <sup>[19]</sup>	70.0	63.7
TransE <sup>[8]</sup>	75.9	81.5
TransH <sup>[9]</sup>	78.8	83.3
TransR <sup>[10]</sup>	85.9	82.5
CTransR <sup>[10]</sup>	85.7	-
TransE-NMM <sup>[13]</sup>	86.8	<b>88.6</b>
TransE-NA	<b>87.0</b>	86.8

## 5 总结

本文提出了聚合邻域信息的联合知识表示模型 TransE-NA，为罕见实体提供更多语义信息，缓解数据稀疏问题。首先，模型提出邻节点选择机制，根据实体的稀疏度决定其邻节点数量。然后，模型为每个关系设置一个投影矩阵，通过对应的投影转换选取邻节点最相关的属性作为实体的邻域信息。实验结果表明该模型性能超过了基准模型，验证了该模型在聚合邻域信息上的有效性。

在未来的工作中，本文将考虑以下几个方向：

(1) 鉴于神经网络在特征提取上取得了有效进展，本文拟尝试运用神经网络模型聚合邻域信息；

(2) 除了邻域信息，知识库中的其他信息如实体描述、实体类别以及关系描述等信息也能辅助知识推理，本文将考虑多源信息的联合表示；

(3) 由于基于 TransE 的模型无法处理 ORC 结构，本文未来将考虑改进模型，使其能够编码 ORC 结构下的实体与关系。

## 参考文献

- [1] Bordes A, Weston J, Usunier N. Open question answering with weakly supervised embedding models[C]//Joint European conference on machine learning and knowledge discovery in databases. Springer, Berlin, Heidelberg, 2014: 165-180.
- [2] Hoffmann R, Zhang C, Ling X, et al. Knowledge-based weak supervision for information extraction of overlapping relations[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011: 541-550.
- [3] Peng M, Huang J, Sun Z, et al. Improving distant supervision of relation extraction with unsupervised methods[C]//International Conference on Web Information Systems Engineering. Springer, Cham, 2016: 561-568.
- [4] Damljanovic D, Bontcheva K. Named entity disambiguation using linked data[C]//Proceedings of the 9th Extended Semantic Web Conference. 2012: 231-240.
- [5] 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. AcM, 2008: 1247-1250.
- [6] Miller G A. WordNet: a lexical database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.
- [7] Suchanek F M, Kasneci G, Weikum G. Yago: a core of semantic knowledge[C]//Proceedings of the 16th international conference on World Wide Web. ACM, 2007: 697-706.
- [8] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data[C]//Advances in neural information processing systems. 2013: 2787-2795.
- [9] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes[C]//Twenty-Eighth AAAI conference on artificial intelligence. 2014.
- [10] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion[C]//Twenty-ninth AAAI conference on artificial intelligence. 2015.
- [11] Ji G, He S, Xu L, et al. Knowledge graph embedding via dynamic mapping matrix[C]//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). 2015, 1: 687-696.
- [12] Feng J, Huang M, Yang Y. GAKE: graph aware knowledge embedding[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016: 641-651.
- [13] Nguyen D Q, Sirts K, Qu L, et al. Neighborhood mixture model for knowledge base completion[J]. arXiv preprint arXiv:1606.06461, 2016.
- [14] Wang C C, Cheng P J. Translating Representations of Knowledge Graphs with Neighbors[C]//The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. ACM, 2018: 917-920.
- [15] Ji G, Liu K, He S, et al. Knowledge graph completion with adaptive sparse transfer matrix[C]//Thirtieth AAAI Conference on Artificial Intelligence. 2016.
- [16] Bordes A, Glorot X, Weston J, et al. Joint learning of words and meaning representations for open-text semantic parsing[C]//Artificial Intelligence and Statistics. 2012: 127-135.
- [17] Bordes A, Weston J, Collobert R, et al. Learning structured embeddings of knowledge bases[C]//Twenty-Fifth AAAI Conference on Artificial Intelligence. 2011.
- [18] Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]//Advances in neural information processing systems. 2013: 926-934.
- [19] Bordes A, Glorot X, Weston J, et al. A semantic matching energy function for learning with multi-relational data[J]. Machine Learning, 2014, 94(2): 233-259.
- [20] Jenatton R, Roux N L, Bordes A, et al. A latent factor model for highly multi-relational data[C]//Advances in Neural Information Processing Systems. 2012:



- 3167-3175.
- [21] Sutskever I, Tenenbaum J B, Salakhutdinov R R. Modeling relational data using bayesian clustered tensor factorization[C]//Advances in neural information processing systems. 2009: 1821-1828.
  - [22] Werbos P J. Backpropagation through time: what it does and how to do it[J]. Proceedings of the IEEE, 1990, 78(10): 1550-1560.
  - [23] Lin Y, Liu Z, Luan H, et al. Modeling relation paths for representation learning of knowledge bases[J]. arXiv preprint arXiv:1506.00379, 2015.
  - [24] Gao H, Shi J, Qi G, et al. Triple Context-Based Knowledge Graph Embedding[J]. IEEE Access, 2018, 6: 58978-58989.
  - [25] Xu J, Chen K, Qiu X, et al. Knowledge graph representation with jointly structural and textual encoding[J]. arXiv preprint arXiv:1611.08661, 2016.
  - [26] Zhang W. Knowledge graph embedding with diversity of structures[C]//Proceedings of the 26th International Conference on World Wide Web Companion. International World Wide Web Conferences Steering Committee, 2017: 747-753.