

面向图神经网络的知识图谱嵌入研究进展

延照耀, 丁苍峰, 马乐荣⁺, 曹璐, 游浩

延安大学 数学与计算机科学学院, 陕西 延安 716000

⁺通信作者 E-mail: mlr@yau.edu.cn

摘要:随着图神经网络的发展,基于图神经网络的知识图谱嵌入方法日益受到研究人员的关注。相比传统的方法,它可以更好地处理实体的多样性和复杂性,并捕捉实体的多重特征和复杂关系,从而提高知识图谱的表示能力和应用价值。首先概述知识图谱的发展历程,梳理知识图谱和图神经网络的基本概念;其次着重讨论基于图卷积、图神经、图注意力以及图自编码器的知识图谱嵌入的设计思路和算法框架;然后描述图神经网络的知识图谱嵌入在链接预测、实体对齐、知识推理以及知识图谱补全等任务中的性能,同时补充图神经网络在常识性知识图谱中的一些研究;最后进行全面性的总结,并针对知识图谱嵌入存在的一些问题和挑战,勾画未来研究方向。

关键词:知识图谱;知识图谱嵌入;图神经网络;表示学习

文献标志码:A **中图分类号:**TP39

Advances in Knowledge Graph Embedding Based on Graph Neural Networks

YAN Zhaoyao, DING Cangfeng, MA Lerong⁺, CAO Lu, YOU Hao

College of Mathematics and Computer Science, Yan'an University, Yan'an, Shaanxi 716000, China

Abstract: As graph neural networks continue to develop, knowledge graph embedding methods based on graph neural networks are receiving increasing attention from researchers. Compared with traditional methods, they can better handle the diversity and complexity of entities, and capture the multiple features and complex relationships of entities, thereby improving the representation ability and application value of knowledge graphs. This paper firstly outlines the development history of knowledge graphs and the basic concepts of knowledge graphs and graph neural networks. Secondly, it focuses on discussing the design ideas and algorithm frameworks of knowledge graph embedding based on graph convolution, graph neural networks, graph attention, and graph autoencoders. Then, it describes the performance of graph neural network knowledge graph embedding in tasks such as link prediction, entity alignment, knowledge graph reasoning, and knowledge graph completion, while supplementing some research on commonsense knowledge graphs with graph neural networks. Finally, this paper makes a comprehensive summary, and future research directions are outlined with respect to some challenges and issues in knowledge graph embedding.

Key words: knowledge graph; knowledge graph embedding; graph neural network; representation learning

基金项目:国家自然科学基金(62262067, 61866038);陕西省级人才项目(YAU202213065, CXY202107);陕西省教育厅自然科学专项(22JK0622);延安大学十四五重大科研项目(2021ZCQ012)。

This work was supported by the National Natural Science Foundation of China (62262067, 61866038), the Talent Project of Shaanxi Province (YAU202213065, CXY202107), the Natural Science Special Project of Shaanxi Provincial Education Department (22JK0622), and the Major Scientific Research Projects of Yan'an University at the 14th Five Year Plan (2021ZCQ012).

收稿日期:2022-12-22 **修回日期:**2023-03-29

“知识”二字早在甲骨文中就已经出现,其中“知”本义为“谈论和传授狩猎作战的经验”,“识”本义为“识辨指认武器”。进入20世纪,“知识”被释义为“人们在社会实践中所获得的认识和经验的综合”。知识对人工智能而言同等重要,谷歌在2012年发布知识搜索产品 Google knowledge graph,提出“Things, Not Strings”的理念,知识图谱(knowledge graph, KG)才被正式提出,将知识图谱定义为一种用图模型来描述知识和构建世界万物之间的关联关系的技术方法^[1]。自此,以知识表示、挖掘和推理等知识图谱技术成为人工智能的研究热点。

知识图谱表示(knowledge representation learning, KRL)^[2]旨在对知识图谱进行建模,从而可以方便知识计算。传统的知识图谱表示使用的是基于离散符号的表示方式,以三元组的格式对知识进行组织,虽然这些符号化的表达能力可以非常有效地将数据结构化,但是在计算机中的表达相关的语义信息和进行语义计算等方面面临着巨大的挑战。为了解决计算效率低下和数据稀疏性强这两个挑战,在词嵌入(word embedding)^[3]的启发下,研究者将注意力集中在将知识图谱中的实体和关系映射到低维稠密实值的向量空间中,并包含一些语义层面的信息。这种将知识图谱中实体和关系的信息映射到低维稠密实值的向量空间的研究方法称为知识图谱嵌入(KG embedding, KGE)。

类似于词向量,知识图谱嵌入也是通过机器学习的方法对模型进行学习的,多数知识图谱嵌入方法依靠知识图谱中的实体、关系的属性信息和图谱的结构信息进行训练。根据有关知识图谱嵌入的综述^[4],将知识图谱嵌入的方法分为基于转移思想、基于语义匹配、考虑附加信息、基于传统深度学习以及基于图神经网络。随着图神经网络(graph neural network, GNN)^[5]的兴起,一种利用深度学习直接学习图的结构化数据的框架,大量的基于图神经网络的知识图谱嵌入方法被提出。由于GNN对图数据结构具有非线性实体拟合的能力,在不同领域的问题上具有更高的精确度和更好的鲁棒性。此外,基于GNN的知识图谱嵌入方法可将知识图谱中基于领域的数据和业务有效地结合起来,展示出越来越丰富的实用价值,典型的下游应用包括问答系统^[6]、信息抽取^[7-8]以及推荐系统^[9]。

为此,本文梳理了基于图神经网络的知识图谱嵌入方法的不同设计思想,并对相应的方法进行了

总结,主要研究框架如图1所示。本文首先阐述了知识图谱和图神经网络的基本概念;然后按照图神经网络的主流算法将不同的知识图谱嵌入方法分为基于图卷积、基于图神经、基于图注意力以及基于图自编码器四种类别的知识图谱嵌入方法,并分析链接预测、实体对齐、知识推理以及知识图谱补全等下游任务中的性能;随后介绍常识性知识图谱的基本概念、开源数据集以及在图神经网络方面的研究;最后讨论了基于图神经网络的知识图谱嵌入方法在分布外泛化、动态知识图谱、多模态知识融入、网络深度与邻居节点数量的矛盾、图神经网络的可解释性以及图谱深层次信息的主要挑战和未来方向。

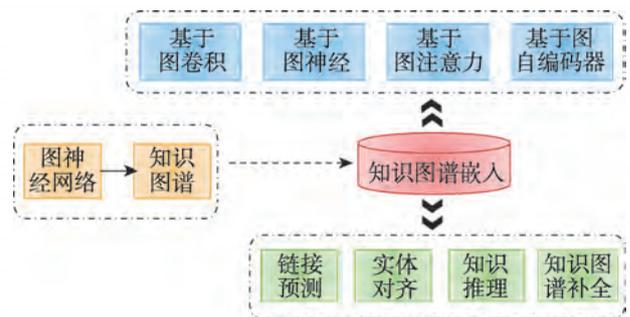


图1 研究内容框架

Fig.1 Research content framework

1 基本概念

本章介绍知识图谱嵌入和图神经网络的基本概念。

1.1 知识图谱嵌入

知识图谱(KG),又被定义为语义网络的知识库,是一种用图模型来描述客观事物和建模事物之间的关联关系的技术方法。一般地,知识图谱定义为 $KG = \{E, R, F\}$,其中, E 表示知识图谱包含的实体集合 $\{e_1, e_2, \dots, e_E\}$, R 表示不同实体之间的关系集合 $\{r_1, r_2, \dots, r_R\}$, F 表示知识图谱中的三元组集合 $\{f_1, f_2, \dots, f_T\}$,每个三元组 f 被定义为 (h, r, t) ,其中 h, r, t 分别表示头实体、关系和尾实体。

为了解决基于离散符号的知识图谱表示不能在计算机中表达相应的语义信息的挑战,在词向量的启发下,将知识图谱中的实体和关系的特征映射到低维的向量空间中,把这种研究领域称为知识图谱嵌入、知识图谱的表示学习。类似于词向量,知识图谱嵌入通过机器学习的方法对模型进行学习,与词袋模型、独热编码不同的是,知识图谱嵌入方法依赖于监督学习进行训练,在训练过程中学习语义信息。

知识图谱嵌入算法的发展经历了多个阶段,从传统的基于矩阵分解的方法到最近的基于深度学习的方法。其中,利用矩阵分解进行知识表示学习中的一个典型模型是 RESCAL^[10],其算法思想是将实体和关系都表示为三维张量,然后对实体和关系张量进行分解,学习低维嵌入表示。近年来,许多基于 RESCAL 的改进算法被提出,为了解决 RESCAL 参数问题,DistMult^[11]模型使用矩阵 W 来表示所有实体之间的关系,矩阵 W 的每一列对应一个实体,每一行对应实体之间的关系,最后实体嵌入向量与关系矩阵 W 相乘得到预测分数。为了解决 DistMult 无法处理非对称关系的问题,ComplEx^[12]在矩阵乘法中引入复数运算,而 HolE (holographic embedding)^[13]将知识图谱中的关系抽象为高阶张量的双线性关系,以此来捕获关系数据中的丰富交互。奠基性的基于转移的知识图谱嵌入方法 TransE (translating embedding)^[14]的设计思想就是将三元组中的头实体 h 经过关系 r 的转移后接近尾实体 t ,使其满足关系 $h+r \approx t$,例如:

$$\text{vec}(\text{罗纳尔多}) + \text{vec}(\text{国籍}) \approx \text{vec}(\text{巴西})$$

其中, vec 表示实体和关系嵌入向量。然而 TransE 在处理对称关系时失去了语义辨识的能力,为了解决该问题, RotatE (rotating embedding)^[15]通过将关系向量表示为旋转矩阵来有效地捕捉关系的方向性信息。在传统深度学习的知识图谱嵌入中, ConvE^[16]将知识图谱中的实体和关系嵌入到低维空间中,并通过卷积网络学习实体和关系之间的语义关联。

1.2 图神经网络

图神经网络将现有的神经网络方法拓展到了图结构数据处理领域,来学习、提取和挖掘图结构数据中的特征和模式^[17]。节点采用其特征和图中与其相关的节点进行描述,根据输入的节点邻居更新节点状态,它可以编码该节点的邻居节点信息和图的拓扑结构;因此,图神经网络的目标是为每个节点学习一个状态表示 $h_i \in \mathbf{R}^S$,利用这个状态表示可以生成模型的输出 o_i ,将所有的状态、输出、特征和节点特征堆叠起来,即可得到基础图神经网络模型的公式表达:

$$\begin{cases} H = F(H, X) \\ O = G(H, X_N) \end{cases} \quad (1)$$

其中, F 是全球转移函数, G 是全球输出函数, X_N 和 X 分别表示通过堆叠所有节点特征和所有特征构建的矩阵, H 和 O 分别表示所有节点的状态和所有节点的输出。根据巴拿赫不动点定理^[18],基础图神经网络

模型使用如下迭代方式更新节点的状态:

$$H^{t+1} = F(H^t, X) \quad (2)$$

其中, H^t 表示 H 的第 t 轮迭代。

图神经网络模型也可以用消息传递网络模型进行描述^[19],该模型在前向传播过程中可以分为消息传递和消息读取两个阶段。其中消息传递阶段又分为目标节点传递语义信息至邻居节点、聚合自身信息与邻居节点信息两个过程,这也是 GNN 的前向传播思想。为此, GNN 的前向传播思想可以统一建模为:

$$h_i^{k+1} = \text{Merge}(\{h_i^k\} \cup \text{aggr}(\{h_j^k\}_{j \in \mathcal{N}(i)})) \quad (3)$$

其中, $\text{Merge}(\cdot)$ 表示特征融合函数, $\text{aggr}(\cdot)$ 表示邻居特征聚合函数, $\mathcal{N}(i)$ 表示节点 i 的邻居集合, h^{k+1} 表示节点在第 k 层网络中输入的特征向量。

2 模型框架

本章将具体介绍不同类型的基于 GNN 的知识图谱嵌入算法的基本思想和模型。按照 GNN 衍生的许多变体,将知识图谱嵌入方法大致分类为:基于图卷积、基于图神经、基于图注意力以及基于图自编码器的方法。其中,图卷积是将卷积神经网络中的卷积运算推广到图数据结构,对实体和关系进行表示学习,进行卷积操作的主要方法为基于谱分解和节点空间变化;与图卷积网络注重于利用邻域信息不同的是,图神经网络采用消息传递机制来学习嵌入表示,更好地捕捉实体和关系之间的交互信息;图注意力是在基础 GNN 的消息传递过程中引入注意力机制,为每个相邻节点分配不同的注意力分数,使神经网络关注较为重要的相邻实体和关系;图自编码器是将输入的实体和关系信息作为学习目标,对实体和关系信息进行表示学习^[20-21],被广泛应用于无监督学习的图节点嵌入方法中。

变体 GNN 的知识图谱嵌入方法都是建立在传统知识图谱嵌入方法之上,利用变体 GNN 的图嵌入方法作为编码器学习图结构信息,然后采用 TransE、ComplEx 和 ConvE 等传统的知识图谱嵌入方法作为模型的解码器,将学习到的实体和关系的嵌入进行解码并在链接预测等下游任务中验证模型的性能。下文将按照上述分类的顺序依次介绍各类方法的思想以及相应的算法实现。

2.1 基于图卷积的知识图谱嵌入

本节从基础 GCN (graph convolutional network) 的思想入手介绍奠基性方法 R-GCN (relational graph convolutional network),然后从该方法的问题角度介

绍该方法的改进版本,理清基于图卷积的知识图谱嵌入模型的发展脉络。

虽然 GCN 在节点分类等下游任务中取得了不错的成绩,但基础 GCN 模型只是针对同构图,为了处理知识图谱这种特殊的异构图, Schlichtkrull 等人^[22]提出关系图卷积网络 R-GCN,与 GCN 不同的是, R-GCN 中边表示为不同的关系,为每个关系 r 引入相关联的权重矩阵 W_r ,在聚合邻居信息的过程中使用相应的关系权重矩阵更新邻居特征,其特征更新过程表示为:

$$h_i^{k+1} = \sigma \left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{C_{i,r}} W_r^{(k)} h_j^{(k)} + W_0^{(k)} h_i^{(k)} \right) \quad (4)$$

可以看出, R-GCN 将基础 GCN 模型扩展到了关系图中,但是每种关系关联相应的权重矩阵,造成参数量过大的问题,尤其是在处理关系类别极大的数据时。针对该问题, R-GCN 使用参数共享和稀疏化矩阵策略减少模型参数并防止过拟合。R-GCN 还提出一种 Encoder-Decoder 的框架,将基础图神经网络模型应用于链接预测任务中。如图 2 所示, R-GCN 作为 Encoder 学习实体特征,传统知识图谱嵌入算法 DistMult 作为 Decoder 计算三元组得分,后续大多数基于图神经网络的知识图谱模型将 Encoder-Decoder 框架作为模型的基础框架。

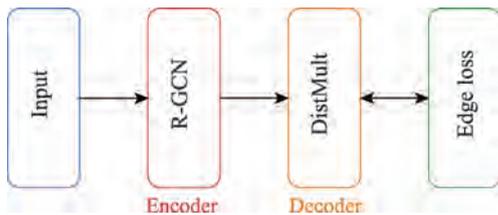


图2 R-GCN用于链接预测任务框架

Fig.2 R-GCN for linking prediction task framework

R-GCN 虽然在聚合邻居信息时考虑了不同关系信息,但是忽略了关系的方向和异构图中不同实体类型可能具有不同的特征空间的问题。为了应对上述问题, Ye 等人^[23]提出一种量化的关系图卷积网络 (variational relational graph convolutional networks, VR-GCN), 将 TransE 的思想引入 R-GCN 模型中,同时考虑关系的类别信息和方向信息,以同时学习实体和关系的嵌入, VR-GCN 的特征更新函数如下:

$$h_i^{l+1} = \sigma \left(\frac{1}{d} \left(\sum_{r \in \mathcal{N}_i^+} (h_i^l - h_r^l) + \sum_{r \in \mathcal{N}_i^-} (h_i^l + h_r^l) \right) + h_i^l \right) W^l \quad (5)$$

TransGCN^[24]在 VR-GCN 的基础上,利用 RotatE 思想对聚合函数进行了扩展,提出一种新的处理异

构关系的方法。除此之外, CompGCN (compositional graph convolutional network)^[25]利用基于语义匹配的知识图谱嵌入算法 DistMult 和 HoE 思想,提出新的聚合函数,将实体和关系联合嵌入到关系图中,并且还根据关系的数量进行扩展,以解决模型的过度参数化问题。但大多数 GCN 模型关注实体嵌入过程中加入关系信息,而缺乏关系嵌入过程中加入实体信息。为了解决这一局限性, Yu 等人^[26]提出一种基于知识嵌入的图卷积网络 (knowledge-enhanced graph convolutional network, KE-GCN), 结合 GCN 的消息传递和知识图谱嵌入方法的优势,将 CompGCN、R-GCN 和 WGCN (weighted graph convolutional network)^[27]方法进行统一,如图 3 所示,与之前基于图卷积的方法相比, KE-GCN 实现简单,性能更好。各模型的实体更新和关系更新函数见表 1,其中 E_h^l 、 E_r^l 、 E_t^l 表示头实体 h 、关系 r 和尾实体 t 在 l 层的嵌入, m_i^{l+1} 表示邻居实体的聚合表示, σ 表示实体更新和关系更新的激活函数, W 表示权重矩阵。

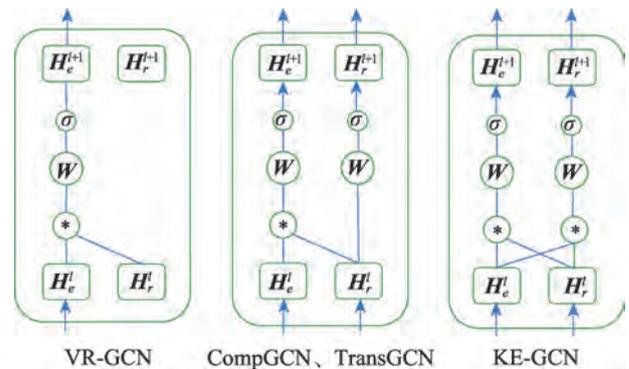


图3 VR-GCN、CompGCN、TransGCN 和 KE-GCN 模型框架

Fig.3 Model frameworks of VR-GCN, CompGCN, TransGCN and KE-GCN

在研究动态知识图谱对齐工作中, Yan 等人^[28]使用图卷积模型提出 DINGAL,并在 DINGAL 的基础上引出一系列知识图谱嵌入方法。主要思想是将 GCN 的参数矩阵看作特征转换算子,并将转换过程与聚合过程进行解耦,在此基础上提出 DINGAL-B,用于静态知识图谱对齐。随之提出两种有效的动态知识图谱对齐方法 DINGAL-O (利用之前的参数矩阵更新实体嵌入) 与 DINGAL-U (利用新的关系微调参数矩阵), DINGAL 系列算法细节如图 4 所示,从而有效解决动态知识图谱的实体嵌入问题,这项工作也被认为是第一个研究动态知识图谱对齐问题的工

表1 R-GCN、WGCN、CompGCN与KE-GCN更新函数描述

Table 1 Description of R-GCN, WGCN, CompGCN and KE-GCN update functions

模型	实体更新	关系更新
R-GCN	$m_i^{l+1} = \sum_{(h,r) \in \mathcal{N}_{in}(i)} W_r^l E_h^l + \sum_{(h,r) \in \mathcal{N}_{out}(i)} W_r^l E_h^l, E_i^{l+1} = \sigma(m_i^{l+1} + W_0^l E_i^l)$	No
WGCN	$m_i^{l+1} = \sum_{(h,r) \in \mathcal{N}_{in}(i)} W_r^l (\alpha_r^l E_h^l) + \sum_{(h,r) \in \mathcal{N}_{out}(i)} W_r^l (\alpha_r^l E_h^l), E_i^{l+1} = \sigma(m_i^{l+1} + W^l E_i^l)$	No
CompGCN	$m_i^{l+1} = \sum_{(h,r) \in \mathcal{N}_{in}(i)} W_r^l \phi_{in}(E_h^l, E_i^l) + \sum_{(h,r) \in \mathcal{N}_{out}(i)} W_r^l \phi_{out}(E_h^l, E_i^l), E_i^{l+1} = \sigma(m_i^{l+1} + W_0^l E_i^l)$	$E_r^{l+1} = W_r^l E_r^l$
KE-GCN	$m_i^{l+1} = \sum_{(h,r) \in \mathcal{N}_{in}(i)} W_r^l \frac{\partial f_{in}(E_h^l, E_i^l, E_i^l)}{\partial E_i^l} + \sum_{(h,r) \in \mathcal{N}_{out}(i)} W_r^l \frac{\partial f_{out}(E_i^l, E_r^l, E_i^l)}{\partial E_i^l}, E_i^{l+1} = \sigma(m_i^{l+1} + W_0^l E_i^l)$	$m_r^{l+1} = \sum_{(u,r) \in \mathcal{N}_{in}(r)} \frac{\partial f_r(E_u^l, E_r^l, E_r^l)}{\partial E_r^l}, E_r^{l+1} = \sigma(W_r^l (m_r^{l+1} + E_r^l))$

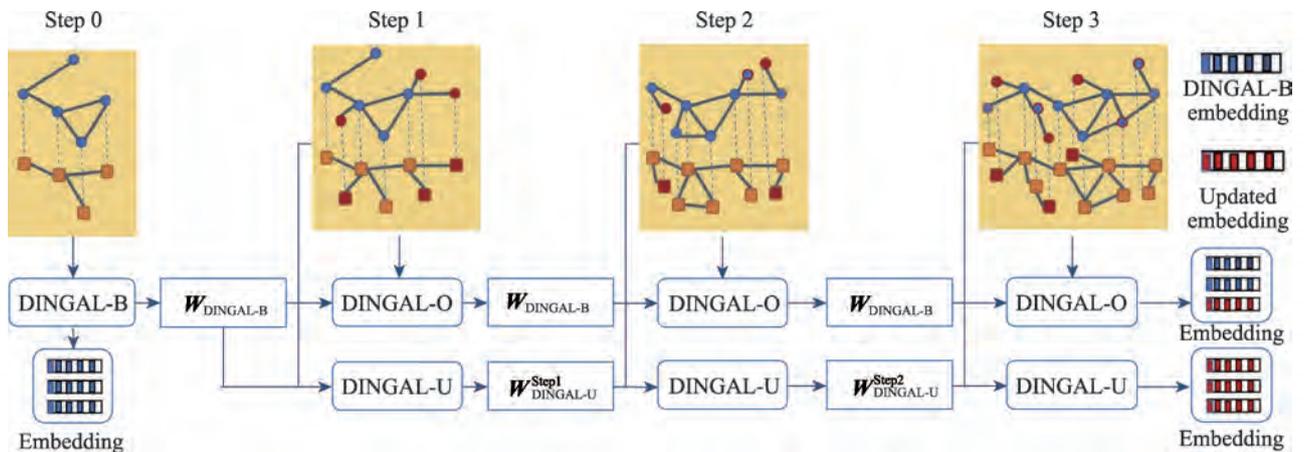


图4 DINGAL系列算法细节

Fig.4 Algorithm details of DINGAL series

基于图卷积的知识图谱嵌入方法由于运算操作简单、参数量适中、学习效率高等优势,成为了知识图谱嵌入领域非常活跃的研究方向之一。然而图卷积网络本身也存在诸多缺点,比如邻居实体重要性分配、动态图处理等,使得后续工作可以针对这些缺点对基于图卷积的知识图谱嵌入方法进一步改进。

2.2 基于图神经的知识图谱嵌入

上述基于图卷积的知识图谱嵌入方法使用卷积操作来捕捉局部邻域信息,核心思想是利用实体和其邻居实体之间的关系来表示实体的特征。而基于图神经网络的知识图谱嵌入方法则将图作为输入,通过对图中节点的消息传递和聚合来学习节点的嵌入向量。它可以充分利用图的全局结构信息,同时也考虑了节点之间的复杂依赖关系。此外,基于图神经网络的知识图谱嵌入方法还可以根据时间信息对实体的嵌入向量进行动态更新,处理时序知识图谱,从而更好地表达知识图谱的演化过程。本节将介绍基于图神经网络的知识图谱嵌入方法。近几年,GNN在知识图谱嵌入的研究发展较为迅速,主要

由于GNN在建模图数据方面所表现出优越的性能。

先前的R-GCN模型还存在着表示能力低、扁平化堆叠和对噪音的鲁棒性差三个缺点,为此,Wang等人^[29]提出一种多级图神经网络(multi-level graph convolutional neural network, M-GCN)来解决上述缺点。R-GCN利用均值聚合器聚合邻居信息,与R-GCN不同的是,M-GCN使用多层感知机设计GNN层进行邻居聚合操作,以提高GNN层的表示能力。M-GCN的更新过程如下:

$$h_v^{(k)} = \text{MLP}^{(k)}((1 + \varepsilon^{(k)}) \cdot h_{r_0,v}^{k-1} + \sum_{u \in \mathcal{N}_v^r} h_{r,u}^{k-1}) \quad (6)$$

其中, \mathcal{N}_v^r 表示关系 r 下节点 v 的邻居集合, $h_{r,u}^{k-1}$ 表示关系 r 下每个邻居节点 u 传递的消息, $h_{r_0,v}^{k-1}$ 表示自连接消息。在此基础上,将知识图谱粗化为一系列图,并以多层方式堆叠GNN层建模知识图谱的层级结构。

前期的一些工作是将知识图谱嵌入单曲率空间,比如欧几里德空间和双曲空间,但忽略了知识图谱特有的异构特性,因此无法捕获其拓扑结构。为了解决这个问题,Wang等人^[30]提出一种混合曲率关

系图神经网络模型(mixed-curvature multi-relational graph neural network, M^2GNN),以解决单曲率空间忽略知识图谱的异构特性的问题。具体而言,将多个单曲率空间(例如球面、双曲线和欧几里德空间)的流形乘积来定义和构造混合曲率空间,以此来建模知识图谱的各种异构结构。然而,混合曲率空间的构造需要手动定义,需要额外的数据分析,定义不当会无法准确地捕捉知识图谱的异构特性。为了解决这个问题, M^2GNN 将混合曲率设置为可训练的参数,利用GNN作为嵌入更新模块,用于聚合实体的邻居和关系特征,并更新每个实体和关系的嵌入表示,以此来提高知识图谱嵌入的性能,基于图神经的更新函数定义如下:

$$e'_i = GNN(e_j, r_k | j \in \mathcal{N}_i, k \in \mathcal{R}_{ij}) \quad (7)$$

其中, e'_i 定义为目标实体 e_i 更新后的嵌入, r_k 表示实体 e_i 和实体 e_j 之间的关系, \mathcal{N}_i 表示实体 e_i 的邻居实体, \mathcal{R}_{ij} 是实体 e_i 和实体 e_j 的关系集合。

于此同时,马尔可夫逻辑网络(Markov logic networks, MLNs)^[31]将逻辑规则和概率图模型结合起来解决知识图谱的问题,但是MLNs中的推理规则计算太过密集,无法应用在工业中。随着GNN的兴起,提出一种结合MLNs和GNN的变体模型ExpressGNN^[32],使用图神经网络在马尔可夫逻辑网络中进行推理,充分利用GNN在图嵌入学习的优点和MLNs在逻辑推理的优势,ExpressGNN的聚合函数如下:

$$\begin{cases} \alpha'_t = MLP_1(h'_t, r; \theta_1) \\ \alpha'_t{}^{+1} = AGG(\{\alpha'_t\}_{t:r(h,t) \in O}) \\ h'_t{}^{+1} = MLP_2(h'_t, \alpha'_t{}^{+1}; \theta_2) \end{cases} \quad (8)$$

其中, α'_t 表示实体 t 的邻居聚合信息, h'_t 表示实体 t 在 l 层的嵌入表示, θ_1 、 θ_2 为GNN参数并在整个图谱中共享, $MLP(\cdot)$ 和 $AGG(\cdot)$ 表示消息传递和消息聚合函数。另一个知识图谱嵌入的路线是逻辑归纳问题,一种试图导出给定知识图谱的概率逻辑规则,其中代表性的框架为GraIL^[33],从包含两个目标实体的子图中预测两个实体的关系,最后利用多关系GNN对三元组进行评分,启发于多关系模型R-GCN, GraIL的聚合函数如下:

$$\begin{cases} \alpha'_t = \sum_{r=1}^R \sum_{s \in \mathcal{N}(t)} \alpha'_{rr, st} W_r^l h_s^{l-1} \\ h'_t = \text{ReLU}(W_{\text{self}}^l h_t^{l-1} + \alpha'_t) \end{cases} \quad (9)$$

其中, α'_t 表示实体 t 的邻居聚合信息, h'_t 表示实体 t 在 l 层的嵌入表示。但是GraIL太过依赖启发式和评

分函数,因此GNN的归纳能力并没有充分利用。与GraIL不同的是,INDIGO^[34]将知识图谱以透明的方式完全编码在GNN中,并且可以直接从GNN的最后一层读取预测的三元组,从而不需要额外的评分函数,以此来解决太过依赖启发式和评分函数的缺点。

上述工作都是在链接预测任务中测试知识图谱嵌入方法的性能,很少考虑实体对齐这一关键技术,实体对齐通常也存在无法处理异构特性的问题。为了解决这一问题,研究者提出了一种多通道图神经网络模型(multi-channel graph neural network, MuGNN)^[35],通过多通道对两个知识图谱进行编码来学习基于对齐的知识图谱嵌入,每个通道使用不同的关系加权策略对知识图谱进行编码,其第 $l+1$ 层与第 i 个通道的隐层状态表示为 $H_i^{l+1} = GNN(A_i, H^l, W_i)$,其中 A 表示图的邻接矩阵, H 是当前实体的嵌入表示, W 为可学习的参数。此外还进行推断和传递并一致性地完成两个知识图谱的规则知识,以此协调两个知识图谱的结构差异性。但MuGNN在低维空间中嵌入实体,然后通过向量表示获得实体对齐,尽管之前的工作取得了持续的改进,但其性能仍然令人不满意。在知识图谱的实体对嵌入(entity-pair embedding approach, EPEA)^[36]方法中引入两个知识图谱的成对联通图(pairwise connectivity graph, PCG),其节点是实体对,边对应关系,然后学习节点(实体对)嵌入,并利用GNN增强实体对的邻居实体之间的消息传递,基于GNN的特征提取公式为 $H^{l+1} = \text{ReLU}(W^{l+1} \otimes H^l + b^{l+1})$, H^l 为第 l 层的输出, \otimes 为卷积算子,以此学习理想的实体对嵌入。另一方面,当前的基于图神经网络的知识图谱嵌入方法要么忽略实体间的差异,要么通过实体或者三元组来表示关系,并没有对关系中的元语义信息进行建模。Mao等人^[37]提出一种新的元关系感知实体对齐方法(meta relation-aware entity alignment, MRAEA),通过关注实体的出度与入度邻居以及连接关系的元语义信息,直接对跨语言知识图谱进行建模。此外,MRAEA还采用双向迭代策略处理基于对齐的不对称性。

2.3 基于图注意力的知识图谱嵌入

本节介绍基于图注意力的知识图谱嵌入方法,基于图注意力的方法是通过引入注意力机制,为每个相邻实体和关系分配不同的注意力分数,使神经网络关注较为重要的邻居实体和关系,以此来学习三元组的特征。

基于图卷积的模型的缺点在于平等对待知识图

谱中的关系,而忽略了每个关系拥有不同的重要性,这可能会限制模型的扩展性。启发于基于GCN的知识图谱嵌入在聚合邻居信息过程中加入权重参数,以实现不同重要性的思想,如图5所示,SACN(structure-aware convolutional networks)^[27]使用加权图卷积网络WGCN,利用知识图谱实体结构、实体属性和关系类型学习关系权重参数。然而SACN只考虑了关系类别,没有考虑邻居实体的关系,为了同时考虑邻居实体和邻居关系拥有不同的重要性,Nathani等人^[38]参考GAT(graph attention network)的思想聚合邻居信息,提出一种新的基于注意力的知识图谱嵌入方法,与之前规则不同的是,Nathani等人直接聚合邻居三元组的信息,而不是聚合邻居实体的信息,聚合函数如下:

$$h_i = \sigma \left(\frac{1}{M} \sum_{m=1}^M \sum_{j \in \mathcal{N}_i^r} \sum_{r \in \mathcal{R}_{ij}} \alpha_{hr}^m c_{hr}^m \right) \quad (10)$$

其中, c_{hr}^m 为三元组的嵌入表示, α_{hr}^m 是三元组的注意力系数。不同的是, r-GAT(relational graph attention network)^[39]通过多通道学习实体的嵌入表示,不同的通道对应不同的实体语义信息,并且利用关系特征聚合邻居信息,很好地处理了复杂的关系图。在实体对齐任务中,一种分散的知识图谱嵌入方法 decent-RL^[40]利用分散注意力网络(decentralized attention network, DAN),依赖目标实体的邻居实体而忽略嵌入要求来计算注意力分数。上述方法虽然在知识图谱嵌入中引入了注意力机制,并没有将知识图谱的异构特性加入模型之中。相比之下,Liu等人^[41]提出一种关系感知的图注意力网络模型(relation aware graph attention network, RAGAT),该模型利用知识图

谱的异构特性,引入特定于关系的网络参数 θ_r ,为不同关系建模相应的消息函数,自适应地学习不同关系下相邻实体的信息,RAGAT的消息函数定义如下:

$$C_{(h,r,i)}^r = \phi_r(e_h, e_r, e_i, \theta_r) \quad (11)$$

其中, ϕ_r 是聚合函数, e_h 和 e_i 表示实体的嵌入, e_r 表示关系的嵌入。

不同关系类别下,并非中心实体的所有邻居实体都具有相同的重要性,为此提出一种层级注意力的图神经网络模型(relational graph neural network with hierarchical attention, RGHAT)^[42]。除了在关系层计算不同关系类别的权重之外,RGHAT还加入了实体层的注意力机制,强调不同关系下的邻居实体具有不同的重要性。同时,异构关系注意力网络(heterogeneous relation attention network, HRAN)^[43]模型也采用一种包含实体级和关系级的层次结构,生成知识图谱嵌入。与RGHAT不同的是,HARN是基于关系路径下为邻居实体和关系分配不同的重要性。在知识推理任务中,Zhao等人^[44]也提出一种分层的注意力机制模型(target relational attention-oriented reasoning, TRAR),实体级注意力使用经典的GAT聚合邻居信息,子图级注意力根据推理任务中的关系决定注意力分数,利用注意力机制使模型更关注与目标实体相匹配的关系,从而获得更好的实体嵌入表示,模型从局部领域到层次结构细节如图6所示。

然而,前期的一些工作主要依赖于节点的一阶邻居信息,忽略了知识图谱丰富的局部结构,如图7所示的循环结构、三角结构和星型结构,因此没能更好地捕获这些结构下的各种语义信息。Ji等人为了

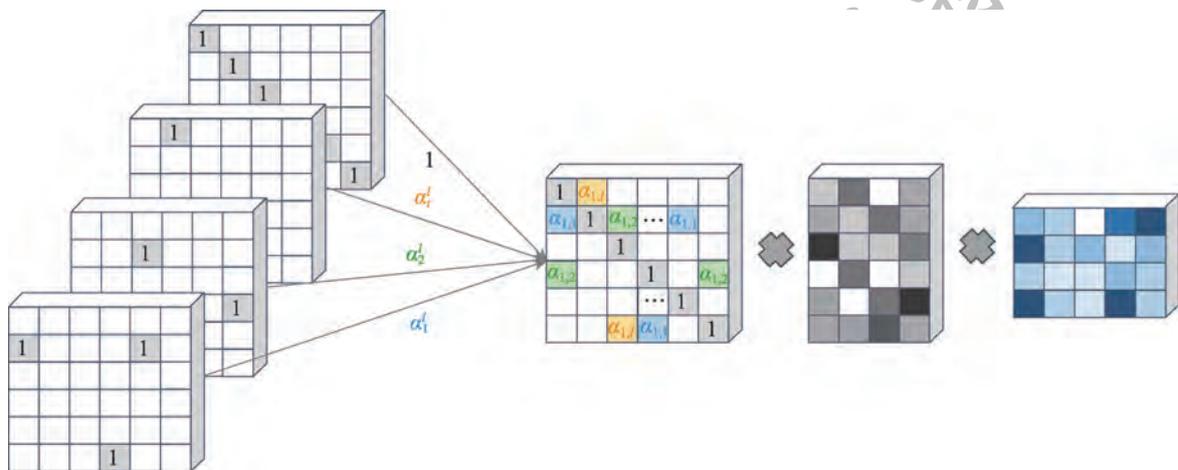


图5 WGCN模型框架
Fig.5 WGCN model framework

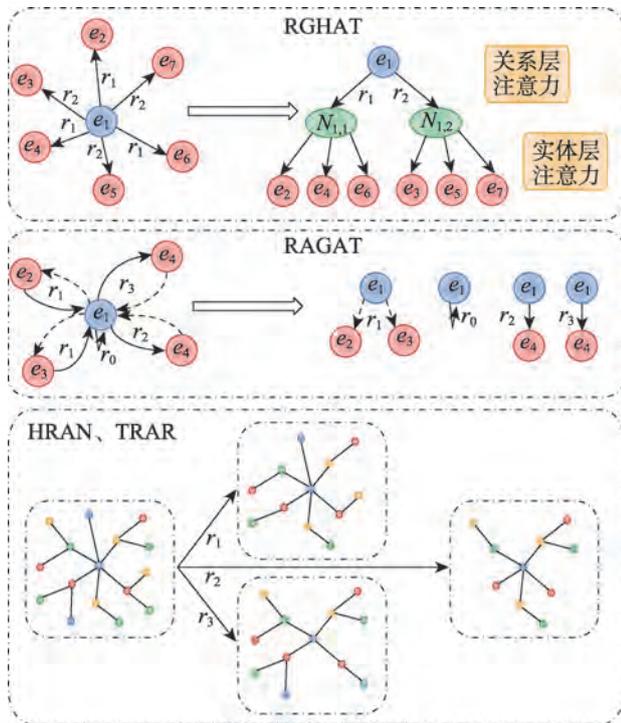


图6 RGHAT、RAGAT、HRAN与TRAR分级细节

Fig.6 Grading details of RGHAT, RAGAT, HRAN & TRAR

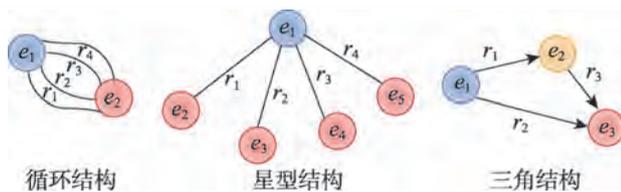


图7 局部结构

Fig.7 Local structure

克服这一缺点,提出一种具有新的邻居聚合策略的图注意力网络模型LSA-GAT(local structure-aware graph attention networks)^[45],该模型考虑了知识图谱丰富的局部结构并且捕获了复杂的结构信息。除知识图谱丰富的同构结构之外,其异构的邻居结构也不可避免,异构的邻居结构也很容易导致GAT产生不同的嵌入表示。为此,提出一种新的知识图谱嵌入方法AliNet(alignment network)^[46],以端到端的方式减轻邻居结构的异构特性。AliNet捕获了多阶邻居的语义信息,使用注意力机制放大邻居结构之间的重叠部分,并限制实体对在每个GAT层中具有相同的隐藏状态。

在大多数现有的知识图谱嵌入方法中,仅仅使用三元组的信息,尚未对知识图谱补全任务中的逻辑规则进行研究。为了解决这个问题,Zhang等人^[47]提出一种关联规则增强的知识图谱注意力方法

(association rules enhanced knowledge graph attention network, AR-KGAT),AR-KAGT利用GAT框架,以端到端的方式捕获目标实体的高阶邻居实体的语义信息和关系信息,AR-KAGT最主要的部分是使用基于关联规则和基于图注意力来估计关系级和邻居级的权重系数,注意力嵌入的传播层函数为:

$$e_h = \sum_{f_i(h,r,t) \in \mathcal{N}_i^L(h)} (\alpha_{(h,r,t)}^{L_i} + \alpha_{(h,r,t)}^{N_i}) C_{(h,r,t)}^1 \quad (12)$$

其中, $\alpha_{(h,r,t)}^{L_i}$ 、 $\alpha_{(h,r,t)}^{N_i}$ 分别表示基于逻辑和基于图的注意力系数, $C_{(h,r,t)}^1$ 为三元组的表示形式。与AR-KGAT不同的是,AR-KGAT的变体模型AR-KACN(association rules enhanced knowledge graph convolutional network)将GCN作为聚合器,而不是利用注意力机制进行邻居聚合。在解决知识图谱补全问题中,另一个挑战是知识图谱的多关系异构特性以及实体和关系之间的多种复杂交互,Dai等人^[48]在可以为邻居分配不同权重的前提下设计了一种多关系图注意力模型(multi-relational graph attention network, MRGAT),该模型可以适应异构特性,在不同实体权重的网络中引入自注意力机制进行优化,然后通过自注意力层计算邻居实体的重要性。

相比图卷积的知识图谱嵌入算法,基于图注意力的思想就是考虑知识图谱的异构特性,对目标实体的邻居实体和关系分配不同的重要性,以解决图卷积不能处理动态图和不易对邻居节点分配不同权重的瓶颈。与图卷积相同的是,二者都是将目标实体的邻居实体特征聚合,学习目标实体的嵌入表示。

2.4 基于图自编码器的知识图谱嵌入

本节简单介绍基于图自编码器的知识图谱嵌入算法,自编码器(autoencoder, AE)^[49-50]和变分自编码器(variational autoencoder, VAE)^[51-52]被广泛应用于无监督学习的图节点嵌入方法中,成功地利用AE和VAE解决知识图谱的链接预测问题。

图自编码器模型同图卷积模型一样侧重于无向图,忽略了图结构中的潜在方向。为此,Salha等人^[53]启发于牛顿万有引力定律,扩展图自编码器和变分编码器,提出一种新的解码器框架Gravity Graph VA和Gravity Graph VAE,从而有效地解决有向图中链接预测问题。值得一提的是,Salha等人在节点嵌入中应用牛顿万有引力定律(任何两个物体都存在通过其连心线方向上的相互吸引的力,该引力 F 大小为 $\frac{Gm_1m_2}{r^2}$,其中 G 为万有引力常量, m_1 、 m_2 为物体

的质量, r 为两个物体之间的距离), 得出有向图中节点 i 与节点 j 相关联的可能性:

$$a_{i \rightarrow j} = \frac{F}{m_i} = \frac{Gm_j}{r^2} \quad (13)$$

然而, 随着知识图谱规模的不断扩大, 将知识图谱中的实体和关系表示为 32 bit 的浮点型向量, 使得知识图谱嵌入在内存消耗方面变得越来越大。为了解决上述问题, Zhang 等人^[54]提出一种关系图自编码器模型 R-GAE, R-GAE 是图自编码器的扩展, 对知识图谱的局部邻域进行操作。R-GAE 过程说明如图 8 所示。首先利用 R-GAT 作为编码器改变实体和关系嵌入的维度, 其中 R-GAT 使用多头注意力机制学习实体和关系的特征表示 Z , 并通过 GCN 进行聚合。然后利用变分自编码器 VAE 和 Gumbel-Softmax^[55] 将浮点型嵌入压缩为二进制嵌入 B , 以获得大规模的关系型数据, 大大减少了数据存储的空间, 同时也保留了知识图谱的内在信息。最后解码器使用编码器的输出重构知识图谱。

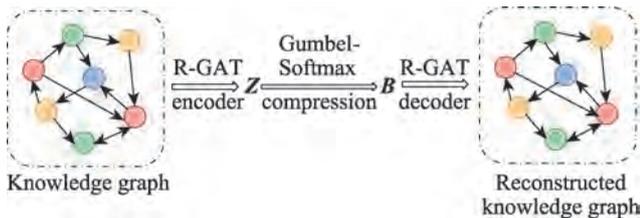


图8 R-GAE 过程说明

Fig.8 R-GAE process description

大多数基于图自编码器模型的工作都有一个共同的问题, 这类模型可能无法捕获知识图谱更复杂的结构, 比如一对多或多对多, 也无法适应不同类型实体和关系的不确定性。为此, Yang 提出了 GCN-VAE^[56], 一种新的用于学习大规模知识图谱中实体和关系嵌入的概率框架, 利用变分自编码器将目标实体的邻域结构编码为潜在嵌入分布, 输出潜在表示 z 的分布参数, 并将其建模为高斯后验分布, 生成 z 的均值和方差 $\mu, \sigma = q_{\phi}(Z|R, E)$, 同时为了将高斯后验分布建模为更灵活的非高斯分布, 而不是简单的多元高斯分布。Yang 还研究了逆自回归流 (inverse autoregressive flow, IAF) 作为 VAE 的解码器, 改进实体嵌入分布的密度估计。于此同时, Hu 等人^[57]提出了一种新的卷积自编码器模型作为知识图谱嵌入的解码器, 其中利用双向关系编码器对实体嵌入进行编码, 以生成实体的双向关系特征, 旨在获取相应的隐藏关系表示。

总体来看, 基于图自编码器的知识图谱嵌入方法的研究工作并不算多。虽然图自编码器能够自动提取目标特征, 有效地解决了传统手动方法提取特征不足的问题, 同时避免过拟合的问题, 也使得这一类模型在开源数据集中取得了不错的性能。但是在处理知识图谱这种大规模的图数据中, 其模型的训练时间变得更长, 并且图自编码器属于无监督学习, 不能很好地描述所学习到的特征表示, 因此图自编码器的问题也随之遗留在知识图谱嵌入中。因此如何设计更好的基于图自编码器的知识图谱嵌入模型, 有待进一步研究。

3 任务分析

本章介绍了知识图谱嵌入模型在链接预测、实体对齐、知识图谱推理以及知识图谱补全下游任务中的性能。需要说明的是, 在不同下游任务中的模型指标均来自现有论文, 考虑到不同方法在嵌入维度、批采样大小、负采样大小、损失函数设计以及训练轮数等都存在差异, 因此任务分析中的数据不完全适用于分析对比不同类别方法之间的性能优劣。

3.1 链接预测

链接预测 (link prediction, LP) 的任务是通过一个已知的实体和关系去预测另一个实体, 或者通过已知的两个实体去预测关系 (如头实体和关系已知, 去预测正确的尾实体)。形式化定义为, 给定查询 $(h, r, ?)(?, r, t)(h, ?, t)$, 分别预测正确的头实体、关系和尾实体集合, 其中 h, r, t 分别表示头实体、关系和尾实体, “?” 表示三元组缺失的实体或关系, 图 9 为链接预测的示例。当知识图谱的嵌入被学习完成之后, 就可以试图预测知识图谱中缺失的信息。例如给定查询 $(h, r, ?)$, 将知识图谱中的每一个实体都放在尾实体的位置上, 通过知识图谱嵌入模型中的得分函数计算不同实体作为该三元组的尾实体的得分, 得分最高的实体会被作为链接预测的最终结果。

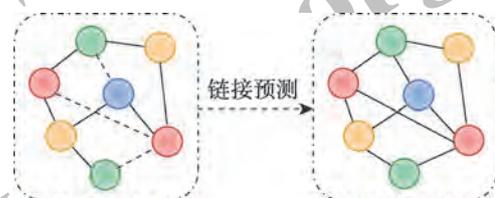


图9 链接预测示例

Fig.9 Example of LP

链接预测通常被用于评估知识图谱嵌入算法的

有效性,将正确的三元组排名记录在有序列表中,查看正确三元组是否在错误三元组之前,因此链接预测任务也被定义为实体排序任务。其链接预测的评估指标通常采用平均排名(mean rank, MR)、平均倒数排名(mean reciprocal rank, MRR)和正确三元组排序在前 n 个的比例 $Hits@n$,评估指标定义如下:

$$\left\{ \begin{array}{l} MR = \frac{1}{2|\mathcal{T}_{test}|} \sum_{x_i \in \mathcal{T}_{test}} (rank_i^h + rank_i^t) \\ MRR = \frac{1}{2|\mathcal{T}_{test}|} \sum_{x_i \in \mathcal{T}_{test}} \left(\frac{1}{rank_i^h} + \frac{1}{rank_i^t} \right) \\ Hits@n = \frac{1}{2|\mathcal{T}_{test}|} \sum_{x_i \in \mathcal{T}_{test}} [I[rank_i^h \leq n] + I[rank_i^t \leq n]] \quad (14) \\ rank_i^h = 1 + \sum_{x_i \in \mathcal{T}_{test}} I[S_{(h,r,t)} < S_{(h',r',t')}] \\ rank_i^t = 1 + \sum_{x_i \in \mathcal{T}_{test}} I[S_{(h,r,t)} < S_{(h',r',t')}] \end{array} \right.$$

其中, \mathcal{T}_{test} 是测试集评估样本的个数, $rank_i^h$ 、 $rank_i^t$ 为第 i 个样本中正确三元组的排序值, $I[\cdot]$ 为示性函数, $I[x]=1$ 当且仅当 x 为真, 否则 $I[x]=0$ 。由于评估指标中 MR 是一种不被认可的指标, 下面将在 MRR 和 $Hits@n$ 中描述模型的性能, 不再阐述 MR。

在链接预测实验中, 主要在 4 个使用广泛的开源数据集 FB15K^[14]、FB15K-237^[58]、WN18^[14] 和 WN18RR^[16] 上评估知识图谱嵌入模型的性能。FB15K 是知识图谱库 FreeBase 的一个子集, 其中包含大量的知识事实; WN18 是 WordNet 的一个子集, 是一个描述词汇关系的知识图谱。研究发现, FB15K 和 WN18 包含了实体之间的逆向关系。为了解决这种情况, 将数据集中的逆向关系删除, 提出一种更具有挑战性的数据集 FB15K-237 和 WN18RR, 数据集的详细概述见表 2。由于数据集 FB15K 和 WN18 存在的逆向关系会影响模型的性能, 将在 FB15K-237 和 WN18RR 数据集中阐明模型的性能指标。

表 2 FB15K、WN18、FB15K-237 和 WN18RR 数据集概述

Table 2 Overview of FB15K, WN18, FB15K-237 and WN18RR datasets

数据集	实体	关系	训练集	验证集	测试集
FB15K	14 951	1 345	483 142	50 000	59 071
WN18	40 943	18	141 442	5 000	5 000
FB15K-237	14 541	237	272 415	17 535	20 466
WN18RR	40 943	11	86 835	3 034	3 134

在图 10 和图 11 中, 对每个基于图神经网络的知识图谱嵌入模型都阐述了在数据集中的各项评估指

标值。在链接预测任务中, RGHAT 和 AR-KGAT 在数据集 FB15K-237 和 WN18RR 上均取得了不错的性能, 这可能取决于模型的注意力机制, 不仅利用了知识图谱的异构特性, 而且采用 GAT 的注意力机制, 对实体和关系分配不同的重要性。

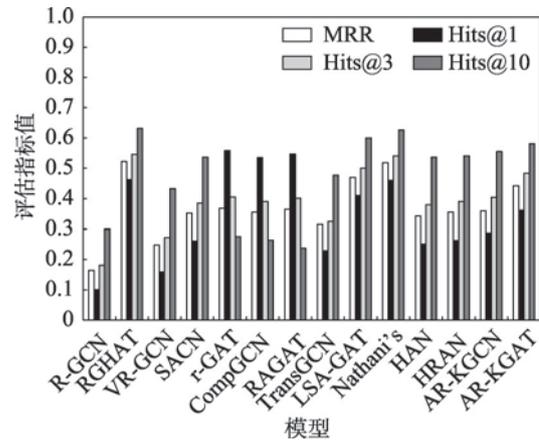


图 10 FB15K-237 数据集上的模型指标

Fig.10 Model metrics on FB15K-237 dataset

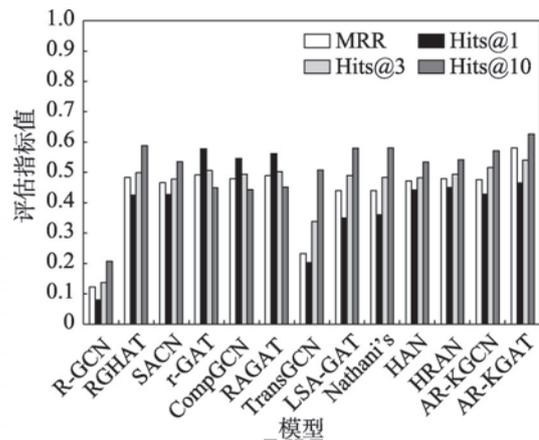


图 11 WN18RR 数据集上的模型指标

Fig.11 Model metrics on WN18RR dataset

3.2 实体对齐

在知识图谱嵌入应用中, 实体对齐也是一项关键的技术。实体对齐(entity alignment, EA), 也称为实体解析, 其任务是验证两个实体是否指代或者引用的是同一个事物或对象, 在处理知识图谱中冗余的实体上很有帮助, 也可以在知识图谱融合的时候从异构的数据集中找相同的实体。实体对齐的定义为: 两个不同的知识图谱 KG_1 、 KG_2 , 其中的实体有一定的对应关系, 在已知一些实体对齐样本的前提下, 预测知识图谱中其他实体的对齐关系, 图 12 为实体对齐的示例。

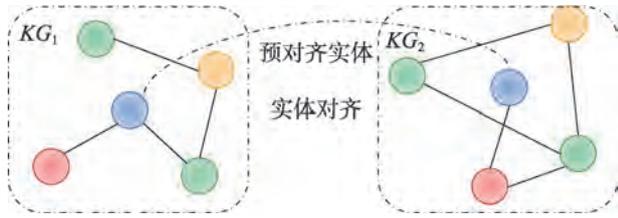


图12 实体对齐示例

Fig.12 Example of EA

传统的实体对齐方法主要从基于相似度和基于关系推理来比较知识图谱中实体之间的关系。随着知识图谱嵌入研究的发展,研究者也随之提出了基于图神经网络的实体对齐方法,由于图神经网络具有较强的鲁棒性和泛化能力,大多数实体对齐方法使用图神经网络进行知识图谱表示学习。基于图神经网络的知识图谱嵌入方法按照利用信息的不同,分为只利用图谱的结构信息、利用图谱结构信息和实体的额外信息(实体的描述信息、属性)、只利用实体的额外信息三种类别。

除了链接预测的评估指标 MR、MRR、Hits@n 之外,个别实体对齐任务还使用 Precision、Recall、F1-measure 作为评估指标。精确率(Precision)被定义为实体对齐的准确程度,定义为:

$$P = \frac{N_{\text{success}}}{N_{\text{total}}} \quad (15)$$

其中, N_{success} 、 N_{total} 分别表示知识图谱嵌入方法对齐的正确实体的数量和实体对总数。

召回率(Recall)表示被正确对齐的实体占全部真实存在实体的比例,定义为:

$$R = \frac{N_{\text{success}}}{R_{\text{total}}} \quad (16)$$

其中, R_{total} 表示所有的真实存在的实体对数量。

F1 值(F1-measure)则反映实体对齐的综合效果,定义为:

$$F1 = \frac{2 \times R \times P}{R + P} \quad (17)$$

在实体对齐实验中,大多数实体对齐数据集都是基于 Wikidata、DBpedia 和 YAGO 等知识图谱构建的,从 DBpedia 构建的实体对齐数据集 DBP15K^[59]使用最为广泛。DBP15K 从 DBpedia 的不同语言版本构建出 3 个版本的子数据集 DBP15K(ZH-EN)、DBP15K(FR-EN)和 DBP15K(JA-EN),以进行实体对齐任务,数据集的详细信息如表 3 所示。

图 13~图 15 描述了每个基于图神经网络的知识图谱嵌入模型在数据集 DBP15K 的 3 个子数据集的评估指标值。不难发现,EPEA 模型在实体对齐方面

表 3 DBP15K 数据集概述

Table 3 Overview of DBP15K dataset

数据集	子集	图	实体	关系	关系三元组	属性三元组
DBP15K	DBP15K	ZH	66 469	2 830	153 929	379 684
	ZH-EN	EN	98 125	2 317	237 674	567 755
	DBP15K	FR	66 858	1 379	192 191	528 665
	FR-EN	EN	105 889	2 209	278 590	576 543
	DBP15K	JA	65 774	2 043	164 373	354 619
	JA-EN	EN	95 680	2 096	233 319	497 230

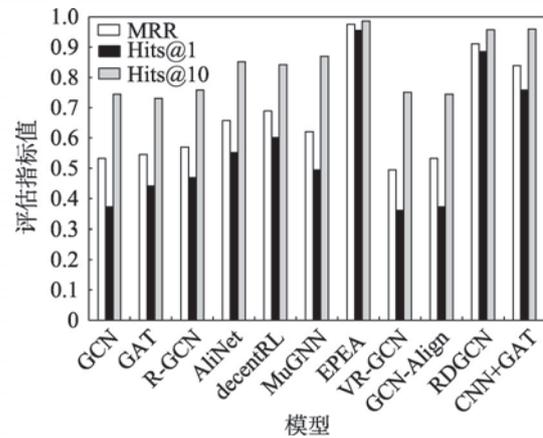


图13 DBP15K(FR-EN)数据集上的模型指标

Fig.13 Model metrics on DBP15K (FR-EN) dataset

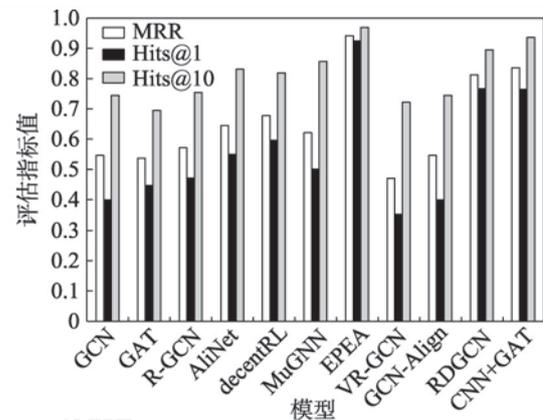


图14 DBP15K(JA-EN)数据集上的模型指标

Fig.14 Model metrics on DBP15K (JA-EN) dataset

表现效果最好。这表明成对连通图(PCG),利用GCN捕获实体对的属性信息,再加上关系感知的图注意力网络传播相似特征,这种嵌入方法聚合邻居实体的特征信息时,可以同时兼顾实体的属性信息和结构信息,最终得到实体的嵌入表示。因此在基于图神经网络的知识图谱嵌入方法中加入结构信息和额外的属性信息,可以更好地捕获实体的语义信息,从而获得更好的区分度。

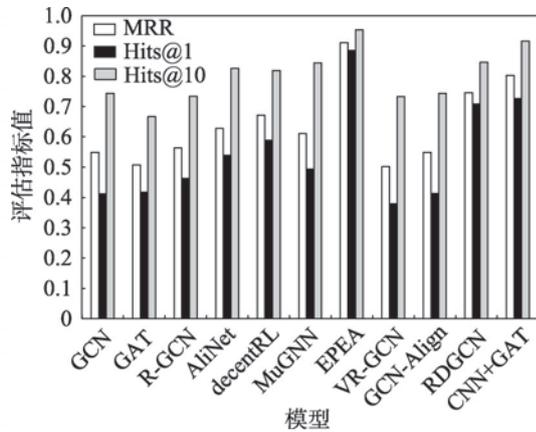


图15 DBP15K(ZH-EN)数据集上的模型指标

Fig.15 Model metrics on DBP15K (ZH-EN) dataset

3.3 知识图谱推理

知识图谱推理(knowledge graph reasoning, KGR), 简称知识推理, 从一个或者多个现有的知识出发, 用特定的方法推断新的或未知的结论, 或者识别错误的信息。例如, 给出已知的三元组信息(X , 出生地, Y), 通过推理可以得到新的三元组信息(X , 国籍, Y)。在某种程度上, 知识推理类似于链接预测, 其核心都是对三元组中的实体和关系进行预测。图16给出了知识推理的示例。

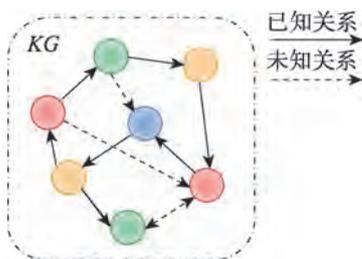


图16 知识图谱推理示例

Fig.16 Example of KGR

知识推理方法又可分为基于逻辑规则、基于分布式表示、基于深度学习和基于图神经网络四大类。随着GNN在知识图谱嵌入的广泛应用, 基于图神经网络的知识推理也成为当下主流的知识推理方法。基于图神经网络的知识推理方法, 可以将知识图谱的异构特性和实体的语义信息有效结合起来, 在信息聚合的时候不仅考虑目标实体的隐层信息, 同时也考虑该实体的邻居实体信息和局部结构信息, 大大提高了知识推理方法的推理能力。

同链接预测任务一样, 验证知识推理方法所使用的数据集除FB15K-237和WN18RR外, 还使用NELL-995^[60]和YAGO-10^[61]作为验证数据集, 具体详情

见表4。其验证指标同链接预测一样使用MR、MRR和Hits@ n , 下面不再赘述。图17描述了部分知识推理方法在FB15K-237和WN18RR数据集的性能指标。

表4 NELL-995、YAGO-10数据集概述

Table 4 Overview of NELL-995 and YAGO-10 datasets

数据集	实体	关系	训练集	验证集	测试集
NELL-995	74 536	200	149 678	543	2 818
YAGO-10	123 188	37	1 079 040	5 000	5 000

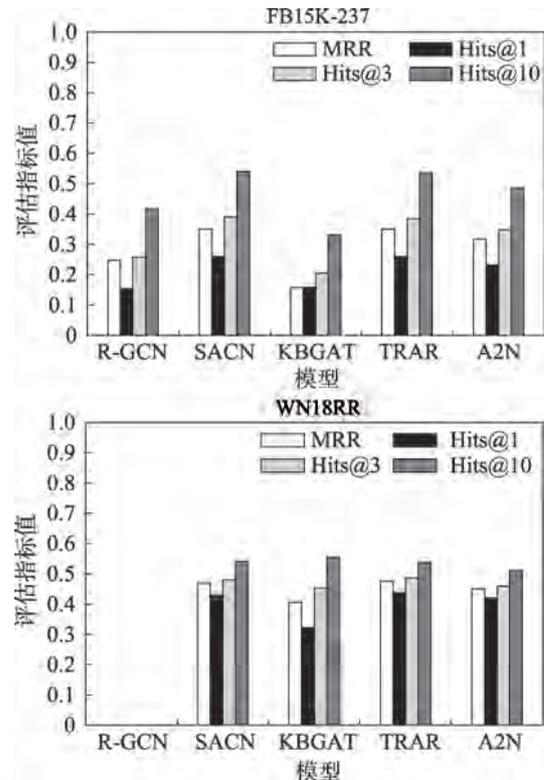


图17 FB15K-237和WN18RR数据集上的性能指标

Fig.17 Performance metrics on FB15K-237 and WN18RR datasets

知识推理作为知识图谱嵌入的重要任务之一, 是通过对三元组中的实体和关系进行预测, 实现对未知知识的推断。链接预测、实体对齐和三元组分类等任务与知识推理密切相关, 但它们是在已知实体和关系的基础上进行预测。虽然大部分知识推理方法将链接预测作为评估其有效性的任务, 但对于知识推理任务需要根据不同的场景和需求选择合适的任务进行分析。

比较新颖的是, 知识图谱推理在军事、医疗和交通领域也被广泛应用。在军事领域, 知识图谱推理帮助指挥官理解复杂多变的战场形势, 其中军事决策知识图谱中不仅包括战斗行为实体和关系, 而且

也包括军队指定的军事规则。Nie 等人^[62]提出一种混合规则和图神经网络学习的知识图谱推理方法 (context-surrounding graph neural networks with numbers, CS-GNN-N), 其中规则学习、规则注入和图神经网络学习在 CS-GNN-N 中迭代完成。在建模药物之间的相互作用和多药副作用中, Zitnik 等人^[63]提出一种 Decagon 的建模方法, Decagon 使用 GCN 来学习药物的低维向量表示, 同时考虑药物之间的相互作用, 利用药物的嵌入向量计算之间的相似度来预测多药副作用。为了应对交通模式的时变性和道路网络的复杂空间依赖性的挑战, Cui 等人^[64]提出一种新颖的深度学习框架 TGC-LSTM (traffic graph convolutional long short-term memory neural network), TGC-LSTM 使用 GCN 从城市交通网络提取特征, 然后使用循环神经网络对交通流量时间序列数据进行建模和预测, 从而有效捕捉城市交通网中的复杂关系。

由于部分方法使用基于领域的数据集进行方法评估, 且链接预测、实体对齐等任务的方法评价指标前面已经介绍, 图 17 只给出部分方法的评估指标。

3.4 知识图谱补全

知识图谱补全 (knowledge graph completion, KGC) 作为知识图谱构建中的一个热门话题, 同链接预测一样, 预测知识图谱中缺失的实体或关系。不同的是, 知识图谱补全通过挖掘未知事实来补全知识图谱的结构, 图 18 给出了知识图谱补全的示例。知识图谱补全的目标是解决知识图谱中缺失的事实或者链接导致知识图谱的不完整性和稀疏性的问题, 是发现新知识的重要手段, 同时也是提高知识图谱质量的必要手段。除知识推理等三种具体任务之外, 知识图谱补全也可以应用于特定任务, 比如属性预测或其他子任务。

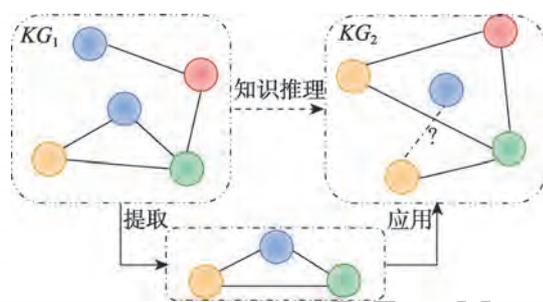


图 18 知识图谱补全示例

Fig.18 Example of KGC

知识图谱补全是指通过自动化方法识别和填补知识图谱中的空缺或错误^[65], 知识图谱补全包括实体

补全^[66]、关系补全^[67]和属性补全^[68], 它们是知识图谱补全的三个重要方面。实体补全是指识别并添加新实体到知识图谱中; 关系补全是指将已知实体之间的未知关系添加到知识图谱中; 属性补全是指根据已知信息自动生成实体的属性信息。除了实体补全、关系补全和属性补全这三方面外, 还有一些其他的补全任务, 例如事件补全^[69]和时间补全^[70], 它们也可以被视为知识图谱补全的子任务。在实际应用场景中, 知识图谱补全可以帮助人们更好地理解 and 利用数据, 提高信息的精度和效率。

根据任务场景的不同, 知识图谱补全可以分为静态知识图谱补全与动态知识图谱补全。具体而言, 如果补全过程中涉及的实体和关系属于原始知识图谱, 则称为静态知识图谱补全, 否则称为动态知识图谱补全。目前大量的知识图谱补全模型都是基于静态知识图谱^[71], 在这种情况下假设所有的实体和关系都属于同一知识图谱中, 而图谱的补全只能挖掘现有实体之间的潜在关系来实现, 而不是向现有的图谱中添加新的关系或对应实体, 这也导致了静态知识图谱补全主要适用于规模小的领域知识图谱, 也不能很好地扩展图谱结构。基于概率数据库理论中的动态假设^[1], 为动态知识图谱补全提供了一种预测新实体和关系的方法。由于替代新知识的范围很广, 在动态知识图谱补全中难以建立本地和外部之间的联系, 但是当扩大知识图谱的规模时, 动态知识图谱补全具有更大的优势。

由于目前的大多数研究集中于静态知识图谱补全, 根据三元组中缺失的部分, 知识图谱补全的任务可分为头实体预测、关系预测、尾实体预测, 且静态知识图谱补全任务所使用的数据集以及评估指标同链接预测任务一样, 因此本文的内容也限制在此范围内, 不再针对知识图谱补全任务进行赘述。

4 常识知识图谱

上文所述模型大部分应用于百科全书知识图谱 (encyclopedia KG), 百科全书知识图谱是指将百科全书中的知识整理出来, 形成一个具有结构化信息的知识图谱, 为读者提供详实全面的知识体系。与之相比, 常识知识图谱 (common sense KG) 主要涵盖日常生活中的各种常识, 为人们提供一种方便快捷的查询工具, 帮助人们解决生活中的各种疑惑和问题。这两种知识图谱在信息的处理和展现方面都具有较为清晰的信息结构和易于查询的特点, 因为它们都采用了一种语义化的表示方式, 使得它们之间

的关系可以被机器自动理解和处理。

针对常识性知识图谱,研究者会挖掘词之间的语言知识,比较注意的关系有 isA Relation、HasProperty、isUsedFor 等,可以将此过程解释为问题理解,所需要的知识称之为常识性知识(common sense knowledge)。对于百科全书知识图谱而言,研究者通常会注意实体与实体之间的事实连接,比如 SpouseOf、DayOf-Birth、LocatedIn 等,通常将此过程解释为问题回答。根据常识知识图谱^[72]已有介绍,表5总结了代表性的常识知识图谱,这些资源涵盖了丰富的知识范围,从日常知识到以事件为中心的知识,再到视觉知识。常识知识图谱已经在自然语言处理、计算机视觉和知识图谱等领域被应用^[81-83]。

常识知识图谱嵌入通常使用 TransE 图嵌入和 BERT 文本嵌入,而利用 GCN 传输不同类别的信息研究较少。随着零样本学习(zero-shot learning)^[84]的兴起,Wang 等人^[85]提出利用常识图谱的结构化信息和复杂关系,而不需要任何类标签来学习视觉分类器的方法 GCNZ(graph convolutional networks with zero-shot learning)。该方法是建立在图卷积网络之上,将具有回归损失的图卷积应用于零样本学习,该图卷积的作用是在不同类别之间传递消息。该方法在训练期间调整整个图结构,但常识知识图谱的规模庞大,并且在整个图上训练图神经网络开销过于昂贵。为了仍然可以利用图谱的图结构特性,一种归纳方法模型 DGP(dense graph propagation)^[86]被提出,

该方法通过额外的关系利用图谱的层次结构,进一步使用加权方案对实体分配不同的重要性,以改进图谱中的消息传递。根据图卷积消息传递规则 $H^{l+1} = \sigma(D^{-1}AH^l\theta^l)$,DGP 的加权消息传递规则变为:

$$H = \sigma \left(\sum_{k=0}^K \alpha_k^a D_k^{a^{-1}} A_k^a \sigma \left(\sum_{k=0}^K \alpha_k^d D_k^d A_k^d X \theta_d \right) \theta_a \right) \quad (18)$$

其中, A_k^a 和 A_k^d 表示为邻接矩阵部分, D_k^a 和 D_k^d 是 A_k^a 和 A_k^d 对应的度矩阵, θ 为权重矩阵。但 DGP 的前提假设是图谱为有向非循环图,比如 WordNet,而常识知识图谱并不是有向无环图结构。为了解决这一局限性,Nayak 等人^[87]引入 ZSL-KG(zero-shot learning with knowledge graph),这是一个利用机器翻译和语言建模任务的非线性模块 Transformers^[88],使用多层感知机和自注意力机制的特性学习非线性聚合器 TrGCN(transformer graph convolutional network),用于学习嵌入表示。与之前 GCNZ、DGP 以及大多数图神经网络方法使用的聚合函数不同的是,ZSL-KG 利用 TrGCN 学习基于转移的聚合器,计算邻居实体的非线性组合,增强嵌入表示的表达能力,从而捕获常识知识图谱中的复杂关系。TrGCN 将特征向量经过非线性层标准化后,利用自注意力机制计算每个查询实体的特征加权组合,表示如下:

$$\begin{cases} h_u^l = \text{LN}\{W_{fh}^{l+1}[\sigma(W_{fh}^{l+1} \cdot h_u^l)] + h_u^l\} \\ \{z_u^{l+1} \forall u \in \mathcal{N}(v)\} = \text{softmax} \left(\frac{QK^T}{\sqrt{d(p)}} \right) V \end{cases} \quad (19)$$

表5 代表性常识知识图谱统计

Table 5 Representative common sense KG statistics

知识库	描述	映射
ConceptNet ^[73]	ConceptNet 是一个开源的语义网络,旨在帮助计算机理解人们使用的单词的含义	WordNet、DBpedia、OpenCyc、Wiktionary
ATOMIC 2020 ^[74]	ATOMIC 2020, 一个拥有 133 万个关于实体和事件的大规模的文本描述常识库,这些文本描述编码了人类日常经验的社会和物理方面,对当前语言模型中编码的常识性知识进行补充	ConceptNet、Cyc
WikiData ^[75]	WikiData 是一个自由和开放的知识库,人类和计算机都可以阅读和编辑,其内容可在自由许可下使用,使用标准格式导出,并可与链接数据网上的其他开放数据集相互链接	various
WordNet ^[76]	WordNet 是一个大型的英语词汇数据库。名词、动词、形容词和副词被分组为 synsets (sets of cognitive synonyms), 每个同义词都表达了一个独特的概念	
VerbNet ^[77]	VerbNet (VN) 是最大的英语动词在线网络,将其句法和语义模式联系起来,是一个分层的、与领域无关的、覆盖面很广的动词词典	FrameNet、WordNet
FrameNet ^[78]	FrameNet, 一个英语词汇数据库,既是人类可读的,也是计算机可读的,其基础是对实际文本中的词汇使用方式进行注释	
Visual Genome ^[79]	Visual Genome 是一个有 100 000 多张图片和上面所有物体的描述的数据集,该数据集旨在用于物体搜索和识别任务,包含了对图像、物体、属性和关系的描述	WordNet
ImageNet ^[80]	ImageNet 是一个根据 WordNet 层次结构(目前只有名词)组织的图像数据库,其中层次结构的每个节点都由成百上千的图像描绘	WordNet

其中, $Q = \mathbf{W}_q^{l+1} \cdot \mathbf{h}'_u$ 是所有邻居查询向量的集合, $\mathbf{W} = \mathbf{W}_k^{l+1} \cdot \mathbf{h}'_u$ 是所有键向量, $\mathbf{V} = \mathbf{W}_v^{l+1} \cdot \mathbf{h}'_u$ 是所有值向量。TrGCN通过聚合局部邻居特征来学习实体的嵌入表示,意味着学习的模型可以用于预测新的图形结构,而不需要再次训练,这也使得ZSL-KG更适合零样本学习。

5 总结与展望

5.1 总结

知识图谱嵌入的思想是将实体和关系的语义信息映射到低维稠密实值的向量空间中,因此,需要对知识图谱的结构信息以及实体和关系的语义信息进行学习,进而将学习到的实体和关系嵌入应用于多种下游任务。GNN将现有的神经网络方法拓展到了图结构数据处理领域,依靠图结构数据进行建模,来学习、提取和挖掘图结构数据中的特征和模式。知识图谱嵌入利用图神经网络强大的图编码能力,可以很好地学习知识图谱中的实体信息、关系信息以及结构信息,从而高效地解决了知识图谱中的图结构数据问题。

本文分别从基于图卷积、图神经、图注意力和图自编码器等方面对知识图谱嵌入进行了综述,介绍了各类知识图谱嵌入的基础模型和核心算法,并就基于图神经网络的知识图谱嵌入在链接预测、实体对齐、知识推理以及知识图谱补全等下游任务中进行了任务分析及归纳。表6~表9总结了这四类知识图谱嵌入方法的优缺点和基础特征。

为了更好地展示不同方法在下游任务中的性能,本文在第3章任务分析中给出了部分方法在开源

数据集FB15K-237和WN18RR数据集上的链接预测任务、DBP15K数据集上的实体对齐任务以及NELL-995和YAGO-10数据集上的知识推理任务的实验指标结果,实验指标采用MRR和Hits@n指标。

虽然基于图神经网络的知识图谱嵌入方法大部分应用于百科全书类知识图谱,但当前的常识知识图谱数据源包含对下游任务整体有益的补充知识,但不同的建模方法阻止了二者之间的交互。未来,可以将图神经网络的图嵌入技术和预训练语言模型的信息嵌入技术结合,以捕获知识图谱的各种形式的结构相关性和文本相似性。

5.2 展望

GNN经过近几年的不断发展,这种有效的方法和框架也在理论和实验上被证实,基于图神经网络的知识图谱嵌入方法也得到了迅速发展,但是仍然存在许多的挑战和待完善的问题,不少已有方向都遇到了瓶颈。根据国内外的研究现状,下面探讨了一些基于图神经网络的知识图谱嵌入方法的未来研究方向和待完善的问题。

(1)如何解决知识图谱的分布外泛化问题。分布外泛化问题OOD泛化(out-of-distribution generalization)是指当测试样本的分布与训练样本的分布不同时,如何提高在新数据(例如未知实体或未知分布)上的泛化性能。目前大部分关于分布外泛化问题的研究主要集中在欧式数据(比如图片),虽然有一定的文献^[89-92]在研究图的分布外泛化问题,但对知识图谱的相关研究还是比较少,如何利用有限观测的数据,学习一个稳健的模型,能够泛化到训练分布有明显差异的测试数据上是分布外泛化问题要解决

表6 图卷积嵌入方法总结

Table 6 Summary of graph convolutional embedding methods

类别	方法	优缺点	特征
基于图卷积	R-GCN	优点:易扩展,参数共享 缺点:稀有或缺失关系的处理不够优秀,可解释性较差	分为空间图卷积和谱分解图卷积,参数共享,但实体数量越多,知识图谱规模越大,训练成本越高
	VR-GCN	优点:自适应学习实体的分布表示 缺点:隐空间的节点表示比较抽象,难以直观地解释实体	
	TransGCN	优点:引入可训练的转移矩阵建模实体和关系的语义关系 缺点:运算复杂度比较高,导致计算效率较低	
	CompGCN	优点:考虑多关系类型的组合,更好地描述实体的复杂关系 缺点:无监督的嵌入方法,无法直接应用于一些需要标注的任务	
	KE-GCN	优点:将语义匹配和拓扑匹配融合起来,实现更全面的知识表示 缺点:算法采用随机梯度下降进行训练,容易陷入局部最优解	
	DINGAL	优点:使用基于交替优化的训练过程,提高对齐的稳定性和鲁棒性 缺点:对知识图谱的动态变化做出简单的假设,不能很好地处理复杂场景	

表7 图注意力嵌入方法总结

Table 7 Summary of graph attention embedding methods

类别	方法	优缺点	特征
基于图注意力	SACN	优点:很容易地与其他模型集成 缺点:需要大量的训练数据	引入注意力机制,为邻居实体和关系分配不同的注意力分数,使其关注较为重要的邻居实体和关系,鲁棒性和可解释性更强,但额外增加了时间和内存消耗
	WGCN	优点:自适应地调整权重,提高模型的泛化能力 缺点:对于边权重分布不均的情况,影响模型的表现	
	Nathani	优点:提出新的损失函数,解决标准交叉熵损失函数中的类别不平衡问题 缺点:只考虑了一阶邻居信息,未考虑高阶邻居信息	
	r-GAT	优点:采用多头注意力机制,可以并行计算不同的注意力权重,提高准确率 缺点:引入关系特征矩阵增加了模型的参数量,可能会导致过拟合的问题	
	decentRL	优点:利用去中心化的结构并行处理大规模的图数据,降低了时间复杂度 缺点:模型结构较为简单,可能会受到模型容量的限制	
	RAGAT	优点:引入关系感知机制,多层注意力机制,支持扩展性 缺点:计算量较大,对于长距离依赖的建模不足,需要合适的参数选择	
	RAHAT	优点:采用较大的滑动窗口,考虑更多的关系三元组 缺点:可解释性相对较差,难以解释模型的决策过程	
	HRAN	优点:学习异构图上的关系特征,提高嵌入质量和知识图谱补全的准确性 缺点:模型需要对异构图谱的实体和关系类型进行预定义	
	TRAR	优点:采用基于目标关系的约束,更好地适应实际任务中的目标关系需求 缺点:计算复杂度较高,需要更多的计算资源	
	LSA-GAT	优点:引入基于LSA的词向量嵌入方法,增强对文本信息的理解能力 缺点:采用词向量嵌入方法在处理长文本时可能存在信息损失的问题	
	AliNet	优点:采用基于标签的负采样和动态交互机制,提高负样本的质量 缺点:未考虑实体和关系的类型信息,无法充分挖掘实体和关系的语义关联	
	AR-KGAT	优点:利用自适应规划的参数更新方式提高模型的泛化能力和鲁棒性 缺点:模型计算量较大,需要在GPU等高性能设备上训练和预测	
MRGAT	优点:采用子网络来捕获不同的信息,更好地利用知识图谱的结构信息 缺点:使用均匀分布初始化参数,可能存在梯度消失或梯度爆炸等问题		

表8 图神经嵌入方法总结

Table 8 Summary of graph neural embedding methods

类别	方法	优缺点	特征
基于图神经	M-GNN	优点:引入元图的概念,动态地学习不同的子图结构,并且可以自适应调整 缺点:使用基于邻居聚合的消息传递方式,存在消息传递的误差累积问题	将现有的神经网络方法拓展到了图结构数据处理领域,学习到更深层次的语义信息
	M ² GNN	优点:自适应学习图层之间的重要性,提升了嵌入的效果 缺点:模型结构相对复杂,难以直观地理解内部的学习过程和机制	
	MLNs	优点:支持不确定性推理,可以捕捉复杂的关系,可以进行自动参数学习 缺点:计算代价高,数据要求高,难以调参	
	ExpressGNN	优点:有效处理大规模的概率逻辑规则和知识图谱,支持复杂的推理任务 缺点:适用于概率逻辑推理任务,对于其他类型的任务可能效果不佳	
	GraIL	优点:采用互动网络来对关系进行编码,有效地利用了关系之间的交互信息 缺点:只考虑二元关系,对于多元关系的处理需要进行改进	
	INDIGO	优点:充分利用实体的语义信息,提高知识图谱补全的准确性 缺点:计算复杂度上较高,对于稀疏知识图谱的处理效果不够理想	
	MuGNN	优点:采用多通道的方式来综合利用不同类型的图 缺点:需要预定义不同类型的图,对于一些新颖的类型无法适应	
	EPEA	优点:将实体对齐任务转化为向量空间中的距离度量,提高计算效率 缺点:未考虑关系类型的信息,无法处理更复杂的实体关系	
	MRAEA	优点:利用跨语言知识图谱中实体的语言信息进行实体嵌入学习和对齐 缺点:算法的训练和推理速度较慢	
	CS-GNN-N	优点:使用上下文信息来丰富实体和关系的表示,解决数据稀疏和噪声问题 缺点:数字特征对结果产生噪声影响,需要对数字特征的质量进行处理	

表9 图自编码器嵌入方法总结

Table 9 Summary of graph autoencoder embedding methods

类别	方法	优缺点	特征
基于图自编码器	Gravity Graph VA	优点:能够处理有向图的重构,在较小的训练集上进行有效的训练 缺点:需要进行大量的超参数调整	编码器将图转为低维稠密表示,解码器重构图的结构和属性,缩减输入量,但需要手动设置隐层维度,多为无监督学习,性能较差
	Gravity Graph VAE	优点:能够生成与输入图相似的图形,能够学习图数据的连续表示 缺点:训练模型需要更多的时间和计算资源,比 Gravity Graph VA 更耗时	
	R-GAE	优点:采用自编码器的思想将嵌入空间进行压缩,提高空间和时间的效率 缺点:处理多种关系类型存在一定的限制,需要对关系类型进行预定义	
	GCN-VAE	优点:能够处理不同大小和形状的图,并生成具有多样性的图 缺点:生成的图缺乏真实性,处理大规模图时,训练和生成时间较长	
	Hu 等人	优点:引入双向关系编码网络,更好地捕捉实体的关系和上下文信息 缺点:在编码网络中使用简单的结构,无法充分表达实体的复杂关系	

的核心问题。

(2)如何学习动态知识图谱的知识图谱嵌入。现有的知识图谱结构是静态的图结构,但现实中知识图谱的图结构往往会随着时间的推移而发生改变,例如部分实体被增加或者删除,而且传统基于静态图谱的嵌入技术不能很好将语义信息表示出来,如何将基于动态知识图谱的嵌入表示进行高效的学习也是未来的研究方向之一。

(3)如何将多模态的知识融入图谱嵌入技术。通过将不同形式表达的知识(比如视频、文本或音频)融入知识图谱嵌入技术中,进一步丰富嵌入信息,进而从不同类型的数据中获得更符合客观世界规律的知识图谱嵌入方法,基于多模态的知识图谱嵌入方法也将应运而生。将多模态数据的特征作为节点或边的属性输入到图神经网络,利用多模态数据的信息更好地学习知识之间的关系和图谱之间的复杂结构,以实现将多模态的知识融入基于图神经的图谱嵌入,这将会成为研究的又一热点。

(4)如何解决网络深度与邻居节点特征聚合数量之间的矛盾。在自然语言处理或者计算机视觉中,神经网络的层数可以进行多层叠加,而且在一定范围内神经网络的增加会更有效地提取语义信息。但是,在知识图谱中,实体的邻居节点数量往往会随着网络深度的增加而呈指数级别增长,这就带来了网络深度与邻居节点特征聚合数量之间的矛盾。图神经网络的核心是邻居节点的消息传递和消息读取,传递和读取过程中邻居节点的阶数会随着网络层数的增加而扩张,进而导致邻居节点的特征聚合数量成指数增加。如果增加网络深度,就必须限制每层的节点数量,但节点之间的消息传递和读取也会随之受阻。需要注意的是,不同的方法适用于

不同的场景和任务,需要根据具体的情况选择合适的方法来解决这个问题。

(5)如何研究图神经网络中的可解释性以提高知识预测的可靠性。图神经网络在知识图谱中的应用面临着可解释性不足的问题,这可能会导致模型在预测时出现不确定性,降低预测的可靠性。此外,可解释性技术可以进一步发现知识图谱中的规律和模式,提高对知识的理解和分析能力。需要注意的是,在研究图神经网络中的可解释性时,需要综合考虑模型的预测性能和可解释性之间的平衡,以达到最优的预测效果。

(6)如何挖掘和探索知识图谱中更深层次的信息。目前基于图神经网络的知识图谱嵌入研究中,诸多的研究者将精力放在了知识图谱的关系信息上。但知识图谱除了关系信息,还蕴含了实体和关系的属性信息、知识图谱的层次信息以及实体信息,如果可以充分挖掘和探索这些信息,必将会提升知识图谱嵌入的性能。

综上所述,基于图神经网络的知识图谱嵌入目前还面临着众多挑战,这也需要研究者投入更多的精力去解决这些问题。

参考文献:

- [1] SINGHAL A. Introducing the knowledge graph: things, not strings[J]. Official Google Blog, 2012, 5: 16.
- [2] 刘知远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展[J]. 计算机研究与发展, 2016, 53(2): 247-261.
LIU Z Y, SUN M S, LIN Y K, et al. Knowledge representation learning: a review[J]. Journal of Computer Research and Development, 2016, 53(2): 247-261.
- [3] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their composition-

- ality[C]//Advances in Neural Information Processing Systems 26, Lake Tahoe, Dec 5-8, 2013: 3111-3119.
- [4] 杨东华, 何涛, 王宏志, 等. 面向知识图谱的图嵌入学习研究进展[J]. 软件学报, 2022, 33(9): 3370-3390.
YANG D H, HE T, WANG H Z, et al. Survey on knowledge graph embedding learning[J]. Journal of Software, 2022, 33(9): 3370-3390.
- [5] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61-80.
- [6] BORDES A, WESTON J, USUNIER N. Open question answering with weakly supervised embedding models[C]//LNCS 8724: Proceedings of the 2014 European Conference on Machine Learning and Knowledge Discovery in Database, Nancy, Sep 15-19, 2014. Cham: Springer, 2014: 165-180.
- [7] DAIBER J, JAKOB M, HOKAMP C, et al. Improving efficiency and accuracy in multilingual entity extraction[C]//Proceedings of the 9th International Conference on Semantic Systems, Graz, Sep 4-6, 2013. New York: ACM, 2013: 121-124.
- [8] HOFFMANN R, ZHANG C, LING X, et al. Knowledge-based weak supervision for information extraction of overlapping relations[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, Jun 19-24, 2011. Stroudsburg: ACL, 2011: 541-550.
- [9] GUO Q, ZHUANG F, QIN C, et al. A survey on knowledge graph-based recommender systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(8): 3549-3568.
- [10] NICKEL M, TRESP V, KRIEGEL H P. A three-way model for collective learning on multi-relational data[C]//Proceedings of the 28th International Conference on Machine Learning, Bellevue, Jun 28-Jul 2, 2011. Madison: Omnipress, 2011: 809-816.
- [11] YANG B, YIH S W, HE X, et al. Embedding entities and relations for learning and inference in knowledge bases[C]//Proceedings of the 3rd International Conference on Learning Representations, San Diego, May 7-9, 2015: 1-12.
- [12] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[C]//Proceedings of the 33rd International Conference on Machine Learning, New York, Jun 19-24, 2016: 2071-2080.
- [13] NICKEL M, ROSASCO L, POGGIO T. Holographic embeddings of knowledge graphs[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence, Phoenix Arizanam, Feb 12-17, 2016: 1955-1961.
- [14] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]//Advances in Neural Information Processing Systems 26, Lake Tahoe, Dec 5-8, 2013: 2787-2795.
- [15] SUN Z, DENG Z H, NIE J Y, et al. Rotate: knowledge graph embedding by relational rotation in complex space[J]. arXiv:1902.10197, 2019.
- [16] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2D knowledge graph embeddings[C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence, and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence, New Orleans, Feb 2-7, 2018. Menlo Park, AAAI, 2018: 1811-1818.
- [17] 吴博, 梁循, 张树森, 等. 图神经网络前沿进展与应用[J]. 计算机学报, 2022, 45(1): 35-68.
WU B, LIANG X, ZHANG S S, et al. Advances and applications in graph neural network[J]. Chinese Journal of Computers, 2022, 45(1): 35-68.
- [18] KHAMSI M A, KIRK W A. An introduction to metric spaces and fixed point theory[M]. New York: John Wiley & Sons, 2011.
- [19] GILMER J, SCHOENHOLZ S S, RILEY P F, et al. Neural message passing for quantum chemistry[C]//Proceedings of the 34th International Conference on Machine Learning, Sydney, Aug 6-11, 2017: 1263-1272.
- [20] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: a review and new perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798-1828.
- [21] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.
- [22] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]//LNCS 10843: Proceedings of the 15th International Conference the Semantic Web, Heraklion, Jun 3-7, 2018. Cham: Springer, 2018: 593-607.
- [23] YE R, LI X, FANG Y, et al. A vectorized relational graph convolutional network for multi-relational network alignment[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, Aug 10-16, 2019: 4135-4141.
- [24] CAI L, YAN B, MAI G, et al. TransGCN: coupling transformation assumptions with graph convolutional networks for link prediction[C]//Proceedings of the 10th International Conference on Knowledge Capture, Marina Del Rey, Nov 19-21, 2019. New York: ACM, 2019: 131-138.
- [25] VASHISHTH S, SANYAL S, NITIN V, et al. Composition-based multi-relational graph convolutional networks[J]. arXiv: 1911.03082, 2019.
- [26] YU D, YANG Y, ZHANG R, et al. Knowledge embedding

- based graph convolutional network[C]//Proceedings of the Web Conference 2021, Ljubljana, Apr 19-23, 2021. New York: ACM, 2021: 1619-1628.
- [27] SHANG C, TANG Y, HUANG J, et al. End-to-end structure-aware convolutional networks for knowledge base completion[C]//Proceedings of the 33rd AAAI Conference on Artificial Intelligence, the 31st Innovative Applications of Artificial Intelligence Conference, the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, Honolulu, Jan 27-Feb 1, 2019. Menlo Park: AAAI, 2019: 3060-3067.
- [28] YAN Y, LIU L, BAN Y, et al. Dynamic knowledge graph alignment[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence, the 33rd Conference on Innovative Applications of Artificial Intelligence, the 11th Symposium on Educational Advances in Artificial Intelligence, Feb 2-9, 2021. Menlo Park: AAAI, 2021: 4564-4572.
- [29] WANG Z, REN Z, HE C, et al. Robust embedding with multi-level structures for link prediction[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, Aug 10-16, 2019: 5240-5246.
- [30] WANG S, WEI X, DOS SANTOS C N, et al. Mixed-curvature multi-relational graph neural network for knowledge graph completion[C]//Proceedings of the Web Conference 2021, Ljubljana, Apr 19-23, 2021: 1761-1771.
- [31] RICHARDSON M, DOMINGOS P. Markov logic networks [J]. *Machine Learning*, 2006, 62(1): 107-136.
- [32] ZHANG Y, CHEN X, YANG Y, et al. Efficient probabilistic logic reasoning with graph neural networks[C]//Proceedings of the 8th International Conference on Learning Representations, Addis Ababa, Apr 26-30, 2020: 1-20.
- [33] TERU K K, DENIS E G, HAMILTON W L. Inductive relation prediction by subgraph reasoning[C]//Proceedings of the 37th International Conference on Machine Learning, Jul 13-18, 2020: 9448-9457.
- [34] LIU S, GRAU B, HORROCKS I, et al. INDIGO: GNN-based inductive knowledge graph completion using pairwise encoding[C]//Advances in Neural Information Processing Systems 34, Dec 6-14, 2021: 2034-2045.
- [35] CAO Y, LIU Z, LI C, et al. Multi-channel graph neural network for entity alignment[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Jul 28-Aug 2, 2019. Stroudsburg: ACL, 2019: 1452-1461.
- [36] WANG Z, YANG J, YE X. Knowledge graph alignment with entity-pair embedding[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, Nov 16-20, 2020. Stroudsburg: ACL, 2020: 1672-1680.
- [37] MAO X, WANG W, XU H, et al. MRAEA: an efficient and robust entity alignment approach for cross-lingual knowledge graph[C]//Proceedings of the 13th International Conference on Web Search and Data Mining, Houston, Feb 3-7, 2020. New York: ACM, 2020: 420-428.
- [38] NATHANI D, CHAUHAN J, SHARMA C, et al. Learning attention-based embeddings for relation prediction in knowledge graphs[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Jul 28-Aug 2, 2019. Stroudsburg: ACL, 2019: 4710-4723.
- [39] CHEN M, ZHANG Y, KOU X, et al. r-GAT: relational graph attention network for multi-relational graphs[J]. *arXiv:2109.05922*, 2021.
- [40] GUO L, WANG W, SUN Z, et al. Decentralized knowledge graph representation learning[J]. *arXiv:2010.08114*, 2020.
- [41] LIU X, TAN H, CHEN Q, et al. RAGAT: relation aware graph attention network for knowledge graph completion [J]. *IEEE Access*, 2021, 9: 20840-20849.
- [42] ZHANG Z, ZHUANG F, ZHU H, et al. Relational graph neural network with hierarchical attention for knowledge graph completion[C]//Proceedings of the 2020 AAAI Conference on Artificial Intelligence, New York, Feb 7-12, 2020: 9612-9619.
- [43] LI Z, LIU H, ZHANG Z, et al. Learning knowledge graph embedding with heterogeneous relation attention networks [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022, 33(8): 3961-3973.
- [44] ZHAO X, JIA Y, LI A, et al. Target relational attention-oriented knowledge graph reasoning[J]. *Neurocomputing*, 2021, 461: 577-586.
- [45] JI K, HUI B, LUO G. Graph attention networks with local structure awareness for knowledge graph completion[J]. *IEEE Access*, 2020, 8: 224860-224870.
- [46] SUN Z, WANG C, HU W, et al. Knowledge graph alignment network with gated multi-hop neighborhood aggregation[C]//Proceedings of the 34th AAAI Conference on Artificial Intelligence, the 32nd Innovative Applications of Artificial Intelligence Conference, the 10th AAAI Symposium on Educational Advances in Artificial Intelligence, New York, Feb 7-12, 2020. Menlo Park: AAAI, 2020: 222-229.
- [47] ZHANG Z, HUANG J, TAN Q. Association rules enhanced knowledge graph attention network[J]. *Knowledge-Based Systems*, 2022, 239: 108038.
- [48] DAI G, WANG X, ZOU X, et al. MRGAT: multi-relational graph attention network for knowledge graph completion [J]. *Neural Networks*, 2022, 154: 234-245.
- [49] BALDI P. Autoencoders, unsupervised learning, and deep architectures[C]//Proceedings of the 2011 ICML Workshop on Unsupervised and Transfer Learning, Bellevue, Jul 2, 2011: 37-50.
- [50] RUMELHART D E, HINTON G E, WILLIAMS R J.

- Learning internal representations by error propagation[M]. Cambridge: MIT Press, 1985.
- [51] KINGMA D P, WELLING M. Auto-encoding variational Bayes[J]. arXiv:1312.6114, 2013.
- [52] TSCHANNEN M, BACHEM O, LUCIC M. Recent advances in autoencoder-based representation learning[J]. arXiv:1812.05069, 2018.
- [53] SALHA G, LIMNIOS S, HENNEQUIN R, et al. Gravity-inspired graph autoencoders for directed link prediction[C]// Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, Nov 3-7, 2019. New York: ACM, 2019: 589-598.
- [54] ZHANG S, ZHANG Z, ZHUANG F, et al. Compressing knowledge graph embedding with relational graph auto-encoder[C]// Proceedings of the 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication, Beijing, Jul 17-19, 2020. Piscataway: IEEE, 2020: 366-370.
- [55] JANG E, GU S, POOLE B. Categorical reparameterization with gumbel-softmax[J]. arXiv:1611.01144, 2016.
- [56] YANG K. GCN-VAE for knowledge graph completion[EB/OL]. [2022-09-14]. <http://snap.stanford.edu/class/cs224w-2019/project/26425038.pdf>.
- [57] HU K R, LIU H, ZHAN C J, et al. Learning knowledge graph embedding with a bi-directional relation encoding network and a convolutional autoencoder decoding network[J]. Neural Computing and Applications, 2021, 33(17): 11157-11173.
- [58] TOUTANOVA K, CHEN D. Observed versus latent features for knowledge base and text inference[C]// Proceedings of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality, Beijing, Jul 26-31, 2015. Stroudsburg: ACL, 2015: 57-66.
- [59] SUN Z, HU W, LI C. Cross-lingual entity alignment via joint attribute-preserving embedding[C]// LNCS 10587: Proceedings of the 16th International Semantic Web Conference, Vienna, Oct 21-25, 2017. Cham: Springer, 2017: 628-644.
- [60] XIONG W, HOANG T, WANG W Y. DeepPath: a reinforcement learning method for knowledge graph reasoning[J]. arXiv:1707.06690, 2017.
- [61] SUCHANEK F M, KASNECI G, WEIKUM G. YAGO: a core of semantic knowledge[C]// Proceedings of the 16th International Conference on World Wide Web, Banff, May 8-12, 2007. New York: ACM, 2007: 697-706.
- [62] NIE K, ZENG K, MENG Q. Knowledge reasoning method for military decision support knowledge graph mixing rule and graph neural networks learning together[C]// Proceedings of the 2020 Chinese Automation Congress, Shanghai, Nov 6-8, 2020. Piscataway: IEEE, 2020: 4013-4018.
- [63] ZITNIK M, AGRAWAL M, LESKOVEC J. Modeling poly-pharmacy side effects with graph convolutional networks[J]. Bioinformatics, 2018, 34(13): i457-i466.
- [64] CUI Z, HENRICKSON K, KE R, et al. Traffic graph convolutional recurrent neural network: a deep learning framework for network-scale traffic learning and forecasting[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(11): 4883-4894.
- [65] CHEN Z, WANG Y, ZHAO B, et al. Knowledge graph completion: a review[J]. IEEE Access, 2020, 8: 192435-192456.
- [66] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion[C]// BONET B, KOENIG S. Proceedings of the 29th AAAI Conference on Artificial Intelligence, Austin, Jan 25-30, 2015. Menlo Park: AAAI, 2015: 2181-2187.
- [67] SHEN Y, DING N, ZHENG H T, et al. Modeling relation paths for knowledge graph completion[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 33(11): 3607-3617.
- [68] ZHANG Y, ZHENG W, WANG H, et al. Joint entity structural and attribute information for knowledge graph completion[C]// Proceedings of the 2021 International Conference on High Performance Computing and Communication, Xiamen, Dec 3-5, 2021: 264-269.
- [69] JUNG J, JUNG J, KANG U. Learning to walk across time for interpretable temporal knowledge graph completion[C]// Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, New York, Aug 14-18, 2021. New York: ACM, 2021: 786-795.
- [70] GARCÍA-DURÁN A, DUMANČIĆ S, NIEPERT M. Learning sequence encoders for temporal knowledge graph completion[J]. arXiv:1809.03202, 2018.
- [71] REITER R. On closed world data bases[M]. San Mateo: Morgan Kaufmann, 1981.
- [72] ILIEVSKI F, SZEKELY P, ZHANG B. CSKG: the commonsense knowledge graph[C]// LNCS 12731: Proceedings of the 18th International Conference the Semantic Web, Jun 6-10, 2021. Cham: Springer, 2021: 680-696.
- [73] SPEER R, CHIN J, HAVASI C. ConceptNet 5.5: an open multilingual graph of general knowledge[C]// Proceedings of the 31st AAAI Conference on Artificial Intelligence, San Francisco, Feb 4-9, 2017. Menlo Park: AAAI, 2017: 4444-4451.
- [74] SAP M, LEBRAS R, ALLAWAY E, et al. ATOMIC: an atlas of machine commonsense for if-then reasoning[J]. arXiv:1811.00146, 2018.
- [75] VRANDEČIĆ D, KRÖTZSCH M. Wikidata: a free collaborative knowledgebase[J]. Communications of the ACM,

- 2014, 57(10): 78-85.
- [76] MILLER G A. WordNet: a lexical database for English[J]. Communications of the ACM, 1995, 38(11): 39-41.
- [77] SCHULER K K. VerbNet: a broad-coverage, comprehensive verb lexicon[D]. Philadelphia: University of Pennsylvania, 2005.
- [78] BAKER C F, FILLMORE C J, LOWE J B. The Berkeley FrameNet project[C]//Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Montreal, 1998: 86-90.
- [79] KRISHNA R, ZHU Y, GROTH O, et al. Visual genome: connecting language and vision using crowdsourced dense image annotations[J]. International Journal of Computer Vision, 2017, 123(1): 32-73.
- [80] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, Jun 20- 25, 2009. Washington: IEEE Computer Society, 2009: 248-255.
- [81] YASUNAGA M, REN H, BOSSELUT A, et al. QA-GNN: reasoning with language models and knowledge graphs for question answering[J]. arXiv: 2104. 06378, 2021.
- [82] SHWARTZ V, WEST P, BRAS R L, et al. Unsupervised commonsense question answering with self-talk[J]. arXiv: 2004.05483, 2020.
- [83] BOSSELUT A, RASHKIN H, SAP M, et al. COMET: commonsense transformers for automatic knowledge graph construction[J]. arXiv:1906.05317, 2019.
- [84] PALATUCCI M, POMERLEAU D, HINTON G E, et al. Zero-shot learning with semantic output codes[C]//Advances in Neural Information Processing Systems 22, Vancouver, Dec 7-10, 2009: 1410-1418.
- [85] WANG X, YE Y, GUPTA A. Zero-shot recognition via semantic embeddings and knowledge graphs[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, Jun 18-22, 2018. Washington: IEEE Computer Society, 2018: 6857-6866.
- [86] KAMPPMEYER M, CHEN Y, LIANG X, et al. Rethinking knowledge graph propagation for zero-shot learning[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, Jun 16-20, 2019. Piscataway: IEEE, 2019: 11479-11488.
- [87] NAYAK N V, BACH S H. Zero-shot learning with common sense knowledge graphs[J]. arXiv:2006.10713, 2020.
- [88] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems 30, Long Beach, Dec 4-9, 2017: 5998-6008.
- [89] ROJAS-CARULLA M, SCHÖLKOPF B, TURNER R, et al. Invariant models for causal transfer learning[J]. The Journal of Machine Learning Research, 2018, 19(1): 1309-1342.
- [90] WU Q, ZHANG H, YAN J, et al. Handling distribution shifts on graphs: an invariance perspective[J]. arXiv:2202.02466, 2022.
- [91] ARJOVSKY M, BOTTOU L, GULRAJANI I, et al. Invariant risk minimization[J]. arXiv:1907.02893, 2019.
- [92] LI H, WANG X, ZHANG Z, et al. OOD-GNN: out-of-distribution generalized graph neural network[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(7): 7328-7340.



延照耀(1999—),男,陕西榆林人,硕士研究生,主要研究方向为知识图谱、图神经网络、自然语言处理。

YAN Zhaoyao, born in 1999, M.S. candidate. His research interests include knowledge graphs, graph neural networks and natural language processing.



丁苍峰(1978—),男,河南唐河人,博士,副教授,硕士生导师,主要研究方向为多层复杂网络、图神经网络、自然语言处理。

DING Cangfeng, born in 1978, Ph.D., associate professor, M.S. supervisor. His research interests include multilayer complex networks, graph neural networks and natural language processing.



马乐荣(1974—),男,陕西神木人,博士,教授,硕士生导师,主要研究方向为人工智能、大数据知识工程、自然语言处理。

MA Lerong, born in 1974, Ph.D., professor, M.S. supervisor. His research interests include artificial intelligence, big data knowledge engineering and natural language processing.



曹璐(2001—),女,陕西咸阳人,硕士研究生,主要研究方向为图神经网络、自然语言处理。

CAO Lu, born in 2001, M.S. candidate. Her research interests include graph neural networks and natural language processing.



游浩(2000—),男,陕西安康人,硕士研究生,主要研究方向为大数据知识工程。

YOU Hao, born in 2000, M.S. candidate. His research interest is big data knowledge engineering.