

# 知识图谱表示学习综述

寇晓宇<sup>1</sup>

<sup>1</sup>(北京大学 信息科学技术学院 北京 北京 250101)

通讯作者: , E-mail: \*\*\*\*@ntu.edu.cn

**摘 要:** 知识图谱表示学习是利用深度学习相关技术将实体和关系映射到连续低维向量空间中, 从而有效的解决数据稀疏性的问题。它可用于知识图谱补全和三元组分类等多种任务, 因此很快得到了广泛的关注。在本文中, 首先系统地介绍当前知识图谱表示学习进展, 包括仅使用结构信息的方法和融入多源信息的方法。之后, 本文讨论了如何将实体和关系的表示应用到各种下游任务中, 如知识图谱补全, 实体分类等。最后, 我们简单展望了该技术的未来发展前景。

**关键词:** 知识图谱; 表示学习; 链接预测; 深度学习

**中图法分类号:** TP311

## State-of-the-Art Survey of Knowledge Graph Embedding

KOU Xiao-Yu<sup>1</sup>

<sup>1</sup>(School of Information Science and Technology, Peking University, Beijing 250101, China)

**Abstract:** Knowledge graph embedding is to embed entities and relationships into continuous low-dimensional vector space by using deep learning technology, so as to alleviate sparsity issues. It can benefit a variety of tasks, such as KG completion and triple classification, and has attracted many attentions. In this paper, we systematically introduce the state-of-the-art advances of knowledge graph embedding, including methods using only structural information and methods incorporating multi-source information. After that, we discuss how to apply these methods to various downstream tasks, such as KG completion, and triple classification, etc. Finally, we briefly looked into the future development prospects of this technology.

**Key words:** Knowledge map; Learning; Link prediction; Deep learning

目前, 知识图谱在许多领域已经被广泛的应用, 如信息抽取<sup>[1]</sup>、自动问答<sup>[2]</sup>、命名实体消歧<sup>[3]</sup>等。知识图谱可以看做是一个图结构的信息网络。典型的知识图谱如 Freebase<sup>[4]</sup>, DBpedia<sup>[5]</sup>和 YAGO<sup>[6]</sup>, 由结构化的事实三元组(head、relation、tail)表示, 其中 head 和 tail 是实体, relation 表示实体之间的关系。例如(周杰伦, 职业, 歌手)三元组表示“周杰伦的职业是歌手”的事实。虽然用 RDF 三元组形式表示知识图谱的方式比较通用, 但是会存在计算效率和数据稀疏性的问题, 从而会导致不容易对知识图谱进行语义推理操作。因此, 如何更好地表示知识图谱逐渐成为一项具有挑战性的任务。

近年来, 以深度学习为代表的知识图谱表示学习技术获得广泛关注<sup>[7] [8] [9]</sup>。该技术将知识图谱中的实体和关系映射到稠密低维向量空间中, 从而在保证不改变知识图谱本身结构的同时, 高效的计算实体之间的语义联系。除此之外, 高质量的实体和关系向量可以进一步用于各种任务, 如知识图谱补全、实体分类等任务。

目前大多数知识图谱表示学习的方法仅根据实体的结构信息。对于给定的知识图谱, 首先将实体和关系映射为低维空间的连续向量, 并根据每个事实定义一个评分函数来度量其合理性。然后, 通过最大化观察到的事实的可信度, 可以获得最终的实体和关系嵌入(embedding)。这个学习的过程是基于封闭世界假设的, 即测试集中的实体都在训练集中出现过, 而通过这种方式所学习到的嵌入对于下游任务来说, 可能是不够充分的。因此, 越来越多的研究者开始进一步利用其他类型的信息, 如实体类型<sup>[10] [11]</sup>、文本描述<sup>[12] [13]</sup>、

逻辑规则<sup>[14]</sup> 等来学习更多的嵌入信息。

在本文中,我们将对当前的知识图谱表示学习技术进行全面的回顾,包括那些仅使用实体结构信息的技术,以及那些进一步融合多源信息的技术。我们还将进一步介绍如何将学到的嵌入应用到各种下游任务中,并使之受益。官赛萍等人[40] 对知识图谱上的知识推理方法进行了研究,包括单步推理和多步推理多种推理方式;刘知远等人[41] 对知识表示学习的研究进行了总结,主要是介绍了知识表示学习技术的优越性、当前进展以及未来发展方向。与他们的工作不同,我们特别关注知识图谱表示学习技术,分别详细介绍仅使用结构信息和融合了多源信息的技术,不仅包括最新的技术,还包括未来的发展趋势。

本文第 1 节介绍知识图谱表示学习的基本概念和相关的基本符号。第 2 节总结了仅使用了实体结构信息的相关技术,主要包括翻译模型、矩阵分解模型和神经网络模型三大类。第 3 节对融合了多源信息的技术进行总结,我们主要关注于实体类型、文本描述等多源信息。第 4 节探讨了 KG 嵌入在下游任务中的应用,如 KG 补全、三元组分类等任务。最后在第 6 节中总结全文,并对未来的研究方向进行简单探讨。

## 1 基本概念和符号介绍

### 1.1 知识图谱表示学习的基本概念

知识图谱表示学习的目标是,将知识图谱中的三元组(实体和关系)通过机器学习的方式表示为低维稠密空间中的实值向量。在该向量空间中,我们可以计算任意事实三元组之间的语义距离,如欧氏距离或者余弦距离等。除此之外,知识图谱通过表示学习得到的低维向量可以用于高效的计算实体、关系之间的语义关系,从而可以进行知识图谱补全等任务。

在介绍知识图谱表示学习的几类代表方法之前,我们首先定义了一些符号。我们使用  $G=(E,R)$  表示知识图谱,其中  $E=\{e_1, e_2, \dots, e_i\}$  是实体集合,共包含  $|E|$  种不同实体;  $R=\{r_1, r_2, \dots, r_i\}$  是关系集合,共包含  $|R|$  种不同关系;事实三元组表示为  $(h, r, t)$ ,其中  $h \in E$  和  $t \in E$  表示头实体和尾实体,  $r \in R$  表示  $h$  和  $t$  实体之间的关系。

## 2 基于实体结构信息的模型

典型的知识图谱表示学习技术一般包括三个步骤:(i)将实体和关系映射到低维空间中;(ii)定义得分函数(score function);(iii)学习实体和关系的向量表示。第一步中是指定实体和关系在连续向量空间中表示的形式,实体通常用向量表示,关系通常表示为向量、矩阵或者张量,代表向量之间的一些操作。在第二步中,得分函数是定义每个事实  $(h, r, t)$  真实存在的概率。当然,能够在训练集中观察到的事实往往比那些没有观察到的有更高的分数。最后,通过最小化损失函数来学习这些实体和关系表示。

我们大致将基于实体结构信息的模型分为三类:翻译模型,矩阵分解模型和神经网络模型。在本节中,我们首先介绍这三种技术中有代表性的模型,然后我们比较了这些表示学习技术的优劣性。

### 2.1 翻译模型

翻译模型通常是利用基于距离的得分函数,在通过关系进行翻译之后,使用两个实体之间的距离来衡量一个事实的可信性。TransE<sup>[8]</sup> 是最具代表性的翻译模型。它将实体和关系都表示为同一空间中的向量,并将关系看做是实体间的某种平移向量。该想法是受到 Mikolov 等人<sup>[15]</sup> 在 2013 年提出的 word2vec 词表示学习模型的启发, Mikolov 等人发现词汇的语义关系和语法关系中普遍存在这种平移不变性。对于给定的三元组  $(h, r, t)$ , TransE 用关系的向量  $r$  作为头实体向量  $h$  和尾实体向量  $t$  之间的转移(翻译)。如图 1(a) 所示, TransE 的主要假设是,如果某三元组成立,则向量  $h$  与  $r$  的和与  $t$  相近,否则远离。因此损失函数定义为  $d = -\|h + r - t\|_{L_1/L_2}$ , 其中距离使用  $L_1$  或者  $L_2$  范数来衡量。

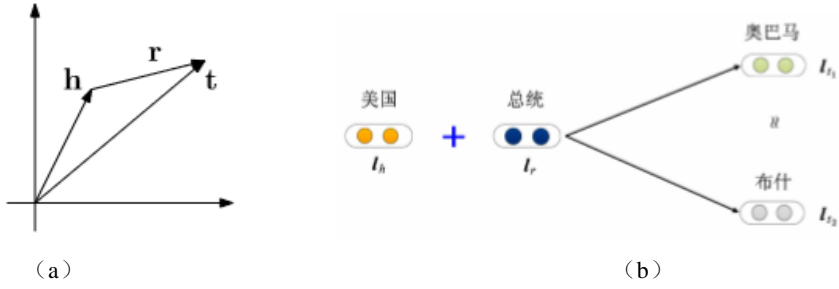


图 1

而在实际学习过程中, TransE 使用了替换头实体和尾实体的方式进行负采样, 来增强向量表示的区分度, 并且采用最大间隔的方法, 最小化公式 (1) 的目标函数:

$$\mathcal{L} = \sum_{(h, \ell, t) \in S} \sum_{(h', \ell, t') \in S'_{(h, \ell, t)}} [\gamma + d(h + \ell, t) - d(h' + \ell, t')]_+ \quad (1)$$

其中  $S$  是正确的三元组,  $S'$  是通过随机替换头实体和尾实体 (不能同时替换) 得到的负三元组,  $\lambda$  为正样本得分和负样本得分之间的间隔距离。TransE 简单有效, 但存在一些不足。比如 TransE 严格要求有效的三元组满足头实体加关系在向量空间中与尾实体足够靠近, 因此可以很好地处理一对一关系 (1-1)。但在处理多映射关系类型时 (1-N, N-1, N-N), 存在多个实体竞争一个点的现象。例如图 1 (b) 所示, 在一对多关系中, 同一个头实体“美国”与多个尾实体“奥巴马”, “布什”存在关系“总统”, 由于  $h + r \approx t_i, i = \{1, 2, \dots, x\}$  的约束, 这些尾实体将竞争空间中的同一个点, 即使它们语义差别很大, 因而会造成向量空间的拥挤和误差。

针对 TransE 在处理多映射关系时的局限性, TransH<sup>[9]</sup> 提出在 TransE 的基础上为每个关系多学一个映射向量, 用于将实体映射到关系指定的超平面, 如图 2 (a) 所示; 然后在该超平面, 与 TransE 一样, 关系表示向量看成映射后的实体之间的转移。映射向量使得同一个实体在不同关系指定的超平面有不同的表示, 一定程度上缓解了不能很好地处理多映射属性关系的问题。但是 TranH 模型仍然假设实体和关系处于相同的语义空间中, 这在一定程度上会限制 TransE 的效果。进一步的, Lin 等人提出 TransR<sup>[16]</sup>, 在单独的实体空间和关系空间建立实体和关系的表示。如图 2 (b) 所示, 每个关系对应一个空间, 有一个映射矩阵, 实体通过对应的映射矩阵映射到关系空间后, 将关系向量看成实体向量间的转移。会发现, 在原本实体空间中相似的实体 (三角形表示), 在关系空间中被区分开了。

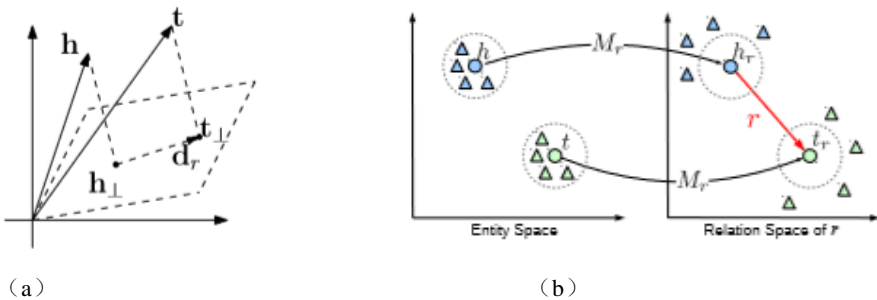


图 2

考虑到关系可以进行更细致的划分, Lin 等人还提出了 CTransR<sup>[16]</sup>, 将关系划分为关系组, 为每个关系组学习一个关系向量和映射矩阵。在 TransR 基础上, Ji 等人提出了 TransD<sup>[17]</sup>, 考虑实体和关系之间的交互; TransSparse<sup>[18]</sup> 提出使用稀疏矩阵代替 TransR 中的稠密矩阵来解决实体和关系的异质性 (不同关系所关联的实体数目不同); TransA<sup>[19]</sup> 提出将损失函数中的 L1/L2 距离替换为马氏距离, 并且为不同维度学习不同的权重; TransG<sup>[20]</sup> 提出用高斯混合模型描述头尾实体之间的关系, 即每种关系的语义用一个高斯分布来表示。

以上多种模型从不同角度尝试了改进 TransE 来解决多映射关系问题,并在相关数据集上的实验表明了方法的有效性。

## 2.2 矩阵分解模型

矩阵分解模型利用基于相似性的评分函数。它们通过匹配在对应空间中的实体向量和关系向量来度量某三元组事实的可信性。首先将 $(h, r, t)$ 三元组看成张量/矩阵中的元素构建张量/矩阵,通过张量/矩阵分解方法进行表示学习,分解得到的向量表示相乘重构成张量/矩阵,元素值即为对应三元组有效与否的得分。可以认为得分大于特定阈值的三元组有效,或按照预测的得分排序,选择得分高的候选作为推理结果。

该类方法的代表是 Nickel 等人提出的 RESCAL<sup>[7]</sup>,基于三阶张量进行表示学习,模型如图 4(a)所示。如果三元组成立,三阶张量上对应的元素值为 1,否则为 0。例如,图 3 中第  $i$  个实体和第  $j$  个实体存在第  $k$  种关系,对应的位置 $(i, j, k)$ 元素值为 1。该三阶张量对每种关系进行切片分解,其中,对第  $k$  种关系的切片可以分解为实体表示矩阵  $\mathbf{A}$ 、第  $k$  种关系对应的非对称矩阵  $\mathbf{R}_k$  与实体表示矩阵的转置  $\mathbf{A}^T$  的连乘,通过最小化重构误差学习实体和关系的表示。得分函数定义为公式(2),其中  $\mathbf{M}_r$  表示关系矩阵。张量分解模型虽然推理准确率高,但内存占用量大,计算速度慢,而且在小数据集上容易过拟合。

$$f_r(h, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t} = \sum_{i=0}^{d-1} \sum_{j=0}^{d-1} [\mathbf{M}_r]_{ij} \cdot [\mathbf{h}]_i \cdot [\mathbf{t}]_j, \quad (2)$$

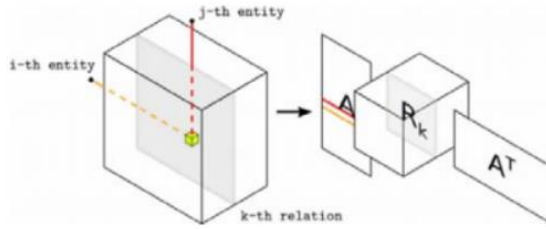


图 3

针对于 RESCAL 的问题, DistMult<sup>[21]</sup> 通过限制关系矩阵  $\mathbf{M}_r$  为对角矩阵来简化 RESCAL 模型,如图 4(2)。对于给定三元组 $(h, r, t)$ ,其得分函数为公式(3)。这个分数使用了  $\mathbf{h}$  和  $\mathbf{t}$  在同一维度上的分量之间的成对交互作用(pairwise interactions),并大大减少了参数的数量。由于 DistMult 使用了对角矩阵,所以只能建模对称关系,而不能建模非对称关系。

$$f_r(h, t) = \mathbf{h}^T \text{diag}(\mathbf{r}) \mathbf{t} = \sum_{i=0}^{d-1} [\mathbf{r}]_i \cdot [\mathbf{h}]_i \cdot [\mathbf{t}]_i. \quad (3)$$

HolE<sup>[22]</sup> 将 RESCAL 的强大的表达能力与 DistMult 的效率和简单相结合。它将实体和关系映射到关系向量空间中。对于 $(h, r, t)$ , HolE 首先采用循环相关运算(可以看作张量乘法的特殊形式),将实体表示组合成  $\mathbf{h} \star \mathbf{t}$ ,然后融合组合向量和关系向量,从而对三元组进行评分。评分函数见公式(4):

$$f_r(h, t) = \mathbf{r}^T (\mathbf{h} \star \mathbf{t}) = \sum_{i=0}^{d-1} [\mathbf{r}]_i \sum_{k=0}^{d-1} [\mathbf{h}]_k \cdot [\mathbf{t}]_{(k+i) \bmod d}. \quad (4)$$

循环相关操作可以压缩两两相互作用的向量,如图 4(c)。因此 HolE 对于每个关系只需要  $O(d)$  的参数量,同时,因为循环相关操作是不可以交换的(不符合交换律),因此可以建模非对称关系。

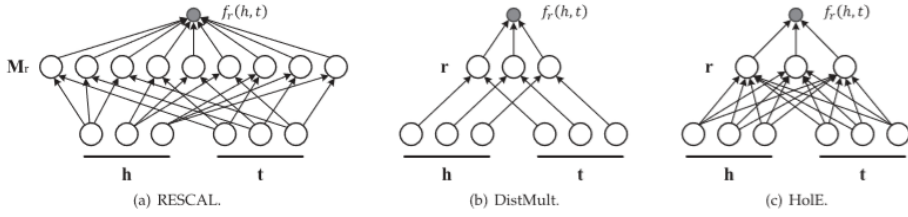


图 4

Complex<sup>[23]</sup> 通过将实体和关系映射到复数空间来拓展了 DistMult 算法, 因此评分函数不再是对称的, 来自非对称关系的事实可以根据涉及实体的顺序得到不同的分数; ANALOGY<sup>[24]</sup> 通过对实体和关系的类比属性进行建模来拓展了 RESCAL, 因为 ANALOGY 将关系表示为矩阵, 这些矩阵可以同时块对角化 (block-diagonalized) 形成稀疏的近似对角矩阵, 因此参数量也是  $O(d)$ 。

近期, Ivana 提出 TuckER<sup>[25]</sup> 模型, 一个基于 Tucker 分解的简单有效的线性张量分解模型。该模型采用 RACSAL 表示知识图谱的方式, 将知识图谱中的三元组表示为张量, 若某  $(h, r, t)$  成立, 则张量对应位置为 1, 否则为 0。如图 5(a)所示, TuckER 将张量分解为一个较小的核张量与 3 个维度上的向量 (即头实体、尾实体、关系的对应表示), 并且对于相同实体出现在头实体位置或者尾实体位置时的向量表示是相同的。因为分解出来的核张量  $W$  是非对称的, 所以可以表示非对称关系, 并且可以认为不同关系之间的知识是共享的。该方法被证明是完全表达(full expressiveness)的, 并且其他矩阵分解方法, 如 RESCAL, DistMult, Complex 和 Simple 等都可以看做是 TuckER 的一种特殊形式。

Sun 等人提出 ROTATE<sup>[26]</sup>, 该方法从建模和表示 3 类关系 (对称/反对称关系, 翻转关系, 合成关系) 入手, 在复数空间中关系定义为从头实体到尾实体的一种旋转操作, 如图 5(b)所示。首先, ROTATE 将实体和关系映射到复数空间中, 并且对于同一实体作为头实体和尾实体时设置不同的向量表示, 从而可以更好的建模对称/非对称关系。然后定义了复数空间的旋转操作 (即向量的内积操作, 图 5(c)), 从而对于翻转关系和合成关系可以很好的表示。

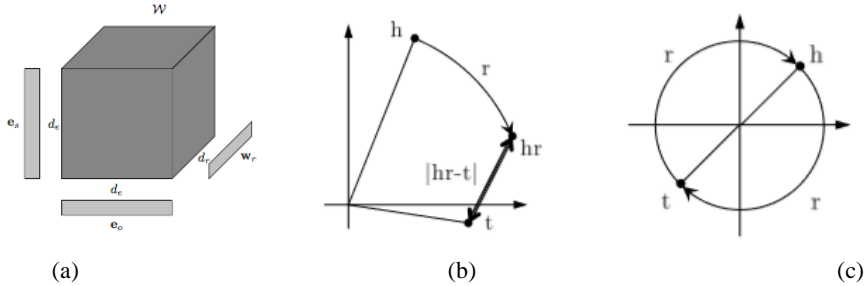


图 3

除此之外, ROTATE 提出了一种自对抗 (self-adversarial) 负采样方式, 通过每次迭代得到的当前模型来负采样, 从而可以得到当前的负样本对模型的贡献程度。模型的目标函数最终定义为公式 (5)

$$L = -\log \sigma(\gamma - d_r(h, t)) - \sum_{i=1}^n p(h'_i, r, t'_i) \log \sigma(d_r(h'_i, t'_i) - \gamma) \quad (5)$$

$$d_r(h, t) = \|h \circ r - t\| \quad p(h'_j, r, t'_i | \{(h_i, r_i, t_i)\}) = \frac{\exp \alpha f_r(h'_j, t'_i)}{\sum_i \exp \alpha f_r(h'_i, t'_i)}$$

其中  $L$  仍然是最大间隔函数, 只是负样本生成部分换成自对抗方式,  $d$  为损失函数,  $p$  为当前负样本对模型的重要程度。通过在数据集上实验发现, 该模型能够很好的建模 3 类关系, 以及多映射关系。

### 2.3 神经网络模型

基于神经网络的模型是利用神经网络直接建模知识图谱事实三元组,得到头实体、关系和尾实体的向量表示。该方法依然是一种基于得分函数的方法,区别于其他方法,整个网络构成一个得分函数,神经网络的输出即为得分值。

Socher 等人提出了神经张量网络 NTN(neural tensor network)<sup>[27]</sup>,用双线性张量层代替传统的神经网络的线性变换层,在不同的维度下,将头实体和尾实体联系起来,刻画实体间复杂的语义联系(如图6所示)。与以往模型不同的是,NTN的实体向量表示是通过该实体中所有单词词向量的平均得到,因为实体单词数量远小于实体数量,所有可以充分重复利用单词向量。具体地,每个三元组用关系特定的神经网络学习,头尾实体作为输入,与关系张量构成双线性张量积,进行三阶交互,同时建模头尾实体和关系的二阶交互。最后,模型返回三元组的置信度,即:如果头尾实体之间存在该特定关系,返回高的得分;否则,返回低的得分。特别地,关系特定的三阶张量的每个切片对应一种不同的语义类型,一种关系多个切片可以更好地建模该关系下不同实体间的不同语义关系。NTN的得分函数定义为公式(6)

$$f_r(h, t) = \mathbf{r}^\top \tanh(\mathbf{h}^\top \mathbf{M}_r \mathbf{t} + \mathbf{M}_r^1 \mathbf{h} + \mathbf{M}_r^2 \mathbf{t} + \mathbf{b}_r), \quad (6)$$

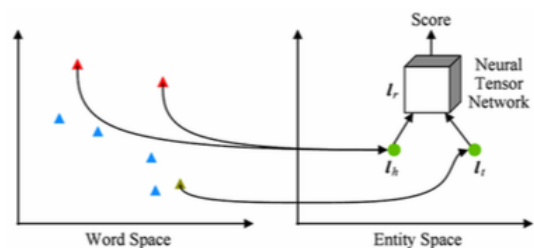


图6

Shi 和 Weninger 提出了共享变量神经网络模型 ProjE<sup>[28]</sup>。ProjE 通过简单的组合操作组合三元组的已知部分(头实体与关系或尾实体与关系)建立目标向量空间,并映射未知部分(尾实体或头实体)的候选集到相同的空间。用基于候选的排序损失学习模型,相比于普遍采用的转移矩阵,组合操作减少了大量参数。进一步通过候选采样,可以处理大规模知识图谱。

近期, Tim 提出 ConvE<sup>[29]</sup> 模型,该模型是首个利用了多层卷积神经网络的方法,如图7所示。首先将实体和关系表示为向量的形式,通过 2D reshape 操作,将一维向量转换为二维的矩阵,目的是希望可以获得嵌入表示之间的隐层信息。然后将头/尾实体和关系拼接起来,输入到卷积网络中,对候选尾/头实体进行排序,得到最终的分值。该方法在多个数据集上验证了 2D 矩阵和卷积神经网络的有效性。

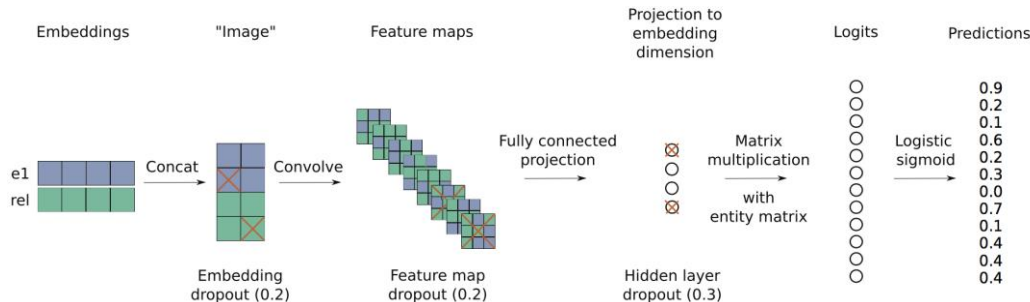


图7

总而言之,基于神经网络模型试图利用网络强大的学习能力建模知识图谱事实三元组,可以获得很好的推理能力和泛化能力。然而,神经网络固有的可解释性问题也依然存在于知识图谱的应用中,如何恰当地解释神经网络的推理能力是一大难点。



### 3 融合多源信息的模型

上面介绍的方法仅使用知识图谱三元组中的结构信息来进行表示学习任务。事实上,有许多与知识有关的多源信息可以用来进一步降低知识图谱的不连通性和稀疏性,从而提高下游任务的效果,例如实体类型、文本描述以及其他信息等。在本节中,我们将讨论如何融合这些多源信息来进行知识图谱表示学习。

#### 3.1 实体类型

我们考虑的第一类多源信息是实体类型,即实体所属的语义类别。例如,“AlfredHitchcock”的类型是“Person”,而“Psycho”的类型是“CreativeWork”。这类信息在大多数知识图谱中是有用的,通常由特定关系表示并以三元组形式存储,比如:(Psycho, IsA, CreativeWork)。Xie 等人<sup>[13]</sup>设计了一种基于实体类型建模(type-embodied)的知识表示学习方法(TKRL),它可以处理层次实体类别和多个类别标签。TKRL 是一种具有特定类型实体投影的平移距离模型。给定三元组(h,r,t),TKRL 首先将 h 和 t 使用与特定类型相关的矩阵映射到向量空间,然后将关系 r 作为两个实体向量之间的翻译(平移)。因此,评分函数定义为公式(7):

$$f_r(h, t) = -\|M_{rh}h + r - M_{rt}t\| \quad M_{rh} = \frac{\sum_{i=1}^{n_h} \alpha_i M_{c_i}}{\sum_{i=1}^{n_h} \alpha_i}, \quad \alpha_i = \begin{cases} 1, & c_i \in \mathbb{C}_{r_h}, \\ 0, & c_i \notin \mathbb{C}_{r_h}, \end{cases} \quad (7)$$

其中  $M_{rh}$  和  $M_{rt}$  是 h 和 t 的对应的映射矩阵。为了能够表示多种类型,  $M_{rh}$  表示为所有类型矩阵的加权平均,  $c_i$  是第 i 种类型,  $M_{c_i}$  是  $c_i$  的映射矩阵,是对应权重。虽然 TKRL 在链接预测、三元组分类等下游任务中取得了较好的性能,但由于它将每个类别与特定的映射矩阵相关联,因此空间复杂度相对较高。

实体类型也可以作为不同关系的头实体和尾实体的约束,例如,关系“DirectorOf”的头实体应该是“Person”类型的,而实体应该是“CreativeWork”类型的。研究<sup>[30]</sup>和<sup>[31]</sup>试图在训练中产生负样本的过程施加这样的约束,将违反实体类型约束的反例被排除在训练之外<sup>[30]</sup>,或者生成的概率非常低<sup>[31]</sup>。

#### 3.2 文本描述

本节讨论融合实体文本信息知识图谱表示学习模型。实体描述信息是实体本身精炼的文字性介绍,本质上是一类文本信息。除了存储在知识图谱中的实体描述之外,还可以扩展到包含更一般的文本信息,比如新闻稿和 Wikipedia 文章等。

将文本信息融入到知识图谱表示学习中可以追溯到 NTN 模型,其中文本信息仅用于初始化实体表示。具体来说,NTN 首先从一个辅助新闻语料库中学习单词向量,然后通过平均每个实体名称中包含的单词向量来初始化每个实体的表示。例如,AlfredHitchcock 的嵌入是由“alfred”和“hitchcock”的平均字向量初始化的。后来,Wang 等人<sup>[32]</sup>提出了第一个联合学习模型,该模型可以更好地利用表示学习过程中的文本信息。关键思想是是基于平移模型<sup>[8]</sup>与 Skip-gram<sup>[33]</sup>模型的思想,对知识空间与文本空间分别进行建模,学习实体向量、关系向量与词向量。同时,该模型设计了对齐模块,使用维基百科锚文本信息或者实体名字信息,实现了知识空间与文本空间的对齐。随后,Zhong<sup>[12]</sup>等研究者基于上面模型做出了改进。在对齐模块中,该模型引入了实体描述文本辅助对齐,认为实体向量应与实体描述中的所有词向量尽可能接近。但是以上和文本结合的知识表示学习模型仅孤立地考虑单词级别的文本信息,忽略了篇章级别重要的语序语义信息。

针对这个问题,Xie 等人提出了 DKRL 模型<sup>[13]</sup>,该模型使用神经网络模型整体考虑实体描述的语义信息,直接基于实体描述构建实体的知识表示,同时能够对新实体构建合理的知识表示。如图 8 所示,DKRL 模型以实体描述为输入,通过实体描述编码器得到实体基于描述的表示。在顶层,基于描述的表示与基于结构的表示被统一在平移模型的框架下进行学习。学习到的信息通过反向传播更新实体描述编码器的参数和所有向量。DKRL 沿用了平移模型的假设,定义了评分函数为公式(8)

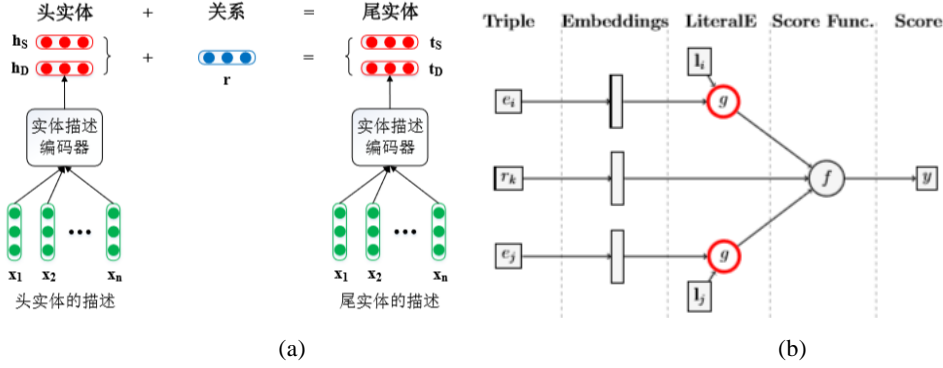


图 8

$$E(h, r, t) = \alpha_1 \|h_S + r - t_S\| + \alpha_2 \|h_S + r - t_D\| + \alpha_3 \|h_D + r - t_S\| + \alpha_4 \|h_D + r - t_D\|. \quad (8)$$

其中  $\alpha_1, \alpha_2, \alpha_3, \alpha_4$  是控制各项权值的超参数。在函数中,  $\|h_S + r - t_S\|$  部分与平移模型的能量函数类似, 而  $\|h_S + r - t_D\|$ ,  $\|h_D + r - t_S\|$  和  $\|h_D + r - t_D\|$  三项则加入了基于描述的实体向量影响。这样, 通过混合项的软限制, DKRL 模型将实体的两种向量映射到了同一语义空间中, 并共享相同的关系向量。实验结果表明, DKRL 算法优于 TransE 算法, 特别是在 zero-shot 的情况下。

近期, 一些新方法提出来更好的利用实体的文本信息。如 Literal 方法<sup>[34]</sup>, 使用可学习的参数化函数(如简单的线性或者非线性变换或多层神经网络)将实体嵌入与它们的文字信息合并, 从而形成新的实体嵌入, 如图 8(b)。主要思想是通过结构信息得到的实体表示之外加上一个矩阵, 该矩阵是每个实体通过文本信息得到的向量组合而成, 即如果某三元组  $(h, r, t)$  成立, 则  $h$  实体对应的向量上  $t$  位置上为 1。Literal 通过这种简单方式扩展了几种常用的基于结构信息的表示学习模型, 并发现在链接预测任务中的性能有一定的提升。Shi 基于开放世界(Open-World)假设提出了 ConMask 方法<sup>[35]</sup> (开放世界假设测试集中可能会出现之前未见到的实体), 该模型学习实体名称及其部分文本描述的向量表示, 以便学习到在训练数据集中不可见的实体。为了减小文本描述的噪声影响, ConMask 使用一个“mask”操作来提取与关系相关的重要文本片段(如图 9), 并引入卷积神经网络来融合提取的内容和实体。

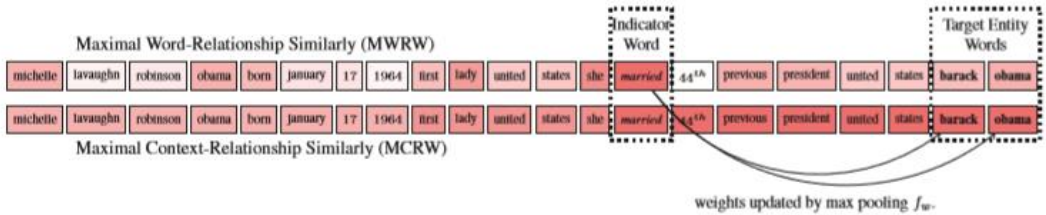


Figure 2: Relationship-dependent Content Masking heat map for the description of Michelle Obama given relationship type spouse. Stop-words are removed. Higher weights show in darker color.

图 9

### 3.3 其他信息

除了上述两种附加信息外, 还有一些研究试图将其他信息纳入知识图谱表示学习中。如 Hatem 提出的一种 multimodal 方法<sup>[36]</sup>, 基于翻译学习的方法, 同时融合了两种不同类型的外部多源信息, 包括文本描述信息和对应实体的图片信息(如图 10)。



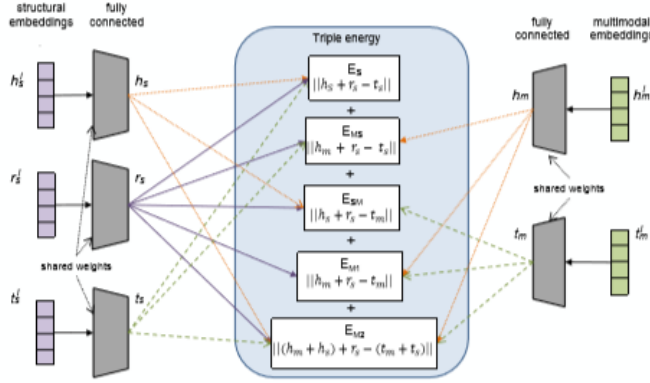


图 10

Ding<sup>[37]</sup> 提出了利用非常简单的约束来提高知识图谱的表示能力, 包括实体表示的非负性约束和关系表示的近似蕴涵约束。前者有助于学习实体的紧凑表示和可解释表示。后者将关系之间逻辑蕴涵的规则进一步编码到它们的分布式表示中。这些约束将先验知识强加于嵌入空间中, 而不会对效率或可伸缩性产生负面影响。该方法非常的简单, 将该方法加入到 ComplEx 上面, 在多个数据集上证明了相应的效果。说明对一些相对复杂的模型加入简单的约束或者逻辑规则是一种提高模型泛化能力的很好的方式。

## 4 下游任务应用

在对当前的知识图谱表示学习技术进行系统回顾之后, 本节将探索如何将学习到的实体和关系嵌入应用于各种下游任务并从中受益, 主要包括知识图谱补全 (链接预测)、三元组分类和其他领域任务等。

### 4.1 知识图谱补全 (链接预测)

现有的知识图谱由于数据源的不全面以及知识获取的遗漏, 不可能构建完备的知识图谱。通常, 代表性的知识图谱中有 69%~99% 的实体缺少至少一个属性信息三元组<sup>[38]</sup>。例如, 在 Freebase 中, 93.8% 的人没有出生地信息, 78.5% 的人没有国籍信息<sup>[39]</sup>。解决办法之一就是通表示学习方法, 利用知识图谱中已有的知识去推断出新的事实(即隐含的知识), 从而尽可能地对知识图谱进行补全, 也叫做链接预测任务。

具体地, 知识图谱补全任务要求模型在给定三元组(h,r,t)的某一实体或关系丢失的情况下, 利用另外两个对象正确预测出丢失的实体或关系。根据具体丢失的对象, 知识图谱补全可以细分为实体预测与关系预测两个任务。知识图谱补全任务旨在验证所学知识表示的准确性, 同时也与现实中问答系统的实际任务相呼应。如需补全三元组(?, 写作, 哈姆雷特), 即相当于回答“谁写了《哈姆雷特》这本书?”这个问题。通过预先学习实体和关系表示, 链接预测可以简单地通过一个排序过程来实现。评分函数给出的分数越低, 说明某三元组越有可能成立。

评价指标通常使用 Mean Rank 和 Hits@N 两个。其中, Mean Rank 表示所有候选集中正确三元组的平均排序, 关注知识表示的整体效果; 而 Hits@N 表示排名前 N 的答案中有正确三元组的比例, 关注在实际任务中模型补全的实用性。除此之外, 由于知识图谱中存在大量一对多、多对一甚至多对多的关系, 仅仅依靠排序直接进行统计可能会存在一些问题。例如, 针对(莎士比亚, 写作, ?)这个三元组补全, 尾实体有多个正确答案, 如哈姆雷特、罗密欧与朱丽叶、仲夏夜之梦等。那么在评测时, 即使最好情况下这三个实体排名前三, 这些样例的 Mean Rank 值也只有 2, 这将与我们的评测初衷相违背。针对这个问题, 构造出“Raw”和“Filter”两个指标。其中“Raw”表示不进行额外的处理, 而“Filter”表示在评测实体时, 不考虑其它所有正确的实体。如果某正确实体排名第 4, 而排序前 3 的实体中有 2 个也是知识图谱中已有的正确答案, 那么在“Raw”的设定下, 这个实体的 Mean Rank 值为 4, 而在“Filter”的设定下, 这个实体的 MeanRank 值为 2。

## 4.2 三元组分类

三元组分类在于验证一个测试三元组(h,r,t)是真是假。例如, (AlfredHitchcock,DirectorOf,Psycho)应该归类为一个真正的事实, 而(JamesCameron, DirectorOf, Psycho)应该分类为假。同样, 这个任务也可以看作是对知识图谱的某种补全形式。三元组分类可以简单在评分函数得到的分数基础上进行。分数高的三元组往往是真实的事实。具体来说, 我们对每个关系  $r$  设置一个阈值  $k$ , 然后对于测试三元组, 如果得分高于阈值  $k$ , 则为真, 否则为假。传统的分类指标可以用来评估这一任务, 如微观和宏观平均精度 (micro- and macro-averaged accuracy)。由于每个三元组评测分数都将与二进制标签一起输出, 所以这里也可以使用排名指标, 例如平均精度 (mean average precision)。

## 4.3 其他领域任务

除了知识图谱补全与三元组分类任务, 知识图谱表示学习在垂直搜索、智能问答、机器翻译等领域也发挥了重要作用, 在疾病诊断、金融反欺诈、数据异常分析等诸多不同的领域已展示出良好的应用前景。

在垂直搜索领域, 在国外, Google 公司提出了 Knowledge Graph 与 Knowledge Vault; Facebook 推出了 Graph Search; 微软推出了 Bing Satori; 在国内, 搜狗提出了知立方; 百度推出了中文知识图谱搜索。知识推理能够更好地理解用户的搜索意图, 提供接近“专、精、深”的垂直搜索, 回答复杂的推理问题。

在智能问答领域, IBM 的 Watson、Google 的 Google Now、苹果公司的 Siri、亚马逊的 Alexa、微软的小娜和小冰以及百度的度秘等是近期代表性的智能问答系统。这些系统基于知识图谱的知识推理, 提供精确、简洁的答案。

## 5 总结与展望

知识图谱表示学习是一种将实体和关系嵌入到连续向量空间的方法, 在各种面向实体的任务中得到了重要的应用, 并迅速得到了广泛的关注。本文系统地回顾了当前可用的技术, 首先介绍了仅使用三元组结构信息进行表示学习的最新技术, 然后描述了融合多源信息的表示学习模型, 着重介绍了实体类型和文本信息两种多源信息。对于当前技术, 我们描述了总体框架, 具体的模型设计, 以及这些技术的优缺点。最后, 本文探讨了知识图谱表示学习技术的应用, 主要是介绍了两种应用, 即知识图谱补全和三元组分类任务。然而, 面向知识图谱的表示学习与实际应用还有很长的距离。未来的研究工作比如: 更好的融合多源信息、基于小样本的知识表示、如何进行动态知识表示等都亟待开展。

## References:

- [1] Hoffmann, Raphael, et al. "Knowledge-based weak supervision for information extraction of overlapping relations." Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011.
- [2] Bordes, Antoine, Jason Weston, and Nicolas Usunier. "Open question answering with weakly supervised embedding models." Joint European conference on machine learning and knowledge discovery in databases. Springer, Berlin, Heidelberg, 2014.
- [3] Damljanovic, Danica, and Kalina Bontcheva. "Named entity disambiguation using linked data." Proceedings of the 9th Extended Semantic Web Conference. 2012.
- [4] Bollacker, Kurt, et al. "Freebase: a collaboratively created graph database for structuring human knowledge." Proceedings of the 2008 ACM SIGMOD international conference on Management of data. AcM, 2008.
- [5] Lehmann, Jens, et al. "DBpedia—a large-scale, multilingual knowledge base extracted from Wikipedia." Semantic Web 6.2 (2015): 167-195.
- [6] Suchanek, Fabian M., Gjergji Kasneci, and Gerhard Weikum. "Yago: a core of semantic knowledge." *Proceedings of the 16th international conference on World Wide Web*. ACM, 2007.
- [7] Nickel, Maximilian, Volker Tresp, and Hans-Peter Kriegel. "A Three-Way Model for Collective Learning on Multi-Relational Data." ICML. Vol. 11. 2011.

- [8] Bordes, Antoine, et al. "Translating embeddings for modeling multi-relational data." *Advances in neural information processing systems*. 2013.
- [9] Wang, Zhen, et al. "Knowledge graph embedding by translating on hyperplanes." *Twenty-Eighth AAAI conference on artificial intelligence*. 2014.
- [10] Guo, Shu, et al. "Semantically smooth knowledge graph embedding." *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)*. Vol. 1. 2015.
- [11] Xie, Ruobing, Zhiyuan Liu, and Maosong Sun. "Representation Learning of Knowledge Graphs with Hierarchical Types." *IJCAI*. 2016.
- [12] Zhong, Huaping, et al. "Aligning knowledge and text embeddings by entity descriptions." *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. 2015.
- [13] Xie, Ruobing, et al. "Representation learning of knowledge graphs with entity descriptions." *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.
- [14] Wang, Quan, Bin Wang, and Li Guo. "Knowledge base completion using embeddings and rules." *Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015.
- [15] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." *Advances in neural information processing systems*. 2013.
- [16] Lin, Yankai, et al. "Learning entity and relation embeddings for knowledge graph completion." *Twenty-ninth AAAI conference on artificial intelligence*. 2015.
- [17] Ji, Guoliang, et al. "Knowledge graph embedding via dynamic mapping matrix." *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)*. Vol. 1. 2015.
- [18] Ji, Guoliang, et al. "Knowledge graph completion with adaptive sparse transfer matrix." *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.
- [19] Xiao, Han, et al. "TransA: An adaptive approach for knowledge graph embedding." *arXiv preprint arXiv:1509.05490* (2015).
- [20] Xiao, Han, Minlie Huang, and Xiaoyan Zhu. "TransG: A generative model for knowledge graph embedding." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vol. 1. 2016.
- [21] Yang, Bishan, et al. "Embedding entities and relations for learning and inference in knowledge bases." *arXiv preprint arXiv:1412.6575* (2014).
- [22] Nickel, Maximilian, Lorenzo Rosasco, and Tomaso Poggio. "Holographic embeddings of knowledge graphs." *Thirtieth Aaai conference on artificial intelligence*. 2016.
- [23] Trouillon, Théo, et al. "Complex embeddings for simple link prediction." *International Conference on Machine Learning*. 2016.
- [24] Liu, Hanxiao, Yuexin Wu, and Yiming Yang. "Analogical inference for multi-relational embeddings." *Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org*, 2017.
- [25] Balažević, Ivana, Carl Allen, and Timothy M. Hospedales. "TuckER: Tensor Factorization for Knowledge Graph Completion." *arXiv preprint arXiv:1901.09590* (2019).
- [26] Sun, Zhiqing, et al. "RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space." *arXiv preprint arXiv:1902.10197* (2019).
- [27] Socher, Richard, et al. "Reasoning with neural tensor networks for knowledge base completion." *Advances in neural information processing systems*. 2013.
- [28] Shi, Baoxu, and Tim Weninger. "ProjE: Embedding projection for knowledge graph completion." *Thirty-First AAAI Conference on Artificial Intelligence*. 2017.
- [29] Dettmers, Tim, et al. "Convolutional 2d knowledge graph embeddings." *Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.
- [30] Krompaß, Denis, Stephan Baier, and Volker Tresp. "Type-constrained representation learning in knowledge graphs." *International semantic web conference*. Springer, Cham, 2015.

- [31] Xie, Ruobing, Zhiyuan Liu, and Maosong Sun. "Representation Learning of Knowledge Graphs with Hierarchical Types." IJCAI. 2016.
- [32] Wang, Zhen, et al. "Knowledge graph and text jointly embedding." Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP). 2014.
- [33] Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.
- [34] Kristiadi, Agustinus, et al. "Incorporating literals into knowledge graph embeddings." arXiv preprint arXiv:1802.00934 (2018).
- [35] Shi, Baoxu, and Tim Weninger. "Open-world knowledge graph completion." Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [36] Sergieh, Hatem Mousselly, et al. "A multimodal translation-based approach for knowledge graph representation learning." Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics. 2018.
- [37] Ding, Boyang, et al. "Improving knowledge graph embedding using simple constraints." arXiv preprint arXiv:1805.02408 (2018).
- [38] Min, Bonan, et al. "Distant supervision for relation extraction with an incomplete knowledge base." Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2013.
- [39] Dong, Xin Luna, et al. "From data fusion to knowledge fusion." Proceedings of the VLDB Endowment 7.10 (2014): 881-892.

#### 附中文参考文献:

- [40] 官赛萍, 靳小龙, 贾岩涛, et al. 面向知识图谱的知识推理研究进展[J]. 软件学报, 2018, 29(10):74-102.
- [41] 刘知远, et al. "知识表示学习研究进展." 计算机研究与发展 53.2 (2016): 247-261.