

用于知识库完善的端到端结构感知卷积网络

尚超,^{1*} 唐云,² 黄静,² 毕金波,¹ 何晓东,² 周博文²

¹康涅狄格大学计算机科学与工程系, 美国康涅狄格州斯托尔市。

²JD 人工智能研究, 美国加州山景城。

{chao.shang.jinbo.bi@uconn.edu, {yun.tang, jing.huang, xiaodong.he, bowen.zhou@jd.com}

摘要

知识图谱嵌入一直是知识库完成的一个活跃的研究课题, 从最初的*TransE*、*TransH*、*DistMult*等到目前最先进的*ConvE*, 都有逐步的改进。*ConvE*使用嵌入的二维卷积和多层非线性特征对知识图进行建模。该模型可以被有效地训练, 并可扩展到大型知识图谱。然而, 在*ConvE*的嵌入空间中并没有结构执行。最近的图卷积网络工作(*GCN*)通过成功利用图连接结构提供了另一种学习图节点嵌入的方法。在这项工作中, 我们提出了一种新型的端到端结构感知卷积网络(*SACN*), 它将*GCN*和*ConvE*的优点结合起来。*SACN*由一个加权图卷积网络(*WGCN*)的编码器和一个称为*Conv-TransE*的卷积网络的去编码器组成。*WGCN*利用知识图的节点结构、节点属性和边缘关系类型。它有可学习的权重, 可以调整局部聚集中使用的邻居的信息量, 从而导致更准确的图节点嵌入。图中的节点属性在*WGCN*中被表示为额外的节点。解码器*Conv-TransE*使最先进的*ConvE*在实体和关系之间进行转换, 同时保持与*ConvE*相同的链接预测性能。我们在标准的F B15k-237和WN18RR数据集上证明了所提出的*SACN*的有效性, 它在HITS@1、HITS@3和HITS@10方面比最先进的*ConvE*有大约10%的相对改进。

简介

近年来, 大规模的知识库(KBs), 如Freebase (Bolacker等人, 2008)、DBpedia (Auer等人, 2007)、NELL (Carlson等人, 2010)和YAGO3 (Mahdisoltani, Biega, and Suchanek, 2013), 已经被建立起来, 以存储有关普通事实的结构化信息。KBs是多关系图, 其节点代表实体, 边代表实体之间的关系, 边上标有不同的关系。关系以(*s*, *r*, *o*)三联体的形式被组织起来(例如, 实体*s* = Abraham Lincoln, relation *r* = DateOfBirth, entity *o* = 02-12-1809)。这些知识库被广泛地用于网络搜索、检索-----。

*在JD AI Research实习期间完成的工作。

Copyright ©2019, Association for the Advancement of Artificial Intelligence (人工智能促进会)

智力(www.aaai.org)。保留所有权利。

ommendation和问题回答。尽管这些知识库已经包含了数以百万计的实体和三要素, 但与现实世界的现有事实和新增加的知识相比, 它们还远远不够完整。因此, 为了在现有的基础上预测新的三要素, 从而进一步扩展知识库, 知识库的完成是很重要的。

最近活跃的知识库完善研究领域之一是知识图谱嵌入: 它将实体和关系的语义编码在一个连续的低维向量空间(称为嵌入)。然后, 这些嵌入被用于预测新的关系。从一个简单有效的方法*TransE* (Bordes等人, 2013)开始, 许多知识图的嵌入方法被提出, 如*TransH* (Wang等人, 2014)、*TransR* (Lin等人, 2015)、*DistMult* (Yang等人, 2014)、*TransD* (Ji等人, 2015)、*ComplEx* (Trouillon等人, 2016)、*STra nsE* (Nguyen等人, 2016)。一些调查(Nguyen 2017; Wang等人, 2017)给出了这些embedding方法的细节和比较。

最近的*ConvE* (Dettmers等人, 2017)模型在嵌入和多层非线性特征上使用二维卷积, 并在知识图谱链接预测的常见基准数据集上实现了最先进的性能。在*ConvE*中, *s*和*r*的嵌入被重新塑造并串联成一个输入矩阵, 并被送至卷积层。 $n \times n$ 的卷积滤波器被用来输出跨越不同维度的特征图。

嵌入条目。因此, *ConvE*并不像*TransE*那样保持反式属性, *TransE*是一个加法嵌入向量操作: $e_s + e_r \approx e_o$ (Nguyen et al. 2017)。在本文中, 我们删除了*ConvE*的重塑步骤, 并对其进行了操作。与原始的*ConvE*相比, 这种修改带来了更好的性能, 并且有一个直观的解释, 即在一个嵌入三元组(e_s, e_r, e_o)中保持*s*、*r*和*o*的全局学习度量相同。我们将这种嵌入命名为*Conv-TransE*。

*ConvE*也没有将知识图中的连接结构纳入嵌入空间。相比之下, 图卷积网络(*GCN*)一直是创建节点嵌入的有效工具, 它聚合了每个节点在图邻域的局部形成(Kipf

and Welling 2016b; Hamilton, Ying, and Leskovec 2017a;

Kipf和Welling 2016a; Pham等人2017; Shang等人。

2018).GCN模型有额外的好处 (Hamilton, Ying, and Leskovec 2017b), 例如利用与节点相关的属性。他们还可以在计算每个节点的卷积时施加相同的ag娱乐平台, 这可以被认为是一种正则化的方法, 并提高效率。虽然可扩展性最初是GCN模型的一个问题, 但最新的数据高效GCN, PinSage (Ying等人, 2018), 能够处理数十亿的节点和边。

在本文中, 我们提出了一个端到端的图结构感知卷积网络 (SACN), 它将GCN和ConvE的所有优点结合起来。SACN由一个加权图卷积网络 (WGCV) 的编码器和一个称为Conv-TransE的卷积网络的解码器组成。WGCV利用知识图谱的节点结构、节点的属性和关系类型。它有可学习的权重来确定用于局部聚合的邻居的信息量, 从而导致更准确的图节点嵌入。节点属性被添加到WGCV中, 以方便整合。WGCV的输出成为解码器Conv-TransE的输入。Conv-TransE与ConvE类似, 但不同的是Conv-TransE保留了实体和关系之间的转换特性。我们表明Conv-TransE比ConvE表现得更好, 而且我们的SACN在标准基准数据集上比Conv-TransE有进一步的提高。我们的模型和实验的代码是公开的,¹。

我们的贡献总结如下。

- 我们提出了一个端到端的网络学习框架SACN同时受益于GCN和Conv-TransE。编码器GCN模型利用了图结构和图节点的属性。解码器Conv-TransE通过特殊的卷积来模拟ConvE, 并保持TransE的平移特性和ConvE的预测性能。
- 我们证明了我们提出的SACN的有效性。在标准的FB15k-237和WN18RR数据集上, 和在HITS@1、HITS@3和HITS@10方面, 比最先进的ConvE有大约10%的相对改进。

相关工作

知识图谱嵌入学习一直是一个活跃的研究领域, 直接应用于知识库完成 (即链接预测) 和关系提取。TransE (Bordes等人, 2013年) 通过将实体和关系投射到同一个嵌入中, 开始了这一工作路线。

的向量空间, 其平移约束为 $e_s + e \approx e_o$ 。后来的工作增强了KG的嵌入模型, 如TransH (Wang等人, 2014)、TransR (Lin等人, 2015) 和TransD (Ji等人, 2015) 引入了关系翻译的新表示, 从而增加了模型的复杂性。这些模型被归类为翻译差异模型 (Wang等人, 2017) 或加法模型, 而DistMult (Yang等人, 2014) 和ComplEx (Trouillon等人, 2014) 则被归类为翻译差异模型。

2016) 是乘法模型 (Sharma, Talukdar, and others 2018), 由于用于计算实体-关系-实体三联体可能性的乘法得分函数。

最新的KG嵌入模型是ConvE (Dettmers等人, 2017) 和ConvKB (Nguyen等人, 2017)。ConvE是第一个在不同嵌入维度的embeddings上使用二维卷积的模型, 希望能提取更多的特征交互。ConvKB用一维卷积取代了ConvE中的二维卷积, 它将卷积压缩为相同的嵌入维度, 并保持TransE的平移特性。ConvKB可以被认为是一个特例, 它只使用宽度等于1的滤波器。尽管ConvKB被证明比ConvE好, 在两个数据集 (FB15k-237和WN18RR) 并不一致, 所以我们将这些结果排除在我们的比较表中。ConvE和ConvKB的另一个主要区别是在模型中使用的损失函数。ConvE使用的是交叉熵损失, 在解码器中可以用1-N评分来加速, 而ConvKB使用的是铰链损失, 从正面例子和负面例子采样计算。我们采用ConvE的解码器是因为我们可以很容易地将GCN的编码器和ConvE的解码器整合到一个端到端的训练框架中, 而ConvKB不适合我们的方法。

这些嵌入模型在完成知识库的效率和扩展能力方面取得了良好的表现。然而, 这些方法只对关系型三联体进行建模, 而忽略了大量与图节点相关的属性, 例如人的年龄或音乐的发行区域。此外, 这些模型在嵌入空间中不执行任何大规模的连接结构, 并且完全忽略了知识图谱结构。本项目 (SACN) 通过使用图形卷积网络 (GCN) 的一个变体作为编码器, 以及ConvE的一个变体作为解码器, 在一个端到端的训练框架中处理这两个问题。

GCN最早是在 (Bruna等人, 2013) 中提出的, 其中图卷积操作被定义在傅里叶域。图Laplacian的eigendecomposition引起了激烈的计算。后来, 平滑的参数化规格过滤器 (Henaff, Bruna, and LeCun 2015; Defferrard, Bresson, and Vandergheynst 2016) 被引入, 以实现空间域的定位并提高计算效率。最近, Kipf等人 (Kipf and Welling 2016b) 通过Chebyshev多项式的一阶近似来简化这些光谱方法。空间图卷积方法 (Hamilton, Ying, and Leskovec 2017a) 直接在图上定义卷积, 利用邻接矩阵对所有空间邻居的节点特征进行汇总。

GCN模型大多因其巨大的内存需求而被批评, 无法扩展到大规模的图。然而, (Ying等人, 2018年) 开发了一种数据高效的GCN算法, 称为PinSage, 它结合了高效的随机漫步和图卷积来生成节点的嵌入, 包含了图结构以及节点特征。在Pinterest数据上的实验是迄今为止最大的深度图嵌入的应用, 有30亿个节点和

¹<https://github.com/JD-AI-Research-Silicon-Valley/SACN>

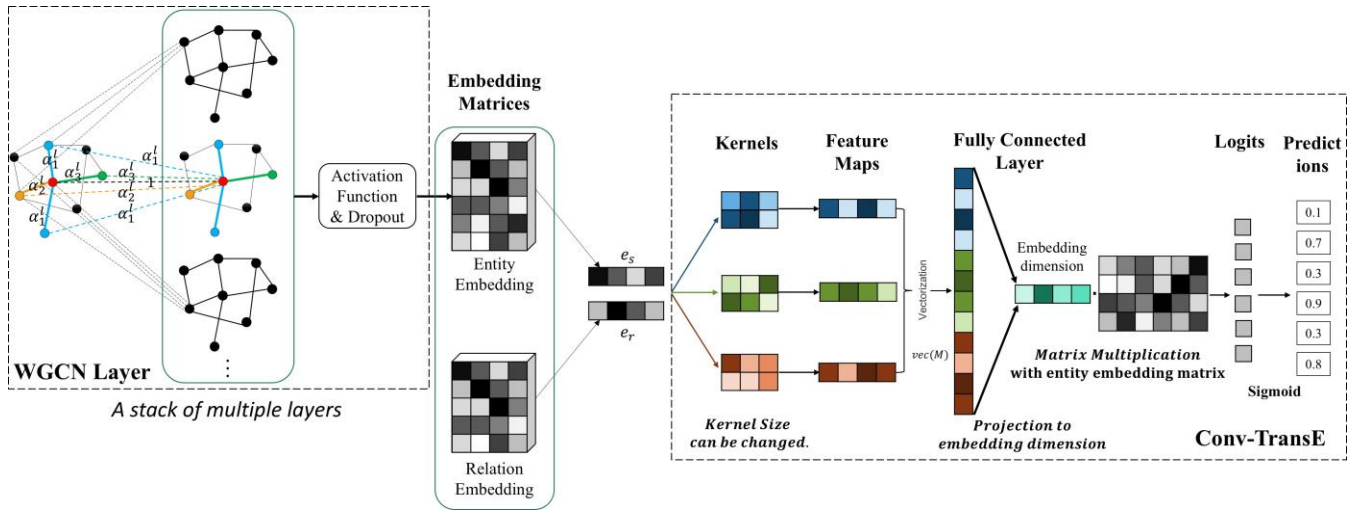


图1：我们的端到端结构感知卷积网络模型的说明。对于编码器，多个WGCN层的堆栈建立了一个实体/节点嵌入矩阵。对于解码器， e_s 和 e_r 被送入Conv-TransE。输出的嵌入被矢量化和投影，并通过内积与所有候选的 e_o 嵌入相匹配。一个logistic sigmoid函数被用来获得分数。

180亿条边（Ying等人，2018）。这一成功为基于GCN的新一代网络规模的推荐系统铺平了道路。因此，我们相信，我们所提出的模型可以利用Conv-TransE的巨大图结构和高计算效率。

方法

在这一节中，我们描述了提议的端到端SACN。编码器WGCN的重点是通过聚合KB中的关系所指定的连接实体来表示实体。以节点嵌入为输入，解码器Conv-TransE网络旨在通过恢复KB中的原始三联体来更准确地表示关系。编码器和解码器都是通过最小化嵌入 e_s 和 e_o 之间的差异（交叉熵）共同训练的。 r 和 e_o ，以保留平移属性 $e_s + e_r \approx e_o$ 。本节中我们考虑一个无向图 $G(V, E)$ ，其中 V 是一个节点集， $|V| = N$ ， $E \subseteq V \times V$ 是一个边集， $|E| = M$ 。

加权图卷积层

WGCN是经典GCN的延伸（Kipf和Welling 2016b），它权衡了不同类型的边。在进行汇总时，对边进行不同的处理，而权重是在网络的训练中自适应地学习。通过这种自适应，WGCN可以控制信息量。在聚合过程中使用来自邻近节点的信息。粗略地说，WGCN将一个多关系的KB图视为多个单关系的子图，其中每个子图需要一种特定类型的关系。WGCN确定了当组合一个节点的GCN嵌入时，要给每个子图多少权重。

第 l 个WGCN层取长度为 F 的输出向量。 l

为前一层的每个节点作为输入，并生成一个由 F^{l+1} 元素组成的新表示。让 h^l 代表第 l 层的节点 v_i 的输入（行）向量，因此 $H^l \in \mathbb{R}^{N \times F}$ 是该层的输入矩阵。初始嵌入 H^0 是随机抽取的高斯的。如果WGCN中共有 L 层，第 L 层的输出 H^{L+1} ，就是最终的嵌入。在一个有 E 条边的多关系KB图中，让边缘类型的总数为 T 。两个相邻节点之间的交互强度由其关系类型决定，这强度由每个边缘类型的参数 $\{a_t, 1 \leq t \leq T\}$ 指定，该参数在神经网络中自动学习。网络。

图1说明了SACN的整个过程。在这个例子中，网络的WGCN层计算了中间图中红色节点的嵌入。这些层聚合了KB关系中规定的相邻实体节点的嵌入。边缘的三种颜色（蓝色、黄色和绿色）表示图中三种不同的关系类型。相应的三个实体节点根据 a_t ，在这一层用不同的权重相加，以获得红色节点的嵌入。具有相同颜色（相同关系类型）的边使用相同的 a_t 。每一层都有自己的关系权重集 a^l 。因此，在这一层中的输出

节点 v 的第 l 层 i 可以写成如下。

$$h_{v,i}^{l+1} = \sigma \sum_{j \in N_v} \text{atg}(h_{v,i}^l, h_j^l) \quad (1)$$

其中 $h^l \in \mathbb{R}^F$ 是节点 v 的输入，而 v 是一个节点 j 在节点 v 的邻居 N_v 。ii
。 g 函数指定如何纳入邻居信息。请注意，这里的行为函数 σ 适用于它的每一个组件。

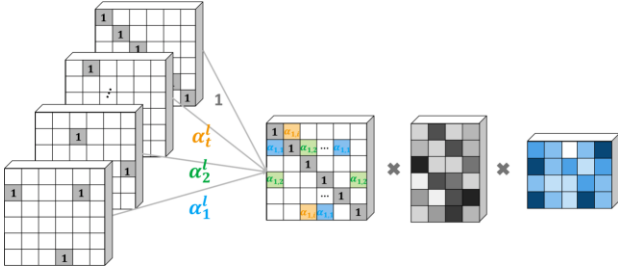


图2：用于实体嵌入的加权图卷积网络（WGCN）。

输入向量。尽管任何适合于KB embedding的函数 g 都可以与提议的框架结合使用，但我们实现了以下的 g 函数。

$$g(h^l_i, h^l_j) = h^l_i W^l_j. \quad (2)$$

其中， $W^l \in \mathbb{R}^{F \times F+1}$ 是连接系数矩阵并用于线性转换 h^l_i 到 h^{l+1}_i 。

在公式（1）中，所有相邻节点的输入向量被加起来，但不包括节点 v_i 本身，因此网络中执行自循环。对于节点 v_i ，传播过程被定义为。

$$h^{l+1}_i = \sigma \left(\alpha^l_i h^l_i W^l_i + h^l_i W^l_i \right). \quad (3)$$

第 l 层的输出是一个节点特征矩阵。 $H^{l+1} \in \mathbb{R}^{N \times F}$ 是 H^l 的第 $l+1$ 行，代表第 $(l+1)$ 层的节点 v_i 的特征。

上述过程可以被组织成一个矩阵多如图2所示，通过一个邻接矩阵同时计算所有节点的嵌入。对于每个关系（边）类型，邻接矩阵 A_t 是一个二进制矩阵，如果存在连接 v_i 和 v_j 的边，其第 ij 项为1，否则为0。最终的邻接矩阵写成如下。

$$A^l = \sum_{t=1}^T (\alpha^l_t A_t) + I. \quad (4)$$

其中 I 是大小为 $N \times N$ 的身份矩阵。基本上， A^l ，就是" A "的邻接矩阵的加权和。

子图加上自我连接。在我们的实现中，我们考虑了每层线性变换中的所有一阶邻居，如图2所示。

$$H^{l+1} = \sigma (A^l H^l W^l). \quad (5)$$

节点属性。在一个知识库图中，节点与几个属性相关，形式为（实体，关系，属性）。例如， $(s = \text{Tom}, r = \text{people.person.gender}, a = \text{male})$ 是一个实例，其中gender是一个与人相关的属性。如果一个向量如果对节点属性使用"表示法"，就会有两个潜在的问题。首先，每个节点的属性数量通常很少，而且各不相同。

因此，属性向量将是非常稀疏的。第二，属性向量中的零值可能有模糊的含义：节点没有特定的属性，或者节点错过了这个属性的值。这些零值会影响嵌入的准确性。

在这项工作中，知识图谱中的实体属性由网络中的另一组节点表示，称为属性节点。属性节点作为"桥梁"连接相关实体。实体嵌入可以通过这些"桥梁"来将实体的属性纳入其嵌入中。因为这些属性在三联体中展示，所以我们在关系三联体中类似于实体 o 的代表来表示属性。请注意，每种类型的属性都对应于一个节点。例如，在我们的例子中，性别由一个节点表示，而不是由"男性"和"女性"两个节点表示。这样，WGCN不仅利用了图的连接结构（关系和关系类型），而且还有效地利用了节点属性（一种图的结构）。这就是为什么我们把

我们的WGCN是一个结构感知的卷积网络。

Conv-TransE

我们将Conv-TransE模型开发为一个解码器，它是

基于ConvE，但具有TransE的平移特性： $e_s + e_r \approx e_o$ 。我们的方法与ConvE的关键区别在于，在堆叠 e_s 和 e_r 之后没有重塑。大小为 $2 \times k$ 的滤波器（或内核）， $k \in \{1, 2, 3, \dots\}$ ，是在卷积中使用。图1中的例子使用了 2×3

核来计算二维卷积。在我们的实证研究中，我们试验了几种这样的设置。

请注意，在SACN的编码器中，关系嵌入的维度通常被选择为与实体嵌入的维度相同，所以换句话说，等于 F^L 。

因此，这两个嵌入可以叠加。对于解码器，输入是两个嵌入矩阵：一个是来自WGCN的所有实体节点的 $\mathbb{R}^{N \times F^L}$ ，另一个是

$\mathbb{R}^{M \times F^L}$ 为关系嵌入矩阵，它被训练为好。因为我们使用的是小批量的随机训练算法。算法，解码器的第一步对嵌入矩阵进行查找操作，以检索小型批次中三联体的输入 e_s 和 e_r 。

更确切地说，给定 C 个不同的内核，其中第 c 个内核的参数为 ω_c ，解码器中的卷积计算方法如下。

$$m_c(e_s, e_r, n) = \sum_{\tau=0}^{K-1} \omega_c(\tau, 0) \hat{e}_s(n + \tau) + \omega_c(\tau, 1) \hat{e}_r(n + \tau). \quad (6)$$

其中， K 是内核宽度， n 索引输出向量中的条目， $n \in [0, F^L - 1]$ ，内核参数 ω_c 是可训练的。 \hat{e}_s 和 \hat{e}_r 是 e 的填充版本。 s 和 r

。如果内核的维度 s 是奇数，则第一个 $\lfloor K/2 \rfloor$ 和最后一个 $\lfloor K/2 \rfloor$ 分量被填充为

0。这里 $\lfloor value \rfloor$ 返回值的下限。否则，前 $\lfloor K/2 \rfloor - 1$ 和最后 $\lfloor K/2 \rfloor$ 的分量被填充为0。其他组件直接从 e_s 和 e_r 中复制。如公式（6）所示，卷积操作相当于

表1：评分函数 $\psi(e_s, e_o)$ 。这里 e_s 和 e_r 表示 e_s 和 e_r 的2D重塑。

模型	评分函数 $\psi(e_s, e_o)$
转换器	$\ e_s + e_r - e_o\ _p$
复杂情	$\langle E_s, E_r, E_o \rangle$
况下, C	$\langle e_s, e_r, e_o \rangle f$
onvEx	$(\text{vec}(f(\text{concat}(e_s, e_r) * \omega))W)e_o$
ConvE	$\text{concat}(g([e_s, e_r, e_o] * \omega))\beta$
ConvKB	$f(\text{vec}(\mathbf{M}(e_s, e_r))W)e_o$
SACN。	

一维卷积后的 e_s 和 e_r 的总和。因此，它保留了 e_s, e_r 嵌入的平移特性。输出形成一个矢量 $M_c(e_s, e_r) = [m_c(e_s, e_r, 0), \dots, m_c(e_s, e_r, F^L - 1)]$ 。将卷积的输出向量与所有内核对齐，得到一个矩阵 $\mathbf{M}(e_s, e_r) \in \mathbb{R}^{C \times F^L}$ 。

最后，非线性卷积后Conv-TransE方法的评分函数定义如下。

$$\psi(e_s, e_o) = f(\text{vec}(\mathbf{M}(e_s, e_r))W)e_o. \quad (7)$$

其中， $W \in \mathbb{R}^{CF^L}$ 是线性变换的矩阵， f 表示一个非线性函数。特征图矩阵被重塑为一个向量 $\text{vec}(\mathbf{M}) \in \mathbb{R}^{CF^L}$ ，并且 pro- 在一个 F^L 维的空间中，使用 W 进行线性转换。形成。然后，计算出的嵌入与 e_o ，

通过一个适当的距离度量进行匹配。在我们实验的训练过程中，我们应用了Logistic sigmoid函数来进行评分。

$$p(e_s, e_r, e_o) = \sigma(\phi(e_s, e_o)). \quad (8)$$

在表1中，我们总结了几个最先进的模型所使用的评分函数。向量 e_s 和 e_o 分别是主语和宾语嵌入， e_r 是关系嵌入，"concat"表示将输入连接起来，"*"表示卷积运算。

综上所述，拟议的SACN模型利用了知识图谱节点的连接性、节点属性和关系类型。WGCN中的可学习权重有助于从相邻的图节点中收集适应性的信息量。实体属性作为网络中的附加节点被添加到WGCN中，并且很容易被集成到WGCN中。Conv-TransE保持了实体和关系之间的平移属性，以学习节点嵌入来进行链接预测。我们还强调，无论是否使用节点属性，我们的SACN都比ConvE有重大改进。

实验

基准数据集

本研究利用三个基准数据集（FB15k-237、WN18RR和FB15k-237-Attr）来评估链接预测的性能。

FB15k-237。FB15k-237（Toutanova和Chen 2015）数据集包含知识库关系三元组和Freebase实体对的文本提及，在（Toutanova和Chen 2015）发表的工作中使用。这些知识库三元组是FB15K（Bordes等人，2013）的一个子集。

表2：数据集的统计。

数据集	FB15k-237	WN18RR	FB15k-237-Attr
实体	14,541	40,943	14,744
关系	237	11	484
火车的边缘	272,115	86,835	350,449
瓦尔。边缘	17,535	3,034	17,535
测试边缘	20,466	3,134	20,466
属性的三倍	-	-	78,334
属性	-	-	203

最初来自于Freebase。在FB15k-

237中，逆向关系被删除。

WN18RR。WN18RR（Dettmers等人，2017）是由WN18（Bordes等人，2013）创建的，它是WordNet的一个子集。WN18由18个关系和40,943个实体组成。然而，许多通过倒置三元组获得的文本三元组从训练集。因此，WN18RR数据集（Dettmers等人，2017）被创建，以确保评估数据集不存在反向关系测试泄漏。总之，WN18RR数据集包含93,003个三元组，有40,943个实体和11种关系类型。

数据建设

以前的方法大多只对实体和关系进行建模，而忽略了丰富的实体属性。我们的方法可以很容易地对大量的实体属性三元组进行建模。为了证明其效率，我们从FB24k（Lin, Liu, and Sun 2016）数据集中提取属性三元组，建立了名为FB15k-237-Attr的评估数据集。

FB24k。FB24k（Lin, Liu, and Sun 2016）是基于Freebase数据集建立的。FB24k只选择构成至少30个三要素的实体和关系。实体的数量是23,634个，关系的数量是673个。此外，反转的关系被从原始数据集中删除。在FB24k数据集中，提供了属性三元组。FB24k包含207,151个属性三元组和314个属性。

FB15k-237-Attr。我们从FB24k中提取了FB15k-237中实体的属性三元组。在映射过程中，从原来的14,541个实体中，有7,589个结点具有结点属性。最后，我们从FB24k中提取了78,334个属性三元组。这些三元组包括203个属性和247个关系。基于这些三元组，我们创建了"FB15k-237-Attr"数据集，其中包括14,541个实体节点，203个属性节点，484个关系类型。所有的78,334个属性三元组都与FB15k-237的训练集相结合。

实验设置

我们的Conv-

TransE和SACN模型中的超参数是在训练期间通过网格搜索确定的。我们手动指定超参数

之三的范围：学习率{0.01, 0.005, 0.003, 0.001}。

辍学率{0.0, 0.1, 0.2, 0.3, 0.4, 0.5}，嵌入大小

{100, 200, 300}，果核数量{50, 100, 200, 300}。

表3：FB15k-237、WN18RR和FB15k-237-Attr数据集的链接预测。

模型	FB15k-237				WN18RR			
	点 击 率				点 击 率			
	@10	@3	@1	MRR	@10	@3	@1	MRR
DistMult (Yang et al. 2014)	0.42	0.26	0.16	0.24	0.49	0.44	0.39	0.43
ComplEx (Trouillon等人, 2016)	0.43	0.28	0.16	0.25	0.51	0.46	0.41	0.44
R-GCN (Schlichtkrull等人, 2018)	0.42	0.26	0.15	0.25	-	-	-	-
ConvE (Dettmers等人, 2017)	0.49	0.35	0.24	0.32	0.48	0.43	0.39	0.46
Conv-TransE	0.51	0.37	0.24	0.33	0.52	0.47	0.43	0.46
SACN	0.54	0.39	0.26	0.35	0.54	0.48	0.43	0.47
使用FB15k-237-Attr的SACN	0.55	0.40	0.27	0.36	-	-	-	-
绩效改进	12.2%	14.3%	12.5%	12.5%	12.5%	11.6%	10.3%	2.2%

和内核大小 $\{2 \times 1, 2 \times 3, 2 \times 5\}$ 。

这里所有的模型都使用了两层的WGCN。对于不同的数据集，我们发现以下设置效果很好：对于FB15k-

237，设置dropout为0.2，核数为100，学习率为0.003，SACN的嵌入大小为200；对于WN18RR数据集，设置dropout为0.2，核数为300，学习率为0.003，SACN的嵌入大小为200。当使用Conv-TransE-单独的模型时，这些设置仍然很有效。

每个数据集被分成三组：训练、验证和测试，这与原始ConvE的设置相同。我们使用自适应矩（Adam）算法（Kingma和Ba 2014）来训练模型。我们的模型由PyTorch实现，在NVIDIA Tesla P40图形处理单元上运行。对于FB15k-237数据集，SACN每个历时的计算时间约为1分钟。对于WN18RR，SACN在一个历时中的计算时间约为1.5分钟。

结果

评估协议

我们的实验使用排名在前1、3和10位的正确实体的比例（Hits@1，Hits@3，Hits@10）和平均互惠排名（MRR）作为衡量标准。此外，由于知识图谱中存在一些被破坏的三元组，我们采用了过滤设置（Bordes等人, 2013），即在排名前过滤掉所有有效的三元组。

链接预测 我们在标准FB15k-237、WN18RR和FB15k-237-

Attr上的结果见表3。表3报告了四个不同基线模型和两个我们的模型在三个知识图谱数据集上的Hits@10、Hits@3、Hits@1和MRR结果。FB15k-237-

Attr数据集被用来证明节点属性的效率。因此，我们在FB15k-237-Attr中运行我们的SACN，与使用FB15k-237的SACN做比较。

我们首先将我们的Conv-TransE模型与四个基线模型进行比较。ConvE在与所有基线的比较中表现最好。在FB15k-237数据集中，我们的Conv-TransE模型比ConvE的Hits@10提高了4.1%，比ConvE

的Hits@3提高了5.7%的测试幅度。在WN18RR数据集中，Conv-TransE比ConvE的Hits@10提高了8.3%，比ConvE的Hits@3提高了9.3%。对于这些结果。

我们得出结论，使用神经网络的Conv-TransE保持了实体和关系之间的平移特性，并取得了更好的性能。

其次，结构信息被添加到我们的SACN模型中。在表3中，与所有基线方法相比，SACN在测试数据集中也得到了最好的表现。在FB15k-237中，与ConvE相比，我们的SACN模型将Hits@10值提高了10.2%，Hits@3值提高了11.4%，Hits@1值提高了8.3%，测试的MRR值提高了9.4%。在WN18RR数据集中，与ConvE相比，我们的SACN模型将Hits@10值提高了12.5%，Hits@3值提高了11.6%，Hits@1值提高了10.3%，MRR值提高了2.2%。因此，我们的方法比没有属性的ConvE有明显的改进。

第三，我们将节点属性添加到我们的SACN模型中，即我们使用FB15k-237-Attr来训练SACN。注意，SACN比没有属性的ConvE有明显的改进。添加属性后，性能再次提高。我们使用属性的模型比ConvE的Hits@10提高了12.2%，Hits@3提高了14.3%，Hits@1提高了12.5%，MRR提高了12.5%。此外，我们使用属性的SACN与没有属性的SACN相比，Hits@10提高了1.9%，Hits@3提高了2.6%，Hits@1提高了3.8%，MRR提高了2.9%。

为了更好地与ConvE进行比较，我们也在ConvE中使用属性。在这里，属性将被视为实体三联体。按照官方ConvE代码的默认设置，FB15k-237-Attr的测试结果是：0.46（Hits@10），0.33（Hits@3），0.22（Hits@1）和0.30（MRR）。与没有属性的性能相比，在ConvE中添加属性并没有提高性能。

。

收敛分析

图3显示了三种模型的收敛情况。我们可以看到，SACN（红线）在七次历时后总是比Conv-TransE（黄线）好。而在120个历时后，SACN的性能不断提高。然而，Conv-TransE在大约120个历时后取得了最佳性能。这两个模型之间的差距证明了结构信息的有用性。当使用FB15k-237-Attr数据集时，"SACN+Attr"的性能优于

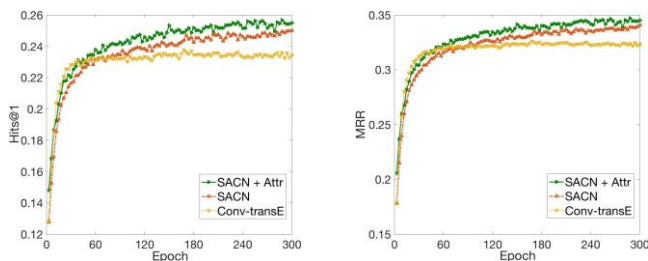


图3：SACN、Conv-TransE的收敛研究在FB15k-237中的模型和FB15k-237-Attr中的SACN (SACN

+ Attr) 使用验证集。由于篇幅限制，这里只报告Hits@1和MRR的结果。

表4：FB15k-237和FB15k-237-Attr数据集的内核大小分析。"SACN+Attr"指使用FB15k-237-Attr数据集的SACN。

模型	内核大小	FB15k-237			
		点 击 率			MRR
		@10	@3	@1	
Conv-TransE	2×1	0.504	0.357	0.234	0.324
Conv-TransE	2×3	0.513	0.365	0.240	0.331
Conv-TransE	2×5	0.512	0.361	0.239	0.329
SACN	2×1	0.527	0.379	0.255	0.345
SACN	2×3	0.536	0.384	0.260	0.351
SACN	2×5	0.536	0.385	0.261	0.352
SACN+Attr	2×1	0.535	0.384	0.260	0.351
"SACN+Attr"	2×3	0.543	0.394	0.268	0.360
SACN+Attr	2×5	0.547	0.396	0.268	0.360

内核大小分析

在表4中，我们的模型考察了不同的内核大小。 2×1 的内核意味着知识或信息在一个at-1之间转换。实体向量的属性和对应的再定位向量的属性。如果我们将核的大小增加到 $2 \times k$ ，其中 $k=\{3, 5\}$ ，信息在一个com-实体向量中的S属性的组合和一个组合的关系向量中的k属性。如表4所示，更大的视图来收集属性信息可以帮助提高性能。在FB15k-237和FB15k-237-Attr数据集中，所有的Hits@1、Hits@3、Hits@10和MRR的值都可以通过增加内核大小来提高。然而，最佳内核大小可能取决于任务。

节点度分析

知识图中节点的度是指与该节点相连的边的数量。度数大的节点意味着它有更多的相邻节点，这种节点比其他度数小的节点能从相邻节点获得更多的information。如表5所示，我们列出了不同度数范围内的不同节点集的结果。我们计算了Hits@10和Hits@3的平均分值。伴随着度数范围的增加，平均数值

表5：使用FB15k-237数据集的节点indegree研究。

学位范围	Conv-TransE		SACN	
	平均点击率		平均点击率	
	@10	@3	@10	@3
[0,100]	0.192	0.125	0.195	0.134
[100,200]	0.441	0.245	0.441	0.253
[200,300]	0.696	0.446	0.705	0.429
[300,400]	0.829	0.558	0.806	0.577
[400,500]	0.894	0.661	0.868	0.663
[500,1000]	0.918	0.767	0.891	0.695
[1000, 最大]	0.992	0.941	0.981	0.922

的Hits@10和Hits@3将被增加。首先，对于具有小度数的节点，它受益于来自SACN的WGCN层的邻居信息的聚合。它的嵌入可以被稳健地估计。其次，对于高度数的节点来说，这意味着有更多的信息通过GCN聚合，其嵌入的估计在邻居中被分阶段地平滑。因此，从SACN学到的嵌入比从Conv-TransE学到的嵌入要差。这个问题的一个解决方案是邻居选择，如（Ying等人，2018）。

结论和未来工作

我们引入了一个端到端的结构感知的卷积网络（SACN）。编码网络是一个加权的图卷积网络，利用知识图连接结构、节点属性和关系类型。具有可学习权重的WGCN的好处是可以从相邻的图节点收集-摄取适应性的信息量。此外，实体属性被添加为网络中的节点，因此属性被转化为知识结构信息，这很容易被整合到节点嵌入中。SACN的评分网络是一个卷积神经网络，称为Conv-TransE。它使用卷积网络将关系建模为转置操作，捕捉实体和关系之间的转置特性。我们还证明，仅Conv-TransE就已经达到了最先进的性能水平。SACN的性能总体上比ConvE等技术水平提高了10%左右。

在未来，我们希望将邻居选择的想法纳入我们的训练框架，例如，（Ying等人，2018）中的importance pooling，在聚合邻居的向量表示时考虑到邻居的重要性。我们还想扩展我们的模型，使其能够在更大的知识图谱中得到扩展（Ying等人，2018）。

鸣谢

这项工作得到了美国国家科学基金会CCF-1514357和IIS-1718738拨款的部分支持，以及美国国立卫生研究院R01DA037349和K02DA043063拨款给毕金波。

参考文献

- Auer, S.; Bizer, C.; Kobilarov, G.; Lehmann, J.; Cyganiak, R.; and Ives, Z. 2007. Dbpedia:一个开放数据网络的核心。In *The semantic web*. Springer. 722-735.
- Bollacker, K.; Evans, C.; Paritosh, P.; Sturge, T.; and Taylor, J. 2008. Freebase: 一个合作创建的用于构建人类知识的图形数据库。In *Proceedings of the 2008 ACM SIGMOD international conference on Management- ment of data*, 1247-1250. ACM.
- Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; and Yakhnenko, O. 2013. 为多关系数据建模的翻译嵌入。In *Advances in neural information processing systems*, 2787-2795.
- Bruna, J.; Zaremba, W.; Szlam, A.; and LeCun, Y. 2013. 谱系网络和图上的局部连接网络。arXiv预印本arXiv:1312.6203.
- Carlson, A.; Betteridge, J.; Kisiel, B.; Settles, B.; Hruschka Jr, E. R.; and Mitchell, T. M. 2010. 迈向永不停息的语言学习的架构。In *AAAI*, volume 5, 3. Atlanta.
- Defferrard, M.; Bresson, X.; and Vandergheynst, P. 2016. 图上的卷积神经网络与快速局部谱系过滤。In *Advances in Neural Information Processing Systems*, 3844-3852.
- Dettmers, T.; Minervini, P.; Stenetorp, P.; and Riedel, S. 2017. 卷积2d知识图嵌入。arXiv预印本arXiv:1707.01476.
- Hamilton, W.; Ying, Z.; and Leskovec, J. 2017a. 大图上的诱导性表征学习。In *Advances in Neural Information Processing Systems*, 1025-1035.
- Hamilton, W. L.; Ying, R.; and Leskovec, J. 2017b. 图上的再现学习。方法和应用。arXiv预印本arXiv:1709.05584.
- Henaff, M.; Bruna, J.; and LeCun, Y. 2015. 图结构数据上的深度对话网络。arXiv预印本arXiv:1506.05163.
- Ji, G.; He, S.; Xu, L.; Liu, K.; and Zhao, J. 2015. 通过动态映射矩阵的知识图谱嵌入。In *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)*, volume 1, 687-696.
- Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Kipf, T. N., and Welling, M. 2016a. 变量图自动编码器。In *Advances in neural information processing systems, Bayesian Deep Learning Workshop*.
- Kipf, T. N., and Welling, M. 2016b. 用图卷积网络进行半监督的分类。arXiv预印本arXiv:1609.02907.
- Lin, Y.; Liu, Z.; Sun, M.; Liu, Y.; and Zhu, X. 2015. 学习实体和关系嵌入以完成知识图谱。In *AAAI*, volume 15, 2181-2187.
- Lin, Y.; Liu, Z.; and Sun, M. 2016. 知识代表含有实体、属性和关系的概念学习。民族性 1:41-52.
- Mahdisoltani, F.; Biega, J.; and Suchanek, F. M. 2013. Yago3: 一个来自多语言维基百科的知识库。在 *C IDR* 中。
- Nguyen, D. Q.; Sirts, K.; Qu, L.; and Johnson, M. 2016. Stranse: 知识库中实体和关系的新型嵌入模型。arXiv预印本arXiv:1606.08140.
- Nguyen, D. Q.; Nguyen, T. D.; Nguyen, D. Q.; and Phung, D. 2017. 基于卷积神经网络的知识库完成的新型嵌入模型。arXiv预印本arXiv:1712.02121.
- Nguyen, D. Q. 2017. 用于知识库完成的实体和关系的嵌入模型概述。arXiv预印本arXiv:1703.08098.
- Pham, T.; Tran, T.; Phung, D.; and Venkatesh, S. 2017. 用于集体分类的柱状网络。
- Schlichtkrull, M.; Kipf, T. N.; Bloem, P.; van den Berg, R.; Titov, I.; and Welling, M. 2018. 用图卷积网络对关系数据进行建模。在 *欧洲语义网会议上*, 593-607. 斯普林格。
- Shang, C.; Liu, Q.; Chen, K.-S.; Sun, J.; Lu, J.; Yi, J.; and Bi, J. 2018. 基于边缘注意力的多关系图卷积网络。arXiv预印本arXiv:1802.04944.
- Sharma, A.; Talukdar, P.; et al. 2018. 争取理解知识图谱嵌入的几何学。In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, 122-131.
- Toutanova, K., and Chen, D. 2015. 用于知识库和文本推理的观察特征与潜在特征。In *Proceedings of the 3rd Workshop on Continuous Vector Space Models and their Compositionality*, 57-66.
- Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, E.; and Bouchard, G. 2016. 用于简单链接预测的复杂嵌入。在 *国际机器学习会议上*, 2071-2080.
- Wang, Z.; Zhang, J.; Feng, J.; and Chen, Z. 2014. 通过超平面上的平移实现Knowledge graph的嵌入。在 *AAAI*, 第14卷, 1112-1119.
- Wang, Q.; Mao, Z.; Wang, B.; and Guo, L. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering* 29 (12): 2724-2743.
- Yang, B.; Yih, W.-t.; He, X.; Gao, J.; and Deng, L. 2014. 嵌入实体和关系, 用于知识库的学习和推理。arXiv预印本arXiv:1412.6575.
- Ying, R.; He, R.; Chen, K.; Eksombatchai, P.; Hamilton, W. L.; and Leskovec, J. 2018. 用于网络规模推荐系统的图卷积神经网络。In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 974-983.