



基于多尺度空洞卷积的知识图谱表示方法

杜昊桐^{1,2}, 王震^{1,2*}, 聂弘毅^{1,2}, 姚权铭^{3*}, 李学龙²

1. 西北工业大学机电学院, 西安 710072

2. 西北工业大学光电与智能研究院, 西安 710072

3. 清华大学电子工程系, 北京 100084

* 通信作者. E-mail: w-zhen@nwpu.edu.cn, qyaoaa@connect.ust.hk

收稿日期: 2021-03-31; 修回日期: 2021-05-28; 接受日期: 2021-08-30; 网络出版日期: 2022-07-14

科技部重点研发计划 (批准号: 2018YFB1403501)、国家杰出青年科学基金 (批准号: 62025602)、国家自然科学基金 (批准号: U1803263, 11931015, 81961138010)、陕西省科技创新团队 (批准号: 2020TD-013) 和霍英东教育基金 (批准号: 171105) 资助项目

摘要 知识图谱嵌入是知识图谱研究中的一项重要课题. 它旨在根据已观测到的三元组, 学习知识图谱中实体与关系的低维向量表示. 知识图谱嵌入在许多下游任务中发挥了巨大作用, 例如知识图谱补全、三元组分类. 如今, 深度模型利用神经网络强大的非线性拟合能力, 在知识图谱嵌入领域展示出了优异的性能. 然而, 现有的大多数方法忽略了实体与关系之间的多尺度特征交互, InceptionE 是目前唯一考虑到了多尺度交互特征的模型, 但由于大量的计算开销导致其很难进行训练. 本文提出了一种全新的知识图谱嵌入模型 MDCE, 它使用多尺度空洞卷积核在不同的尺度空间捕捉丰富的交互特征. 同时, MDCE 相比于 InceptionE 方法的计算开销更小. 我们在多个基准数据集上进行了大量实验. 在链接预测任务上的结果表明, MDCE 不仅在性能方面超过了已有的工作, 而且更加高效、稳健.

关键词 知识图谱, 知识图谱嵌入, 深层模型, 多尺度特征, 链接预测, 人工智能

1 引言

知识图谱^[1,2]是一种特殊的图结构, 图上的节点代表实体, 节点间的有向边表示实体间的关系. 知识图谱由大量事实组成, 每个事实都可以用(头实体、关系、尾实体)这样的三元组形式来描述, 用符号表示为 (h, r, t) . 这意味着头实体 h 与尾实体 t 通过某个特定的关系 r 形成了关联, 例如(Bill Gates, founded, Microsoft), 如图1所示. 知识图谱相较于传统的图数据, 包含了丰富的语义信息, 它作为一种全新的数据组织形式, 在搜索引擎^[3]、问答系统^[4]和推荐系统^[5]等领域中扮演了关键角色. 因此, 人们花费了大量精力构建了许多知识图谱, 例如 WordNet^[6], Freebase^[7]和 YAGO^[8].

然而, 现有的知识图谱中存在许多缺失事实, 人工补全的成本过于巨大, 如何利用图谱中已有的事实对缺失关系进行预测, 成为我们关心的问题. 基于知识图谱嵌入的链接预测方法因为其优异的性能

引用格式: 杜昊桐, 王震, 聂弘毅, 等. 基于多尺度空洞卷积的知识图谱表示方法. 中国科学: 信息科学, 2022, 52: 1204–1220, doi: 10.1360/SSI-2021-0111

Du H T, Wang Z, Nie H Y, et al. Multi-scale dilated convolutional network for knowledge graph embedding (in Chinese). Sci Sin Inform, 2022, 52: 1204–1220, doi: 10.1360/SSI-2021-0111

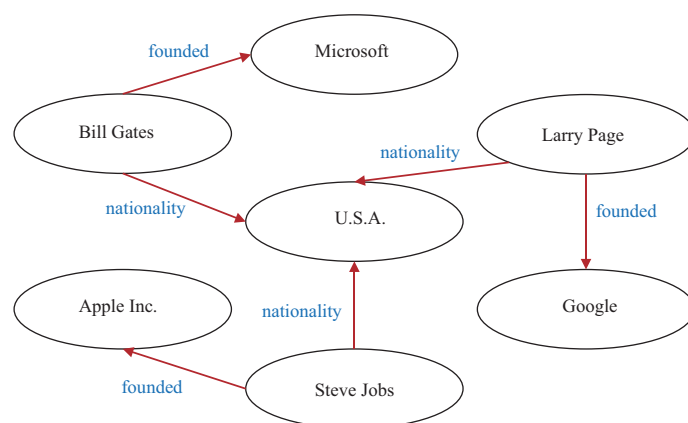


图 1 (网络版彩图) 一个知识图谱的简单示例

Figure 1 (Color online) An example of the knowledge graph

与良好的扩展性受到了广泛关注. 知识图谱嵌入技术旨在学习知识图谱中的实体与关系的潜在特征, 这些信息可以帮助我们对缺失关系进行预测^[9~12].

基于浅层模型的知识图谱嵌入方法, 例如平移距离模型^[13~17]或双线性模型^[18~20], 采用简单的结构或是基础的矩阵运算操作, 提取知识图谱中的潜在语义信息与结构特征. 这些模型的表达能力较弱, 泛化能力不足, 利用这些模型训练得到的特征向量进行链接预测, 不能得到很好的效果^[21, 22]. 为了提高模型的表达能力, 出现了许多复杂的多层架构模型方法, 特别是基于神经网络架构的嵌入方法, 依靠其深层、复杂的结构展示出强大的非线性表达能力与良好的泛化性能. 基于神经网络的知识表示方法受到研究者的广泛关注, 代表性的模型有 ConvE^[21], ConvR^[22], InceptionE^[23]等.

其中, ConvR 将关系嵌入参数 r 重塑为若干个卷积核, 将这些卷积核与实体嵌入 h 进行卷积操作, 这相当于使用关系特定的卷积核与实体进行卷积运算, 使得实体与关系的不同维度都进行了充分的特征交互, 卷积核产生的所有特征图都能够捕捉到这样的交互特征. 但 ConvR 只关注了实体与关系之间的单一尺度特征交互, 没有考虑在不同尺度上的特征交互情形, 这限制了模型的表达能力. InceptionE 创新地利用 Inception 网络^[24], 使用不同大小的静态卷积核, 捕捉不同范围内的实体与关系的交互特征. 这些特征是由实体与关系在不同的尺度范围内进行交互得到的. 这种非单一尺度的特征交互建模方式, 提升了生成的嵌入质量, 这意味着多尺度信息在知识图谱嵌入表示问题中是需要被关注与考虑的.

基于上述观察, 本文结合了 ConvR 与 InceptionE 的优点, 提出了基于多尺度空洞卷积的知识图谱表示方法 (multi-scale, dilated convolutional network for knowledge graph embedding, MDCE). 在 MDCE 中, 我们创新地将空洞卷积引入到知识图谱表示问题中, 对关系特定的卷积核添加多种比例的空洞, 使其获得不同尺度的特征交互与提取能力, 在参数利用率更高的情况下获得了实体与关系的多尺度交互特征, 然后通过多尺度的信息融合, 获得知识图谱中实体与关系的合理表示, 利用得到的嵌入表示对实体间可能存在的链接缺失进行预测.

本文的主要贡献如下:

- 我们创新性地将在空洞卷积应用到知识图谱嵌入表示方法中, 在计算量保持不变的情况下, 对更大区域进行特征交互建模;
- 我们提出了基于多尺度空洞卷积的知识图谱表示方法, 获得了实体与关系的多尺度交互特征, 通

表 1 本文相关符号描述
Table 1 Symbols and notations

Symbol	Description
\mathcal{E}, \mathcal{R}	The set of entities and set of relations
$h, t \in \mathcal{E}, r \in \mathcal{R}$	Head and tail entity, relation
$\mathbf{h}, \mathbf{t} \in \mathbb{R}^{d_e}$	Embedding of head entity and tail entity
$\mathbf{r} \in \mathbb{R}^{d_r}$	Embedding of relation
$\psi(h, r, t)$	The score function of the triplet (h, r, t)
$\hat{\mathbf{h}}$	2D reshaping of vector
ω_r^n	The n -dilated convolution kernel generated by reshaping

表 2 常见嵌入表示方法对应的评分函数
Table 2 Definitions of some popular scoring functions

Model type	Model	Score function	Multi-scale method
Shallow model	TransE [13]	$-\ \mathbf{h} + \mathbf{r} - \mathbf{t}\ _1$	\times
	RotatE [17]	$-\ \mathbf{h} \odot \mathbf{r} - \mathbf{t}\ _1$	\times
	DistMult [19]	$\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle$	\times
	ComplEx [20]	$\text{Re}(\langle \mathbf{h}, \mathbf{r}, \text{conj}(\mathbf{t}) \rangle)$	\times
Deep model	ConvE [21]	$f(\text{vec}(f([\hat{\mathbf{h}}; \hat{\mathbf{r}}] * \omega))\mathbf{W})\mathbf{t}$	\times
	HypER [25]	$f(\text{vec}(\mathbf{h} * \text{vec}^{-1}(\mathbf{r}\mathbf{H}))\mathbf{W})\mathbf{t}$	\times
	ConvR [22]	$f(\text{vec}(f(\hat{\mathbf{h}} * \omega_r))\mathbf{W})\mathbf{t}$	\times
	InteractE [26]	$f(\text{vec}(f([\hat{\mathbf{h}}; \hat{\mathbf{r}}] * \omega))\mathbf{W})\mathbf{t}$	\times
	ParamE-CNN [27]	$\mathbf{W}(\text{vec}(f(f(\hat{\mathbf{h}} * \omega_{r_0}) * \omega_{r_1})))\mathbf{t}$	\times
	InceptionE [23]	$f(\text{vec}(\text{concat}(f([\hat{\mathbf{h}}; \hat{\mathbf{r}}] * \omega_{1 \times 1}); f([\hat{\mathbf{h}}; \hat{\mathbf{r}}] * \omega_{3 \times 3}); f(f([\hat{\mathbf{h}}; \hat{\mathbf{r}}] * \omega_{3 \times 3}) * \omega_{3 \times 3})))\mathbf{W})\mathbf{t}$	\checkmark
	MDCE	$f(\text{vec}(\text{concat}(\hat{\mathbf{h}} * \omega_r^1; \hat{\mathbf{h}} * \omega_r^2; \hat{\mathbf{h}} * \omega_r^3)\mathbf{W}_F)\mathbf{W}_P)\mathbf{t}$	\checkmark

过多尺度的信息融合, 获得知识图谱中实体与关系的合理表示;

- 我们在多个数据集上, 比较了 MDCE 与其他方法的链接预测效果与运行效率, 大量的实验表明, 我们的方法相比于目前最先进的模型在性能与效率方面均获得了很大的提升.

符号定义. 本文中使用较多的符号及其描述如表 1 所示. 本文用小写粗体字母表示向量, 用大写粗体字母表示矩阵. $\text{Re}(\cdot)$ 表示复数的实部, $\text{conj}(t)$ 代表复数 t 的共轭复数, $\langle \mathbf{a}, \mathbf{b}, \mathbf{c} \rangle = \sum_{i=1}^d \mathbf{a}_i \cdot \mathbf{b}_i \cdot \mathbf{c}_i$ 代表内积, \odot 表示哈达玛积.

2 相关工作

第 2.1 小节介绍几类具有代表性的知识图谱嵌入表示的相关工作, 主要可以分为两大类, 浅层模型与深层模型. 其中, 浅层模型包括平移距离模型与双线性模型, 深层模型包括神经网络模型. 表 2 [13, 17, 19~23, 25~27] 展示了一些常见的嵌入表示方法对应的评分函数. 第 2.2 小节主要介绍空洞卷积的相关内容.

2.1 知识图谱嵌入

2.1.1 浅层模型

在知识图谱嵌入表示方法中,浅层模型主要包括平移距离模型和双线性模型,下面简要介绍这两种模型.

平移距离模型是受“词类比”现象^[28]启发而产生的. Bordes 等^[13]提出了第 1 个平移距离模型 TransE,他认为知识库中的关系对应着实体间的某种平移作用,这与“词类比”现象是相似的.具体来说,对于知识库中的每个三元组,TransE 希望 $\mathbf{h} + \mathbf{r}$ 尽量与 \mathbf{t} 接近,评分函数表示为 $\psi(\mathbf{h}, \mathbf{r}, \mathbf{t}) = -\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_1$. 在 TransE 的基础上,TransH^[14], TransR^[15], TransD^[16] 等模型通过引入不同的映射矩阵,使平移距离模型可以更好地建模复杂关系,从不同方面提升模型性能.

然而,平移距离模型大多形式简单,模型表达能力不强,对复杂关系进行建模的能力有限.

双线性模型采用基于点积的相似度函数来衡量一个元组的合理性.这类模型通常都会采用如下的形式:

$$\psi(\mathbf{h}, \mathbf{r}, \mathbf{t}) = \mathbf{h}^T \mathbf{M}_r \mathbf{t}, \quad (1)$$

其中 \mathbf{M}_r 表示与关系 \mathbf{r} 相关的方阵. 在 DistMult^[19] 中,令 $\mathbf{M}_r = \text{diag}(\mathbf{r})$ 以防止模型的过拟合. ComplEx^[20] 在这种形式基础上更进一步,将实体与关系嵌入扩展到复空间,增强模型的表达能力.这些模型通过不同形式的 \mathbf{M}_r 建模复杂的语义关系.

然而,对于这类模型,想要提升其表达能力,唯一的方法就是扩大实体与关系的嵌入维度^[21],但增加嵌入维度意味着极高的参数消耗,模型极易出现过拟合的情形,这限制了此类模型对于大规模知识图谱的可扩展性.

2.1.2 深层模型

随着近年来海量数据涌现和深度学习的发展,神经网络展示出强大的非线性拟合能力与良好的泛化性能,基于神经网络深层结构的知识图谱嵌入表示模型受到研究者的广泛关注.

MLP^[29], NTN^[30] 等方法使用多层感知机结构,加深了模型的层数,引入了更复杂的设计思路,尽管这样有助于提高模型的表现力,但此类方法通常也需要大量参数,模型容易过拟合而导致性能下降.

ConvE^[21] 利用多层卷积神经网络提取实体与关系之间的局部交互特征,这种基于卷积神经网络的方法,利用了卷积操作参数效率高、计算速度快等优势,克服了传统多层感知机结构的缺点,增强了模型的表达能力.基于卷积神经网络的知识图谱嵌入表示方法也成为当前研究热点. InteractE^[26] 对 ConvE 进行了改进,使用特征排列和圆周卷积等方法,进一步增强特征的局部交互能力,提升知识表示效果. CapsE^[31] 利用胶囊网络可以捕捉到不同特征映射在同一维度上的深层特征这一特点,首次针对实体与关系嵌入表示的同一维度信息进行深度结构建模. InceptionE^[23] 将计算机视觉领域中的 Inception 网络^[24, 32] 首次用于知识图谱嵌入表示中,使用不同大小的卷积核捕捉多个尺度的交互特征,提升了生成的嵌入质量,这也是首个在知识图谱嵌入表示问题中,考虑多尺度特征交互的模型.

而 ConvR^[22] 模型不同于上述的卷积神经网络模型,没有单纯地将实体嵌入与关系嵌入进行堆叠,再进行静态卷积运算,而是将关系嵌入重塑为卷积核,将关系嵌入特征作为卷积神经网络的参数,这相当于使用关系特定的卷积核与实体进行动态自适应卷积运算,使得实体与关系的不同维度都进行了充分的特征交互,产生的所有特征图都能够包含丰富的交互特征.知识图谱是一种多关系图,所以关系

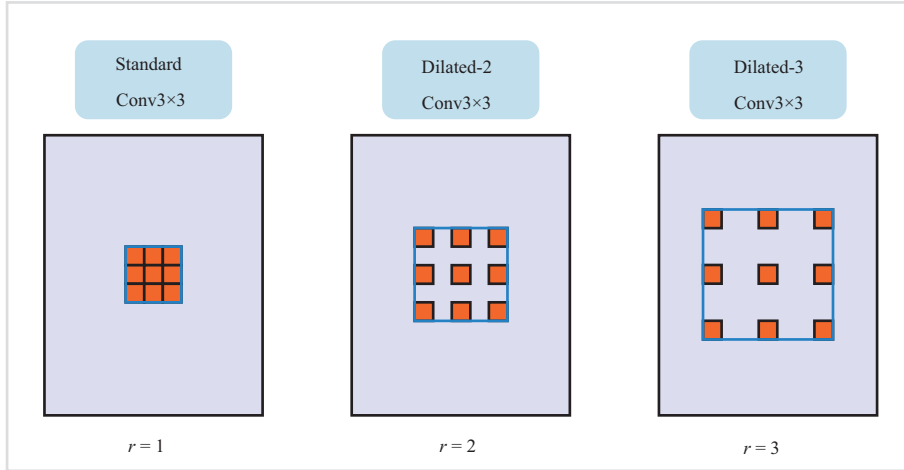


图 2 (网络版彩图) 不同空洞率的 3×3 空洞卷积示意图

Figure 2 (Color online) An illustration of 3×3 convolutions with different dilation rates

自适应方法更适合用来处理知识图谱嵌入表示问题. ParamE-CNN [27] 参考 ConvR 方法, 堆叠了两层关系自适应卷积核, 提升了性能但产生了大量冗余参数, 且模型运行效率低下.

2.2 空洞卷积

空洞卷积, 又名扩张卷积或膨胀卷积 [33]. 它的主要设计思想是在卷积核的每个维度上, 在两个连续的滤波器值之间插入若干个空洞 (0 值), 以此来增加卷积核的感受野. 对于一个二维信号 \mathbf{X} , 其通过空洞卷积算子运算得到的输出可以表示为如下形式:

$$Y(m, n) = \sum_{a=1}^{k_h} \sum_{b=1}^{k_w} \mathbf{X}(m + r \cdot a, n + r \cdot b) \times \omega^r(a, b), \quad (2)$$

其中, Y 表示输出信号, k_w 和 k_h 代表卷积核的宽度与高度, r 对应着我们对 \mathbf{X} 进行采样所使用的卷积核 ω^r 的空洞率. 我们可以通过改变 r , 得到具有不同大小的感受野的空洞卷积核. 如图 2 所示, 对于一个 $k \times k$ 大小的二维标准卷积核, 其对应的空洞率为 r 的空洞卷积核的边长为 $k_d = k + (k - 1) \cdot (r - 1)$. 特别地, 标准离散卷积可以看作是 $r = 1$ 的空洞卷积.

空洞卷积最早用于小波变换 [34] 中, 用来提升计算效率. 如今, 空洞卷积已经被广泛应用于语义分割 [35]、语音合成 [36] 等领域. 通过调节空洞卷积的空洞率, 获得不同大小的卷积核, 利用这些卷积核对同一个信号进行采样, 可以捕捉到不同粒度的多尺度信息. 就我们所知, MDCE 是第 1 个通过多尺度空洞卷积捕捉多尺度信息的知识图谱嵌入表示方法.

3 本文提出的模型

首先在 3.1 小节对要解决的问题进行形式化描述, 然后在 3.2 小节, 对基于多尺度空洞卷积的知识表示方法进行介绍.

3.1 问题描述

定义1 (知识图谱) 一个知识图谱 $\mathcal{G} = \{(h, r, t) | (h, r, t) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}\}$ 通常被定义为一系列三元组

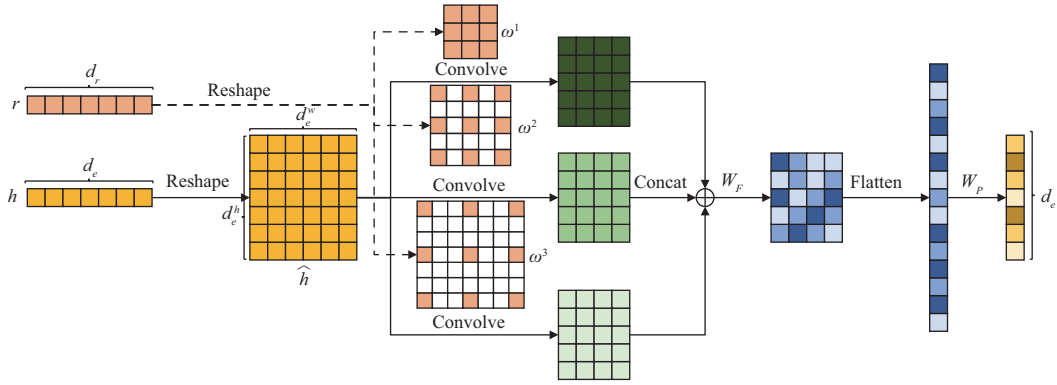


图 3 (网络版彩图) MDCE 模型的示意图

Figure 3 (Color online) An illustration of the proposed MDCE

的集合, 其中 \mathcal{E} 代表实体集合, \mathcal{R} 代表关系集合, $h, t \in \mathcal{E}$ 分别代表头实体和尾实体, $r \in \mathcal{R}$ 代表从头实体链接到尾实体的特定关系.

定义 2 (知识图谱嵌入) 知识图谱嵌入旨在学习知识图谱中所有实体与关系的低维、实值表示, 形式化定义为学习一个映射函数 $\phi: \mathcal{E} \times \mathcal{R} \times \mathcal{E} \mapsto \mathbb{R}^{|\mathcal{E}| \times d_e} \times \mathbb{R}^{|\mathcal{R}| \times d_r}$, 利用学习到的嵌入表示进行下游任务.

定义 3 (链接预测) 链接预测任务是指利用知识图谱中已存在的实体与关系去推断某两个实体间是否有关系缺失, 它属于知识图谱补全中的一个子任务. 该问题可以形式化定义为学习一个评分函数 $\psi: \mathcal{E} \times \mathcal{R} \times \mathcal{E} \mapsto \mathbb{R}$. 给定一个输入元组 $x = (h, r, t)$, $\psi(x)$ 的分值越高, 这个元组为有效元组的概率就越大.

3.2 基于多尺度空洞卷积的知识图谱表示方法

MDCE 的模型架构如图 3 所示. 给定一个元组 (h, r, t) , 将 h 对应的向量 $\mathbf{h} \in \mathbb{R}^{d_e}$ 重塑为 $\hat{\mathbf{h}} \in \mathbb{R}^{d_e^h \times d_e^w}$ (在这里 $d_e = d_e^h d_e^w$), $\mathbf{r} \in \mathbb{R}^{d_r}$ 重塑为不同大小的空洞卷积核 $\omega^1, \omega^2, \omega^3$, 并与 $\hat{\mathbf{h}}$ 进行卷积运算, 生成包含不同尺度信息的特征图; 再对多尺度特征图进行特征融合和拉直操作, 并通过全连接层投影到实体 t 的嵌入; 最后, 基于点积相似度计算给定元组的得分 $\psi(h, r, t)$.

3.2.1 嵌入表示

对于一个知识图谱 \mathcal{G} , 我们将其中每个实体初始化为一个唯一的 $\mathbf{h} \in \mathbb{R}^{d_e}$, 每个关系初始化为一个唯一的 $\mathbf{r} \in \mathbb{R}^{d_r}$. 给定一组三元组 (h, r, t) , 头实体 h 对应向量 $\mathbf{h} \in \mathbb{R}^{d_e}$ 被切分成 d_e^h 个长度为 d_e^w 的向量 ($d_e = d_e^h d_e^w$), 通过按行堆叠的方式将这些向量重塑成矩阵 $\hat{\mathbf{h}} \in \mathbb{R}^{d_e^h \times d_e^w}$ 参与卷积运算. 将向量重塑为矩阵的原因是 Dettmers 等^[21]指出, 二维卷积在实体-关系交互建模方面优于一维卷积, 可以有效提高嵌入之间的交互效率.

将关系 r 对应的嵌入 $\mathbf{r} \in \mathbb{R}^{d_r}$ 分割成 k_c 个相同大小的块 $\mathbf{r}^{(1)}, \dots, \mathbf{r}^{(k_c)}$, 每个 $\mathbf{r}^{(l)} \in \mathbb{R}^{d_r/k_c}$ 被切分成 k_h 个长度为 k_w 的向量, 通过按行堆叠的方式将这些向量重塑为二维标准卷积核 $\omega_r^{(l)} \in \mathbb{R}^{k_h \times k_w}$. 其中, k_c 表示参与运算的卷积核数目, k_h 与 k_w 代表每个卷积核的高和宽. 因此, d_r 的值应该设置为 $k_c k_h k_w$, 以保证关系嵌入能够被准确分割为卷积核.

通过调节卷积核的空洞率参数 i , 生成具有不同大小的空洞卷积核 $\omega^i \in \mathbb{R}^{k_h^i \times k_w^i}$, 其中, $k_h^i = k_h + (k_h - 1) \times (i - 1)$, $k_w^i = k_w + (k_w - 1) \times (i - 1)$. 如图 3 所示, 这些具有不同大小的空洞卷积核

$\omega^1 \in \mathbb{R}^{3 \times 3}$, $\omega^2 \in \mathbb{R}^{5 \times 5}$, $\omega^3 \in \mathbb{R}^{7 \times 7}$, 用于和 $\hat{\mathbf{h}}$ 进行卷积运算, 产生多个尺度的交互特征.

3.2.2 特征交互与融合

在特征交互模块中, 我们设计了多尺度的特征交互结构. 对于 $\hat{\mathbf{h}}$ 与卷积核 ω_r^i 运算得到的特征输出 \mathbf{C}_i , 其第 mn 项可以由下列公式计算:

$$\mathbf{C}_i(m, n) = \hat{\mathbf{h}} * \omega_r^i = \sum_{a=1}^{k_h} \sum_{b=1}^{k_w} \hat{\mathbf{h}}(m+i \cdot a, n+i \cdot b) \times \omega_r^i(a, b), \quad (3)$$

其中, \mathbf{C}_i 代表与空洞率为 i 的卷积核 ω_r^i 交互得到的特征输出, $*$ 代表卷积操作.

每一个分支得到的特征图, 代表了不同尺度下实体与关系的交互特征, 为了聚合这些特征, 我们使用拼接操作将这些特征图堆叠在一起, 通过一个全连接层对其进行特征融合, 得到包含多尺度特征信息的特征图, 最后通过整平操作将特征图转变为特征向量.

对应图 3, 通过特征融合得到的特征向量 \mathbf{o} 的公式表示如下:

$$\mathbf{o} = \text{vec}(\text{concat}(\mathbf{C}_1; \mathbf{C}_2; \mathbf{C}_3)\mathbf{W}_F), \quad (4)$$

其中, vec 表示整平操作, concat 表示拼接操作, \mathbf{W}_F 代表特征融合矩阵.

3.2.3 评分函数

为了计算 (h, r, t) 的得分, 我们将 \mathbf{o} 利用全连接层投影到 \mathbb{R}^{d_e} , 再计算该向量与尾向量的内积相似度来衡量 (h, r, t) 这个元组的合理性. 根据式 (3) 和 (4), MDCE 的评分函数 $\psi(h, r, t)$ 可以表示为如下形式:

$$\psi(h, r, t) = f(\mathbf{o}\mathbf{W}_P)^T \mathbf{t} = f(\text{vec}(\text{concat}(\hat{\mathbf{h}} * \omega_r^1; \hat{\mathbf{h}} * \omega_r^2; \hat{\mathbf{h}} * \omega_r^3)\mathbf{W}_F)\mathbf{W}_P)^T \mathbf{t}, \quad (5)$$

其中, f 代表非线性激活函数, \mathbf{W}_P 表示特征投影矩阵.

3.2.4 模型训练

本小节主要针对模型训练的相关细节进行阐述, 包括损失函数、更新模型参数所使用的优化方法等.

在 MDCE 的训练中, 我们希望模型对于正确元组的输出为 1, 对于错误元组的输出为 0, 那么模型的输出就符合多变量的伯努利 (Bernoulli) 分布, 在给定知识图谱 \mathcal{G} 的情况下, 关于参数 Θ 的似然函数可以写成如下形式:

$$p(\mathcal{G}|\Theta) = \prod_{(h,r,t) \in \{\mathcal{S} \cup \mathcal{S}'\}} p^y (1-p)^{1-y}, \quad (6)$$

其中,

$$y = \begin{cases} 1, & \text{if } (h, r, t) \in \mathcal{S}, \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

\mathcal{S} 是正样本集合, \mathcal{S}' 是通过替换正样本的尾实体得到的负样本集合, p 为模型判断 (h, r, t) 为正确元组的概率, 计算公式为 $p = \text{sigmoid}(\psi(h, r, t))$.

因此, MDCE 的损失函数可以定义为

$$\mathcal{L} = -\log p(\mathcal{G}|\Theta) = - \sum_{(h,r,t) \in \{\mathcal{S} \cup \mathcal{S}'\}} \left(y \log p + (1-y) \log(1-p) \right). \quad (8)$$

表 3 实验数据集基本信息

Table 3 Detailed information of the datasets used in KG embedding experiments

Dataset	Entity	Relation	Training set	Validation set	Test set
UMLS	135	46	5216	652	661
KINSHIP	104	25	8544	1068	1074
WN18RR	40943	11	86835	3034	3134
FB15k-237	14541	237	272115	17535	20466

在参数优化的过程中, 我们使用 Adam 优化器^[37] 最小化损失函数, 与其他优化器相比, Adam 通过计算梯度的一阶矩估计和二阶矩估计, 可以为不同的参数设计独立的自适应性学习率, 相比于 SGD 优化器^[38] 可以更快地使损失函数收敛, 减少模型训练时间. 同时, 我们使用批规范化技术 (batch normalization)^[39] 对每层的输出进行处理, 保证模型能够稳定、快速地收敛. 我们也使用了 dropout^[40] 与标签平滑技术^[32], 这些技术可以有效防止模型过拟合, 提高模型的泛化能力.

4 实验与分析

4.1 实验设置

4.1.1 实验环境

本文进行的所有实验均在同一台设备上, 其 CPU 型号为 Intel(R) Core(TM) i9-9900K 3.60 Hz, 配有 24 G 显存的 NVIDIA TITAN RTX 显卡, 操作系统为 Ubuntu 20.04.1 LTS. 本文所有的实验程序都是基于深度学习框架 Pytorch^[41] 实现的.

4.1.2 数据集

为了证明所提出方法的有效性, 本文在几个被广泛应用的基准数据集上进行了实验, 表 3 给出了这几组数据集的详细统计指标.

UMLS^[42]: UMLS 是由一组描述生物医学概念 (例如疾病及相关联的治疗方法与症状) 的 RDF 三元组构成的数据集.

KINSHIP^[43]: KINSHIP 数据集描绘的是澳大利亚中部地区的阿尔亚瓦拉部落内部的亲属关系.

FB15k-237^[44]: FB15k-237 是 FB15k 数据集的一个子集, 由于 FB15k 的训练集和测试集中存在反转关系的元组, 这就产生了测试数据泄露的问题. 为了避免这种情况, FB15k-237 中将反转关系全部移除, 使链接预测的难度提高.

WN18RR^[21]: WN18RR 是 WN18 数据集的一个子集, 由于 WN18 的训练集和测试集中存在反转关系的元组, 这就产生了测试数据泄露的问题. 为了避免这种情况, WN18RR 中将反转关系全部移除, 使链接预测的难度提高.

4.1.3 对比方法

本文选择了目前比较具有代表性的方法与近期提出的知识表示学习方法进行对比, 这些方法大致可以分为两类:

- **浅层方法:** 包括平移距离模型和双线性模型在内的, 利用简单向量与矩阵操作计算评分函数的方法, 如 TransE^[13], RotatE^[17], DistMult^[19], ComplEx^[20].

表 4 本文提出方法的最优参数设置

Table 4 Optimal configurations of MDCE on the four datasets

Dataset	Number of convolution kernel per branch	Input dropout	Feature dropout	Hidden dropout
UMLS	8	0.3	0.6	0.4
KINSHIP	8	0.5	0.6	0.3
WN18RR	32	0.3	0.3	0.4
FB15k-237	32	0.3	0.3	0.5

• **深层方法:** 使用多层与非线性结构进行知识图谱嵌入表示的方法, 其中包括 ConvE [21], ConvKB [45], ConvR [22], HypER [25], CapsE [31], InteractE [26], ParamE-CNN [27], InceptionE [23].

4.1.4 评价指标

在链接预测任务的测评过程中, 针对测试集 \mathcal{T} 中每一个测试元组 x_i , 我们首先利用数据集中剩下的实体依次替换它的头实体, 生成 $\tilde{x}_i^h \notin \mathcal{T}$, 或依次替换它的尾实体, 生成 $\tilde{x}_i^t \notin \mathcal{T}$. 然后观察模型对于 x_i 是否相较于 \tilde{x}_i^h 或 \tilde{x}_i^t 有更高的得分. 具体来说, 在替换头实体后的全部测试样本中, 正确样本的评分排名 rank_i^h , 和在替换尾实体后的全部测试样本中, 正确样本的评分排名 rank_i^t , 由如下形式定义:

$$\begin{aligned}\text{rank}_i^h &= 1 + \sum_{\tilde{x}_i^h \notin \mathcal{T}} \mathbb{I}[\psi(x_i) < \psi(\tilde{x}_i^h)], \\ \text{rank}_i^t &= 1 + \sum_{\tilde{x}_i^t \notin \mathcal{T}} \mathbb{I}[\psi(x_i) < \psi(\tilde{x}_i^t)],\end{aligned}\quad (9)$$

其中 $\mathbb{I}[\cdot]$ 代表指示函数.

接下来, 我们介绍两个评测链接预测任务的指标: 平均倒数排名 (mean reciprocal rank, MRR) 和 k 命中率 (Hits@ k). 两个指标的定义如下:

$$\begin{aligned}\text{MRR:} & \frac{1}{2|\mathcal{T}|} \sum_{x_i \in \mathcal{T}} \frac{1}{\text{rank}_i^h} + \frac{1}{\text{rank}_i^t}, \\ \text{Hits@}k: & \frac{1}{2|\mathcal{T}|} \sum_{x_i \in \mathcal{T}} [\mathbb{I}[\text{rank}_i^h \leq k] + \mathbb{I}[\text{rank}_i^t \leq k]].\end{aligned}\quad (10)$$

MRR 是在全部测试样本中, 预测正确样本的得分排名倒数的平均值. Hits@ k 是在全部测试样本中, 预测正确样本的得分排名高于 k 名或等于第 k 名的比例, 一般取 k 为 1, 3, 10. MRR 与 Hits@ k 越高, 代表链接预测的效果越好.

4.1.5 参数设置

本文所提方法的参数设置, 是在验证集上通过网格搜索方法得到的. 网格搜索的超参数范围如下: batch size 的范围为 {128, 256}, 输入层、特征层和隐藏层的 dropout 的范围为 {0.1, 0.2, 0.3, 0.4, 0.5}.

下列参数设置在所有数据集中都保持一致: batch size 设置为 128, 实体嵌入维度为 200, 多尺度分支数目设为 3, 第 i 个分支的空洞率设置为 i , 标准卷积核的尺寸为 3×3 , 标签平滑率设置为 0.1, 学习率为 0.001. 其余超参数设置如表 4 所示.

4.2 链接预测结果

在不同数据集上的链接预测实验结果如表 5 和 6 所示. 其中最优结果用粗体标识, 次优结果用下

表 5 在 UMLS 和 KINSHIP 数据集上的链接预测实验结果
 Table 5 Results of the link prediction by MRR and Hits@ k on UMLS and KINSHIP datasets

Method	UMLS				KINSHIP			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
DistMult	0.927	89.4	95.2	99.1	0.673	54.8	74.5	93.1
ComplEx	<u>0.947</u>	92.1	96.8	98.9	0.865	79.1	93.1	97.8
ConvE	0.942	91.1	96.5	98.6	0.833	73.8	91.7	98.1
ConvKB	0.755	65.2	83.9	93.3	0.614	43.6	75.5	95.3
ConvR	0.921	85.6	98.4	99.5	0.869	79.6	93.7	98.2
HypER	0.926	87.1	97.7	99.2	<u>0.879</u>	<u>81.0</u>	<u>94.2</u>	<u>98.6</u>
CapsE	0.841	73.2	94.1	98.6	0.719	71.9	81.5	95.8
InteractE	0.944	90.4	98.0	99.5	0.867	79.2	93.2	<u>98.6</u>
ParamE-CNN	0.915	85.0	97.8	<u>99.4</u>	0.871	79.7	93.6	98.4
InceptionE	0.940	89.9	<u>98.2</u>	<u>99.4</u>	0.851	77.0	92.3	98.2
MDCE	0.949	<u>91.2</u>	98.4	99.5	0.887	82.1	95.0	98.8

表 6 在 FB15k-237 和 WN18RR 数据集上的链接预测实验结果
 Table 6 Results of the link prediction by MRR and Hits@ k on FB15k-237 and WN18RR datasets

Method	FB15k-237				WN18RR			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
TransE	0.257	17.4	28.4	42.0	0.182	2.7	29.5	44.4
RotatE	0.336	24.0	37.2	53.0	0.471	42.0	<u>48.5</u>	<u>55.4</u>
DistMult	0.241	15.5	26.3	41.9	0.430	39.0	44.0	49.0
ComplEx	0.247	15.8	27.5	42.8	0.440	41.0	46.0	49.0
ConvE	0.325	23.7	35.6	50.1	0.430	40.0	44.0	52.0
ConvKB	0.243	15.5	37.1	42.1	0.249	5.6	41.7	52.4
ConvR	0.350	26.1	38.5	52.8	0.467	43.7	48.0	52.4
HypER	0.336	24.8	36.7	51.4	<u>0.468</u>	<u>43.8</u>	48.2	52.6
CapsE	0.160	7.3	16.8	35.6	0.415	33.7	42.3	55.9
InteractE	<u>0.354</u>	<u>26.3</u>	<u>38.8</u>	<u>53.6</u>	0.466	43.5	47.9	52.5
ParamE-CNN	0.347	25.7	37.9	52.9	0.467	43.4	47.8	52.7
InceptionE	0.334	24.6	37.8	51.8	0.451	42.2	46.1	56.3
MDCE	0.357	26.5	39.5	54.3	0.473	44.3	48.6	53.6

画线标识. 为了确保公平对比, 这些结果均是通过对应工作提供的实验代码运行得到的¹⁾, 并采用了完全一致的数据预处理与测评方法^[46].

从表 5 和 6 中可以看到, MDCE 在多个数据集上都取得了最优的结果. 相比于双线性模型, 深度模型在处理像 FB15k-237 这样大规模、复杂的知识图谱上具有明显优势. 相比于 ConvR 模型, MDCE 在 UMLS 和 KINSHIP 数据集上, MRR 指标分别提升了 3% 与 2%; 在 FB15k-237 数据集上, MRR 指标提升了 2%, 这说明了我们的模型的有效性, 通过使用多个尺度的卷积核, 同时提取实体与关系

1) 实验代码可以在 <https://github.com/thunlp/KRLPapers> 中找到.

表 7 在 FB15k-237, WN18RR, KINSHIP 与 UMLS 数据集上的关系预测实验结果
 Table 7 Results of the relation prediction by precision (%) on FB15k-237, WN18RR, KINSHIP and UMLS datasets

Method	Dataset			
	FB15k-237	WN18RR	KINSHIP	UMLS
ConvE	35.9	42.0	12.7	10.0
ConvR	<u>59.0</u>	34.3	17.1	<u>30.0</u>
HypER	56.8	32.8	15.4	29.5
InteractE	56.4	30.8	17.8	16.3
InceptionE	55.8	<u>46.8</u>	20.3	11.3
MDCE	59.6	47.6	<u>19.8</u>	31.2

的多尺度交互特征, 能够更好地提升知识表示的效果。

4.3 关系预测结果

表 7 展示了在不同数据集上, 模型对实体间关系的预测准确率 (precision (%)). 本文选择 Micro 模式计算评价指标, 以减弱测试集数据分布不平衡带来的影响. 由于在 Micro 模式下, 整体的预测准确率 (precision), 召回率 (recall) 与 F1 值的计算结果相同, 故本文只展示了准确率指标. 其中最优结果用粗体标识, 次优结果用下划线标识.

MDCE 在多个数据集上的关系预测任务中, 性能表现都优于其他深度模型, 此外, ConvR 与 InceptionE 模型在这项任务上也取得了较优的成绩, 说明特定关系的卷积核与多尺度信息可以帮助模型更好地学习实体与关系嵌入, 提升关系预测的准确率.

4.4 多尺度信息的重要性

传统的嵌入特征交互方式存在许多问题. 基于加减操作的模型^[13]形式过于简单, 特征之间无法进行复杂交互. 基于张量积操作的模型^[18,30]虽然可以捕捉复杂的交互特征, 但张量积的成对交互形式使得交互特征表示空间过大, 模型极易产生过拟合. 而利用卷积操作的特征交互方式可以解释为成对特征交互的局部压缩, 卷积将定长的成对交互特征进行求和表示, 保留复杂交互特征的同时, 压缩了参数规模, 取得了优异性能^[21,22], 并且易于高效计算. 但本文通过实验发现, 单一尺度的卷积核在交互建模方面仍然存在局限性.

图 4(a) 展示了 /m/06sy4c 实体对应的嵌入表示, 图 4(b) 展示了 FB15k-237 数据集中的 (/m/06sy4c, /people/person/gender, /m/05zppz) 元组输入到多尺度模型时, 不同尺度卷积核生成的特征图可视化结果. 可以看到, 不同尺度的卷积核生成的特征图存在着明显差异, 说明多尺度卷积核相比于单尺度卷积核捕捉复杂的交互特征的能力更强.

图 5 展示了在 FB15k-237 数据集上, 对于不同关系元组, 使用不同尺度特征信息进行链接预测的结果. 图中每一行是从测试集中随机挑选出的关系, 每列从左至右代表不同尺度的卷积核. 色块颜色深浅代表对应关系元组在对应尺度的卷积核上进行链接预测的 Hits@10 结果. 可以看到, 不同关系元组在使用不同的单一尺度卷积核进行链接预测时, 性能存在明显差异, 且不存在唯一的尺度可以使不同关系上链接预测的性能都取得最优.

由此说明, 对于不同关系的链接预测任务, 其所需要的特征尺度是不同的. 该实验验证了多尺度信息对于知识图谱嵌入表示的重要性.

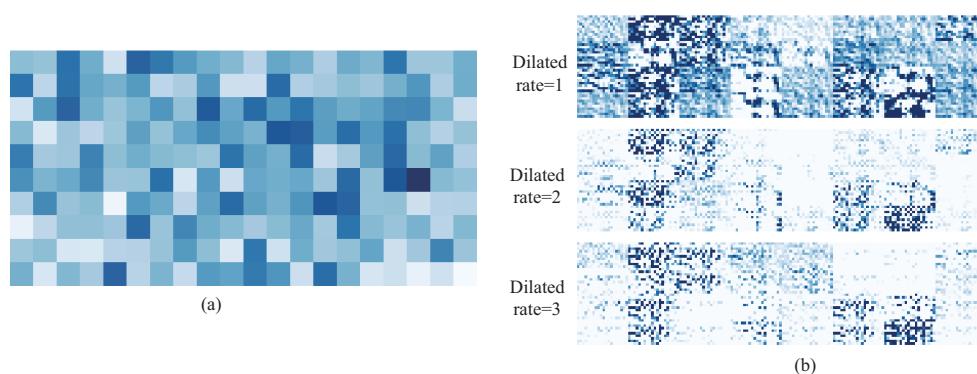


图 4 (网络版彩图) (a) /m/06sy4c 实体对应的嵌入可视化表示; (b) 图 (a) 对应嵌入通过不同空洞率卷积核产生的特征图

Figure 4 (Color online) (a) The visual embedding of entity named /m/06sy4c; (b) Feature maps generated by filters of different dilation rates

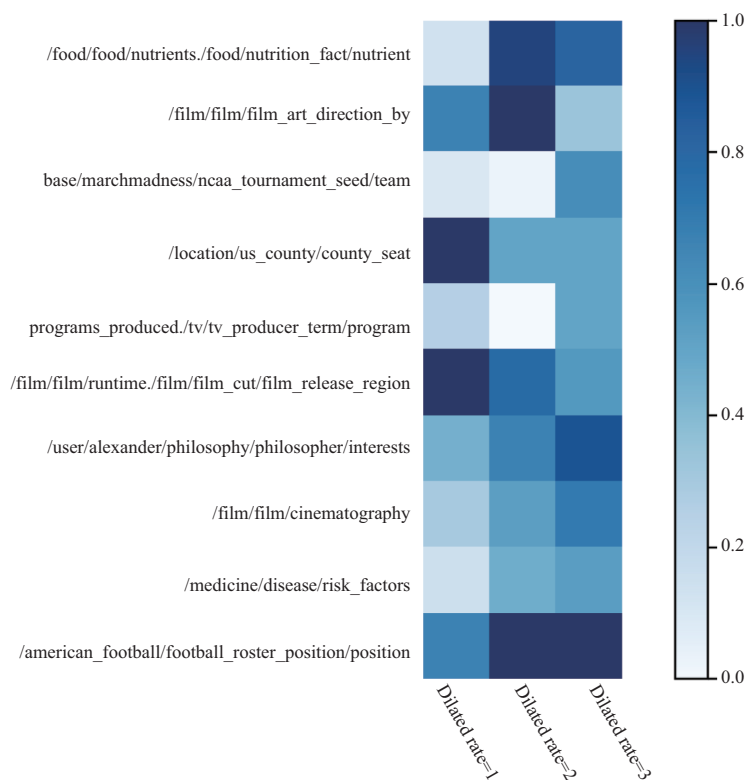


图 5 (网络版彩图) 在 FB15k-237 数据集上对于不同关系元组, 使用不同尺度特征信息进行链接预测的实验结果
 Figure 5 (Color online) Results of the link prediction about different relation-aware triples by using different scale features on FB15k-237 datasets

4.5 收敛性分析

我们选取了一些性能优异的深度模型, 包括 HypER, ConvR, InteractE, InceptionE 和 ParamE-CNN 方法, 绘制了它们与 MDCE 方法在不同数据集上的学习曲线. 如图 6 所示. 可以观察到, MDCE

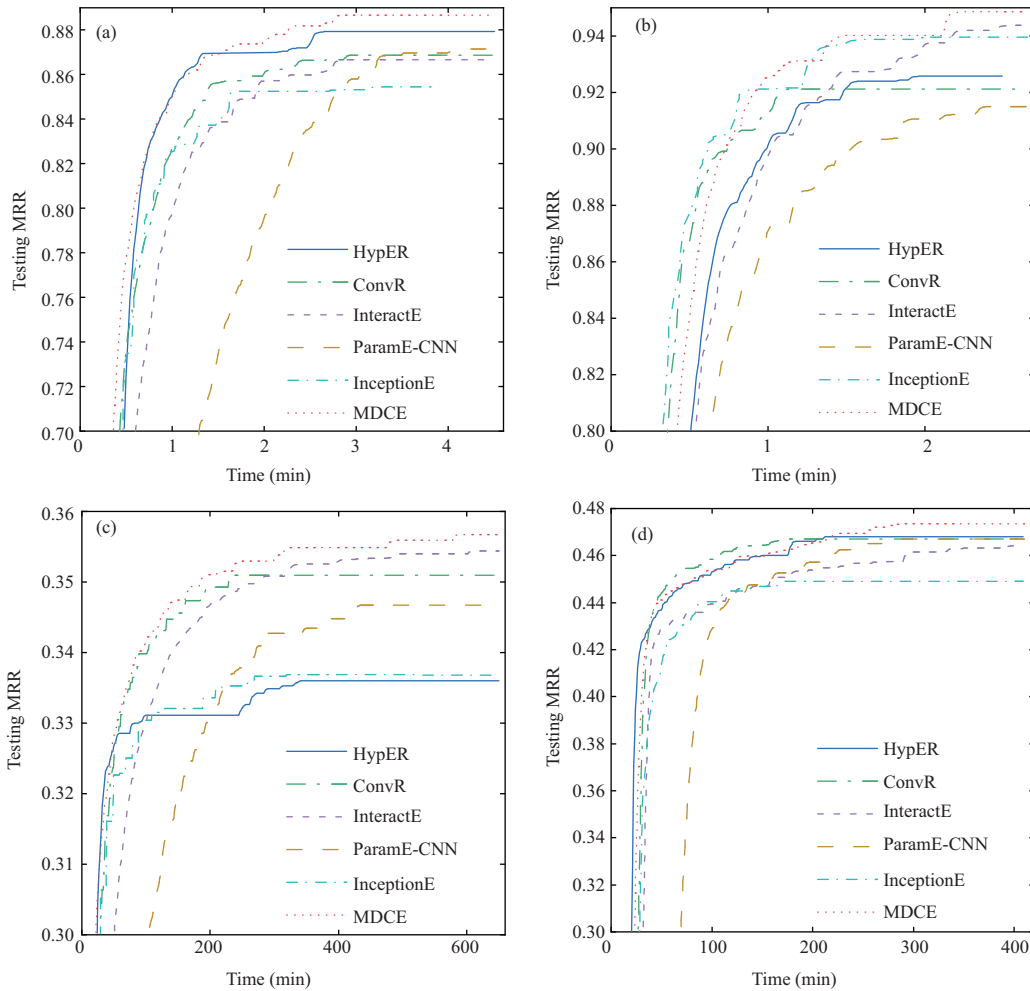


图 6 (网络版彩图) MDCE 与一些基于神经网络的最先进模型的算法收敛时间对比

Figure 6 (Color online) Comparison on clock time of model training vs. testing MRR between MDCE and some state-of-the-art neural network models. (a) KINSHIP; (b) UMLS; (c) FB15k-237; (d) WN18RR

不仅在性能上优于现在的方法, 而且收敛速度相对更快. 在模型训练的前期, MDCE 与 ConvR 的收敛速度相近, 到了训练后期, MDCE 的收敛速度高于 ConvR, 这是由于 MDCE 在模型训练过程中逐步提取到多尺度交互特征, 这些特征相对于单一尺度特征可以帮助模型更快地收敛.

4.6 参数敏感度分析

4.6.1 实体嵌入维度

为了分析提出的模型对实体嵌入维度参数的敏感性, 我们将 HypER, ConvR, InteractE, ParamE-CNN 和 MDCE 模型的超参数都调整到对应数据集的最优参数, 将实体嵌入维度设置为 $\{50, 75, 100, 150, 200, 250, 400, 500\}$, 绘制了在不同实体嵌入维度下, 上述模型在 FB15k-237 与 KINSHIP 数据集上的 MRR 值. 结果如图 7 所示. 可以看到, 相较于其他模型, MDCE 模型对实体嵌入维度参数的敏感程度较低. 这是由于多尺度特征交互中尺度较大的卷积核可以帮助捕获范围更大的交互信息, 小尺度卷积核负责捕捉小范围的交互特征. 这样对于不同嵌入维度, MDCE 都可以取得很好的性能.

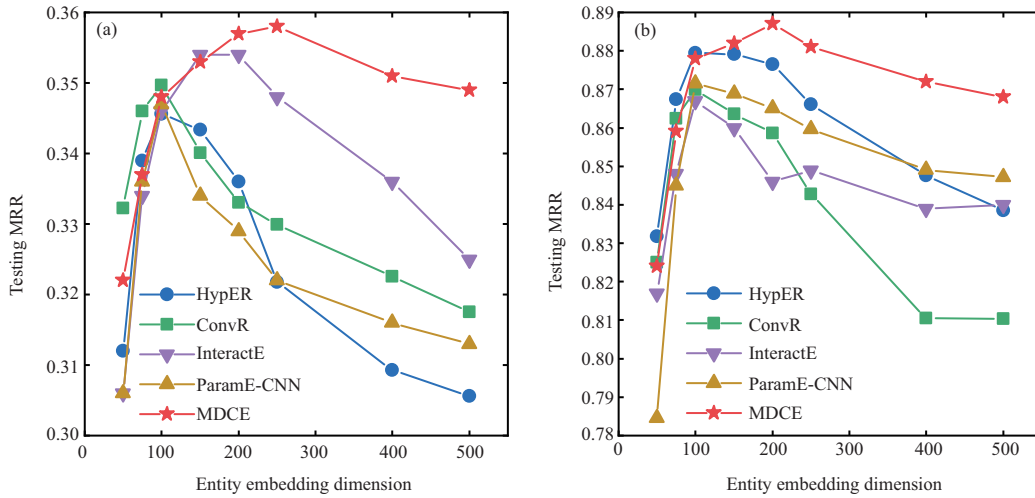


图 7 (网络版彩图) 不同方法在不同实体嵌入维度下的算法性能对比

Figure 7 (Color online) Comparison on performance of MRR with different entity embedding dimensions on four datasets. (a) FB15k-237; (b) KINSHIP

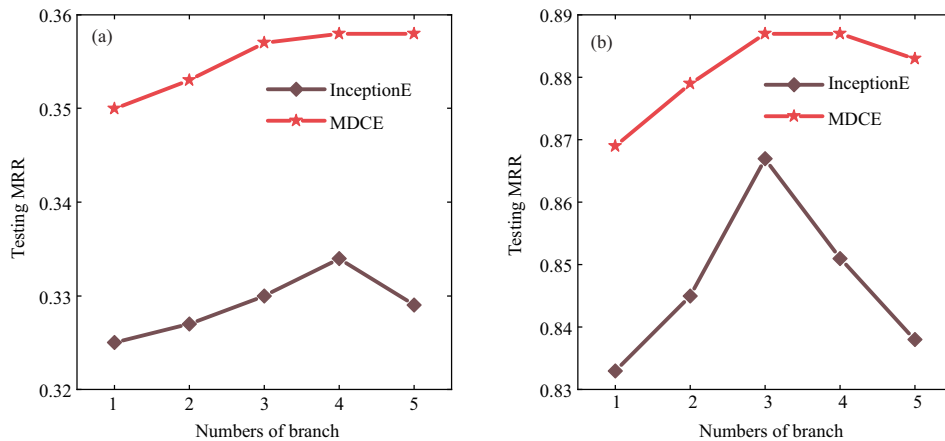


图 8 (网络版彩图) MDCE 与 InceptionE 在不同多尺度分支数目下的算法性能对比

Figure 8 (Color online) Comparison on performance of MRR with different numbers of branch on four datasets. (a) FB15k-237; (b) KINSHIP

4.6.2 多尺度分支数目

为了分析提出的多尺度模型对多尺度分支数目的敏感性, 我们将 InceptionE 与 MDCE 方法中的每个分支均设置为 32 个卷积核, 在 MDCE 中, 第 n 个分支的卷积核空洞率设置为 n , 尺寸为 $(2n + 1) \times (2n + 1)$. 而在 InceptionE 中, 第 i 个分支的卷积核尺寸也设置为 $(2n + 1) \times (2n + 1)$, 以保证两个模型在每个分支上具有相同尺寸的感受野.

图 8 展示了 MDCE 与 InceptionE 方法在 FB15k-237 与 KINSHIP 数据集上, 设置不同多尺度分支数目对性能的影响. 观察结果可以看到, 相较于 InceptionE, MDCE 对多尺度分支数目的敏感度更低, 多尺度分支数目的变化未对性能产生较大影响. 这是由于在 MDCE 中, 空洞卷积核可以在不引入过多计算量与参数的情况下, 获得多尺度信息, 提升模型的表达能力. 而 InceptionE 在不断增加多尺

度分支的情况下, 参数量上升易导致过拟合问题.

5 结论

本文提出了一种基于多尺度空洞卷积的知识表示方法——MDCE, 该方法使用多尺度空洞卷积, 保留了实体与关系的多尺度交互特征, 同时空洞卷积的引入可以使得在不增加参数、不引入额外计算量的情况下, 扩大卷积核的感受野, 提升实体与关系的表示效果, 在一系列数据集上的链接预测实验证明了所提方法的有效性与优越性.

目前该工作的一个局限性在于, 提出的神经网络结构需要人工设计. 在未来的工作中, 我们将考虑使用神经网络架构搜索方法^[47, 48], 针对特定的数据集搜索出最优的卷积神经网络结构, 这将会是一个值得探索的方向.

参考文献

- 1 Nickel M, Murphy K, Tresp V, et al. A review of relational machine learning for knowledge graphs. *Proc IEEE*, 2016, 104: 11–33
- 2 Ji S X, Pan S R, Cambria E, et al. A survey on knowledge graphs: representation, acquisition and applications. 2020. ArXiv:2002.00388
- 3 Mottin D, Lissandrini M, Velegrakis Y, et al. Exemplar queries: a new way of searching. *VLDB J*, 2016, 25: 741–765
- 4 An B, Han X P, Sun L. Knowledge-representation-enhanced question-answering system. *Sci Sin Inform*, 2018, 48: 1521–1532 [安波, 韩先培, 孙乐. 融合知识表示的知识库问答系统. *中国科学: 信息科学*, 2018, 48: 1521–1532]
- 5 Qin C, Zhu H S, Zhuang F Z, et al. A survey on knowledge graph-based recommender systems. *Sci Sin Inform*, 2020, 50: 937–956 [秦川, 祝恒书, 庄福振, 等. 基于知识图谱的推荐系统研究综述. *中国科学: 信息科学*, 2020, 50: 937–956]
- 6 Miller G A. WordNet: a lexical database for English. *Commun ACM*, 1995, 38: 39–41
- 7 Bollacker K, Evans C, Paritosh P, et al. Freebase: a collaboratively created graph database for structuring human knowledge. In: *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, 2008. 1247–1250
- 8 Suchanek F M, Kasneci G, Weikum G. Yago: a core of semantic knowledge. In: *Proceedings of the 16th International Conference on World Wide Web*, 2007. 697–706
- 9 Wang S, Du Z J, Meng X F. Research progress of large-scale knowledge graph completion technology. *Sci Sin Inform*, 2020, 50: 551–575 [王硕, 杜志娟, 孟小峰. 大规模知识图谱补全技术的研究进展. *中国科学: 信息科学*, 2020, 50: 551–575]
- 10 Zhang Z P, Xia C Y, Chen S Y, et al. Reachability analysis of networked finite state machine with communication losses: a switched perspective. *IEEE J Sel Areas Commun*, 2020, 38: 845–853
- 11 Wang Q, Mao Z D, Wang B, et al. Knowledge graph embedding: a survey of approaches and applications. *IEEE Trans Knowl Data Eng*, 2017, 29: 2724–2743
- 12 Rossi A, Barbosa D, Firmani D, et al. Knowledge graph embedding for link prediction. *ACM Trans Knowl Discov Data*, 2021, 15: 1–49
- 13 Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2013. 2787–2795
- 14 Wang Z, Zhang J W, Feng J L, et al. Knowledge graph embedding by translating on hyperplanes. In: *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014. 1112–1119
- 15 Lin Y K, Liu Z Y, Sun M S, et al. Learning entity and relation embeddings for knowledge graph completion. In: *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, 2015. 2181–2187
- 16 Ji G L, He S Z, Xu L H, et al. Knowledge graph embedding via dynamic mapping matrix. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 2015. 687–696

- 17 Sun Z Q, Deng Z H, Nie J Y, et al. Rotate: knowledge graph embedding by relational rotation in complex space. In: Proceedings of the 7th International Conference on Learning Representations, 2019
- 18 Nickel M, Tresp V, Kriegel H P. A three-way model for collective learning on multi-relational data. In: Proceedings of the 28th International Conference on International Conference on Machine Learning, 2011. 809–816
- 19 Yang B S, Yih W, He X D, et al. Embedding entities and relations for learning and inference in knowledge bases. In: Proceedings of the 3rd International Conference on Learning Representations, 2015
- 20 Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning, 2016. 2071–2080
- 21 Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2D knowledge graph embeddings. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence, 2018. 1811–1818
- 22 Jiang X T, Wang Q, Wang B. Adaptive convolution for multi-relational learning. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019. 978–987
- 23 Xie Z W, Zhou G Y, Liu J, et al. ReInceptionE: relation-aware inception network with joint local-global structural information for knowledge graph embedding. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020. 5929–5939
- 24 Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015. 1–9
- 25 Balazević I, Allen C, Hospedales T M. Hypernetwork knowledge graph embeddings. In: Proceedings of the 28th International Conference on Artificial Neural Networks, 2019. 553–565
- 26 Vashishth S, Sanyal S, Nitin V, et al. InteractE: improving convolution-based knowledge graph embeddings by increasing feature interactions. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020. 3009–3016
- 27 Che F H, Zhang D W, Tao J H, et al. Parame: regarding neural network parameters as relation embeddings for knowledge graph completion. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020. 2774–2781
- 28 Mikolov T, Sutskever I, Chen K, et al. Distributed representations of words and phrases and their compositionality. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, 2013. 3111–3119
- 29 Dong X, Gabrilovich E, Heitz G, et al. Knowledge vault: a web-scale approach to probabilistic knowledge fusion. In: Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2014. 601–610
- 30 Socher R, Chen D Q, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, 2013. 926–934
- 31 Nguyen D Q, Vu T, Nguyen T D, et al. A capsule network-based embedding model for knowledge graph completion and search personalization. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019. 2180–2189
- 32 Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 2818–2826
- 33 Wang P Q, Chen P F, Yuan Y, et al. Understanding convolution for semantic segmentation. In: Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision, 2018. 1451–1460
- 34 Shensa M J. The discrete wavelet transform: wedding the a trous and Mallat algorithms. *IEEE Trans Signal Process*, 1992, 40: 2464–2482
- 35 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. In: Proceedings of the 4th International Conference on Learning Representations, 2016
- 36 Oord A, Dieleman S, Zen H, et al. Wavenet: a generative model for raw audio. 2016. ArXiv:1609.03499
- 37 Kingma D P, Ba J. Adam: a method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations, 2015
- 38 Sutskever I, Martens J, Dahl G, et al. On the importance of initialization and momentum in deep learning. In: Proceedings of the 30th International Conference on Machine Learning, 2013. 1139–1147
- 39 Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015. ArXiv:1502.03167
- 40 Srivastava N, Hinton G, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*, 2014, 15: 1929–1958

- 41 Paszke A, Gross S, Massa F, et al. PyTorch: an imperative style, high-performance deep learning library. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems, 2019. 8026–8037
- 42 Kok S, Domingos P. Statistical predicate invention. In: Proceedings of the 24th International Conference on Machine Learning, 2007. 433–440
- 43 Lin X V, Socher R, Xiong C M. Multi-hop knowledge graph reasoning with reward shaping. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018. 3243–3253
- 44 Toutanova K, Chen D Q, Pantel P, et al. Representing text for joint embedding of text and knowledge bases. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015. 1499–1509
- 45 Nguyen D Q, Nguyen T D, Nguyen D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018. 327–333
- 46 Sun Z Q, Vashishth S, Sanyal S, et al. A re-evaluation of knowledge graph completion methods. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020. 5516–5522
- 47 Liu H X, Simonyan K, Yang Y M. DARTS: differentiable architecture search. In: Proceedings of the 7th International Conference on Learning Representations, 2019
- 48 Yao Q M, Xu J, Tu W W, et al. Efficient neural architecture search via proximal iterations. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020. 6664–6671

Multi-scale dilated convolutional network for knowledge graph embedding

Haotong DU^{1,2}, Zhen WANG^{1,2*}, Hongyi NIE^{1,2}, Quanming YAO^{3*} & Xuelong LI²

1. *School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, China;*

2. *School of Artificial Intelligence, OPTics and ElectroNics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China;*

3. *Department of Electronic Engineering, Tsinghua University, Beijing 100084, China*

* Corresponding author. E-mail: w-zhen@nwpu.edu.cn, qyaoaa@connect.ust.hk

Abstract It aims to learn how to represent the low-dimensional vectors of entities and relations using observed triplets. KGE can benefit a variety of downstream tasks, such as KG completion and triplet classification. Deep models achieve state-of-the-art performance by leveraging the powerful nonlinear fitting ability of neural networks. However, most existing methods ignore multi-scale interaction features between entities and relations except InceptionE, which is hard to train because of high computation costs. In this paper, we propose a new KGE model called MDCE, that uses multi-scale dilated convolution to capture rich interaction features at different scales. Meanwhile, MDCE has lower computation costs than InceptionE. We perform extensive experiments on multiple benchmark datasets; results on the link prediction task show that the proposed model MDCE not only significantly outperforms existing state-of-the-art models but is also efficient and robust.

Keywords knowledge graph, knowledge graph embedding, deep model, multi-scale feature, link prediction, artificial intelligence