

面向知识图谱的图嵌入学习研究进展*

杨东华^{1,2}, 何 涛¹, 王宏志¹, 王金宝¹



¹(哈尔滨工业大学 计算学部, 黑龙江 哈尔滨 150001)

²(哈尔滨工业大学 分析测试与计算中心, 黑龙江 哈尔滨 150001)

通讯作者: 王金宝, E-mail: wangjinbao@hit.edu.cn

摘要: 知识图谱是一种用网络结构存储知识的知识库,在知识图谱中,单条知识被表示成三元组的形式,即(头实体,关系,尾实体).得力于知识图谱在各个领域的广泛应用,面向知识图谱的图嵌入学习也得到越来越多研究人员的关注.面向知识图谱的图嵌入学习任务旨在为图谱中的实体与关系学习低维且稠密的向量,通过图嵌入向量表达实体与关系的语义信息以及度量实体之间、关系之间、实体与关系之间的语义联系,已有许多研究证明图嵌入模型在下游任务中的有效性.近年来,越来越多研究人员开始关注知识图谱的图嵌入学习,并取得大量的研究成果,本文尝试将图嵌入算法分成了基于转移思想、基于张量分解、基于传统深度学习模型、基于图神经网络以及融入额外信息的图嵌入学习共五大类,梳理、介绍各类图嵌入算法的设计思路、算法特征以及优缺点,以帮助指导初步接触该领域的研究人员快速学习了解该研究领域的相关模型和算法.

关键词: 知识图谱;图嵌入学习;表示学习;链接预测

中图法分类号: TP311

中文引用格式: 杨东华, 何涛, 王宏志, 王金宝. 面向知识图谱的图嵌入学习研究进展. 软件学报.

英文引用格式: Yang DH, He T, Wang HZ, Wang JB. Survey on knowledge graph embedding learning. Journal of Software,

Survey on Knowledge Graph Embedding Learning

YANG Dong-Hua^{1,2}, HE Tao¹, WANG Hong-Zhi¹, WANG Jin-Bao¹

¹(Department of Computing, Harbin Institute of Technology, Harbin 150001, China)

²(Center of Analysis, Measurement and Computing, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Knowledge Graphs (KGs) serve as a kind of knowledge base by storing facts with network structure, representing each piece of fact as a triple, i.e. (head, relation, tail). Thanks to the general applications of KGs in various of fields, the embedding learning of Knowledge Graph has also quickly gained massive attention. In this article, we try to classify the existing embedding algorithms as five types: translation-based models, tensor factorization-based models, traditional deep learning-based models, graph neural network-based models and models by fusing extra information. Then we introduce and analyze the key ideas, algorithm features, advantages and disadvantages of different embedding models to give the first-time researchers a guide article that can be referenced to help researchers quickly get started.

Key words: knowledge graph; graph embedding learning; representation learning; link prediction

1 引言

近年来,随着互联网、物联网等技术的不断发展,以及其他研究领域的需求发展,各类应用应运而生,也相应地产生了海量的数据资源.其中,图数据作为一种能够广泛建模诸多场景的数据类型,吸引了大量且深入的相关研究,如社交网络、蛋白质网络、知识图谱等领域.其中,知识图谱作为一种异质图网络,将知识表示成三元组的

* 基金项目: 国家自然科学基金(61772157, 61832003, U1866602)

收稿时间: 2021-05-13; 修改时间: 2021-06-29; 采用时间: 2021-07-29; jos 在线出版时间: 2021-10-20

形式(头实体,关系,尾实体),知识中的实体(即头实体与尾实体)表示成图的节点,知识中的关系表示成图的连边,利用网络化结构更加直观地表征与存储知识.

面向知识图谱的学习任务主要包含知识图谱的构建与维护,以及基于知识图谱的下游任务,如基于知识图谱的问答系统^[1]、基于知识图谱的信息抽取^[1]任务等.其中,无论是属于面向知识图谱的知识补全、实体对齐、关系抽取任务还是其他基于知识图谱的下游任务,都能够基于知识图谱的图嵌入表示解决问题.因此,伴随着知识图谱研究热度的上升,面向知识图谱的表示学习也得到越来越多的关注.

面向知识图谱的表示学习,又称为知识图谱的图嵌入学习,旨在将知识图谱的实体与关系映射到低维且稠密的向量或矩阵,相比于 one-hot 向量,图嵌入对存储要求更低,且更加能够反映实体、关系的语义信息,以及实体之间、关系之间、实体与关系之间的语义联系.除此以外,相较于传统的图学习算法,如基于规则学习的知识推理算法^{[3][4]}、基于本体推理的知识推理算法^{[5][6]},图嵌入学习具有较低的计算复杂性,并且得利于并行运算能力与框架的发展,图嵌入学习的计算效率得以不断提升.

相比于其他领域的图嵌入学习,例如面向无向图的图嵌入学习,面向社交网络、蛋白质结构等异质图的图嵌入学习任务,知识图谱因为自身的数据特点使得在图嵌入学习方面与其它类型的异质图有所差异.首先,常见的异质图包含的关系类别很少,如 DBLP 网络只包含论文-作者、论文-会议和论文-题目用词三种关系类别,IMDB 网络只包含电影-演员、电影-导演两种关系,极少的关系类别使得这类异质图学习更长、更丰富的上下文模式更加方便.基于元路径(meta-path)和元图(meta-graph)的图嵌入学习方法是两种异质图表示学习的常见路线^[7],元路径和元图可以视为需要专家人工制定的子图模式.而极少的关系类别也使得人工构建元路径和元图在效率上成为可能,这也使得异质图基于元路径和元图的图嵌入学习方法能够得到有效发展.而知识图谱往往具有很多关系类别,例如 FB15k-237 数据集^[8]包含 237 种关系,想要人工设计元路径和元图模式则不可避免地需要面临组合爆炸的问题;其次,对于常见的异质图,因为结构简单、实体类别较少,每个节点往往能够分配一个特定的类别,这种节点类别信息在异质图的图嵌入学习过程中能够起到不可忽略的作用,例如 HAN 模型^[9]在聚合信息前会首先将邻居信息通过类别相关的线性变换映射到节点类别空间中.而知识图谱实体标签过于复杂,没有显示为每个实体分配特定的类别,同时因为结构过于复杂,知识图谱的实体类别往往具有层次性,可能实体与实体的类别本身都存在图谱内,所以想要现实表达实体的类别信息并不容易.所以,实体类别信息往往作为一种隐式知识被利用在图嵌入学习任务中;除此以外,相比于常见的异质图,知识图谱的节点(即实体)往往具有较为丰富的文本信息,如实体名称、实体描述,这些文本信息能够配合知识图谱本身的图结构信息一起更好地学习实体的语义信息.综上所述,知识图谱虽然作为一种异质图,但因为自身的一系列特点,使得面向知识图谱的图嵌入学习在学习思路、可利用的信息等方面与其它类型的异质图有较大差异,丰富的文本信息和复杂多变的网络结构使得面向知识图谱的图嵌入学习增加了更多的挑战和研究空间.

为此,本文梳理了面向知识图谱的图嵌入学习算法的不同设计思想,并对相应方法进行了总结.具体来说,本文首先给出图嵌入学习的任务描述以及相关符号定义,然后按照设计思路、信息利用程度的不同将图嵌入学习方法分成如下五种类别:基于转移思想的图嵌入算法、基于矩阵/张量分解的图嵌入算法、基于传统深度学习的图嵌入算法、基于图神经网络的图嵌入算法以及融入额外信息的图嵌入算法;最后,本文总结了以上不同类别图嵌入的特征与优缺点,以及部分有待深入研究的问题.

相较于现有的研究进展报告^{[10][11]},本文主要通过分析之前模型的不足来梳理和介绍当前的研究进展,这样有利于从问题出发有的放矢地介绍不同领域方法的发展脉络和不同研究成果的出发点和优缺点,方便读者的理解;而且相较于[10][11]的工作,本文将基于 GNN 的图嵌入模型从基于深度学习模型的图嵌入算法中抽取出来说单独列为一类,一方面是为了结合了现阶段图神经网络模型的快速发展趋势,另一方面是考虑到基于 GNN 的图嵌入模型从编码思想上与其他几类方法有较大的差异,其主要利用 GNN 模型较为强大的图结构编码能力,因此有必要单独列为一类,所以本文在基于 GNN 模型的图嵌入方法上介绍也会较多;除此以外,本文还额外介绍了使用额外信息的一系列图嵌入算法,如基于文本信息、路径信息等,为面向知识图谱的图嵌入算法的后续研究介绍更多可思考和研究的方向;最后,本文列举了面向知识图谱的图嵌入学习的部分难点和有待解决的问

题.

2 面向知识图谱的图嵌入学习任务描述与符号定义

2.1 面向知识图谱的图嵌入学习任务定义

本节我们给出面向知识图谱的图嵌入学习的任务描述以及相关符号定义.给定一个知识图谱 $KG = (E, R, T)$, 其中 E 表示 KG 包含的实体集合, R 表示 KG 涉及的关系集合, T 表示 KG 中包含的三元组集合, 面向知识图谱的图嵌入学习, 简称为图嵌入学习任务, 旨在为每个实体 $e \in E$ 学习相应的低维表示向量 $v_e \in R^{d_e}$, 其中 d_e 表示实体嵌入的维度, 以及为每个关系 $r \in R$ 学习相应的低维表示向量 $v_r \in R^{d_r}$ 或者表示矩阵 $M_r \in R^{d_r \times d_r}$, 其中 d_r 表示关系表示的维度, 通常来说 $d_e = d_r = d$. 旨在通过学习实体和关系的表示向量或矩阵, 有效建模实体和关系的语义信息, 以及实体与实体之间、关系与关系之间的语义联系, 同时有利于下游任务对知识图谱的发掘和应用.

需要注意的是, 图嵌入只是学习结果, 而为了学习图嵌入, 通常需要经过特定的学习任务, 如链接预测任务(即知识补全任务)、实体对齐任务等. 以链接预测任务为例, 该任务要求给定三元组中的两个元素, 去预测第三个元素(如知道头实体和关系, 去预测正确的尾实体), 形式化定义就是给定查询 $(h, r, ?)$ 或者 $(?, r, t)$, 分别预测正确的尾实体和头实体集合, 其中 h 、 r 、 t 分别代表头实体、关系和尾实体. 通常通过学习图嵌入来计算三元组的得分函数, 通过梯度反向传播算法在最大化正确三元组的得分函数的同时也学习了图嵌入表示. 相对于其它任务, 链接预测任务需要的条件较少且研究意义深远, 因此研究人员通常使用链接预测任务来判断图嵌入学习算法的有效性. 下文中除特殊说明外, 默认图嵌入学习的目标任务为面向知识图谱的链接预测任务.

链接预测的评估指标通常采用 MRR、MR 和 Hits@k 这三种排序指标,之所以采用排序指标,是因为链接预测任务通常将学习任务视为排序任务,在预测前,需要为每个候选实体进行打分,并按照打分情况对候选实体进行排序. 对于评估样本 (h, r, t) , 假设其正确的标签集合为 $S = \{e | (h, r, e)\}$ 成立. 为了计算上述指标, 需要统计当前评估样本 (h, r, t) 的正确标签 t 在候选实体中的排序值 $rank$. 在统计排序值时, 有原始排序值和过滤排序值两种统计方式. 原始排序值即直接统计 t 在候选实体中的排序值作为最终 $rank$, 但是考虑到正确的标签集合 S 不仅仅包含标签 t , 还包括其它实体, 如果模型只统计 t 在候选实体中的排序值, 而忽略了排在 t 之前的所有正确标签, 则计算出的排序值偏大, 因此过滤排序值将所有排在 t 之前的所有正确标签过滤掉, 再计算 t 的排序值作为最终的 $rank$. 通常来说给出的指标默认是指过滤后的实验指标. 基于以上统计量, MR 的计算方式为:

$$MR = \frac{1}{N} \sum_{i=1}^N rank_i$$

其中 N 是评估样本的个数, $rank_i$ 表示第 i 个样本中正确标签的排序值. 相应的 MRR 的计算方式为:

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{rank_i}$$

Hits@k 指标统计了所有评估样本中正确标签排序在前 k 个的比例, 计算方式为:

$$Hits @ k = \frac{1}{N} \sum_{i=1}^N I(rank_i \leq k)$$

其中 $I(x)$ 为示性函数, 当参数 x 为真时该函数取值为 1, 否则为 0.

2.2 本文使用的部分符号定义与说明

为了方便下文介绍, 表 1 定义一些本文常用的符号, 并给出了相应的解释说明.

3 面向知识图谱的图嵌入学习模型研究框架

下面我们将具体介绍不同类型图嵌入学习算法的基本思想与代表模型. 我们按照图嵌入学习核心思想、基础模型使用以及信息利用程度的不同, 本文将面向知识图谱的图嵌入算法分成了: 基于转移思路、基于张量分

解、基于传统深度学习、基于图神经网络以及融入额外信息的图嵌入学习算法.其中前三种算法从距离度量、语义相似度和特征抽取三个角度分别考察三元组的评估策略.前三类方法的学习对象都是单个三元组,而基于GNN的模型建立在前三类方法的基础之上,首先利用GNN模型对图谱的全局结构进行编码,然后采用前三类方法作为解码器做链接预测任务.因此基于GNN的图嵌入学习方法学习到的图结构信息理论上更加全面有效,这是从图结构信息的挖掘深度角度进行分类.最后,前四类算法都只是涉及到知识图谱的结构信息,而知识图谱作为一种特别的异质图,其实体与关系都可能关联了名称、描述、类别等信息,如何在学习图嵌入的同时考虑这些信息是第五类算法的关键思想.下文将按照上述分类的顺序依次介绍各类算法的思想以及相应的算法实现.

表1 本文使用的符号

符号定义	解释
G	表示知识图谱
h, r, t	分别表示头实体、关系与尾实体
v_e, v_r, M_r	分别表示实体 e 、关系 r 的表示向量,以及关系 r 的表示矩阵
d_e, d_r, d	分别指代实体表示与关系表示的维度,如果两者维度相同,则用 d 统一表示
V, M	V 表示全体实体的表示向量组成的表示矩阵, V_i 表示第 i 个实体的嵌入表示; M 表示全体关系的表示向量(矩阵)组成的嵌入矩阵(张量), M_k 表示第 k 个实体的嵌入表示
E, R, R_{inv}	分别表示知识图谱 G 的实体集合、关系集合与逆关系集合,因而 $ E , R $ 分别表示实体集合 E 与关系集合 R 的大小
e_i, r_j	分别表示实体集合 E 的第 i 个实体与关系集合 R 的第 j 个关系
$\phi(h, r, t)$	表示三元组 (h, r, t) 的得分函数
$r(h, t)$	三元组 (h, r, t) 另一种表达方式:该值为 1 表示三元组成立;为 0 则说明该三元组不存在

3.1 基于转移思想的图嵌入学习

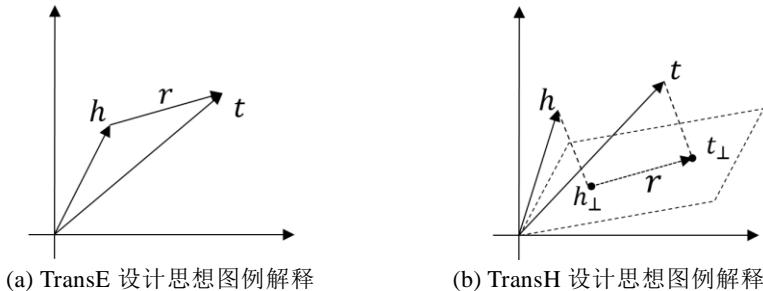


图1 TransE与TransH的核心思想示意图

首先介绍基于转移思想(translation-based)的图嵌入学习方法,我们将从基本思想入手开始介绍奠基性工作,然后从该工作的问题角度分别介绍后续对该工作的改进与优化,依次理清该类模型的发展脉络.

基于转移思想的图嵌入算法最早由Bordes^[12]等人提出,该工作提出了第一个基于转移的表示学习模型TransE. TransE 算法认为如果三元组(头实体,关系,尾实体)正确,则头实体 h 、关系 r 与尾实体 t 的表示向量 v_h 、 v_r 和 v_t 需满足关系 $v_h + v_r \approx v_t$,即头实体在经过关系的转移后需要接近尾实体.TransE 的核心思想如图 1(a)所示,

头、尾实体均被建模成表示空间中的一个点,而关系建模成表示空间中的向量位移, $v_h + v_r \approx v_t$, 表示头实体 h 经过关系 r 的位移后在表示空间中接近尾实体 t , 例如 $v_{\text{哈尔滨}} + v_{\text{方言}} \approx v_{\text{东北话}}$, $v_{\text{广东}} + v_{\text{方言}} \approx v_{\text{粤语}}$. 因此, 对于三元组 (h, r, t) , TransE 设计的优化目标函数为 $f(h, r, t) = \|v_h + v_r - v_t\|_p$, 当三元组正确时最小化该目标函数, 反之则最大化该函数.

需要提及的是, 虽然于 TransE 模型提出之前 Bordes 等人也提出了 Unstructured Model(UM)^[13] 和 Structured Embedding(SE)^[14] 两个模型, 其中 UM 模型的目标函数设计为 $\|v_h - v_t\|_2$ (因此 UM 模型可以视为 TransE 约束关系为零向量时的特例), 而 SE 模型的目标函数设计为 $\|M_{h,r}v_h - M_{t,r}v_t\|_2$, 从目标函数的设计不难看出关系的重要性并不明显, 没有起到对头实体表示向量进行转移操作的作用, 因此本文还是认为最早提出转移思想的图嵌入算法的工作由 TransE 最早提出. TransE 建模简单但效果颇佳, 并且相较于传统的知识推理方法, 学习效率高, 能够很快部署在大型知识图谱上, 因此很快引发了 Trans 系列模型的研究热潮. 尽管 TransE 算法表现不错, 学习效率高, 但依然存在一些不足有待优化:

- 首先, 在处理多对一、一对多以及多对多关系(统称为多映射关系)时, 由于简单约束 $v_h + v_r \approx v_t$ 的作用, 会使得不同实体的分布式表示在表示空间中聚集在一起, 即使这些实体具有不同的语义, 进而造成实体语义的混乱. 例如通过简单约束 $v_{\text{哈尔滨}} + v_{\text{位于}} \approx v_{\text{中国}}$ 与 $v_{\text{北京}} + v_{\text{位于}} \approx v_{\text{中国}}$, 使得哈尔滨与北京的语义发生混淆. 为了解决该问题, TransH^[15] 引入了关系超平面的概念, 在转移头实体之前, 先将头尾实体分别投影到关系所在的超平面上:

$$v_{h,r} = v_h - w_r^T v_h w_r, \quad v_{t,r} = v_t - w_r^T v_t w_r$$

其中 w_r 表示关系 r 相关的超平面的法向量. 在超平面上计算头实体经过关系转移后与尾实体的距离: $\|v_{h,r} + v_r - v_{t,r}\|_p$. 这使得相同的实体在不同的关系下具有不同的表示, 在一定程度上缓减了 TransE 在多映射关系上的弱表现问题, 该算法的设计思想参考图 1(b) 所示. 不同于 TransH 的超平面假设, TransR^[16] 假设每个关系都关联了一个表示空间, 在转移操作前, TransR 通过线性变换将实体表示映射到相应的关系空间中, 然后执行转移操作, 具体来说, TransR 定义了关系关联的映射矩阵 M_r , 在执行转移操作之前, 分别通过 $v_{h,r} = M_r v_h$ 和 $v_{t,r} = M_r v_t$ 获得关系空间内的头实体与尾实体表示, 最后通过计算 $\|v_{h,r} + v_r - v_{t,r}\|_p$ 得到优化目标. TransD^[17] 继承了 TransR 的思路, 但该算法观察到 TransR 为不同的关系类别 r 关联不同的映射矩阵 M_r , 容易造成参数量过大的问题, 因此对 M_r 进行了分解表示. TransD 认为实体空间到关系空间的映射矩阵不应只受限于关系, 还与实体相关. 为了便于构造同时由实体与关系决定的映射矩阵, TransD 为每个实体(关系)引入了两个向量, 其中一个是实体(关系)的真实语义表示 $v_e(v_r)$, 另一个向量 $v_{ep}(v_{rp})$ 用于构建映射矩阵, 对于三元组 (h, r, t) , 通过 $M_{rh} = v_{rp} v_{hp}^T + I$ 与 $M_{rt} = v_{rp} v_{tp}^T + I$ 分别构建头实体与尾实体的映射矩阵. TransR 与 TransD 算法的设计思想可以参见图 2 所示. TranSparse^[18] 同样继承了 TransR 的思想, 为了缓解 TransR 参数量过大的问题, TranSparse 算法考虑使用稀疏矩阵建模 M_r . 同时该算法注意到语义越复杂的关系, 往往需要用更多的参数量来学习, 因此该算法为每个关系类别自适应地学习稀疏度:

$$\theta_r^l = 1 - \frac{1}{N_{r^*}^l} (1 - \theta_{min}) N_r^l \quad (l = h, t)$$

其中 N_r^l 表示关系 r 连接的实体对数目, $N_{r^*}^l$ 表示所有关系中连接实体对数目的最大值. 这里 TranSparse

算法假设连接的实体对数目越多则该关系的语义越复杂. 计算完每个关系的稀疏度 θ_r^l 后, 基于该稀疏

度生成稀疏矩阵 $M_r^l(\theta_r^l)$, 对于头实体和尾实体分别生成映射矩阵 $M_r^h(\theta_r^h)$ 和 $M_r^t(\theta_r^t)$, 最后仿照 TransR 模型计算距离函数:

$$\| M_r^h(\theta_r^t)v_h + v_r - M_r^t(\theta_r^t)v_t \|_p$$

STransE^[19]为每个关系 r 关联了两个独立的矩阵 W_r^h 、 W_r^t , 在转移操作前, 使用 W_r^h 将头实体映射到关系空间, 同时用 W_r^t 将尾实体映射到关系空间。除此以外, CrossE^[20]采用类似 TransE 的思想, 首先计算头实体与关系的合并, 之后希望合并后的向量与尾实体尽可能相似, 为了交互头实体与关系的信息, 引入了交互矩阵 C_r , 使得实体以及关系的表示能够根据不同的三元组自主选择有用的信息, 进而实现图嵌入的动态变化。以上工作都是从表示空间角度来考虑解决多映射关系的问题, TransM^[21]从三元组的重要性角度提出新的思路, 该算法考虑了不同属性关系映射对三元组的贡献程度不同, 认为包含一对一映射属性关系的三元组更重要, 因而赋予相应三元组的损失函数以更高的权重。TransF^[22]则将方向作为度量三元组正确性的标准, 具体来说, 该算法不要求 $v_h + v_r \approx v_t$, 而是只要求 v_t 与 $v_h + v_r$ 在方向上保持一致, 同理, 也要求 v_h 与 $v_t + v_r$ 保持方向一致, 为此该算法最终的得分函数为:

$$(v_h + v_r)^T + (v_t - v_r)^T v_h$$

ManifoldE 算法^[23]观察到对于 1-N 和 N-N 类型的关系, 之所以会造成语义混乱的现象, 主要由于将尾实体集合建模成空间中的一个点, 使得所有正确尾实体在表示空间都簇拥在一起。因此该算法从流形角度出发, 认为在学习链接预测任务 $(h, r, ?)$ 的过程中, 正确的尾实体集合不应该由单个点表示, 而是用流形来建模。例如分布在中心为 $v_h + v_r$, 半径为 D_r 的球面, 这样对于相同的头实体与关系, 球面上的所有点都可以代表正确尾实体的嵌入表示。

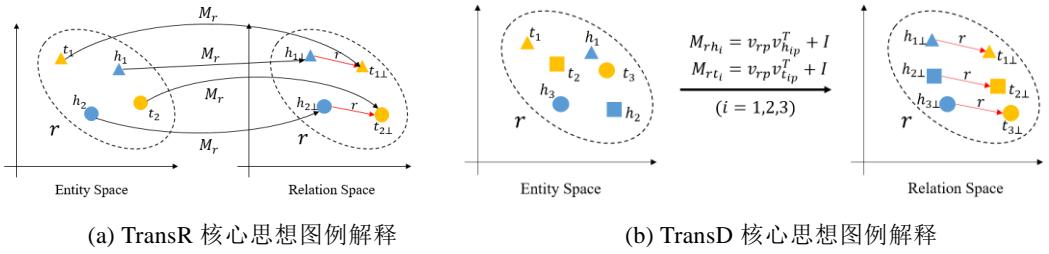


图 2 TransR 与 TransD 的核心思想图例解释

- TransE 在处理对称关系时也具有天然的劣势, 对于对称类关系, 由于三元组(头实体, 关系, 尾实体)与(尾实体, 关系, 头实体)同时成立, 使得这类关系的分布式表示均接近于零向量, 因此失去了语义辨识的能力。为了解决该问题, Sun 等人^[24]提出 RotatE, 利用复数空间中的旋转操作代替欧式空间的加和操作, 考虑到两次旋转 $k\pi$ 弧度总能回到原来的位置, 因此 RotatE 能够较好地学习并区分对称关系的性质。MDE^[25]则针对对称类型关系将优化的损失函数修改为 $\|v_h + v_r - v_r\|$, 同时, 该模型还面向非对称、自反、组合关系针对性地提出了四类损失函数, 旨在利用不同的损失函数学习不同的关系类型。除此以外, ATTH^[26]也利用反射操作一定程度上缓减了对称关系的学习问题, 不过另一方面, 反射操作可以视为旋转操作在旋转角度为 π 时的特殊情况。类似于 RotatE, QuatE^[27]同样采用旋转的思想, 但是 QuatE 没有使用复数空间, 而是采用了四元数向量来表示实体与关系, 利用四元数的内在属性, QuatE 能够自然建模对称与非对称关系。
- 除此以外, TransE 在建模关系的组合模式能力上还略显不足。所谓关系的组合模式, 是指对于一组关系 r_1, r_2, \dots, r_n 和 r , 使得对于任意实体序列 e_1, e_2, \dots, e_{n+1} , 如果满足 $r_1(e_1, e_2) \wedge r_2(e_2, e_3) \wedge \dots \wedge r_n(e_n, e_{n+1})$, 则 $r(e_1, e_{n+1})$ 一定会成立, 那么称关系 r 能够由关系序列 r_1, r_2, \dots, r_n 组合而成, 而 $r_1(e_1, e_2) \wedge r_2(e_2, e_3) \wedge \dots \wedge r_n(e_n, e_{n+1})$

$e_{n+1}) \Rightarrow r(e_1, e_{n+1})$ 则被称为关系的组合模式. TransE 能够捕获简单的组合模式, 假设对于 $\forall x, y, z$, 存在组合模式 $r_1(x, y) \wedge r_2(y, z) \Rightarrow r(x, z)$, 那么在图嵌入的学习过程中, 根据约束 $v_x + v_{r_1} \approx v_y$, $v_y + v_{r_2} \approx v_z$ 和 $v_x + v_r \approx v_z$, 自然能够学习到等式关系 $v_{r_1} + v_{r_2} \approx v_r$. 但 TransE 在学习关系的组合模式时一方面没有考虑关系序列的有序性, 另一方面也没有考虑实体在组合模式中扮演的角色, 虽然 RotatE 将 TransE 的加和操作修改为旋转操作, 但依然没有解决这两个问题. 为此, Dense^[28] 提出基于非阿贝尔群的图嵌入学习方法, 该工作将每个关系分解为基于非阿贝尔群的旋转操作和 3 维空间中的缩放操作, 有效解决了 TransE 和 RotatE 模型存在的问题, 并在 WN18RR^[29] 公开数据集上取得了不错的提升.

- 其次, TransE 没有显式建模实体的语义层级信息, 已有工作^[31]证明实体层次信息有助于处理长尾实体的问题. 为此, Hu 等人^[30] 提出了实体的层级表示, 将实体作为叶节点, 上层概念作为内部节点, 构成有向无环图. Zhang 等人^[31] 则提出了 HAKE, 从模长和弧度两个角度表示图嵌入, 模长用于建模不同层级实体之间的联系, 而弧度用于度量相同语义层级实体之间的转移关系. 上述工作都是在欧式空间内学习分布式表示, 而 MuRP^[32] 与 ATTH^[26] 直接从表示空间的类别选择入手, 借助双曲空间本身内在的层次性, 在双曲空间中学习图嵌入, 有效建模了知识图谱的层次性质.

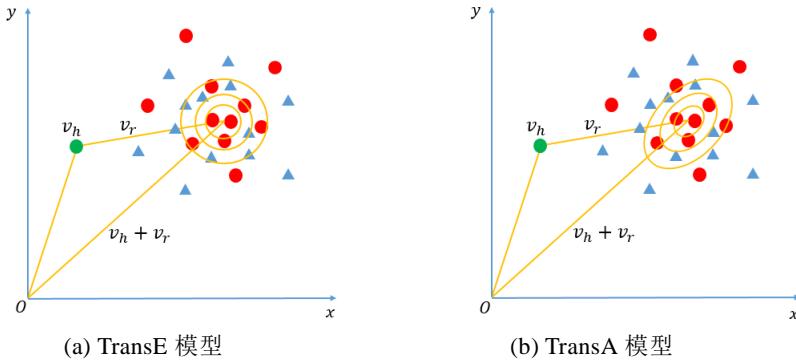


Fig 3 An illustration of equipotential hyper-surfaces for TransE and TransA

图 3 TransE 模型的超球面图例解释与 TransA 模型的超椭球面图例解释

- TransE 模型将每个实体建模成低维空间中的一个确定的点, 将每个关系建模成一个确定的矢量, 这意味着模型被希望能够学习到实体和关系的确定语义, 然而语义的学习还受数据分布、实体与关系本身的含义影响. 例如对于出现频率很低的实体与关系(又称为长尾实体与长尾关系), 由于能够刻画其语义信息的三元组较少, 想要充分学习其语义信息较为困难; 除此以外, 很多实体与关系在不同上下文中(即不同三元组中)也往往具有不同但近似的含义, 仅仅使用表示空间中的点来表示实体和关系显然是欠合理的. 为此, He 等人提出了 KG2E 模型^[33], 用相互独立的高斯分布表示每个实体和关系类别. 该工作将实体(关系)建模成高斯分布 $N(\mu, \Sigma)$ 的形式, 其中 μ 用来建模实体(关系)的中心位置, 而方差 Σ 度量了该实体(关系)的语义不确定程度, 直观上来看, 如果方差越大, 则表明该实体(关系)的语义越不明确. 为了学习高斯嵌入, 给定三元组 (h, r, t) , 该工作基于 TransE 假设, 建模出 $t - h$ 的高斯分布形式, 然后计算 $t - h$ 与关系 r 代表的高斯分布之间的距离(论文中使用了 KL 散度), 对于正确的三元组, 最小化分布之间的距离, 反之则最大化分布之间的距离.
- 除此以外, Xiao 等人^[34] 观察到 TransE 及其后续工作(如 TransH、TransD、TransR 等)都考虑计算 $h + r$ 与尾实体表示 t 的欧式距离. 以 TransE 为代表, 其目标函数为 $\|v_h + v_r - v_t\|_p$, 在为相同的查询预测时, 模型根据所有实体与 $v_h + v_r$ 的欧式距离对候选实体进行排序, 因此其等势面为超球面, 如图 1(a) 所示. 图中以二维空间为例, 红色圆圈表示正确尾实体, 蓝色三角形代表错误尾实体, TransE 通过 $v_h + v_r$ 或者预测尾实体的空间位置, 然后根据欧式距离找出距离 $v_h + v_r$ 最近的一系列实体作为尾实体的预测, 而所有距离 $v_h + v_r$ 相同大小的实体都位于半径相等的圆环上. Xiao 等人发现在计算向量距离时, 不同维度的重要性

并不相同,而 TransE 计算的欧氏距离将所有维度视为同等重要,为此他们提出了 TransA 模型^[34],该模型基于马氏距离的距离度量方法: $(v_h + v_r - v_t)^T W_r (v_h + v_r - v_t)$,其中 W_r 是关系 r 对应的对称半正定矩阵。基于马氏距离,TransA 的等势面不再是超球面,而是超椭球面,对应于图 1(a)的例子,在图 1(b)展示了相同情况下 TransA 模型可能学习到的等势面,不难看出,通过调整等势面,TransA 可以更好地避开错误实体。

基于转移思想的图嵌入算法由于运算操作简单,参数量适中,学习效率高等优势,成为了图嵌入领域非常活跃的研究方向之一。然而由于 TransE 本身存在的诸多问题,使得后续研究需要从表示空间、运算类型等方面对 TransE 进一步改进。表 2 总结了部分 Trans 系列算法设计的得分函数。

表 2 部分 Trans 系列模型的得分函数设计

模型	得分函数
TransE ^[12]	$-\ v_h + v_r - v_t\ _p$
TransH ^[15]	$-\ (v_h - W_r^T v_h w_r) + v_r - (v_t - W_r^T v_t w_r)\ _2$
TransR ^[16]	$-\ M_r v_h + v_r - M_r v_t\ _2$
TransD ^[17]	$-\ (v_{rp} v_{hp}^T + I)v_h + v_r - (v_{tp} v_{tp}^T + I)v_t\ _2$
STransE ^[19]	$-\ M_{rh} v_h + v_r - M_{rt} v_t\ _2$
CrossE ^[20]	$\sigma(\tanh(c_r \circ v_h + c_r \circ v_h \circ v_r + b) v_t^T)$
TransF ^[22]	$(v_h + v_r)^T v_t + (v_t - v_r)^T v_h$
RotatE ^[24]	$-\ v_h \circ v_r - v_t\ _2 \quad v_h, v_r, v_t \in C^d \text{ and } \ v_r\ =1$
HAKE ^[31]	$-\ v_{hm} \circ v_{rm} - v_{tm}\ _2 - \lambda \ \sin(\frac{v_{hp} + v_{rp} - v_{tp}}{2})\ \quad v_h, v_r, v_t \in [0, 2\pi]^d$
MuRP ^[32]	$-d_B(\exp_0^c(M_r \log_0^c(v_h)), v_t \oplus_c v_r)^2 + b_h + b_t$
QuatE ^[27]	$Q_h \otimes \frac{W_r}{ W_r } \cdot Q_t \quad e.t. \quad Q_h, W_r, Q_t \in H^d$
KG2E ^[33]	$-\frac{1}{2} \left\{ \text{tr}(\Sigma_r^{-1} (\Sigma_h + \Sigma_t)) + \mu^T \Sigma_r^{-1} \mu - \log \frac{\det(\Sigma_h + \Sigma_t)}{\det(\Sigma_r)} \right\} \quad \mu = \mu_h - \mu_t - \mu_r$
TransA ^[34]	$-(\ v_h + v_r - v_t\)^T W_r (\ v_h + v_r - v_t\) \quad W_r \text{ 是半正定矩阵}$

3.2 基于张量分解的图嵌入学习

本节我们介绍基于张量分解的图嵌入算法,同样在介绍完代表性算法后,从分析以及解决代表性算法的缺点入手介绍后续的相关工作。

基于张量分解的图嵌入方法,又称为基于语义匹配的图嵌入学习方法.其基本思想认为知识图谱能够被三维张量 $T \in \{0,1\}^{|E|*|R|*|E|}$ 建模,其中元素 $T_{ijk} = 1$ 表明三元组 (e_i, r_j, e_k) 正确,相反 $T_{ijk} = 0$ 则表明三元组 (e_i, r_j, e_k) 不正确,通过张量分解方法将 T 分解为实体矩阵 $V \in R^{d*|E|}$ 、关系张量 $M \in R^{d*|E|*d}$ 以及实体矩阵的转置 V^T 的矩阵乘积形式: $T = V \cdot M \cdot V^T$, 其中 V_i 表示第 i 个实体的表示向量, M_j 表示第 j 个关系的表示矩阵.

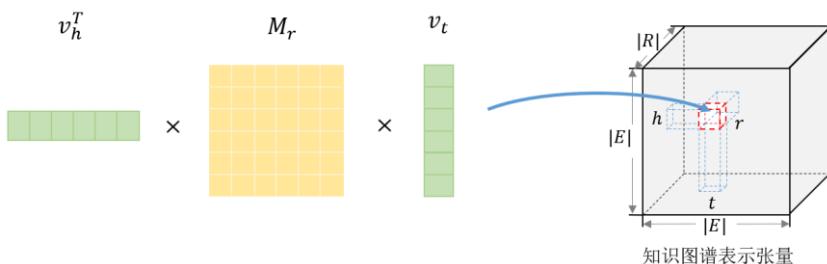
基于张量分解的方法最早由 RESCAL^[35] 提出,对于正确的三元组 (h, r, t) , 该工作通过优化头实体向量 v_h 、关系矩阵 M_r 与尾实体向量 v_t 的乘积 $v_h^T M_r v_t$ 应满足: $v_h^T M_r v_t = T_{hrt}$, 因此该算法的损失函数定义为:

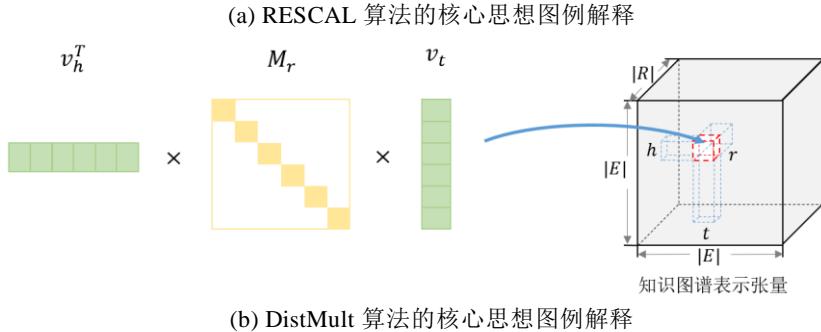
$$\frac{1}{2} \sum_{ijk} (T_{ijk} - v_h^T M_r v_t)^2$$

由于矩阵 T 由图谱知识本身决定,因此最小化上式在优化效果上与直接最大化 $v_h^T M_r v_t$ 类似,而 $v_h^T M_r v_t$ 可以视为计算头实体、关系与尾实体的语义相似度,因此基于张量分解思想与基于语义匹配的学习思路其实是同一种学习思想的不同看待角度.理论上可以证明 RESCAL 算法能够完全建模知识图谱蕴含的事实,但是 RESCAL 算法的缺陷在于其为每个关系都设置了相关矩阵,造成参数量过大,不论是防止过拟合还是为减少对硬件的要求都存在待优化的余地.

为了解决参数问题,如图 4(b) 所示,DistMult^[36] 约束关系矩阵为对角矩阵 $M_r = diag(v_r)$,使得关系矩阵的参数量由 $O(d * |R| * d)$ 减少为 $O(d * |R|)$, 学习效率大幅提升.同时,DistMult 直接优化 $v_h^T M_r v_t$,使得正确的三元组该乘积更大,错误的三元组乘积值更小.但 DistMult 的缺陷也很明显,对角矩阵的引入使得所有关系都被建模成对称关系,因为如果三元组 (h, r, t) 成立,考虑到 $\phi(h, r, t) = v_h^T M_r v_t = v_t^T M_r v_h = \phi(t, r, h)$, 三元组 (t, r, h) 同样成立,这后者往往并不成立.尽管如此,已有工作表明[37]经过改进后 DistMult 仍然能够取得不错的效果.

为了解决 DistMult 无法处理非对称关系的问题,ComplEx^[38] 算法尝试在矩阵乘法中引入复数运算.具体来说,ComplEx 继承了 DistMult 约束关系矩阵为对角矩阵的思想,同时设置矩阵元素类型为复数,即 $V \in C^{|E|*d}$ 、 $M \in C^{|E|*|R|*d}$, 且关系 r 的表示 $M_r = diag(v_r)$, 最终该算法选择的得分函数为 $\phi(h, r, t) = \text{Re}(\langle v_r, v_h, v_t \rangle) \cdot \overline{\text{Hole}}^{[39]}$ 从向量运算的角度出发,对于三元组 (h, r, t) ,首先计算头实体与尾实体的循环相关值 $v_h \star v_t$,然后计算 $v_h \star v_t$ 与关系表示 v_r 的乘积得到最终的得分函数: $\phi(i, j, k) = v_r^T (v_h \star v_t)$.作者在其工作中指出循环相关操作可以视为矩阵乘积的压缩形式,使得 Hole 算法在学习效率和空间利用上更有优势.SimplE^[40] 算法通过为实体、关系引入两个分离的分布式表示来解决 DistMult 无法建模非对称关系的问题:(1)为每个实体 e_i 关联两个不相关的嵌入表示 $v_{i,h}$ 、 $v_{i,t}$,分别表示实体 e_i 充当头实体与尾实体时的向量表示;(2)同时为每个关系 r_i 关联两个嵌入表示 M_i 和 M_{-i} ,分别表示关系 r_i 的正向以及其反方向关系 r_{-i} 的嵌入表示.因为约束关系矩阵为对角矩阵,SimplE 相比于 RESCAL 算法拥有更少的参数量,同时 SimplE 也具有充分表达图谱知识的能力.





基于 RESCAL 的张量分解思想,Liu 等人^[41]希望在图嵌入学习的过程中引入类比推理,提出了 Analogy 算法.通过对平行四边形范式的研究以及针对维持关系矩阵族的要求,该工作在 RESCAL 算法的基础上引入了另外两个约束:(1)任意关系对应的关系矩阵需符合正规矩阵;(2)任意两个关系 r_i, r_j 对应的关系矩阵均满足等式 $M_i \cdot M_j = M_j \cdot M_i$. 同时,该工作还通过关系矩阵的不同设计统一了 DistMult、ComplEx 与 HolE 算法,具有较强的理论意义.

与 RESCAL 算法不同,TuckER^[42]并没有将知识图谱张量 T 分解为实体矩阵与关系张量的乘积形式,而是采用 Tucker 分解算法,将张量 T 分解为实体矩阵、关系矩阵、实体矩阵转置以及核张量 W 的乘积形式,对于正确的三元组 (h, r, t) 最大化得分函数 $\phi(h, r, t) = W \times_1 v_h \times_2 v_r \times_3 v_t$,其中核张量 W 是共享参数,如图 5 所示.与之前工作不同,由于 Tucker 分解的性质,TuckER 算法中实体嵌入与关系嵌入的维度可以是不同的.最后的实验结果表明该算法也能够取得了相当不错的效果.

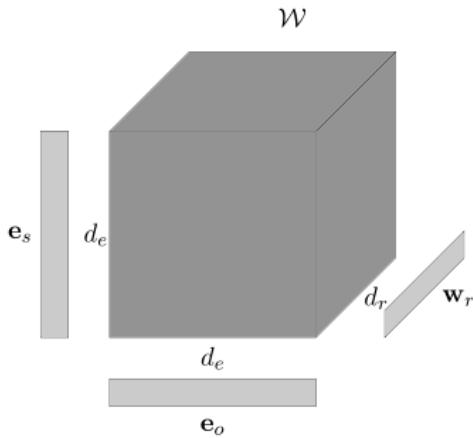


图 5 TuckER 算法示意图^[42]

之前的工作,如 ComplEx、DistMult 算法等,在计算得分函数时使用向量内积方式,该乘积方式使得实体嵌入表示与关系嵌入表示在相同维度上进行作用,以 DistMult 算法为例,其得分函数为 $v_h^T \cdot M_r \cdot v_t$,考虑到关系矩阵 M_r 是对角阵,所以 DistMult 算法的得分函数等价于向量 v_h, v_r, v_t 对位相乘后求和,其中 v_r 满足 $M_r = \text{diag}(v_r)$.而 SEEK^[43]提出除了让嵌入的相同维度元素作用,不同维度元素之间的相互作用有利于增强信息的交互融合.为此,该工作将实体表示与关系表示分成相同数量的片段,表示向量之间不同位置的片段同样可以相互作用,并设计了能够同时建模对称与反对称的向量乘积函数 f .除此以外,该工作还证明了通过设计不同的乘积函数 f ,SEEK 能够统一 DistMult 和 ComplEx 等算法.

相比于 Trans 系列的图嵌入算法,基于张量分解的思路因为自设计之初就是从建模整张图谱的表示张量 T

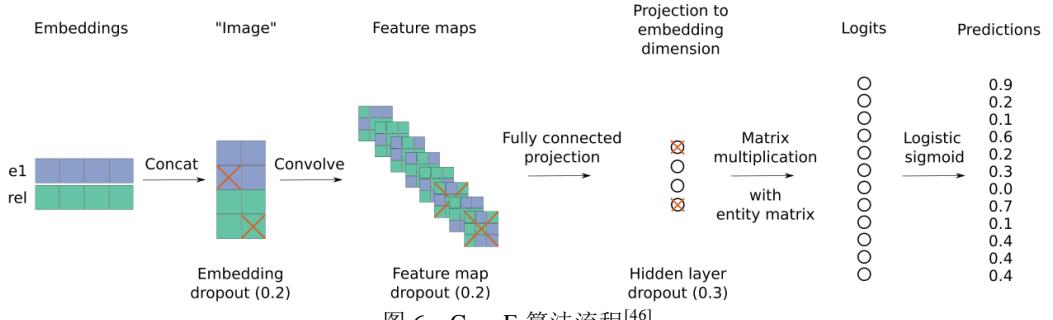
的角度来考虑,所以该类型的许多算法拥有充分表达知识图谱的能力^{[35][40][42]},然而这类模型的设计往往需要较高的数学基础,因此相关研究工作少于Trans系列模型的研究.表3列出了部分基于张量分解算法的得分函数设计.

表3 部分基于张量分解的图嵌入算法的得分函数设计

模型	得分函数
RESCAL ^[35]	$v_h^T M_r v_t$
DistMult ^[36]	$v_h^T \text{diag}(v_r) v_t$
ComplEx ^[38]	$\text{Re}(< v_r, v_h, \bar{v}_t >), \quad v_r, v_h, v_t \in C^d$
HolE ^[39]	$v_r^T (v_h \star v_t)$
Analogy ^[41]	$v_s^T W_r v_o, \quad W_r W_r^T = W_r^T W_r$
SimpleE ^[40]	$v_h \circ v_r v_t + v_t \circ v_{-r} v_h$
TuckER ^[42]	$\phi(h, r, t) = \mathbf{W} \times_1 v_h \times_2 v_r \times_3 v_t$

3.3 基于传统深度学习模型的图嵌入学习

下面我们将简单介绍基于传统深度学习的图嵌入学习方法.基于传统深度学习的方法通过使用或改进现有的的一些深度学习模型来学习三元组的特征,按照使用的深度学习模型类别,大致可以分为基于卷积神经网络^[44]、基于循环神经网络以及基于胶囊网络^[45]的三类模型.

图6 ConvE 算法流程^[46]

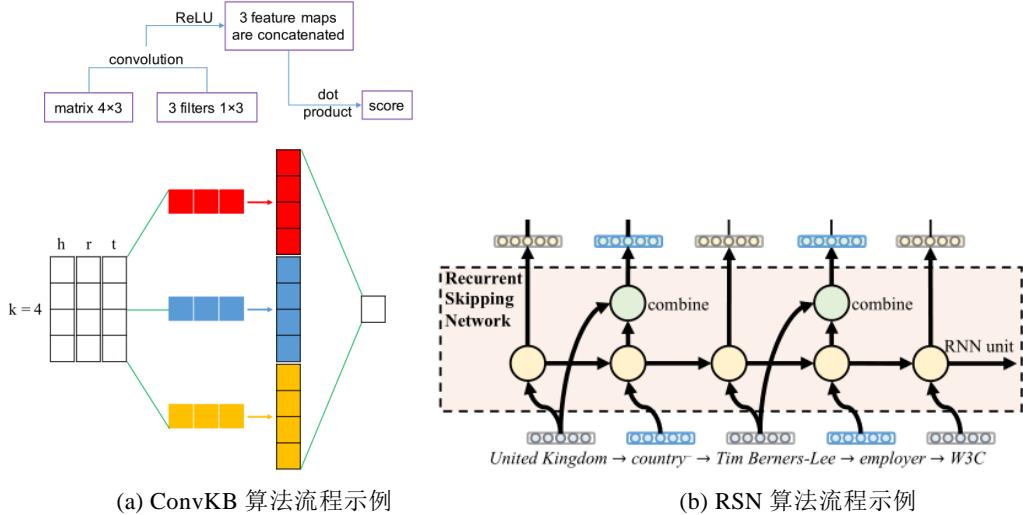


图 7 ConvKB^[47]和 RSN^[51]算法流程

基于卷积神经网络(CNN)的图嵌入模型利用卷积神经网络对三元组或三元组部分元素进行编码,从而实现对三元组进行打分的目的。如图 6 所示,ConvE^[46]对三元组(h , r , t)中的头实体 h 与关系 r 进行拼接,得到矩阵 $[v_h; v_r] \in R^{d^*2}$ 后对该矩阵的形状重新编辑;然后使用卷积神经网络对其进行编码,得到的编码结果经过全连接层输出最终的编码表示 v_{hr} ,计算 v_{hr} 与尾实体表示 v_t 的相似度,实现对三元组进行打分。与 ConvE 不同的是,ConvKB^[47]直接利用卷积神经网络对三元组进行编码:以图 7 为例,对于三元组(h , r , t),ConvKB 首先将其拼接为 $[v_h; v_r; v_t] \in R^{4^*3}$ 的形式,其中 4 表示初试图嵌入表示的维度;然后利用卷积神经网络对其进行编码,使用 3 个 1×3 的卷积核对其进行卷积操作,每个卷积核能够得到 4×1 的卷积结果,最终将 3 个不同卷积核的卷积结果拼接;编码结果最后经过全连接层直接进行打分(该全连接层的输出维度为 1)。ConvR^[48]考虑使用卷积对头实体进行编码,ConvE 与 ConvR 模型的区别可以从图 8 中得到解释:ConvE 首先拼接头实体和关系表示,然后使用卷积操作对拼接结果进行卷积编码;而 ConvR 直接将关系建模成卷积核,直接针对头实体嵌入进行卷积操作。与 ConvE 相同,ConvR 使用卷积编码结果与尾实体嵌入求内积计算三元组的得分。InteractE^[49]观察到在学习图嵌入的过程中,实体表示与关系表示之间的交互计算越多,图嵌入中蕴含的语义信息能够更深层次被利用,也因此更有利子目标任务的学习(如链接预测任务)。为此,InteractE 提出通过特征排列(feature permutation)、特征重塑(checkered feature reshaping)和循环卷积(circular convolution)三步操作增强图嵌入的表达能力。

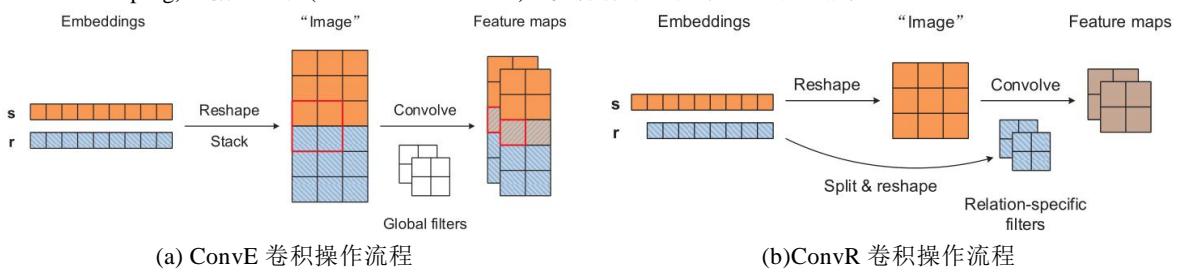
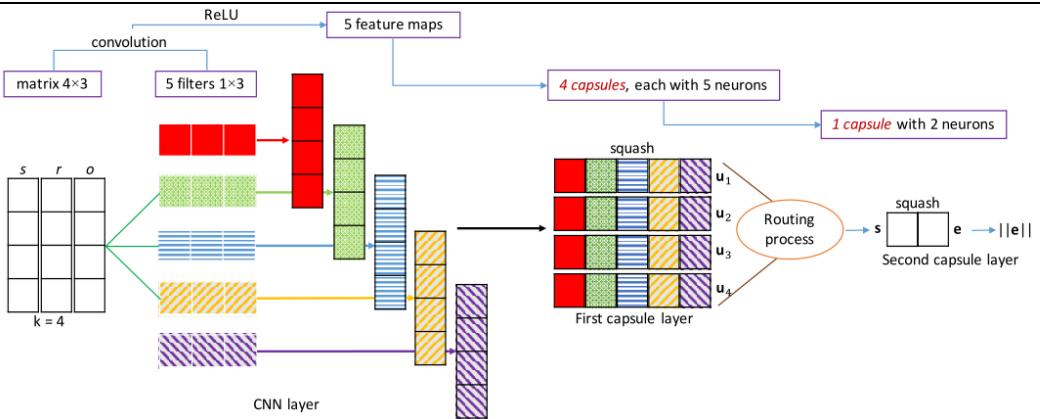


图 8. ConvE 与 ConvR 的卷积操作对比^[48]

图 9 CapsE 算法流程^[50]

类似 ConvKB, CapsE^[50]采用胶囊网络(CapsNet)对三元组进行编码,以图 9 展示为例,算法具体流程如下:首先将三元组 (h, r, t) 拼接成输入矩阵 $[v_h; v_r; v_t] \in R^{d \times 3}$,然后利用 5 个 1×3 大小的卷积核对输入矩阵进行编码,得到大小为 $d \times 5$ 的特征图;该特征图被输入进第一层 CapsNet 层,经过 5 个 capsule 处理,每个 capsule 提取某一方面事实;最后将路由结果输入到第二层 CapsNet,用于计算该三元组的得分。

相比于 CNN 和 CapsNet,循环神经网络(RNN)在建模序列数据上的能力更加突出。前面介绍的 ConvE、ConvKB 以及 CapsE 等工作都只建模单独的三元组,而知识图谱除了组成元素外,三元组之间的连接信息也是构成图网络的重要知识,其中,对应于其它类型异质图中的元路径(meta-path),首尾相连的三元组路径是建模这种深层次图结构知识的可用信息之一。为了更好地挖掘知识图谱的长距离结构信息,Guo 等人基于三元组路径的图嵌入学习模型 RSN^[51]。Guo 等人试图通过 RNN 对长距离结构信息进行编码,然而三元组路径中包含实体和关系两种元素,在编码实体和关系时需要区别对待。为此,Guo 等人对 RNN 进行改进,提出了 Skipping 机制:

$$h'_t = \begin{cases} h_t & x_t \in E \\ S_1 h_t + S_2 x_{t-1} & x_t \in R \end{cases}$$

即如果当前输入为实体,则按照 RNN 的逻辑进行运行;如果当前输入为关系,则输出单元在 RNN 的基础上还要加上前一次模型实体输入 x_{t-1} 。如图 8(b)所示,对于三元组路径 $United\ Kingdom \rightarrow country \rightarrow Tim\ Berners-Lee \rightarrow employer \rightarrow W3C$,当学习实体 $United\ Kingdom$ 的表示时,计算逻辑与原始的 RNN 模型相同,而在编码关系时,则同时使用 $country$ 和 $United\ Kingdom$ 的信息。通过编码三元组序列,RSN 将深层次的结构知识融入到为实体与关系学习的图嵌入表示中。作者后续又论证了 RSN 与残差网络的联系,并在实体对齐任务中取得了相当不错的成绩。

总体来看,基于传统深度学习的图嵌入学习方法的研究内容并不算多。虽然传统神经网络较为强大的学习能力使得这类模型在一些数据集上也拥有不错的表现,但是传统深度学习的可解释性问题也随之遗留在图嵌入学习研究中。因此如何基于传统深度学习设计更优良的模型,以及关于神经网络可解释性问题有待进一步研究开展。

3.4 基于图神经网络模型的图嵌入学习

本小节我们介绍基于图神经网络的图嵌入学习算法,从代表性工作的设计思想、短板分析以及后续改进的思路梳理发展脉络。

图神经网络(GNN)^{[52][53]}是近期研究较为广泛的领域之一,很多研究表明 GNN 在建模图数据方面表现出非常突出的性能,目前图神经网络模型已被成功应用到诸多研究领域,如社交网络^[54]、医药研究、晶体性能预测^[55]、知识图谱^[56]等。

图神经网络模型可以用 Message Passing Neural Networks(MPNN)^[57]框架来描述,该框架在前向传播过程

可分为两个阶段:信息传递和信息读取阶段.其中信息传递阶段又分为传递信息至邻居节点、聚合自身信息与邻居传递的信息两个过程,这也正是 GNN 的前向传播逻辑.为此, GNN 每一层的前向传播过程可以用统一建模成:

$$h_i^{k+1} = \text{Merge}(\{h_i^k\} \cup \text{aggr}(\{h_j^k\}_{j \in N(i)}))$$

其中 Merge 表示特征融合函数, aggr 表示邻居特征聚合函数, $N(i)$ 表示节点 i 的邻居集合, 而 h_i^k 就是节点 i 在第 k 层网络中输入的特征向量.

图卷积神经网络(GCN)^{[52][53][58]}是图神经网络模型研究中得到的一项主要成果, 成功将卷积操作应用到无向图这类非欧空间数据.GCN 的特征学习逻辑可以用下式表示:

$$H^{k+1} = f(\hat{A}H^kW^k)$$

其中 $\hat{A} = D^{-\frac{1}{2}}(A + I)D^{-\frac{1}{2}}$, A 是无向图的邻接矩阵, I 是单位矩阵, 为 D 是无向图的度矩阵. 这里的 H^k 表示第 k 层网络层学习到的节点特征矩阵, f 是可选择的激活函数.

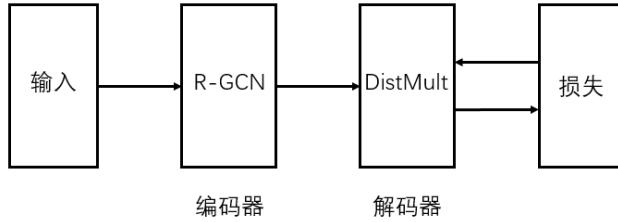


图 10 R-GCN 用于链接预测任务框架

虽然 GCN 在如节点分类、图分类等众多任务上均取得了不错的成绩,但是原始 GCN 模型只针对无向无权图,为了将其应用于知识图谱,需要在 GCN 模型的基础上施以一定的调整.Schlichtkrull^[56]提出 R-GCN,为每个关系 r 引入关联的权重矩阵 W_r ,在 GCN 在聚集邻居信息前先用相应关系权重矩阵特化邻居特征,之后参与聚合过程,其特征更新过程可以用下式表示:

$$h_i^{(k+1)} = \sigma \left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_r^{(k)} h_j^{(k)} + W_0^{(k)} h_i^{(k)} \right)$$

其中 $W_r^{(k)}$ 表示第 k 层关系 r 的权重矩阵.

可以看出,R-GCN 虽然将 GCN 模型扩展到了关系图中,但是仍然存在三方面问题:

- (1) R-GCN 只考虑了关系的类别,却忽略了关系的方向;
- (2) 由于需要为每个关系都关联一个权重矩阵,对于关系类别较多的知识图谱,存在参数量过大的问题;
- (3) R-GCN 只是引入了关系权重矩阵,为每条边赋予了相应的权重,但没有显示学习关系表示.

针对第二个问题,R-GCN 考虑对关系矩阵进行约束,提出了两种策略: (1)共享参数; (2)稀疏化矩阵. 共享参数方法通过设置特定数目的基矩阵 $\{V_b^{(k)}\}_{1 \leq b \leq B}$,其中 B 是基矩阵的数目,然后通过线性组合的方式得到不同关系

关联的权重矩阵 $W_r^{(k)} = \sum_{b=1}^B a_{rb}^{(k)} V_b^{(k)}$. 稀疏化矩阵方案约束 $W_r^{(k)}$ 为对角块状矩阵,即 $W_r^{(k)} = \text{diag}(Q_{1r}^{(k)}, \dots, Q_{Br}^{(k)})$,其中

$Q_{br}^{(k)} \in R^{d^{(k+1)}/B \times d^{(k)}/B}$,这里的 $d^{(k)}$ 与 $d^{(k+1)}$ 分别表示第 k 层网络以及第 $k+1$ 层网络的输出维度.

除了将 GCN 扩展到关系图领域,R-GCN 还提出了将 GNN 系列模型用于链接预测任务的框架范式,即 Encoder-Decoder 框架,如图 10 所示.其中 R-GCN 充当编码器,用于学习节点特征,最后一层输出节点特征后作为解码器的输入,解码器可以选择前面介绍过的传统的图嵌入学习算法,如图 1 中使用的 DistMult 算法,对于三元组(e_i, r, e_j),基于 R-GCN 学习到实体表示 h_i, h_j ,同时在编码器部分为每个关系引入关系矩阵 R_r ,基于 DistMult 算法计算三元组得分 $\phi(e_i, r, e_j) = h_i^T R_r h_j$,根据得分高低判断三元组的正确性.

R-GCN 虽然在收集邻居信息时考虑了不同关系信息,但是忽略了关系的方向,除此以外,R-GCN 直接聚合了邻居的特征信息,而忽略了异质图中不同节点因为实体类型不同可能拥有不同特征域的问题.为此,VR-GCN^[59]通过将 TransE 引入 R-GCN 框架,在聚合邻居信息前,首先基于 TransE 的假设: $h + r \approx t$,将邻居转移到中心实体(也就是待更新特征表示的实体)所在的特征域中,从而解决了头、尾实体表示从属不同特征域的问题.例如对于中心实体 e_i ,具体做法如下:(1)对于与实体 e_i 通过关系 r 相邻且充当尾实体的实体 e_j ,通过组合函数为 $c(h_j^l, h_r^l) = h_j^l - h_r^l$ 将实体 e_j 转移到实体 e_i 的特征域中;(2)对于与实体 e_i 通过关系 r 相邻且充当头实体的实体 e_j ,则利用组合函数为 $\hat{c}(h_j^l, h_r^l) = h_j^l + h_r^l$.VR-GCN 的层特征更新函数如下:

$$h_i^{l+1} = \sigma\left(\frac{1}{d_i}\left(\sum_{r \in N_i} \sum_{t \in N_r^l} (h_t^l - h_r^l) + \sum_{r \in N_i} \sum_{t \in N_r^l} (h_t^l + h_r^l)\right) + h_i^l\right)W^l$$

通过引入 TransE 的思想,VR-GCN 成功在聚集邻居信息的同时既考虑了关系的类别信息,又考虑了关系的方向信息.在 VR-GCN 的基础上,Trans-GCN^[60]与 CompGCN^[61]对其使用的组合函数进行了扩展,分别另外提出了使用 RotatE、DistMult 和 HolE 思想的组合函数,均取得了不错的效果.除此以外,CompGCN 还显示考虑了关系的方向信息,具体的特征更新函数如下式所示:

$$h_v = f\left(\sum_{(u,r) \in N(v)} W_{\lambda(r)} \phi(h_u, h_r)\right)$$

其中 $\lambda(r)$ 表示关系 r 的方向,该工作按照关系的方向类别将关系分成三大类:出节点关系、入节点关系以及自循环关系,权重矩阵 $W_{\lambda(r)}$ 根据关系 r 方向的不同使用不同的参数矩阵.

图注意力网络(GAT)^[62]是图神经网络模型研究得到的另一项重要成果,该模型成功将注意力机制引入到图神经网络的聚合过程中,使得聚合邻居信息时不同邻居拥有不同的重要性,这也启发了基于 GNN 的知识图谱表示学习工作在聚集邻居信息的过程中加入权重参数进行重要性选择.SCAN^[63]通过直接设置与关系类别相关的权重参数,实现了邻居信息的聚集过程中不同关系对于中心实体具有不同重要性的思想.然而这种重要性的计算只与关系类别有关,与邻居实体没有关系,为了同时考虑邻居实体与邻居关系在计算权重过程中的影响,Nathani 等人^[64]在聚合邻居信息的过程中参考了 GAT 的思想,然而不过与之前工作不同,Nathani 等人并没有聚合邻居实体信息,而是直接聚合邻居三元组的信息,以实体 v_i 为例,其聚合邻居信息的工作原理如下:(1)首先考虑将所有实体 v_i 参与的三元组编码成向量形式,如三元组(v_i, r_k, v_j)编码成向量形式 $c_{ijk} = W_1[h_i||h_j||g_k]$,其中 h_i 与 h_j 分别代表实体 v_i 与 v_j 的嵌入表示,而 g_k 代表关系 r_k 的嵌入表示;(2)计算三元组(v_i, r_k, v_j)的权重: $\alpha_{ijk} = softmax(LeakyReLU(W_2 c_{ijk}))$;(3)为实体 v_i 聚合所有邻居信息,得到 $h'_i = \sigma\left(\sum_{j \in N(v_i)} \sum_{k \in R_{ij}} \alpha_{ijk} c_{ijk}\right)$,这里用 R_{ij} 表示连接实体 v_i 与 v_j 的关系集合.

虽然工作[64]引入了 attention 机制,但没有满足同一特征域内融合特征的思想,相比之下,RAGAT^[65]既采用了 GAT 的注意力机制,又继承了[59][60][61]等工作利用组合函数将邻居转移到中心实体特征域的做法.需要说明的是,RAGAT 认为相同实体与不同关系关联时应具有不同表示,在[59][60][61]等工作的基础上提出了与关系相关的组合函数.以基于 TransE 的组合函数为例,实体 v 为中央实体,实体 u 通过关系 r 与实体 v 组成三元组(v, r, u),Trans-GCN 定义的组合函数为 $e_u - e_r$,而 RAGAT 定义的组合函数为 $W_r e_u - e_r$,如果仔细观察,不难发现 RAGAT

选择的组合函数更像是基于 TransR 算法的思想.

以上工作都只考虑每次聚合节点的一阶邻居信息,虽然经过 L 次聚合操作,每个节点都能获取其 L -阶邻居的信息,但是 Sun 等人^[66]指出这种方式容易加剧噪声传播的影响,除此以外,Sun 等人还证明了仅使用一阶邻居信息在特殊情况下可能无法分辨不同节点邻居结构的异质性.为此,考虑多跳邻居信息是能够更好聚合邻居信息的一种手段, Sun 等人提出 AliNet,该模型直接聚合多跳邻居实体的特征信息,并通过门机制控制不同阶邻居信息对组成中心实体最终特征表示的重要性.Nathani 等人^[64]通过直接加入新“边”来聚合多跳邻居的信息,以二阶邻居为例,对于实体 v_i ,实体 v_j 通过路径 $(v_i, r_1, v_k) \wedge (v_k, r_2, v_j)$ 与实体 v_i 相连,该工作通过引入新关系 $r' = r_1 + r_2$ 来加入新的三元组知识 (v_i, r', v_j) ,如此 v_j 通过“关系” r' 成为实体 v_i 的一阶邻居,进而实现聚合多跳邻居信息的目的.目前关于如何在聚合邻居信息的过程中引入多跳邻居信息的工作还可以进一步挖掘.

总而言之,图神经网络凭借其强大的建模图结构数据的能力,能够更加深入地挖掘学习图谱的结构信息,对图谱的结构信息进行深度编码.结合图神经网络的发展趋势,面向知识图谱的基于图神经网络的图嵌入研究有待进一步的发展.表 4 总结了部分基于 GNN 的图嵌入算法的特征更新公式.

表 4 部分基于图神经网络的图嵌入算法的特征更新公式

模型	特征更新公式
R-GCN ^[56]	$h_i^{l+1} = \sigma \left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_r^l h_j^l + W_0^l h_i^l \right)$
VR-GCN ^[59]	$h_i^{l+1} = \sigma \left(W^l \left(\sum_{r \in R} \sum_{j \in N_i^r} (h_j^l - r_k^l) + \sum_{r \in R} \sum_{j \in N_{in}^r} (h_j^l + r_k^l) \right) \right)$ $h_i^{l+1} = \sigma \left(\sum_{(r_k, e_j) \in N(e_i)} W_{\lambda(r_k)}^l \phi(h_j^l, r_k^l) \right)$
CompGCN ^[61]	$W_{\lambda(r)}^l = \begin{cases} W_O^l, & r \in R \\ W_I^l, & r \in R_{inv} \\ W_S^l, & r = self \end{cases}$
Trans-GCN ^[60]	$h_i^{l+1} = \sigma \left(\frac{1}{c_i} W^l \left(\sum_{(e_j, r_k, e_i) \in T_{in}(e_i)} (h_j^l \circ r_k^l) + \sum_{(e_i, r_k, e_j) \in T_{out}(e_i)} (h_j^l * r_k^l) \right) + h_i^l \right)$ $h_i^{l+1} = f \left(\frac{1}{H} \sum_{h=1}^H \sum_{(r_k, e_j) \in N(e_i)} \alpha_{k,j}^h m_{(e_i, r_k, e_j)}^h \right)$
RAGAT ^[65]	$m_{(e_i, r_k, e_j)}^h = W_{dir(r_k)} \phi^r(e_i, e_j)$ $\alpha_{k,j}^h = softmax(LeakyReLU(W_{att} m^h(e_i, r_k, e_j)))$

3.5 引入额外信息的图嵌入学习方法

本小节我们从信息利用的角度考虑图嵌入学习问题,介绍引入额外信息的图嵌入学习方法.之前介绍的工作理论上均专注于学习图谱的结构特征,无论是 TransE 以及 DistMult 等工作每次的学习单元为单个三元组,还是 RAGAT 等工作利用图神经网络通过聚合实体的邻居信息来学习特征,都只是通过学习结构信号来完成下游任务.然而除了结构信号,知识图谱作为一种特殊的异质图,每个实体与关系均能够关联具体的文本概念或类别信息,这些信号都能够进一步补充结构信息,使得图嵌入能够学习到更丰富的语义信息.

通过在图嵌入学习的过程中融入实体与关系的名称、实体描述等信息,有利于丰富图嵌入表示学习到的

语义信息.给定实体名称序列,NTN^[67]直接使用实体名称所有字符的词向量表示的均值作为图嵌入的初始值.Zhang 等人^[68]通过多任务学习联合学习图谱结构与图谱的文本信息,该工作学习实体的文本表示时依然采用 NTN 的思路,用实体名称包含的所有字符的词向量的均值作为实体的文本表示.与上述工作不同,DKRL^[69]通过 CNN 或 CBOW 模型对头、尾实体的描述信息进行编码,得到实体的文本表示 v_{hd} 和 v_{td} ,同时基于 TransE 思想,在优化 $E_S = \|v_h + v_r - v_t\|$ 学习实体的结构表示的同时,又通过将其中 v_h 和 v_t 分别替换成基于实体描述信息编码出的实体表示向量 v_{hd} 、 v_{td} ,得到新的损失 $E_{DD} = \|v_{hd} + v_r - v_{td}\|$ 、 $E_{DS} = \|v_{hd} + v_r - v_t\|$ 和 $E_{SD} = \|v_h + v_r - v_{td}\|$,通过混合训练结构表示和文本表示实现文本信息与结构信息的融合.Xu 等人^[70]通过结构表示与文本表示相加的方式向结构信息中添加图谱的文本信号,同时考虑到信号融合过程中不同信号拥有不同的权重,该工作采取门机制自动平衡两类信号的重要性.Wang 等人^[71]同样分别学习结构表示和文本表示,在联合学习的过程中通过互注意力机制交互结构信息与文本信息,使两类信号能够互相学习,互相弥补.由于近年来预训练语言模型的发展,文本信息的抽取和编码能力得到长足提升,同时考虑到预训练语言模型经过海量文本的自监督学习,预训练语言模型本身的内部参数存储了语法和常识知识,于是开始出现使用预训练语言模型编码图嵌入的工作.Daza 等人提出 blp 框架^[72],该框架使用 Bert 模型^[73]编码实体描述信息,作为传统图嵌入学习算法(如 TransE)的图嵌入向量初始化值,利用 Bert 模型学习实体描述的文本知识,同时利用基于结构信息的图嵌入学习算法学习知识图谱中蕴含的结构信息.

除却实体名称、实体描述等文本信息,类别层级信息同样有利于丰富图嵌入的语义信息.TKRL^[74]通过实体类别信息构造类别的层次树形结构,树的根节点为设置的最一般的实体类型,通过层层细化,叶子节点表示图谱中划分最细致的实体类别,每个叶子节点到树的根节点对应了唯一一条路径,而这条路径可以用来全面描述该叶子实体类型的层次信息,层次树中的每个节点都对应了特定的映射矩阵,其中叶子节点的映射矩阵可以通过 RHE 与 WHE 两种方式获取.同时该工作认为每个实体可能拥有多种类别,而相同实体在不同关系下自然应当考虑不同的类型信息.基于 TransE, TKRL 提出了新的度量函数 $\|M_{rh}v_h + v_r - M_{rt}v_t\|$,其中 M_{rh} 是与头实体 h 以及关系 r 的类别相关的投影矩阵,同理 M_{rt} 是与尾实体 t 以及关系 r 的类别相关的投影矩阵.TKRL 是第一篇将类别信息用于知识图谱表示学习的工作,在一定程度上缓减了 TransE 无法处理多映射关系的问题.MKRL^[75]则统一使用了上述介绍的各种信息,同时考虑了实体描述信息、实体类别层次类别信息以及关系的文本信息.Zhang 等人^[75]从关系的层次信息出发,将关系按照语义分成三个层次:聚类关系层、关系层和次关系层.聚类关系层通过对所有关系进行聚类,这是考虑了同一聚类关系具有一定的语义相似度;关系层建模关系的直接语义信息;而次关系层考虑了相同关系在不同语境下可能具有不同的含义,为此按照关系所参与的三元组进行聚类:对于每个包含关系 r 的三元组 (h, r, t) ,计算得到在该语境下关系的表示 $v_r = v_t - v_h$,如此操作后,对于关系 r 会得到一系列不同语境下的向量表示 $\{\hat{v}_r\}$,对该集合进行聚类,每个聚类对应了该关系在特定语境下的次关系表示.

相比于挖掘结构信息的图嵌入算法,引入额外信息的图表示学习算法通过知识图谱自身蕴含的文本知识和层次信息来补充欠缺的结构信息,进而学习到更有效、建模知识图谱更佳的图嵌入.至于如何同时充分学习知识图谱结构信息和其它类型的信息,还有待进一步的研究和挖掘.

4 总结

4.1 五类图嵌入学习算法的简单总结

本文从算法核心、基础模型使用以及信息利用等角度梳理了五类图嵌入算法的原理和设计特点,并针对某些奠基性的工作给予了着重介绍.我们在表 5 中总结了这五类图嵌入算法的基础特征性质以供参考.

基于转移思想的图嵌入算法具有较为高效的计算效率,适中的参数量,也具有较为优秀的表达能力.除此以外也拥有灵活的设计空间,可以从不同向量空间、不同转移操作等角度进行考虑设计.然而由于其基本思想比较简单,奠基性算法 TransE 在处理对称、1-N、N-1 以及 N-N 等关系类型上具有很大的劣势,即这类算法在充分表达知识图谱的能力上有可提升的空间,因此后续基于转移思想的很多工作都需要着重考虑如何提升模型

的建模能力上.

表 5 不同类型图嵌入算法的基础特征总结

类型	特征
基于转移思想的图嵌入算法	建模思想直观,可设计改进的空间较大,参数量少,计算效率高;但是表达能力有限,需要进一步的设计
基于张量分解的图嵌入算法	理论表达能力强,但现有算法的实验效果仍有待提高;但是通常参数较多,不利于应用与大规模图谱
基于传统深度学习的图嵌入算法	特征学习能力强;然而可解释性不强,部分算法的稳定性不强
基于图神经网络的图嵌入算法	利用图神经网络较强的图编码能力,建模能力强;然而计算效率低,不利于大规模工业领域应用
引入额外信息的图嵌入算法	充分利用知识图谱的包含的其他类型信息,对知识图谱的图结构特征进行补充;目前研究工作不多,有待进一步的研究探索

基于张量分解的图嵌入算法因为其基本思想旨在完整建模知识图谱蕴含的信息,因此理论上具有很强的模型表达能力,例如其奠基性算法 RESCAL 具有完全表达知识图谱的知识的能力,但根据现有的模型的实验结果来看,基于该思想的模型性能有待进一步挖掘.除此以外 RESCAL 拥有过多的参数,使得该算法在泛化能力以及应用在大规模知识图谱上具有较大的劣势.为此 RESCAL 的很多后续算法都旨在解决如何保留表达能力的同时减少参数量,提高模型的计算效率.

基于传统深度学习的图嵌入算法利用 CNN、RNN 以及 CapsNet 等传统深度学习算法在抽取特征上的优势进行表示学习,也取得了不错的成绩.然而,也正是因为深度学习模型的使用使得这类算法的可解释性不足.除此以外,[76]的工作表明,ConvKB、CapsE 算法在不同数据集实验效果浮动很大,这也侧面反映了这类算法的不稳定.

基于图神经网络的图嵌入算法基于知识图谱的图结构特性对原本建模于无向图的图神经网络模型结构进行了适当调整,使得图嵌入学习过程中能够利用图神经网络强大的图编码能力.然而考虑到这类算法在解码器端要使用之前介绍的传统图嵌入算法作为解码器,所以解码器拥有的缺陷也可能同样存在于基于图神经网络的图嵌入模型中.除此以外,图神经网络的计算效率也是阻止其大规模工业化应用的原因之一,如何在保证表达能力的同时提高该类图嵌入模型的计算效率也是需要研究的问题.同时,如何更好地学习图谱蕴含的结构信息也是基于图神经网络的图嵌入算法仍然有待完善的地方.

引入额外信息的图嵌入算法考虑图结构信息以外的信息来提升学习效果,考虑了名称、描述、类别等额外信息作为图结构信息的补充,是从信息来源的角度考虑图嵌入学习问题,在归纳假设上可能仍然采用之前提及的不同图嵌入算法的基本思想,例如 TKRL 虽然使用了实体的描述信息,但该模型学习的基本假设还是采用了 TransE 的基本思想.至于如何更充分地利用以上信息,也是有待进一步挖掘探索的工作.

为了更好地展示不同算法的性能,本文在表 6 中给出了部分算法在标准数据集 FB15k-237^[8]和 WN18RR^[29]数据集上的链接预测任务的实验指标结果,实验指标采取 MRR 和 Hits@k 指标.考虑到 FB15k-237 和 WN18RR 两个数据集提出时间较晚,很多工作在论文中没有基于此二者进行实验,为此本文从其它较新的工作中寻找他人复现后的实验结果,即使这样,还是存在很多算法没有或者缺少实验结果数据,为此对于没有实验结果的算法我们在表 6 中不予展示,对于缺少部分实验指标的算法我们用“——”代替.表 6 中展示的所有结果都通过算法名称右边括号内的索引给出了相应的结果出处,其中 DKRL 算法的实验结果由我们自己复现得到.

需要说明的是,列出的指标来自于现有的论文,仅供参考,考虑到不同算法在图嵌入维度、训练轮数、批采样大小、负采样大小以及损失函数设计等都存在差异,因此该数据不完全能够用于分析对比不同算法之间的性能优劣.除此以外,根据 Sun 等人^[77]的研究发现,部分模型的指标计算存在问题,部分算法如 ConvKB、CapsE 在

为三元组进行打分时常常会为不同的三元组赋予相同的分值.因此这部分算法本文采取 Sun 等人重新评估后的实验结果.

表 6 不同类型图嵌入算法的基础特征总结

模型	FB15k-237		WN18RR	
	MRR	Hits@10	MRR	Hits@10
基于转移思想	TransE ^[12] ([83])[83])	0.207	0.377	—
	TransH ^[15] ([83])	0.211	0.386	—
	CrossE ^[20] ([20])	0.299	0.474	—
	MuRP ^[32] ([32])	0.336	0.521	0.481 0.566
	QuatE ^[27] ([27])	0.366	0.556	0.488 0.582
	RotatE ^[24] ([24])	0.338	0.533	0.476 0.571
	ATTH ^[26] ([26])	0.348	0.540	0.486 0.573
基于张量分解思想	DistMult ^[36] ([32])	0.241	0.419	0.430 0.490
	ComplEx ^[38] ([32])	0.247	0.428	0.440 0.510
	TuckER ^[42] ([42])	0.358	0.544	0.470 0.526
基于传统深度学习	ConvE ^[46] ([46])	0.316	0.491	0.460 0.480
	ConvR ^[48] ([48])	0.350	0.528	0.475 0.537
	ConvKB ^[47] ([77])	0.243	0.421	0.249 0.524
	CapsE ^[50] ([77])	0.150	0.356	0.415 0.559
基于图神经网络模型	R-GCN ^[56] ([56])	0.248	0.417	—
	CompGCN ^[61] ([61])	0.355	0.535	0.479 0.546
	Trans-GCN ^[60] ([60])	0.315	0.477	0.233 0.508
	Rotat-GCN ^[60] ([60])	0.356	0.555	0.485 0.578
	RAGAT ^[65] ([65])	0.365	0.547	0.489 0.562
引入额外信息	DKRL ^[69] (our)	0.225	0.385	0.204 0.550
	Zhang 等人 ^[83] ([83])	0.315	0.496	—

通过表 6 可以观察到,通过引入复数空间、双曲空间、旋转操作和四元数等设计思想,Trans-系列方法的实验性能得到大幅度提升,在 FB15k-237 和 WN18RR 数据集上得到很好的实验效果;基于张量分解方法的实验指标稍微低一些,这可能与该类方法的设计难度较高存在关联,也可以参考 Trans-系列方法的设计思路,从表示空间、相似度计算函数等方面进一步进行算法设计;基于传统深度学习的方法表现不够突出,而且除 ConvE 与 ConvR 的表现相对稳定,ConvKB 与 CapsE 在不同数据集上表现波动较大,例如 CapsE 模型在 WN18RR 数据集上表现良好,但是在 FB15k-237 数据集上效果却不如 TransE,波动的原因也值得进一步研究探索;基于 GNN 的方法表现也相当不错,其中 Rotat-GCN 和 RAGAT 的解码器部分分别基于 RotatE 和 CrossE 模型,因此不难猜想将解码器部分换成表现更好的 QuatE、ATTH 等算法是否会进一步提升模型性能;引入额外信息部分且存有可靠实验结果的相关工作能够留下结果的并不多,通过对比 DKRL 和 TransE 的实验结果,引入实体描述信息确实有利于进一步提升模型性能.

4.2 面向知识图谱的图嵌入学习领域有待进一步解决的问题和挑战

近些年,虽然面向知识图谱的图嵌入学习领域发展迅速,但是仍然存在许多待解决的问题.为了方便以后研究人员的研究,我们总结了一些面向知识图谱的图嵌入学习领域的研究热点和有待解决的问题:

- 如何更好地学习 long-tail 类型的实体?long-tail 类型的实体和关系是指出现频率很少的实体与关系,由于训练样本不足,往往学习出的图嵌入不足以建模该类实体的语义信息,从而影响基于图嵌入的下游

任务.而已有工作[78]分析了常见的知识图谱中 long-tail 类型的实体与关系占有相当的比例.因此,如何在 long-tail 的条件下充分学习实体与关系的表示是一件很有价值的工作.

- 如何更好地将文本信息、类别信息融入图嵌入学习?通过将知识图谱本身拥有的文本知识和类别知识,进一步丰富基于结构知识的图嵌入的信息.除此以外,考虑到实体关联的文本信息与类别信息与实体出现的频率无关的特征,因此结合文本和类别信息也是解决 long-tail 问题的方案之一;
- 如何更好地挖掘知识图谱本身蕴含的层次信息?层次信息作为知识图谱本身蕴含的潜在结构信息,如果能够充分挖掘,对于建模知识图谱的结构具有一定的研究意义.除此以外,利用层次树结构中同一父实体的子实体具有相似的语义信息这一特点,层次信息同样也有利于解决 long-tail 问题;
- 除了文本与类别信息,图片、语音等多模态信息也是知识获取的途径.在学习图嵌入的过程中,考虑融入多模态信息,在面向知识图谱的下游任务中可能取到不错的效果;
- 研究图嵌入算法与其他推理算法的结合.图嵌入算法在面向知识图谱的单步推理任务中具有不错的效果,且相对于其他种类的推理方法又具有学习效率高的优点,但考虑到其主要依赖浅层结构约束来学习图表示,在复杂推理、多跳推理任务中表现不佳,因此,结合图嵌入算法的优点与其他推理算法的优势^[79],例如一阶逻辑推理方法的可解释性,在保证推理能力与可解释性的同时提高推理算法的计算效率;
- 如何在保证建模能力的同时提高图嵌入的学习效率?现有的某些图嵌入算法虽然表达能力出众,但是参数量相对较多,在学习大规模知识图谱表示的任务上具有低效率的缺点,因此如何在充分保证模型学习能力的同时,提高模型的学习效率,这也是现有的图嵌入算法是否能够大规模应用于工业领域的关键问题之一;
- 如何结合上下文动态地学习图嵌入表示?在知识图谱中,实体的邻居相当于该实体的上下文,在某些的任务,例如基于知识图谱的链接预测任务中,对于不同的 query ($h, r, ?$),实体 h 的邻居可能起到不同的作用,为此如何针对不同的任务为实体 h 动态地学习相应的表示^{[80][81]}也是很有意义的研究问题;
- 如何学习动态知识图谱(又称时序知识图谱)的图嵌入表示?常见的知识图谱是静态的图结构,但是现实使用中,知识图谱的图结构往往伴随着时间而发生变化,例如增加了部分实体或者删除了部分实体之间的关系,虽然短时间的局部结构变化对全局表示影响不大,但伴随时间积累下的结构差异也会使得重新学习图嵌入非常重要,然而重新学习的成本通常较大.为此,如何高效学习动态知识图谱的图嵌入表示也是目前需要解决的研究问题之一^[82];
- 除此以外,图嵌入算法的设计还需要考虑特定的下游任务,如何在基于知识图谱的具体下游任务中设计并利用好图嵌入也是需要考虑的问题.

综上所述,面向知识图谱的图嵌入学习目前还存有很多需要解决的问题,这也正需要更多研究人员的努力和投入.

References:

- [1] Xu K, Reddy S, Feng Y, Huang S, Zhao D. Question answering on freebase via relation extraction and textual evidence. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016 - Long Papers, 4, 2016. 2326–2336. [doi: 10.18653/v1/P16-1220]
- [2] Hoffmann R, Zhang C, Ling X, Zettlemoyer L, Weld DS. Knowledge-based weak supervision for information extraction of overlapping relations. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, 2011. 541–550.
- [3] Carlson A, Betteridge J, Kisiel B, Settles B, Hruschka ER., Mitchell TM. Toward an architecture for never-ending language learning. In: Proceedings of the National Conference on Artificial Intelligence, 2010. 1306–1313.

-
- [4] Wang WY, Mazaitis K, Cohen WW. Programming with personalized pagerank: A locally groundable first-order probabilistic logic. In: Proc. of the 22nd ACM Int'l Conf. on Information and Knowledge Management. New York: ACM Press, 2013. 2129-2138. [doi: 10.1145/2505515.2505573]
 - [5] Jiang S, Lowd D, Dou D. Learning to Refine an Automatically Extracted Knowledge Base Using Markov Logic. 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 2012, pp. 912-917. [doi: 10.1109/ICDM.2012.156]
 - [6] Böthmann L, Lehmann J. Pattern based knowledge base enrichment . Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8218 LNCS(PART 1), 2013. 33-48. [doi: 10.1007/978-3-642-41335-3_3]
 - [7] Wang X, Bo D, Shi C, Fan S, Ye Y, Yu PS. A survey on heterogeneous graph embedding: methods, techniques, applications and sources . arXiv preprint arXiv:2011.14867, 2020.
 - [8] Toutanova K, Chen D, Pantel P, Poon H, Choudhury P, Gamon M. Representing text for joint embedding of text and knowledge bases. In Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing, September, 2015. 1499-1509. [doi: 10.18653/v1/D15-1174]
 - [9] Wang X, Ji H, Cui P, Yu P, Shi C, Wang B, Ye Y. Heterogeneous graph attention network. In Proceedings Of the 2019 World WideWeb Conference (WWW'19), 2019. 2022-2032. [doi: 10.1145/3308558.3313562]
 - [10] Wang Q, Mao Z, Wang B, Guo L. Knowledge graph embedding: A survey of approaches and applications . In IEEE Transactions on Knowledge and Data Engineering (Vol. 29, Issue 12, pp. 2017. 2724-2743). [doi: 10.1109/TKDE.2017.2754499]
 - [11] Dai Y, Wang S, Xiong NN, Guo W. A survey on knowledge graph embedding: Approaches, applications and benchmarks . Electronics (Switzerland), 2020,9(5):1-29. [doi: 10.3390/electronics9050750]
 - [12] Bordes A, Usunier N, Garcia-Duran A, Weston J, Yakhnenko O. Translating Embeddings for Modeling Multi-relational Data. In Proceedings of the Neural Information Processing Systems (NIPS), Dec 2013, South Lake Tahoe, United States. pp.1-9.
 - [13] Bordes A, Glorot X, Weston J, Bengio Y. Joint learning of words and meaning representations for open-text semantic parsing. In International Conference on Artificial Intelligence and Statistics, 2012. 127-135.
 - [14] Bordes A, Weston J, Collobert R, Bengio Y. Learning structured embeddings of knowledge bases. In Proceedings of the Twenty-fifth AAAI Conference on Artificial Intelligence, 2011.
 - [15] Wang Z, Zhang J, Feng J, Chen Z. Knowledge graph embedding by translating on hyperplanes. Proceedings of the National Conference on Artificial Intelligence, 2014. 1112-1119.
 - [16] Lin Y, Liu Z, Sun M, Liu Y, Zhu X. Learning entity and relation embeddings for knowledge graph completion. Proceedings of the National Conference on Artificial Intelligence, 2015, 2181-2187.
 - [17] Ji G, He S, Xu L, Liu K, Zhao J. Knowledge graph embedding via dynamic mapping matrix. ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference, 2015. 687-696. [doi: 10.3115/v1/P15-1067]
 - [18] Ji G, Liu K, He S, Zhao J. Knowledge graph completion with adaptive sparse transfer matrix. In: Proc. of the 30th AAAI Conference on Artificial Intelligence, 2016. 985-991.
 - [19] Nguyen DQ, Sirts K, Qu L, Johnson M. STransE: A novel embedding model of entities and relationships in knowledge bases. 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference, 2016. 460-466. [doi: 10.18653/v1/N16-1054]
 - [20] Zhang W, Paudel B, Zhang W, Bernstein A, Chen H. Interaction embeddings for prediction and explanation in knowledge graphs. WSDM 2019 - Proceedings of the 12th ACM International Conference on Web Search and Data Mining, 2019. 96-104. [doi: 10.1145/3289600.3291014]
 - [21] Fan M, Zhou Q, Chang E, Zheng TF. Transition-based knowledge graph embedding with relational mapping properties. In Proceedings of the 28th Pacific Asia Conference on Language, Information and Computation, PACLIC 2014, 328-337.
 - [22] Feng J, Zhou M, Hao Y, Huang M, and Zhu X. Knowledge graph embedding by flexible translation . In arXiv:1505.05253, 2015.
 - [23] Xiao Han, Huang Minlie, Zhu Xiaoyan. From One Point to a Manifold: Knowledge Graph Embedding for Precise Link Prediction. Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-16), 2016. 1315-1321.

-
- [24] Sun Z, Deng ZH, Nie JY, Tang J. Rotate: Knowledge graph embedding by relational rotation in complex space. In International Conference on Learning Representations. 2019.
 - [25] Sadeghi A, Graux D, Yazdi HS, Lehmann J. MDE : Multiple Distance Embeddings for Link Prediction in Knowledge Graphs . arXiv preprint arXiv:1905.10702, 2019.
 - [26] Chami I, Wolf A, Juan D, Sala F, Ravi S, Ré C. Low-Dimensional Hyperbolic Knowledge Graph Embeddings . arXiv preprint arXiv:2005.00545, 2020.
 - [27] Zhang S, Tay Y, Yao L, Liu Q. Quaternion knowledge graph embeddings. Proceedings of the 33rd Conference on Neural Information Processing Systems, 2019. 1–14.
 - [28] Lu H, Hu H. DensE : An Enhanced Non-Abelian Group Representation for Knowledge Graph Embedding . arXiv preprint arXiv:2008.04548, 2020.
 - [29] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2D Knowledge Graph Embeddings. 32nd AAAI Conference on Artificial Intelligence, 2018. 1811–1818.
 - [30] Hu Z, Huang P, Deng Y, Gao Y, Xing EP. Entity hierarchy embedding. ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference, 2015. 1292–1300. [doi: 10.3115/v1/P15-1125]
 - [31] Zhang Z, Cai J, Zhang Y, Wang J. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction. Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020. 3065–3072. [doi: 10.1609/aaai.v34i03.5701]
 - [32] Balažević I, Allen C, Hospedales T. Multi-relational poincaré graph embeddings. Advances in Neural Information Processing Systems, 2019.
 - [33] He S, Liu K, Ji G, Zhao J. Learning to represent knowledge graphs with Gaussian embedding. International Conference on Information and Knowledge Management, Proceedings, 2015. 623–632. [doi: 10.1145/2806416.2806502]
 - [34] Xiao H, Huang M, Hao Y, Zhu XY. TransA: An adaptive approach for knowledge graph embedding . arXiv preprint arXiv:1509.05490, 2015.
 - [35] Nickel M, Tresp V, Kriegel HP. A three-way model for collective learning on multi-relational data. Proceedings of the 28th International Conference on Machine Learning, ICML 2011, 809–816.
 - [36] Yang B, Yih WT, He X, Gao J, Deng L. Embedding entities and relations for learning and inference in knowledge bases. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015.
 - [37] Kadlec R, Bajgar O, Kleindienst J. Knowledge base completion: Baselines strike back. In Proceedings of the 2nd Workshop on Representation Learning for NLP, 2017. 69–74.
 - [38] Trouillon T, Welbl J, Riedel S, Ciaussier E, Bouchard G. Complex embeddings for simple link prediction. In Proceedings of the 33rd International Conference on Machine Learning, ICML 2016. 3021–3032.
 - [39] Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs. In Proceedings of the 30th AAAI Conference on Artificial Intelligence, AAAI 2016. 1955–1961.
 - [40] Kazemi SM, Poole D. Simple embedding for link prediction in knowledge graphs. Advances in Neural Information Processing Systems, 2018. 4284–4295.
 - [41] Liu H, Wu Y, Yang Y. Analogical Inference for Multi-relational Embeddings. In Proceedings of the 34rd International Conference on Machine Learning, 2017.
 - [42] Balažević I, Allen C, Hospedales TM. Tucker: Tensor factorization for knowledge graph completion. EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, 2020. 5185–5194.
 - [43] Xu W, Zheng S, He L, Shao B, Yin J, Liu TY. SEEK: Segmented Embedding of Knowledge Graphs . arXiv preprint arXiv:2005.00856, 2020.
 - [44] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. In Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998. doi: 10.1109/5.726791. [doi: 10.1109/5.726791]
 - [45] Sabour S, Frosst N, Hinton GE. Dynamic routing between capsules. Advances in Neural Information Processing Systems, 2017. 3857–3867.

-
- [46] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2D knowledge graph embeddings. Proceedings of the 32nd AAAI Conference on Artificial Intelligence, AAAI 2018, 1811–1818.
 - [47] Nguyen DQ, Nguyen TD, Nguyen DQ, Phung D. A novel embedding model for knowledge base completion based on convolutional neural network. NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018. 327–333. [doi: 10.18653/v1/N18-2053]
 - [48] Jiang X, Wang Q, Wang B. Adaptive convolution for multi-relational learning. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 2018. 978–987. [doi: 10.18653/v1/N19-1103]
 - [49] Vashishth S, Sanyal S, Nitin V, Agrawal N, Talukdar P. InteractE: Improving convolution-based knowledge graph embeddings by increasing feature interactions. 34st AAAI Conference on Artificial Intelligence, AAAI 2020. [doi: 10.1609/aaai.v34i03.5694]
 - [50] Nguyen DQ, Vu T, Nguyen TD, Nguyen DQ, Phung D. A Capsule Network-based Embedding Model for Knowledge Graph Completion and Search Personalization. Proceedings of NAACL-HLT 2019, 2180–2189. [doi: 10.18653/v1/N19-1226]
 - [51] Guo L, Sun Z, Hu W. Learning to Exploit Long-term Relational Dependencies in Knowledge Graphs. Proceedings of the 36th International Conference on Machine Learning, 2019. 4441–4450.
 - [52] Bruna J, Zaremba W, Szlam A, LeCun Y. Spectral networks and deep locally connected networks on graphs. In Proceedings of the 2nd International Conference on Learning Representations, ICLR 2014.
 - [53] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. Advances in Neural Information Processing Systems, 2016. 3844–3852.
 - [54] Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 2017. 1025–1035.
 - [55] Sanyal S, Balachandran J, Yadati N, Kumar A, Rajagopalan P, Sanyal S, Talukdar P. MT-CGCNN: Integrating crystal graph convolutional neural network with multitask learning for material property prediction . arXiv preprint arXiv:1811.05660, 2018.
 - [56] Schlichtkrull M, Kipf TN, Bloem P, van den Berg R, Titov I, Welling M. Modeling Relational Data with Graph Convolutional Networks . Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2018, 10843(1):593–607. [doi: 10.1007/978-3-319-93417-4_38]
 - [57] Gilmer J, Schoenholz SS, Riley PF, Vinyals O, and Dahl GE. Neural message passing for quantum chemistry. In Proceedings of the 34th International Conference on Machine Learning - Volume 70, ICML 2017, pp. 1263–1272.
 - [58] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In Proceedings of the 5th International Conference on Learning Representations, ICLR 2017, 1–14.
 - [59] Ye R, Li X, Fang Y, Zang H, Wang M. A vectorized relational graph convolutional network for multi-relational network alignment. IJCAI International Joint Conference on Artificial Intelligence, 2019. 4135–4141. [doi: 10.24963/ijcai.2019/574]
 - [60] Cai L, Yan B, Mai G, Janowicz K, Zhu R. TransGCN:coupling transformation assumptions with graph convolutional networks for link prediction. In Proceedings of the 10th International Conference on Knowledge Capture, K-CAP 2019. [doi: 10.1145/3360901.3364441]
 - [61] Vashishth S, Sanyal S, Nitin V, Talukdar P. Composition-based multi-relational graph convolutional networks [C] In Proceedings of the 7th International Conference on Learning Representations, ICLR 2019.
 - [62] Veličković P, Casanova A, Liò P, Cucurull G, Romero A, Bengio Y. Graph attention networks. In Proceedings of the 6th International Conference on Learning Representations, ICLR 2018.
 - [63] Shang C, Tang Y, Huang J, Bi J, He X, Zhou B. End-to-end structure-aware convolutional networks for knowledge base completion. Proceedings of the 33th AAAI Conference on Artificial Intelligence, 2019. 3060–3067. [doi: 10.1609/aaai.v33i01.33013060]
 - [64] Nathani D, Chauhan J, Sharma C, Kaul M. Learning attention-based embeddings for relation prediction in knowledge graphs. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, ACL 2019. 4710–4723. [doi: 10.18653/v1/P19-1466]
 - [65] Liu X, Tan H, Chen Q, Lin G. RAGAT: Relation Aware Graph Attention Network for Knowledge Graph Completion in IEEE Access, 2021, vol. 9, pp. 20840-20849. [doi: 10.1109/ACCESS.2021.3055529]

-
- [66] Sun Z, Wang C, Hu W, Chen M, Dai J, Zhang W, Qu Y. Knowledge Graph Alignment Network with Gated Multi-hop Neighborhood Aggregation. Proceedings of the 34th AAAI Conference on Artificial Intelligence, AAAI 2020. [10.1609/aaai.v34i01.5354]
 - [67] Socher R, Chen D, Manning CD, Ng AY. Reasoning with neural tensor networks for knowledge base completion. In Advances in Neural Information Processing Systems, 2013.
 - [68] Zhang D, Yuan B, Wang D, Liu R. Joint Semantic Relevance Learning with Text Data and Graph Knowledge. Proceedings of the 3rd Workshop on Continuous Vector Space Models, 2015. 32–40. [doi: 10.18653/v1/W15-4004]
 - [69] Xie R, Liu Z, Jia J, Luan H, Sun M. Representation Learning of Knowledge Graphs with Entity Descriptions. Proceedings of the 30th AAAI Conference on Artificial Intelligence, AAAI 2016. 2659–2665.
 - [70] Xu J, Qiu X, Chen K, Huang X. Knowledge graph representation with jointly structural and textual encoding. IJCAI International Joint Conference on Artificial Intelligence, 2017. 1318–1324. [doi: 10.24963/ijcai.2017/183]
 - [71] Wang Y, Zhang H, Shi G, Liu Z, Zhou Q. A Model of Text-Enhanced Knowledge Graph Representation Learning with Mutual Attention . IEEE Access, 2020,8:52895–52905. [doi: 10.1109/ACCESS.2020.2981212]
 - [72] Daza D, Cochez M, Groth P. Inductive Entity Representations from Text via Link Prediction. In Proceedings of the Web Conference 2021 (WWW '21), Ljubljana, Slovenia, 2021. [doi: 10.1145/3442381.3450141]
 - [73] Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019.
 - [74] Xie R, Liu Z, Sun M. Representation learning of knowledge graphs with hierarchical types. IJCAI International Joint Conference on Artificial Intelligence, 2016. 2965–2971.
 - [75] Tang X, Chen L, Cui J, Wei B. Knowledge representation learning with entity descriptions, hierarchical types, and textual relations . Information Processing and Management, 2019,56(3):809–822. [doi: 10.1016/j.ipm.2019.01.005]
 - [76] ROSSI A, FIRMANI D, MATINATA A, MERALDO P, BARBOSA D. Knowledge Graph Embedding for Link Prediction: A Comparative Analysis . ACM Transactions on Knowledge Discovery from Data, 2021,15(1):1-49. [doi: 10.1145/3424672]
 - [77] Sun Z, Vashisht S, Sanyal S, Talukdar P, Yang Y. A re-evaluation of knowledge graph completion methods. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, pp. 5516–5522. [doi: 10.18653/v1/2020.acl-main.489]
 - [78] Wang Z, Lai KP, Li P, Bing L, Lam W. Tackling long-tailed relations and uncommon entities in knowledge graph completion. EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, 2020. 250–260. [doi: 10.18653/v1/D19-1024]
 - [79] Zhang W, Chen J, Paudel B, Zhu H, Wang L, Zhang W, Bernstein A, Chen H. Iteratively learning embeddings and rules for knowledge graph reasoning. The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019, 2366–2377. [doi: 10.1145/3308558.3313612]
 - [80] Sheng J, Guo S, Chen Z, Yue J, Wang L, Liu T, Xu H. Adaptive Attentional Network for Few-Shot Knowledge Graph Completion. EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, 2020. [doi: 10.18653/v1/2020.emnlp-main.131]
 - [81] Zhang C, Yao H, Huang C, Jiang M, Li Z, Chawla NV. Few-shot knowledge graph completion. Proceedings of the 34th AAAI Conference on Artificial Intelligence, AAAI 2020.
 - [82] Kazemi SM, Goel R, Jain K, Kobyzhev I, Sethi A, Forsyth P, Poupart P. Relational representation learning for dynamic (knowledge) graphs: A survey . arXiv preprint arXiv:1905.11485, 2019.
 - [83] Zhang Z, Zhuang F, Qu M, Lin F, He Q. Knowledge graph embedding with hierarchical relation structure. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018, 3198–3207. [doi: 10.18653/v1/D18-1358]