



计算机科学

COMPUTER SCIENCE

一种文本-图像增强的多模态知识图谱嵌入方法

肖桂阳, 王立松, 江国华

引用本文

肖桂阳, 王立松, 江国华. 一种文本-图像增强的多模态知识图谱嵌入方法[J]. 计算机科学, 2023, 50(8): 163-169.

XIAO Guiyang, WANG Lisong, JIANG Guohua. [Multimodal Knowledge Graph Embedding with Text-Image Enhancement](#) [J]. Computer Science, 2023, 50(8): 163-169.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多核CPU的无锁并行Semi-naive算法](#)

Lock-free Parallel Semi-naive Algorithm Based on Multi-core CPU

计算机科学, 2023, 50(6): 29-35. <https://doi.org/10.11896/jsjcx.220800050>

[知识图谱嵌入模型中的损失函数研究综述](#)

Comprehensive Survey of Loss Functions in Knowledge Graph Embedding Models

计算机科学, 2023, 50(4): 149-158. <https://doi.org/10.11896/jsjcx.211200175>

[一种基于深度强化学习的无人小车双层路径规划方法](#)

Bi-level Path Planning Method for Unmanned Vehicle Based on Deep Reinforcement Learning

计算机科学, 2023, 50(1): 194-204. <https://doi.org/10.11896/jsjcx.220500241>

[结合多目标优化算法的模糊聚类有效性指标及应用](#)

Fuzzy Clustering Validity Index Combined with Multi-objective Optimization Algorithm and Its Application

计算机科学, 2021, 48(10): 197-203. <https://doi.org/10.11896/jsjcx.200900061>

[基于残差的端对端图像超分辨率](#)

End-to-end Image Super Resolution Based on Residuals

计算机科学, 2019, 46(6): 246-255. <https://doi.org/10.11896/j.issn.1002-137X.2019.06.037>

一种文本-图像增强的多模态知识图谱嵌入方法

肖桂阳 王立松 江国华

南京航空航天大学计算机科学与技术学院 南京 210000

(xiaogy@nuaa.edu.cn)

摘要 大多传统的知识表示学习方法只关注三元组中的结构化信息,无法很好地利用实体图像、关系路径、文本描述等附加信息来学习知识表示或只融合一种附加信息。因此,提出同时融合实体描述和图像的多模态知识图谱嵌入方法,通过文本、图像相互增强,可以提供更加全面的外部信息来弥补单个信息源的不完整性给知识表示学习带来的不足。首先进行实体描述和图像建模,得到实体的文本表示和图像表示,并把它们作为 TransE 中结构表示的补充,最后通过 3 种实体表示的联合训练实现知识图谱和文本、图像的统一空间表示,提高实体和关系预测的准确性。实验结果表明,该模型实体预测的命中率比不融合附加信息的方法提高了 3.09%,比只融合实体描述的方法提高了 0.97%,比只融合实体图像的方法提高了 1.32%。

关键词: 知识表示学习; 实体描述; 实体图像; Text-CNN; 联合训练

中图法分类号 TP183

Multimodal Knowledge Graph Embedding with Text-Image Enhancement

XIAO Guiyang, WANG Lisong and JIANG Guohua

School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210000, China

Abstract Most traditional knowledge representation learning methods only focus on the structured information in triples, and cannot make good use of the additional information such as entity images, relation path and text descriptions to learn knowledge representation or fuse only one additional information. Therefore, a multimodal knowledge graph embedding method combining entity descriptions and images is proposed. Through mutual enhancement of text and images, more comprehensive external information can be provided to make up for the deficiency of knowledge representation learning caused by the incompleteness of a single information source. Firstly, text representation and image representation of entities are obtained by modeling entity descriptions and images. Then, they are used as a supplement to the structural representation in TransE. Finally, through the joint training of three entity representations, the unified spatial representation of knowledge graph, text and image is realized to improve the accuracy of entity and relation prediction. Experimental results show that the hit rate of entity prediction of this model improves by 3.09% compared with the method of without additional information, improves by 0.97% compared with the method of fusing only entity descriptions, and improves by 1.32% compared with the method of fusing only entity images.

Keywords Knowledge representation learning, Entity descriptions, Entity images, Text-CNN, Joint training

1 引言

知识图谱(Knowledge Graphs, KG)提供了有效的实体和关系的结构化信息,已被成功应用于多个领域,大量的 KGs,如 Freebase^[1], DBpedia^[2], YAGO^[3] 和 NELL, 已经被创建并实现了许多实际应用,例如知识推理^[4]、Web 搜索和问答^[5]。KG 是由实体(节点)和关系(不同类型的边)组成的多关系图,每条边都代表形如(头实体、关系、尾实体)即 (h, r, t) 的事实,表示两个实体通过特定的关系连接在一起。

虽然三元组在表示结构化数据方面很有效,但是其底层

符号特性通常使 KGs 很难操作。为了解决这个问题,知识图谱嵌入^[6]被提出,关键是将实体和关系嵌入到连续的低维向量空间中^[7],从而简化操作,同时保留 KG 的原有结构。例如 TransE^[8]为知识图谱实现了最先进的表示学习性能,尤其是对于那些大规模和稀疏的 KG。

但大多数传统的知识表示学习方法只关注三元组中的结构化信息,仅通过训练事实三元组来进行知识图谱嵌入。事实上,可以通过融合附加信息来改进知识图谱嵌入,例如实体图像、关系路径、文本描述以及逻辑规则。不同的附加信息通常包含同一对象不同方面的知识,这些附加信息之间的知识

到稿日期:2022-07-21 返修日期:2022-11-24

基金项目:基础加强计划重点项目(2019JCJQZD33800)

This work was supported by the Key Projects of Foundation Strengthening Plan(2019JCJQZD33800).

通信作者:王立松(wangls@nuaa.edu.cn)

是互补的,如果能充分挖掘并关联不同附加信息中的知识,则使得不同附加信息之间相互增强。现有的融合附加信息的知识图谱嵌入方法只实现了一种附加信息和三元组结构信息的融合,来提高知识图谱表示的准确性,而没有注意到不同附加信息之间相互补充可以提供更加全面的外部信息来帮助知识图谱的表示学习。

例如从知识图谱的附加信息实体文本描述和实体图像的角度来看,实体刘梅、刘星都有文本描述和图像两种附加信息,实体文本描述中包含关于实体的丰富语义信息,然而文本描述的语义特征只对三元组(刘梅,是…的母亲,刘星)的表示学习有帮助,对三元组(刘梅,长相相似,刘星)无效。实体图像提供的重要视觉信息给出了实体外观和行为的描述,但光从图像入手,获取到的视觉信息即刘梅和刘星的外貌特征仅能帮助学习三元组(刘梅,长相相似,刘星),对三元组(刘梅,是…的母亲,刘星)没有帮助。同时利用文本描述和实体图像中丰富的外部信息可以更好地帮助知识表示学习,因此本文提出了一种能够实现事实三元组和文本描述、实体图像融合的新的知识图谱嵌入方法,即文本-图像增强的多模态知识图谱嵌入方法(Multimodal Knowledge Graph Embedding with Text-Image Enhancement, MKGE)。

在该模型中,实体嵌入有3种,即基于结构的表示、基于文本的表示和基于图像的表示。首先提出文本编码器,它由卷积神经网络实现,用于对实体文本描述建模以获得描述的语义特征,并把它作为实体基于文本的表示;其次利用图像编码器对实体图像进行建模,具体地,它使用 AlexNet^[9]作为特征提取器,并通过投影矩阵把图像特征从图像空间投影到实体空间,得到实体基于图像的表示;然后使用经典的 RL (Representation Learning)方法基于平移的模型(如 TransE)对事实三元组进行建模,把关系看作是头实体到尾实体的平移操作,融合实体的文本表示和图像表示,得到实体基于结构的表示以及关系嵌入;最终联合学习知识表示。实验结果证明,该方法获取到的实体描述文本信息和实体图像视觉信息有助于学习到更加准确的知识图谱嵌入,即知识表示。

2 相关工作

2.1 融合事实信息的知识图谱嵌入

基于平移的方法在知识表示学习(Knowledge Representation Learning, KRL)方面既是有效的又是高效率的,因此近年来受到了广泛关注。Mikolov 等^[10]通过 word2vec 词表示学习模型发现词向量空间存在平移不变现象。受到该平移不变现象的启发,Border 等提出了 TransE 模型,将每个事实三元组(h, r, t)中的实体和关系编码到同一低维连续向量空间中,把关系 r 视为低维空间中从 h 到 t 的平移操作,即 $\mathbf{h} + \mathbf{r} = \mathbf{t}$,其中 \mathbf{h} 和 \mathbf{t} 是实体嵌入, \mathbf{r} 是关系嵌入。TransE 具有良好的性能与可扩展性,但是 TransE 不能对复杂的关系进行建模。

为了解决 TransE 模型在处理一对多、多对一、多对多复杂关系时的局限性,TransH^[11]模型提出将关系建模为超平面,并将头、尾实体投影到关系特定的超平面,让实体在不同

的关系下拥有不同的表示。TransR^[12]使用关系特定的投影矩阵将实体投影到对应的关系空间中,然后建立从头实体到尾实体的平移关系。TransM^[13]模型允许实体在涉及不同关系时具有不同的嵌入,并为每个事实三元组分配特定的关系权重。

然而,这些方法只关注三元组中的结构化信息,而没有考虑到丰富的附加信息可以帮助 KG 的表示学习。我们在 TransE 的基础上提出的文本-图像增强的知识图谱嵌入方法同时考虑了实体文本描述和实体图像,实现了结构信息、文本信息和图像信息的融合。

2.2 融合附加信息的知识图谱嵌入

与使用 KG 中观察到的事实来执行嵌入任务的研究方法(如基于平移的方法)不同,融合附加信息的方法可以通过合并附加信息来改进知识表示任务,例如实体类型、关联路径、实体文本描述、实体图像以及时序信息等。

实体类型即实体所属的语义类别,语义平滑嵌入^[14](SSE)模型要求相同类型的实体在嵌入空间中彼此邻近。关联路径即实体之间的多跳关系,PTTransE^[15]将 TransE 扩展到对关系路径建模。在大多数 KGs 中对实体都有简明的描述,DKRL^[16]将基于平移的嵌入方法扩展为文本增强模型,以便更好地融合实体描述。Jiang 等^[17]观察到 KGs 事实通常对时间是敏感的,基于此他们提出了时间感知嵌入模型,在时间敏感的关系对上加入时间顺序约束。

但是这些方法只融合了一种附加信息,而 MKGE 充分利用文本描述和实体图像互补的知识来提高知识表示的准确性。

3 系统框架

对于给定的 $KG = (E, R, T)$,其中 E 是实体集合, R 是关系集合, T 代表事实三元组集。每个三元组 $(h, r, t) \in T$ 表示在 $h \in E$ 和 $t \in E$ 之间存在一个关系 $r \in R$ 。对于实体、关系、文本描述和实体图像,我们使用 R^d 来表示其相应的低维向量。

为了同时利用事实三元组结构信息、实体描述文本信息和实体图像视觉信息,我们提出了3种实体表示类型:结构表示、文本表示和图像表示。对于给定的知识三元组 (h, r, t) , \mathbf{h}_s 和 \mathbf{h}_t 分别代表头实体和尾实体基于结构的表示,这种表示通常来自现有的基于平移的模型(如 TransE)。 \mathbf{h}_d 和 \mathbf{t}_d 分别代表头实体和尾实体基于实体描述构建的文本表示,下文会介绍构造这种表示的文本编码器。 \mathbf{h}_i 和 \mathbf{t}_i 分别代表头尾实体利用图像编码器训练对应的实体图像所得的图像表示。结构表示可以更好地捕获知识图谱中事实三元组的结构信息,文本表示和图像表示则可以相互补充,更加完整地表示出实体的语义信息。

如图1所示,MKGE 模型包含4个关键部分:1)文本编码器,对实体描述建模生成实体的文本表示;2)图像编码器,对实体图像建模生成实体的图像表示;3)实体/关系表示建模,通过合并文本嵌入和图像嵌入,表示出文本-图像增强的实体/关系嵌入;4)表示的联合训练,使用基于平移的优化

方法来联合训练实体/关系嵌入。

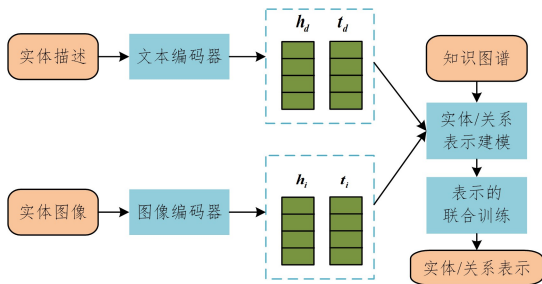


图1 多模态知识图谱嵌入的结构

Fig. 1 Structure of multimodal knowledge graph embedding

3.1 文本编码器

文本编码器借鉴 Kim^[18]提出的用于文本分析的 CNN 即 Text-CNN 来对实体的描述性文本进行编码,从而获得实体基于文本的表示。

文本编码器的整体结构如图 2 所示。该结构以随机初始化的词向量或 word2vec 预训练得到的单词嵌入作为输入,在经过卷积和最大池化后,最终输出实体的文本表示向量。

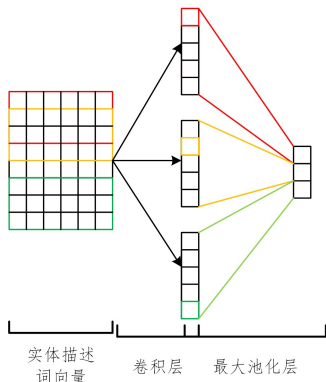


图2 文本编码器的结构

Fig. 2 Structure of text encoder

3.1.1 输入层

首先从 WordNet 语料库的注解文件中提取出实体的文本描述,然后去除实体文本描述中所有的停用词,最后使用预训练的 word2vec 得到所有词向量。

3.1.2 卷积层

将文本描述经过预处理后得到的词向量 $\{x_1, x_2, \dots, x_n\}$ 作为卷积层的输入,首先通过一个 h 词的滑动窗口实现 h 个词向量的连接,滑动处理的过程如下:

$$x_{i:i+h-1} = x_i \oplus x_{i+1} \oplus \dots \oplus x_{i+h-1} \quad (1)$$

其中, \oplus 是连接运算符, $x_i \in R^k$ 是文本中第 i 个单词的 k 维词向量。将卷积操作应用于句子中的每个滑动窗口,以生成新特征。

$$c_i = f(w \cdot x_{i:i+h-1} + b) \quad (2)$$

其中, c_i 表示从窗口 $x_{i:i+h-1}$ 中生成的特征, $w \in R^{hk}$ 表示卷积核, $b \in R$ 表示偏置项, f 是非线性函数如 tanh 函数。最后,所有滑动窗口 $\{x_{1:h}, x_{2:h+1}, \dots, x_{n-h+1:n}\}$ 生成的特征组成特征图。

$$c = [c_1, c_2, \dots, c_{n-h+1}] \quad (3)$$

其中, $c \in R^{n-h+1}$, c 中的这些特征具有相同的权重,这大大减少了参数的数量。上文描述了使用一个卷积核提取一个特征向量的过程,我们使用多个卷积过滤器来获取多个特征向量 (c_1, c_2, \dots, c_k) 。

3.1.3 池化层

CNN 中,卷积层后的池化操作可以有效地减少参数数量,加快计算速度并防止过拟合。为了捕捉每个特征向量中最重要的特征,我们在特征图上使用最大池化操作。

$$m_i = \max\{c_i\} \quad (4)$$

其中, m_i 是特征图 c_i 中最显著的特征值,也是最终对应于第 i 个卷积核的特征。最后得到对应于 k 个卷积核的特征向量 $m = \{m_1, m_2, \dots, m_k\}$ 作为实体的文本表示,其维度与句子长度无关。

3.2 图像编码器

图像编码器使用经典的卷积神经网络 AlexNet 为实体图像构造图像特征表示,以进一步获得实体基于图像表示。

图 3 给出了图像编码器的结构框架。首先使用 AlexNet 作为实体图像的特征提取器,再通过投影矩阵得到实体空间的图像特征,最后计算所有图像特征的平均值,并把它作为实体的图像表示。

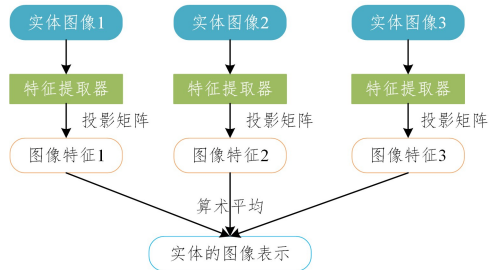


图3 图像编码器的结构

Fig. 3 Structure of image encoder

图像编码器的输入是每个实体的多个图像 $\{img_1, img_2, \dots, img_n\}$ 。首先,使用深度学习框架 PyTorch^[19]和预训练的 AlexNet 来提取实体图像的特征。AlexNet 是由 5 个卷积层、2 个全连接层和 1 个 softmax 层组成的深度 CNN。借鉴 Shutova 等^[20]的做法,我们将 softmax 之前的 4096 维嵌入即第二个全连接层(fc7)的输出作为图像特征表示。

在通过特征提取器获得每个图像的特征表示之后,使用可训练的共享投影矩阵将图像特征表示从图像空间投影到实体空间,得到实体空间的图像特征。第 i 个图像实体空间的图像特征为:

$$p_i = M \cdot f(img_i) \quad (5)$$

其中, f 表示特征提取函数即上述的卷积网络, $M \in R^{d_i \times d_i}$ 是可训练的投影矩阵, d_i 表示图像空间中图像特征的维度, d_i 表示实体的维度。注意,在训练期间不学习特征提取函数,只学习编码器的参数 M 。

最后,为了构造实体基于图像表示,我们为每个实体挑选出最多 10 个图像。由于从 ImageNet 中提取的图像质量非常高,存在的噪声较少,通过对实体相应的图像在实体空间的特征表示进行算术平均,来获取实体的图像表示。

3.3 实体/关系表示建模

TransE 把一对多/多对一关系中的尾/头实体表示为嵌入空间中非常接近的点,而这些实体本身并不相近,因此无法为这些实体创建合适的表示,而文本表示和图像表示会让语义特征和视觉特征相似的实体在嵌入空间相互靠近,从而比单独的结构嵌入更具鲁棒性。

因此,基于学习到的文本表示和图像表示,实体/关系表示模型将文本信息和图像视觉信息融入到知识图谱的表示学习中。该表示模型基于传统的基于平移的方法,以 TransE 为例,评分函数的定义如下:

$$f(h, r, t) = E_s + \mu E_d + \eta E_i \quad (6)$$

其中, μ 和 η 是超参数, E_s 代表结构表示的评分函数,其与 TransE 的评分函数 $\|h + r - t\|$ 相同,头尾实体都是基于结构的表示,即:

$$E_s = \|h_s + r - t_s\| \quad (7)$$

其中, E_d 代表文本表示的评分函数,为使 E_d 的学习过程与 E_s 兼容, E_d 的定义如下:

$$E_d = d(h_s, r, t_d) + d(h_d, r, t_s) + d(h_d, r, t_d) \quad (8)$$

$$d(h_s, r, t_d) = \|h_s + r - t_d\| \quad (9)$$

$$d(h_d, r, t_s) = \|h_d + r - t_s\| \quad (10)$$

$$d(h_d, r, t_d) = \|h_d + r - t_d\| \quad (11)$$

其中, $d(h_s, r, t_d)$ 和 $d(h_d, r, t_s)$ 表示头实体和尾实体中一个使用文本表示,另一个使用结构表示, $d(h_d, r, t_d)$ 中 h 和 t 都使用基于文本的表示。

同理,将图像表示的评分函数 E_i 定义为:

$$E_i = d(h_s, r, t_i) + d(h_i, r, t_s) + d(h_i, r, t_i) \quad (12)$$

其中, E_s, E_d, E_i 共享关系嵌入,并把基于结构、文本和图像的实体表示投影到同一个连续的低维向量空间中,通过联合学习使这 3 种表示之间相互影响。

在实验中,为了方便训练并避免过拟合,我们对嵌入 h, r, t 的 L2 范数施加约束来对向量做归一化处理,即 $\|h_s\|_2 \leq 1, \|t_s\|_2 \leq 1, \|h_d\|_2 \leq 1, \|t_d\|_2 \leq 1, \|h_i\|_2 \leq 1, \|t_i\|_2 \leq 1$ 。

另外,该实体/关系表示模型很容易扩展到其他基于平移的方法,例如 TransH 和 TransR。TransH 把实体嵌入 h 和 t 投影到 w_r 的超平面,因此将模型评分函数中的 h_s 变换成 $h_s - w_r^T h_s w_r, t_s$ 变换成 $t_s - w_r^T t_s w_r$ 。同样,TransR 通过投影矩阵 M_r 将实体嵌入 h 和 t 投影到关系空间中,将评分函数中的 h_s 和 t_s 变为 $h_s M_r$ 和 $t_s M_r$ 即可。

3.4 表示的联合训练

本文模型可以表示为参数集 $\theta = (E, R, X, W^{(1)}, W^{(2)}, M)$,其中 X, E, R 分别代表单词、实体和关系的嵌入, $W^{(1)}, W^{(2)}$ 代表文本 CNN 中不同层的卷积核, M 代表图像编码器中的投影矩阵。 X 通过随机初始化或由 word2vec 预训练得到, E 和 R 通过随机初始化或由现有的 RL 模型预训练得到, $W^{(1)}, W^{(2)}$ 和 M 通过随机初始化得到。

为了优化参数,共同学习实体、关系、文本和图像的表示。基于上述评分函数,将训练目标定义为最小化如下基于间隔的损失函数:

$$L = \sum_{(h, r, t) \in S} \sum_{(h', r, t') \in S'} [\gamma + f(h, r, t) - f(h', r, t')]_+ \quad (13)$$

其中, $\gamma > 0$ 是间隔超参数, $[x]_+$ 表示保持 x 的正数部分, S 是正确三元组的集合, S' 是 S 的负采样集即错误三元组的集合。

$$S' = \{(h', r, t) | h' \in E\} \cup \{(h, r, t') | t' \in E\} \quad (14)$$

S' 的构造方法是通过基于概率的实体替换破坏正确的三元组 (h, r, t) ,来构造错误的三元组 (h', r, t') ,简单地说,就是根据“unif”或“bern”策略使用另一个实体随机替换正确三元组中的头实体或尾实体。特别地,如果替换后的三元组属于 S ,则不能把它加入到 S' 中。由于 h 和 t 有 3 种类型的表示,因此损失函数中的实体也有 3 种表示,即基于结构的表示、基于文本的表示和基于图像的表示。

损失函数希望达到的理想情况是正确 triple 的 $f(h, r, t)$ 尽可能小,而错误 triple 的 $f(h', r, t')$ 尽可能大,这样才能让总体的 loss 趋向于 0。因此,根据随机梯度下降更新三元组对,保证正确三元组的距离越来越小,而错误三元组的距离越来越大,不断优化实体和关系的向量表示。

4 实验

4.1 实验数据集

在实验中,由于需要把两种附加信息即实体文本描述和实体图像融合到知识图谱嵌入任务中,因此需要构造既带有实体文本描述又带有实体图像的多模态知识图谱数据集。

在图像方面,我们借鉴了 IKRL^[21] 中的数据集 WN9-IMG,其三元组部分是 WN18 数据集的子集,图像部分是根据实体相应的 wnid(ImageNet^[22] 中的类别)从 ImageNet 中获取的图像,因为 ImageNet 是根据 WordNet^[23] 组织的图像数据库。但由于在 ImageNet 下载图像时部分实体的图像下载失败且少量下载的图像是单通道的,因此最终数据集的实体、图像和三元组数量与 WN9-IMG 有偏差。

在实体描述方面,我们从 WordNet 中收集数据。WordNet 是一个基于认知语言学的大型英语词汇数据库,每个实体都是一个由多个单词组成的同义词集合(synset),代表不同的词义,关系定义了它们之间的词汇关系。从 WordNet 的 Prolog 数据库中可以获取到 synset 的简要定义,并将其作为 WN9 中实体的文本描述。表 1 列出了我们在 WN9-IMG 基础上构造的同时包含文本描述和实体图像的知识图谱数据集 WN-I-D 的统计情况,包括实体、关系和三元组的数量。

表 1 实验数据集的统计情况

Table 1 Statistics of experimental datasets

数据集	实体	关系	训练集	验证集	测试集
WNID	5825	9	10548	1222	1180

4.2 实验设置

我们按照相关论文中的实验设置在 WN-I-D 数据集上实现 TransE 和 TransR,把它们作为基线,把实体和关系嵌入的维度 d_e 设置为 50,词嵌入的维度 d_w 为 100。从 $\{1, 2, 4\}$ 中选择间隔超参数 γ ,从 $\{0.0002, 0.005, 0.001\}$ 中选择学习率 λ 。为了使文本编码器达到最佳性能,我们把卷积层的滑动窗口大小 h 设置在 $\{1, 2, 3\}$ 中。根据验证集上的实验结果确定最佳配置: $\lambda = 0.001, \gamma = 4, h = 2, \mu = 1, \eta = 1$ 。

4.3 评价指标

本文在两项任务上对模型进行评估:链接预测和关系预测。链接预测旨在根据关系和另一个实体的嵌入来预测三元组中缺失的实体。对于每个缺失的实体,该任务是从知识图谱中给出候选实体的排名列表,而不仅仅是给出最佳答案。对于每个测试三元组 (h, r, t) ,用知识图谱中的每个实体替换 h 或 t ,并将这些实体按照评分函数 f 计算出的分数升序排列。我们希望 (h, r, t) 的得分比任何被替换后的三元组都小。基于实体的排名列表,通过聚合所有的测试三元组采用如下两种评价指标:1)正确实体的平均排名,表示为 Mean Rank (MR);2)正确实体排名前10的比例,表示为 Hits@10。值得注意的是,被替换后的三元组可能有一些存在于知识图谱中,这样的预测应该被认为是正确的,因此在得到排名列表前,过滤掉存在于训练集、验证集或测试集中的被替换后的三元组。将未过滤的情况命名为“Raw”,将经过过滤的情况命名为“Filter”。在这两种情况下,MR 越低或 Hits@10 越高,表示模型的性能越好。

与链接预测相似,关系预测的目的是根据头尾实体的嵌入来预测知识三元组中缺失的关系,即在给定一个三元组 $(h, ?, t)$ 时我们需要预测 h 和 t 之间的关系。该任务根据评分函数计算出的分数进行关系排序,但由于关系的数量比实体的数量少得多,因此使用正确关系排名第1的比例即 Hits@1 作为评价指标。

4.4 实验结果与分析

4.4.1 链接预测

MKGE 模型的实体有结构表示、文本表示和图像表示,因此基于这3种表示报告了5种预测结果;SR 表示实体使用结构表示,TR 和 IR 分别表示使用文本表示和图像表示,STR 表示通过加权结构表示和文本表示进行预测,SIR 表示使用结构表示和图像表示的加权平均。除了将这5种预测结果与基线模型进行比较外,我们还将 MKGE 和 TKGE,IKGE 相同表示类型下的预测结果做了对比,TKGE 和 IKGE 分别

表示模型只融合了实体描述和图像。

链接预测的结果如表2—表5所列,可以观察到如下结果。1)MKGE 模型的5种表示类型在 MR 和 Hits@10 上的评估结果均显著优于基线模型,其中 MKGE(SIR)在 MR 上表现最好。这表明模型已经成功地把实体描述的文本信息和图像的视觉信息融合到知识表示中,提高了知识图谱嵌入的准确性。2)MKGE 中学习的结构表示与 TransE 相似,但预测结果 SR 却远远超出了 TransE。这说明本文模型利用附加信息不仅可以构造实体的另外两种表示,还提高了结构表示的性能。3)相同表示类型下,MKGE 在 MR 和 Hits@10 方面的表现均比 TKGE 和 IKGE 更好。这意味着同时融入两种附加信息,为知识图谱的表示学习提供更加全面的补充是有效的。

表2 unif 设置下链接预测的结果

Table 2 Results of link prediction with unif setting

实验设置	模型	Mean Rank		Hits@10/%	
		Raw	Filter	Raw	Filter
unif	TransE	165	159	77.54	89.32
	TransR	148	142	77.54	90.13
	MKGE(SR)	26	20	81.27	92.84
	MKGE(TR)	24	19	81.65	93.22
	MKGE(STR)	24	19	81.99	93.09
	MKGE(IR)	21	15	80.64	92.80
	MKGE(SIR)	20	14	81.27	92.97

表3 bern 设置下链接预测的结果

Table 3 Results of link prediction with bern setting

实验设置	模型	Mean Rank		Hits@10/%	
		Raw	Filter	Raw	Filter
bern	TransE	165	159	78.31	89.96
	TransR	147	141	78.05	90.34
	MKGE(SR)	28	22	81.19	92.84
	MKGE(TR)	27	22	81.99	92.80
	MKGE(STR)	25	20	81.57	93.01
	MKGE(IR)	22	17	80.93	91.91
	MKGE(SIR)	22	17	81.69	92.46

表4 unif 设置不同附加信息下链接预测的结果

Table 4 Results of link prediction with different additional information under unif setting

实验设置	模型	Mean Rank		Hits@10/%	
		Raw	Filter	Raw	Filter
unif	SR(TKGE/IKGE/MKGE)	51/36/26	45/30/20	80.21/80.51/81.27	91.78/92.25/92.84
	TR(TKGE/MKGE)	47/24	41/19	80.76/81.65	91.86/93.22
	STR(TKGE/MKGE)	48/24	43/19	80.93/81.99	92.12/93.09
	IR(IKGE/MKGE)	22/21	16/15	76.65/80.64	85.13/92.80
	SIR(IKGE/MKGE)	22/20	16/14	80.89/81.27	91.65/92.97

表5 bern 设置不同附加信息下链接预测的结果

Table 5 Results of link prediction with different additional information under bern setting

实验设置	模型	Mean Rank		Hits@10/%	
		Raw	Filter	Raw	Filter
bern	SR(TKGE/IKGE/MKGE)	58/33/28	52/27/22	80.25/80.76/81.19	91.44/92.42/92.84
	TR(TKGE/MKGE)	51/27	46/22	80.38/81.99	91.48/92.80
	STR(TKGE/MKGE)	53/25	48/20	80.55/81.57	91.40/93.01
	IR(IKGE/MKGE)	25/22	20/17	76.27/80.93	84.24/91.91
	SIR(IKGE/MKGE)	24/22	19/17	80.89/81.69	90.68/92.46

表6列出了每个关系头/尾实体预测的 Hits@10 结果。数据集中一共有9个关系,全部都属于 N-to-N 类型。从

表中可以看出,MKGE 在所有关系上的预测结果都明显优于基线,在大多数关系中也优于 TKGE 和 IKGE。这表明本文

模型能够处理 N -to- N 复杂关系的低性能问题,也证明了在知识图谱嵌入任务中同时融合文本描述和实体图像的必要性。

表6 unif/Filter 设置下每个关系头/尾实体预测的 Hits@10 结果

Table 6 Hits@10 results of head/tail prediction on each relation under unif/Filter settings

关系	头实体预测					尾实体预测				
	TransE	TransR	TKGE	IKGE	MKGE	TransE	TransR	TKGE	IKGE	MKGE
derivationally_related_form	0.00	0.00	0.00	100.00	100.00	100.00	100.00	0.00	100.00	100.00
member_meronym	100.00	100.00	100.00	100.00	100.00	50.00	100.00	50.00	100.00	100.00
has_part	81.94	83.33	87.50	87.50	87.50	84.72	83.33	90.28	90.28	90.28
member_holonym	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
part_of	80.26	82.89	88.16	85.53	88.16	80.26	81.58	84.21	84.21	85.53
hyponym	90.60	91.17	93.80	95.68	95.68	89.47	90.60	91.73	90.23	92.48
member_of_domain_topic	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
hypernym	90.87	91.49	92.74	91.49	92.95	91.29	91.70	94.40	95.85	96.27
synset_domain_topic_of	77.78	77.78	88.89	88.89	88.89	77.78	77.78	88.89	88.89	88.89

4.4.2 关系预测

与链接预测类似,我们报告了 MKGE 的关系预测评估结果,如表 7 和表 8 所列。可以看出,MKGE 在 MR 和 Hits@1 方面优于基线、TKGE 以及 IKGE。这些结果进一步说明了 MKGE 可以充分利用实体描述和图像中的外部信息,且文本信息和图像信息相互补充,缺一不可。

表7 unif 设置下关系预测的结果

Table 7 Results of relation prediction with unif setting

实验设置	模型	Mean Rank		Hits@1/%	
		Raw	Filter	Raw	Filter
unif	TransE	1.186	1.185	91.44	91.44
	TransR	1.126	1.125	93.81	93.81
	TKGE	1.059	1.058	96.69	96.78
	IKGE	1.064	1.064	96.27	96.27
	MKGE	1.065	1.064	96.78	96.86

表8 bern 设置下关系预测的结果

Table 8 Results of relation prediction with bern setting

实验设置	模型	Mean Rank		Hits@1/%	
		Raw	Filter	Raw	Filter
bern	TransE	1.165	1.164	92.37	92.37
	TransR	1.124	1.123	93.90	93.90
	TKGE	1.067	1.066	96.02	96.10
	IKGE	1.071	1.070	96.53	96.61
	MKGE	1.062	1.061	96.86	96.95

结束语 本文提出了知识图谱嵌入方法 MKGE,该方法能够实现事实三元组和实体描述、实体图像的融合,旨在同时利用多种附加信息提高知识表示学习能力。

该方法中的实体嵌入除了传统的结构表示外,还包括由 CNN 对实体描述和图像编码生成的文本表示和图像表示,融合这两种表示使用 TransE 对三元组建模并进行联合训练可以得到实体和关系嵌入。

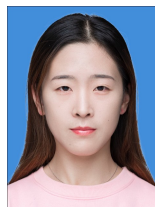
实验结果表明,MKGE 模型优于基线模型以及只融合一种附加信息的模型。这证明本文模型能够将文本信息和图像信息同时编码到知识表示中,从而学习到更加准确的实体和关系嵌入。

在未来的工作中,我们将尝试在知识表示学习中融合更多的附加信息,如除实体描述外其他来源的文本信息、实体类型等。此外,为了使模型更好地利用实体描述和图像,我们将考虑更有效的深度学习模型来对文本和图像进行编码。

参考文献

- [1] 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. 2008:1247-1250.
- [2] AUER S, BIZER C, KOBILAROV G, et al. Dbpedia: A nucleus for a web of open data[C]// The Semantic Web: 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference (ISWC 2007 + ASWC 2007). Busan, Korea: Springer, 2007:722-735.
- [3] SUCHANEK F M, KASNECI G, WEIKUM G. Yago: a core of semantic knowledge[C]// Proceedings of the 16th International Conference on World Wide Web. 2007:697-706.
- [4] YANG B, YI H W, HE X, et al. Embedding entities and relations for learning and inference in knowledge bases[J]. arXiv: 1412.6575, 2014.
- [5] YIN J, JIANG X, LU Z, et al. Neural generative question answering[J]. arXiv:1512.01337, 2015.
- [6] WANG Q, MAO Z, WANG B, et al. Knowledge graph embedding: A survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.
- [7] DONG X, GABRILOVICH E, HEITZ G, et al. Knowledge vault: A web-scale approach to probabilistic knowledge fusion [C]// Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014: 601-610.
- [8] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]// Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 2. 2013:2787-2795.
- [9] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// Proceedings of the 25th International Conference on Neural Information Processing Systems-Volume 1. 2012:1097-1105.
- [10] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]// Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 2. 2013:3111-3119.

- [11] WANG Z,ZHANG J,FENG J,et al. Knowledge graph embedding by translating on hyperplanes[C]// Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. 2014;1112-1119.
- [12] 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.
- [13] FAN M,ZHOU Q,CHANG E,et al. Transition-based knowledge graph embedding with relational mapping properties[C]// Proceedings of the 28th Pacific Asia Conference on Language, Information Cnd computing. 2014;328-337.
- [14] GUO S,WANG Q,WANG B,et al. Semantically smooth knowledge graph embedding[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;84-94.
- [15] LIN Y,LIU Z,LUAN H,et al. Modeling relation paths for representation learning of knowledge bases[J]. arXiv:1506. 00379, 2015.
- [16] XIE R,LIU Z,JIA J,et al. Representation learning of knowledge graphs with entity descriptions[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016; 2659-2665.
- [17] JIANG T,LIU T,GE T,et al. Encoding temporal information for time-aware link prediction[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016;2350-2354.
- [18] KIM Y. Convolutional Neural Networks for Sentence Classification[J]. arXiv:1408. 5882,2014.
- [19] PASZKE A,GROSS S,MASSA F,et al. PyTorch:an imperative style,high-performance deep learning library[C]// Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019;8026-8037.
- [20] SHUTOVA E,KIELA D,MAILLARD J. Black holes and white rabbits;Metaphor identification with visual features[C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics; Human Language Technologies. 2016;160-170.
- [21] XIE R,LIU Z,LUAN H,et al. Image-embodied knowledge representation learning[J]. arXiv:1609. 07028,2016.
- [22] DENG J,DONG W,SOCHER R,et al. Imagenet; A large-scale hierarchical image database[C]// 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009; 248-255.
- [23] MILLER G A. WordNet; a lexical database for English[J]. Communications of the ACM,1995,38(11);39-41.



XIAO Guiyang, born in 1998, master. Her main research interests include knowledge graph embedding and deep learning.



WANG Lisong, born in 1969, Ph.D., professor, is a member of China Computer Federation. His main research interests include natural language processing and formal method.

(责任编辑:喻葵)