

知识图谱推理：现代的方法与应用

王文广
达而观信息科技(上海)有限公司, 上海 201203

摘要
知识图谱推理技术旨在根据已有的知识推导出新的知识,是使机器智能具有和人类一样的推理和决策能力的关键技术之一。系统地研究了知识图谱推理的现代方法,以统一的框架介绍了向量空间中进行知识图谱推理的模型,包括基于几何运算嵌入欧几里得空间和双曲空间的方法,基于卷积神经网络、胶囊网络、图神经网络等深度网络模型的方法。同时,系统地梳理了知识推理技术在各技术领域和各行业的应用情况,指出了当前存在的挑战以及其中蕴含的机会。

关键词
知识推理 ; 双曲空间嵌入 ; 几何运算 ; 胶囊网络 ; 图神经网络

中图分类号 :TP391.1 文献标识码 :A doi: 10.11959/j.issn.2096-0271.2021025

Knowledge graph reasoning: modern methods and applications

WANG Wenguang
DataGrand Inc., Shanghai 201203, China

Abstract
Knowledge reasoning over knowledge graph aims to discover new knowledge according to the existing knowledge. It is a pivotal technology to realize the human reasoning and decision-making ability of machine. The modern methods of knowledge reasoning over knowledge graph were studied systematically. And the methods based on vector representations with a unified framework were introduced, including the methods based on embedding into Euclidean space and hyperbolic space, and based on deep learning methods such as convolution neural network, capsule network, graph neural network, etc. Simultaneously, the applications of knowledge reasoning in various technical fields and industries were presented, and the existing challenges and opportunities were pointed out as well.

Key words
knowledge reasoning, hyperbolic space embedding, geometric operation, capsule network, graph neural network

1 引言

近年来,知识图谱技术得到极大的发展,大量的知识图谱被构建出来,并被广泛应用在各种场景下。从语义解析、实体消歧、信息抽取、智能问答、推荐系统、个性化搜索等技术方向到金融、军工、制造业、生物医药等行业,都可以看到知识图谱的应用,同时知识图谱的应用也在促进知识图谱技术的发展。

知识图谱是事实或知识的结构化表示,是由实体和实体间的关系组成的网状结构。实体是指独立的、拥有清晰特征的、能够区别于其他事物的事物。在知识图谱中,用来描述这些事物的信息即实体。实体在属性图中用顶点来表示,实体关联的类型即实体类型,在属性图中用顶点标签来表示。关系表达了两个实体之间的某种语义关系,通常以语义标签来表示,在属性图中表示为有向的边。也就是说,知识图谱 G 由一系列的三元组 $\langle h, r, t \rangle$ 组成,其中 h 和 t 分别表示头实体和尾实体, r 表示由 h 到 t 的有向关系。

推理是一种人类逻辑思维,让机器能够像人类一样拥有推理能力一直是人工智能发展的目标。符号推理和专家系统是早期的尝试,在知识图谱被提出和发展起来之后,基于知识图谱的推理技术^[1-2]也随之发展,并成为人工智能中非常热门的领域之一,也被认为是人工智能迈向具备和人类一样的推理和决策能力的关键技术。

在知识图谱推理中,知识图谱本身提供了人类知识和经验的总结,推理技术则实现了基于知识图谱中已有的知识来发现潜在的、未知的知识,极大地扩展了诸如知识问答、个性化搜索和智能推荐等能力。同时,在行业应用中,将领域知识图谱

和推理技术结合,从而实现辅助分析和决策支持。

本文给出了知识推理的定义,并介绍了知识图谱推理技术;然后分别介绍了现代的基于几何运算和基于深度学习的知识图谱推理技术,并分别从技术领域和行业领域两个角度介绍了知识图谱推理技术的应用;最后给出了知识图谱推理存在的挑战以及值得关注的研究方向。

2 知识图谱推理

知识图谱推理旨在从已有的知识中发现新的知识。对于知识图谱来说,新的知识可以分为两种:新的实体和新的关系。新的实体涉及的技术领域通常是实体抽取、实体消歧、实体融合等相关的自然语言处理或知识图谱技术。新的关系涉及的技术领域则有关系抽取和知识推理等。知识图谱推理,或称知识推理,指在既定的知识图谱中通过推理技术推导出实体间潜在的或者新的关系,发现新的知识。在图数据库、图论等相关领域往往又被称为链接预测。

知识图谱推理技术是伴随着人工智能、自然语言处理、语义网等技术发展起来的。早期有基于规则的方法,著名知识图谱NELL^[3]利用手写规则的推理方法来不断扩充规模。一阶逻辑(first order logic, FOL)是早期的一种符号推理系统,也被用于知识图谱推理^[4]。针对手写规则烦琐的问题,可采用将规则与统计学习结合的马尔可夫逻辑网(Markov logic network, MLN)^[5-7],MLN是经典的推理方法。知识图谱的层次结构和逻辑结构往往被表示为本体(ontology)或模式(schema),基于本体的推理方法是知识

图谱推理的经典方法之一,也是现实应用中很有意义的方法^[8]。随机游走(random walk)是概率统计中经典的随机过程,将其应用于知识图谱推理中产生了著名的路径排序算法(path ranking algorithm, PRA)^[9]及其与深度强化学习结合的深度路径(deep path)方法^[10]。近年来,随着以深度学习为基础的人工智能技术的蓬勃发展,将知识图谱嵌入低维空间的方法逐渐成为主流,TransE^[11]是其中的先锋。

本文关注的知识图谱推理方法是从TransE开始的,它源自自然语言处理领域的word2vec^[12],是当前效果最好的、主流的、契合深度学习潮流的方法,也被称为现代的方法。这类方法可学习出知识图谱的稠密向量表示,是知识图谱领域的表示学习。对于学习出的稠密向量,既可以直接进行推理应用,也便于使用各种深度学习模型和算法来实现下游任务,如知识问答或辅助决策等。

知识图谱 $G=\{<h,r,t>\} \subseteq E \times R \times E$,其中 E 为实体集合, $h,t \in E$; R 为关系集合, $r \in R$ 。现代知识图谱推理方法的目标是通过定义一个打分函数 $f_r(h,t)$ 来学习出表示实体和关系的向量 $h,t \in \mathbb{R}^{d_e}$, $r \in \mathbb{R}^{d_r}$,如式(1)所示,其中 \rightarrow 表示映射:

$$f_r(h,t): \mathbb{R}^{d_e} \times \mathbb{R}^{d_r} \times \mathbb{R}^{d_e} \rightarrow \mathbb{R} \quad (1)$$

对于符合知识图谱的正样本来说,期望打分函数计算出的分数无限接近于0。其中, d_e 和 d_r 表示实体和关系向量的维度,在大多数模型中, $d_e=d_r$ 。在进行知识图谱推理时,对于任意给定的实体对 $<h,t>$,遍历知识图谱中的关系 $r \in R$,并计算 $<h,r,t>$ 的分数,若低于某个阈值,则表示 h 和 t 之间存在关系 r 。在某些时候需要从 $<h,r>$ 中推断 t ,此时可以遍历所有的实体 $t \in E$,使用式(1)来计算分数,若分数低于某个阈值,则表明实体 h 在关系 r 的作用下,得到了实

体 t 。本文使用式(1)所表示的模型框架来介绍各种现代的知识推理方法。

3 基于几何运算的方法

基于几何运算的模型是从word2vec延伸出来的,将知识图谱通过平移或旋转等几何运算嵌入低维的几何空间中(通常是欧几里得空间,也可以是双曲空间等)。其中平移表现为向量加法,旋转表现为哈达玛积(Hadamard product),嵌入则是一个数学中与流形相关的概念,表达一个数学结构的实例通过映射包含到另一个实例中。

将知识图谱嵌入几何空间时,解决知识图谱中不同特点的关系的推理问题,从而推进基于几何运算的方法的发展,这些特点包括一对一、一对多、多对一、多对多^[13]、对称性(symmetry)、反对称性(anti-symmetry)、反向性(inversion)和组合性(composition)^[14]等。

3.1 欧几里得空间嵌入

知识图谱推理的现代方法的雏形是word2vec,并从TransE模型开始逐渐发展起来。图1表示了word2vec模型学习出来的词向量满足 $w_{\text{广东省}} - w_{\text{广州市}} = w_{\text{浙江省}} - w_{\text{杭州市}}$,其隐含的关系“省会”(图1中虚线)没有被明确表示出来。TransE将word2vec中隐含的关系用向量明确地表示出来,并应用到知识图谱中。

基于几何运算的知识图谱推理在TransE^[11]的基础上持续发展。TransE把实体间的关系用向量明确地表示出来,并用几何平移来解释实体间的关系,如图2(a)所示。继TransE后大量平移或旋转的方法被提出来,这些方法和TransE一样使用欧氏

距离来计算打分函数,使用基于能量的方法来定义损失函数,并用随机梯度下降来优化模型。

TransE将实体和关系嵌入同一个空间中,并定义打分函数为:

$$f_r(\mathbf{h}, \mathbf{t}) = \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_2^2 \quad (2)$$

其中, $\|\cdot\|_2$ 表示 L2 范数 (L2 norm)。TransE 模型对仅有一对一关系的知识图谱非常友好,能够学习出各种具有反对称性、反向性和组合性的关系。其结构简单、运算量小,是某些现实场景的首选方法。

TransH^[13]拓展了TransE模型,为每个关系学习嵌入空间的一个超平面 \mathbf{W}_r ($\|\mathbf{W}_r\|_2=1$),并将 $\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle$ 三元组解释为实体在关系超平面上的平移变换。如图2(b)所示,TransH先将实体向量 \mathbf{h} 和 \mathbf{t} 映射到关系超平面 \mathbf{W}_r 上,得到 $\mathbf{h}' = \mathbf{h} - \mathbf{W}_r^\top \mathbf{h} \mathbf{W}_r$ 和 $\mathbf{t}' = \mathbf{t} - \mathbf{W}_r^\top \mathbf{t} \mathbf{W}_r$,并将关系表示为超平面上的平移变换,即向量加法 $\mathbf{h}' + \mathbf{r} = \mathbf{t}'$ 。由此TransH的打分函数为:

$$f_r(\mathbf{h}, \mathbf{t}) = \|\mathbf{h}' + \mathbf{r} - \mathbf{t}'\|_2^2 = \|\left(\mathbf{h} - \mathbf{W}_r^\top \mathbf{h} \mathbf{W}_r\right) + \mathbf{r} - \left(\mathbf{t} - \mathbf{W}_r^\top \mathbf{t} \mathbf{W}_r\right)\|_2^2 \quad (3)$$

TransH模型通过关系特定的超平面,实现了自反、一对多、多对一和多对多的嵌入表示。TransR模型^[15]将实体和关系分别嵌入不同的几何空间,使得相同的实体在不同的关系下能够表示不同的语义,进一步增强了知识图谱的推理能力。如图2(c)所示,TransR通过映射矩阵 $\mathbf{M}_r \in \mathbb{R}^{d_e \times d_r}$ 将实体向量 \mathbf{h} 、 \mathbf{t} 映射到关系空间得到 $\mathbf{h}_r = \mathbf{h} \mathbf{M}_r$ 和 $\mathbf{t}_r = \mathbf{t} \mathbf{M}_r$,并在关系空间进行平移变换,即 $\mathbf{h}_r + \mathbf{r} = \mathbf{t}_r$ 。由此,将TransR的打分函数定义为:

$$f_r(\mathbf{h}, \mathbf{t}) = \|\mathbf{h}_r + \mathbf{r} - \mathbf{t}_r\|_2^2 = \|\mathbf{h} \mathbf{M}_r + \mathbf{r} - \mathbf{t} \mathbf{M}_r\|_2^2 \quad (4)$$

TransD模型^[16]用实体映射向量 $\mathbf{h}_p, \mathbf{t}_p \in \mathbb{R}^{d_r}$ 和关系映射向量 $\mathbf{r}_p \in \mathbb{R}^{d_r}$ 来构造映射矩阵 $\mathbf{M}_{r_1} = \mathbf{r}_p \mathbf{h}_p^\top + \mathbf{I}^{d_r \times d_e}$ 和 $\mathbf{M}_{r_2} = \mathbf{r}_p \mathbf{t}_p^\top + \mathbf{I}^{d_r \times d_e}$

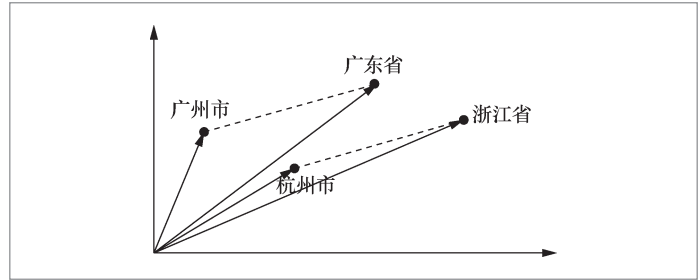


图1 word2vec 的示例

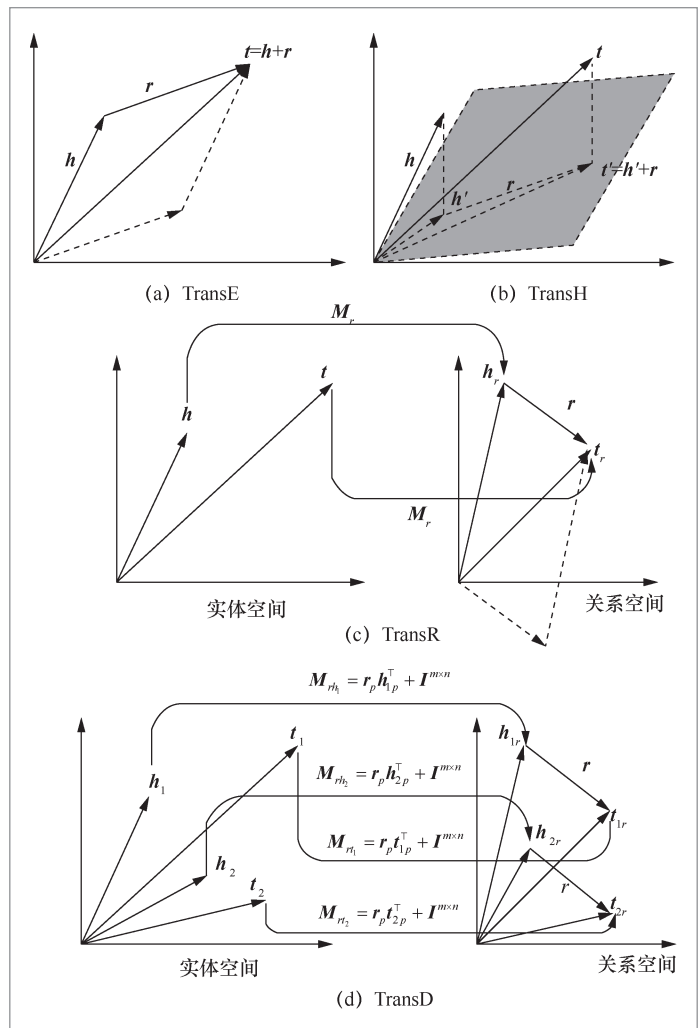


图2 基于欧几里得空间的平移运算的知识图谱推理模型

(\mathbf{I} 表示单位矩阵),并代替TransR中的映射矩阵 \mathbf{M}_r ,使得模型能够表达实体在不同关系中潜在的多个语义关系。由于向量乘法效率高于矩阵乘法,将TransD模型应用于大规模知识图谱中具有计算效率上的优

势。类似地, TransD的打分函数为:

$$\begin{aligned} f_r(\mathbf{h}, \mathbf{t}) &= \|\mathbf{h}_r + \mathbf{r} - \mathbf{t}_r\|_2^2 = \\ &= \|\mathbf{h}\mathbf{M}_{rh} + \mathbf{r} - \mathbf{t}\mathbf{M}_{rt}\|_2^2 = \\ &= \|\mathbf{h}(\mathbf{r}_p\mathbf{B}_p^\top + \mathbf{I}^{d_r \times d_e}) + \mathbf{r} - (\mathbf{r}_p\mathbf{t}_p^\top + \mathbf{I}^{d_r \times d_e})\|_2^2 \end{aligned} \quad (5)$$

图2(d)直观地表示了TransD的模型。从图2和式(2)~式(5)可以看出, TransE、TransH、TransR和TransD 4个模型一脉相承, 都是用欧几里得空间中的平移来解释实体间的关系, 并用欧氏距离来表示分数。进一步地, TransG模型^[17]对这一模式进行了泛化建模, 并用贝叶斯参数无限混合模型(Bayesian non-parametric infinite mixture model)^[18]和中国餐馆过程(Chinese restaurant process, CRP)来解决关系的多语义表达问题。

TransG的打分函数是:

$$f_r(\mathbf{h}, \mathbf{t}) = \sum_{i=1}^{m_r} \pi_{r,i} \exp\left(-\frac{\|\mathbf{u}_h + \mathbf{u}_{r,i} - \mathbf{u}_t\|_2^2}{\sigma_h^2 + \sigma_t^2}\right) \quad (6)$$

其中, $\mathbf{u}_h \sim N(0, 1)$ 和 $\mathbf{u}_t \sim N(0, 1)$ 是先验地从标准正太分布中抽样获得的平均向量, σ_h 和 σ_t 是相应的方差; m_r 是关系的语义成分数量, 通过中国餐馆过程能够从训练语料中学习出关系的不同的语义成分; $\pi_{r,i} \sim \text{CRP}(\beta)$ 是关系 r 的语义成分的混合因子(mixing factor)^[17], 可以通过中国餐馆过程的吉布斯采样获得; $\mathbf{u}_{r,i}$ 表示关系 r 的第 i 个语义成分的平移向量。

$\mathbf{u}_{r,i} = \mathbf{t} - \mathbf{h} \sim N(\mathbf{u}_t - \mathbf{u}_h, (\sigma_h^2 + \sigma_t^2)\mathbf{E})$ 是由头实体向量 $\mathbf{h} \sim N(\mathbf{u}_h, \sigma_h^2\mathbf{E})$ 和尾实体向量 $\mathbf{t} \sim N(\mathbf{u}_t, \sigma_t^2\mathbf{E})$ 运算得到的。

总体来说, TransG模型能够表示关系的多种语义。对其进行主成分(primary component)分析可知, TransE模型是TransG模型针对主成分的特例, 而TransG则是TransE的泛化, 图3展示了两个模型。

旋转和平移一样是几何空间的基本操作, RotatE^[14]使用旋转代替平移对关

系进行建模, 图4展示了RotatE与TransE的区别。同时, 为了表达旋转, 实体和关系的嵌入向量从实数扩展到复数向量空间。旋转在复向量空间的运算表示为向量的哈达玛积, 即每个元素分别相乘的运算, 这个与平移操作的加法是类似的。对于 $\langle \mathbf{h}, \mathbf{r}, \mathbf{t} \rangle$, RotatE将实体和关系嵌入复向量空间 $\mathbf{h}, \mathbf{t}, \mathbf{r} \in \mathbb{C}^d$, 并定义打分函数为:

$$f_r(\mathbf{h}, \mathbf{t}) = \|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\| \quad (7)$$

其中, \circ 表示哈达玛积, 模型将向量限制在单位圆中, 并解释为逆时针的旋转, 作用于复向量的相位部分。与平移的方法相比, RotatE能更加高效地实现对称和反对称、反向和组合这些关系类型的建模。例如, $\mathbf{r} = \mathbf{e}^{0/i\pi} = \pm \mathbf{1}$ 可以表示关系 r 是对称的, \mathbf{r}_1 和 \mathbf{r}_2 是共轭的, 可以表示两个关系 \mathbf{r}_1 和 \mathbf{r}_2 是反向的, $\mathbf{r}_3 = \mathbf{r}_1 \circ \mathbf{r}_2$ 可以表示 \mathbf{r}_3 是 \mathbf{r}_1 和 \mathbf{r}_2 的组合。RotatE对具有如上特点的知识图谱推理能够事半功倍。

3.2 双曲空间嵌入

双曲空间(hyperbolic space)是具有常数负曲率的齐次空间, 而欧几里得空间则是零曲率的。双曲几何提供了高效的方法来学习层次数据的低维嵌入, 特别地, 仅仅用二维的双曲空间就能够以任意低的失真度嵌入树形数据^[19]。也就是说, 将类似知识图谱这样具有丰富层次结构的数据嵌入双曲空间中, 比嵌入欧几里得空间更加高效。

双曲空间嵌入通常使用 d 维庞加莱球 $\mathcal{B}_c^d = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\|_2 < \frac{1}{c}\}$ 来表示, 其中 d 表示庞加莱球的维度, $-c$ 表示曲率。庞加莱球的原点所对应的切线空间 \mathcal{T}_0^c 是从原点离开的有可能路径的方向的 d 维向量空间。切线空间 \mathcal{T}_0^c 到 \mathcal{B}_c^d 的映射通过指数映射(exponential map) $\exp_0^c(\mathbf{v})$ 实现, 而从 \mathcal{B}_c^d 到 \mathcal{T}_0^c 则通过对数映射(logarithmic

map) $\log_0^c(v)$ 实现^[20], 这里的0表示原点:

$$\begin{aligned}\exp_0^c(v) &= \frac{\tanh(\sqrt{c}v)}{\sqrt{c}\|v\|}v \\ \log_0^c(v) &= \frac{\operatorname{artanh}(\sqrt{c}\|v\|)}{\sqrt{c}\|v\|}v\end{aligned}\quad (8)$$

在双曲空间中, 向量 \mathbf{x} 、 \mathbf{y} 的加法通过莫比乌斯加法 \oplus_c 实现, 向量 \mathbf{y} 和矩阵 \mathbf{M} 的乘法通过莫比乌斯矩阵-向量乘法 \otimes_c 实现, 分别定义如下:

$$\mathbf{x} \oplus_c \mathbf{y} = \frac{(1 + 2c\langle \mathbf{x}, \mathbf{y} \rangle + c\|\mathbf{y}\|^2)\mathbf{x} + (1 - c\|\mathbf{x}\|^2)\mathbf{y}}{1 + 2c\langle \mathbf{x}, \mathbf{y} \rangle + c^2\|\mathbf{x}\|^2\|\mathbf{y}\|^2} \quad (9)$$

$$\mathbf{M} \otimes_c \mathbf{y} = \exp_0^c(\mathbf{M} \log_0^c(\mathbf{y})) \quad (10)$$

其中, $\langle \mathbf{x}, \mathbf{y} \rangle$ 表示两个向量的内积, 其结果是个标量。类似欧几里得空间的欧氏距离, 双曲空间距离^[21-22] 定义如下:

$$d_{\mathbb{B}}^c(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{c}} \operatorname{artanh}(\sqrt{c}\|\mathbf{x} \oplus_c \mathbf{y}\|) \quad (11)$$

在双曲空间中, 三元组 $\langle h, r, t \rangle$ 对应的 d 维向量 \mathbf{h} 、 \mathbf{t} 、 $\mathbf{r} \in \mathbb{B}_c^d$, 其中 c 表示双曲空间的曲率。和欧几里得空间一样, 模型通过打分函数 $f_r(\mathbf{h}, \mathbf{t})$ 来实现知识图谱推理。

MuRP^[23] 模型采用曲率为-1的庞加莱球模型 ($c=1$) 对图谱的实体关系进行建模, 使用式 (11), 类似 TransE, MuRP 的打分函数定义为:

$$\begin{aligned}f_r(\mathbf{h}, \mathbf{t}) &= -d_{\mathbb{B}}^c(\mathbf{h}^r, \mathbf{t}^r)^2 + b_h + b_t \\ &= -d_{\mathbb{B}}^c(\mathbf{R} \otimes_c \mathbf{h}, \mathbf{t} \oplus_c \mathbf{r})^2 + b_h + b_t\end{aligned}\quad (12)$$

其中, $\mathbf{R} \in \mathbb{B}_c^{d \times d}$ 是关系对角矩阵, b_h 、 b_t 是两个偏置。

双曲嵌入与欧几里得空间嵌入不同的是, 模型会为每个实体额外学习一个偏置 b , b 表示实体在双曲空间中的影响范围, 图5是MuRP模型的几何直观解释。即头实体 h 和尾实体 t 经过同一关系 r 调整为双曲空间的两个球体, 如果存在重叠, 则表示 $\langle h, r, t \rangle$ 成立, 反之则表示 h 和 t 不存在关系 r 。

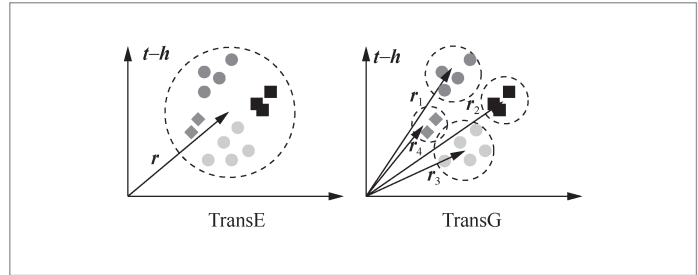


图3 TransG 和 TransE 对关系 r 进行不同表达, 在 TransE 中只能学习出一种向量表示 r , 而 TransG 能够学习出有细微语义差别的 4 个关系向量 r_1 、 r_2 、 r_3 和 r_4

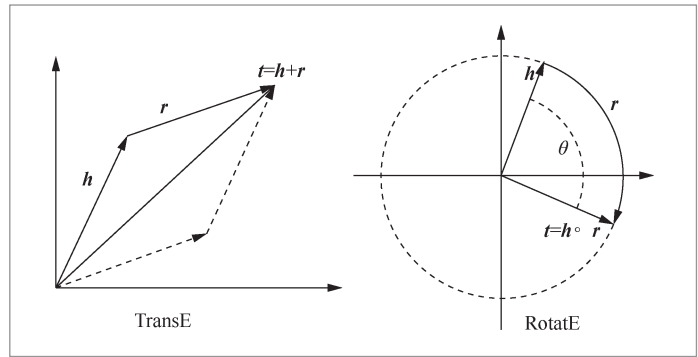


图4 RotatE 和 TransE 模型的比较

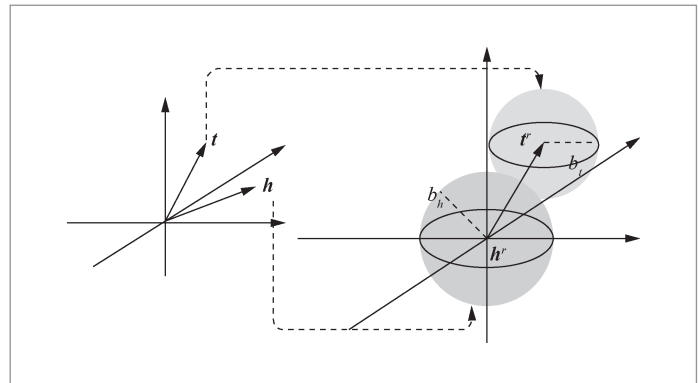


图5 MuRP 模型的几何直观解释

和 MuRP 使用平移来建模实体间的关系不同, ATTH^[24] 模型使用吉文斯变换表示旋转和反射, 从而捕捉不同语义的关系, 并使用双曲注意力机制来融合几何运算, 同时归纳出更为复杂的逻辑模式。此外,

考虑到不同关系蕴含的语义信息不一样, ATTH采用可学习的变量为不同的关系自动学习出最佳的曲率 c_r 。

类似 TransR、TransD 定义出双曲空间中的旋转运算 $\mathbf{h}'_{\text{rot}} = \mathbf{M}'_{\text{rot}} \mathbf{h}$ 和反射运算

$$\mathbf{h}'_{\text{ref}} = \mathbf{M}'_{\text{ref}} \mathbf{h}, \text{ 这里的 } \mathbf{M}'_{\text{rot}} = \text{diag}\left(G^+(\theta_{r,1}), \dots, G^+\left(\theta_{r,\frac{d}{2}}\right)\right) \text{ 和 } \mathbf{M}'_{\text{ref}} = \text{diag}\left(G^-(\phi_{r,1}), \dots, G^-\left(\phi_{r,\frac{d}{2}}\right)\right)$$

是由关系特定的参数 $\Theta_r := \{\theta_{r,i}\}, i \in \left\{1, 2, \dots, \frac{d}{2}\right\}$

和 $\Phi_r := \{\phi_{r,i}\}, i \in \left\{1, 2, \dots, \frac{d}{2}\right\}$ 构造的块对角矩阵,

其中, $G^\pm(\theta) := \begin{bmatrix} \cos(\theta) & \mp \sin(\theta) \\ \sin(\theta) & \pm \cos(\theta) \end{bmatrix}$ 是 2×2 的吉文斯变换矩阵。

ATTH模型中注意力机制是在庞加莱球的切线空间(一种欧几里得空间)中进行的。通过式(8)的对数映射将双曲空间嵌入向量 \mathbf{h}'_{rot} 和 \mathbf{h}'_{ref} 映射到切线空间 $\ddot{\mathbf{h}}'_{\text{rot}} = \log_0^c(\mathbf{h}'_{\text{rot}})$ 和 $\ddot{\mathbf{h}}'_{\text{ref}} = \log_0^c(\mathbf{h}'_{\text{ref}})$, 并计算注意力得分 $\alpha_{\text{rot}} = \text{softmax}(\mathbf{a}^\top \ddot{\mathbf{h}}'_{\text{rot}})$ 和 $\alpha_{\text{ref}} = \text{softmax}(\mathbf{a}^\top \ddot{\mathbf{h}}'_{\text{ref}})$ (\mathbf{a} 是注意力机制的环境参数), 然后通过式(8)映射回双曲空间, 得到头实体经过关系调整后的向量 $\tilde{\mathbf{h}} = \exp_0^c(\alpha_{\text{rot}} \ddot{\mathbf{h}}'_{\text{rot}} + \alpha_{\text{ref}} \ddot{\mathbf{h}}'_{\text{ref}})$ 。最后类似MuRP, ATTH的打分函数定义为:

$$f_r(\mathbf{h}, \mathbf{t}) = -d_{\mathbb{B}}^c(\tilde{\mathbf{h}} \oplus_{c_r} \mathbf{r}, \mathbf{t})^2 + b_h + b_t \quad (13)$$

4 基于深度学习的方法

近年来, 深度学习被广泛地应用在计算机视觉、图像处理、语音识别和自然语言处理等人工智能领域。同样地, 深度学习也被应用于知识图谱推理, 本文介绍了主流的和应用非常广泛的若干方法。

首先介绍基于深度学习的知识图谱推理方法, 然后将相应的神经网络转化成打

分函数。深度神经网络更加复杂, 且包含了大量的非线性运算, 从而能够学到更多的特征, 但也因此存在“黑盒”的不直观问题, 影响了推理的解释性。

4.1 卷积神经网络模型

卷积神经网络将卷积应用于神经网络, 通过局部感知、参数共享等特点, 实现了深层的特征学习, 成为深度学习中非常重要的网络之一。ConvE^[25]将卷积神经网络应用于知识图谱推理, 其网络结构如图6(a)所示, 打分函数定义为:

$$f_r(\mathbf{h}, \mathbf{t}) = \text{ReLU}(\text{flatten}(\text{ReLU}(\text{concat}(\bar{\mathbf{r}}, \bar{\mathbf{h}}) * \omega)) \mathbf{W}) \mathbf{t} \quad (14)$$

其中, $\bar{\mathbf{r}}$ 、 $\bar{\mathbf{h}}$ 是由 \mathbf{r} 和 \mathbf{h} 重组成的二维矩阵, 维数关系为 $d = d_w \times d_h$, d_w 、 d_h 是 $\bar{\mathbf{r}}$ 、 $\bar{\mathbf{h}}$ 维度的长和宽。把 $\bar{\mathbf{r}}$ 、 $\bar{\mathbf{h}}$ 拼接成二维矩阵, 与滤波器 ω 进行卷积后, 得到 k 个 $m \times n$ 特征图, 式(14)中 $*$ 表示卷积。使用非线性激活函数ReLU将特征图拉平为 kmn 维的向量, 然后与矩阵 $\mathbf{W} \in \mathbb{R}^{kmn \times d}$ 相乘得到 d 维向量(图6(a)中使用全连接网络实现), 再次ReLU后与尾实体嵌入向量 \mathbf{t} 相乘求内积, 得到分数值。

ConvE模型通常使用小的滤波器进行卷积, 这样实体和关系间的交互仅在二维矩阵 $\bar{\mathbf{r}}$ 和 $\bar{\mathbf{h}}$ 的拼接处, 能够很好地被学习, 而远离拼接处的特征则较少被学习。ConvR^[26]将关系转化为滤波器来对实体向量进行卷积运算, 从而获得更好的实体-关系间交互的特征表达。通过比较图6(a)和图6(b), 可以直观地理解ConvE和ConvR的异同, 并理解各自的特点和优劣。类似地, ConvR的打分函数为:

$$f_r(\mathbf{h}, \mathbf{t}) = \text{ReLU}(\text{concat}(\text{flatten}(\text{ReLU}((\bar{\mathbf{h}} * \omega^{(i)}))) \mathbf{M}) \mathbf{t} \quad (15)$$

和ConvE一样, $\bar{\mathbf{h}}$ 由 \mathbf{h} 重组得到, 不同

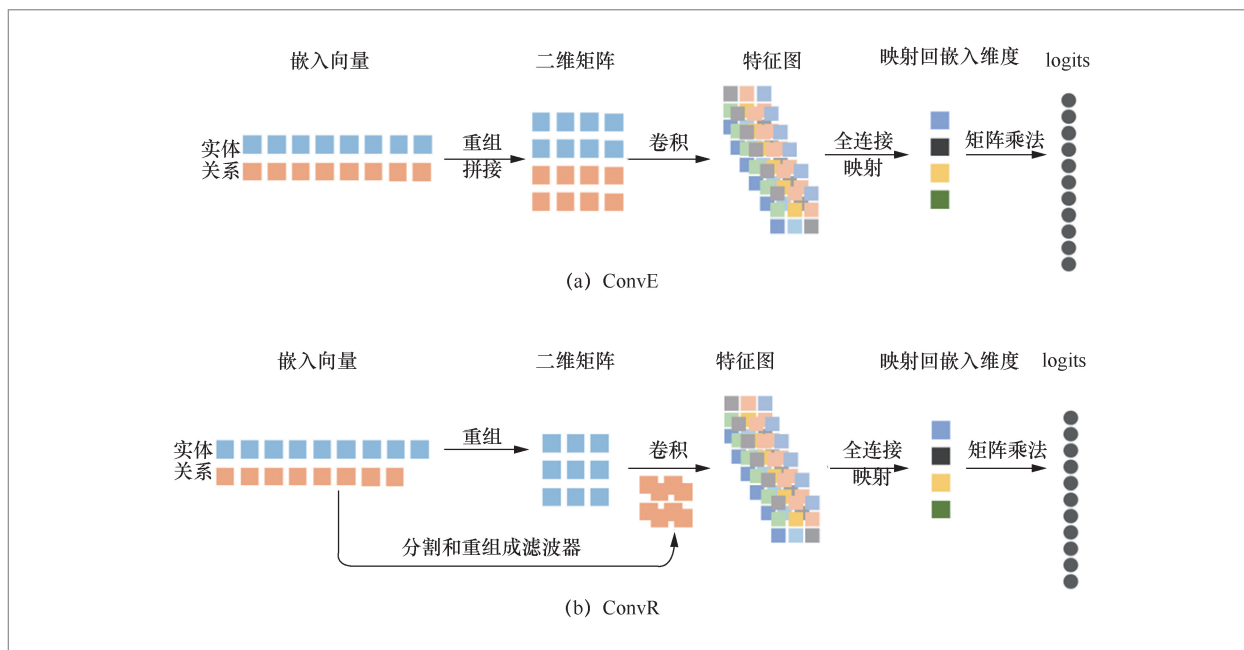


图6 深度卷积网络用于知识图谱推理的 ConvE 和 ConvR 模型

的是, 关系向量 \mathbf{r} 被分割为 k 个 $\frac{d_r}{k}$ 维的向量 $\mathbf{r}^{(1)}, \mathbf{r}^{(2)}, \dots, \mathbf{r}^{(k)}$, 每个 $\mathbf{r}^{(i)}$ 都被重组为 $h \times w$ 的滤波器 $\omega^{(i)}$ (h 、 w 为滤波器的长和宽), 其中 $d_r = khw$ 。每个滤波器 $\omega^{(i)}$ 与 $\bar{\mathbf{h}}$ 进行卷积得到特征图 $\mathbf{K}^{(i)} \in \mathbb{R}^{(d_h-h+1) \times (d_w-w+1)}$, 将所有特征图 $\mathbf{K}^{(i)}$ 拉平后拼接成向量, 并与矩阵 $\mathbf{W} \in \mathbb{R}^{k(d_h-h+1)(d_w-w+1) \times d}$ 相乘得到 d_e 维向量。再次 ReLU 后与 \mathbf{t} 相乘, 得到分数。

4.2 胶囊网络模型

CapsE^[27] 是将胶囊网络 (capsule

network, CapsNet)^[28] 用于知识图谱推理的一种模型。其基本原理是将三元组的3个嵌入向量组装成3列的矩阵, 用卷积网络来学习出相应的特征, 并通过两层的胶囊网络为三元组打分。CapsE模型如图7所示, 由于滤波器同时作用于 \mathbf{h} 、 \mathbf{r} 、 \mathbf{t} , 因此可以捕捉到更多的特征。其打分函数为:

$$f_r(\mathbf{h}, \mathbf{t}) = \|\text{capsnet}(\text{ReLU}([\mathbf{h}, \mathbf{r}, \mathbf{t}] * \Omega))\| \quad (16)$$

其中, capsnet 表示胶囊网络, Ω 表示 n 个滤波器 $\omega \in \mathbb{R}^{1 \times 3}$, 滤波器 ω 作用于输入矩阵 $[\mathbf{h}, \mathbf{r}, \mathbf{t}] \in \mathbb{R}^{d \times 3}$, 形成 n 个 $d \times 1$ 维的特征图。

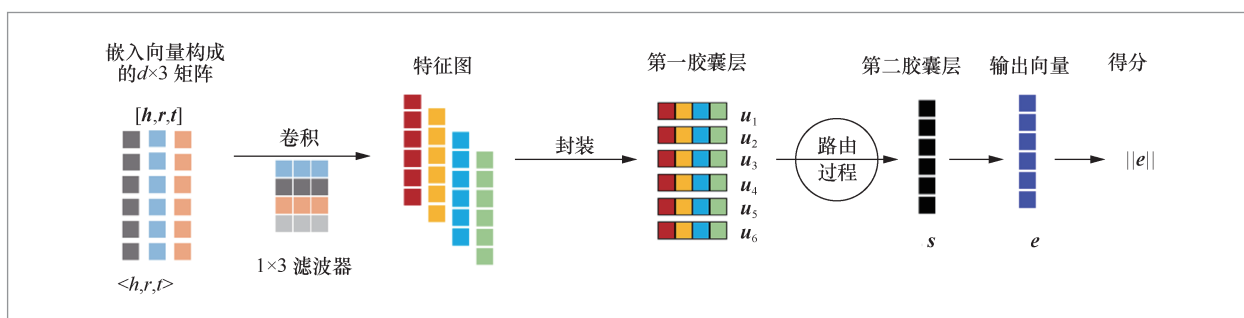


图7 胶囊网络用于知识图谱推理的 CapsE 模型

特征图被封装成第一个胶囊层的 d 个胶囊 $\mathbf{u}_i \in \mathbb{R}^{n \times 1}$ ，通过路由过程路由到第二个胶囊层后生成连续的输出向量 \mathbf{e} ，输出向量的L1范数 $\|\mathbf{e}\|$ 就是得分。

路由过程^[28]在第一个胶囊层的 $s = \sum_i c_i \mathbf{W}_i \mathbf{u}_i$ ($\mathbf{W}_i \in \mathbb{R}^{d \times n}$ 是权重矩阵, c_i 是耦合系数) 和第二胶囊层的非线性压缩 $\mathbf{e} = \text{squash}(s) = \frac{\|s\|^2}{1 + \|s\|^2} \frac{s}{\|s\|}$ 中迭代 m 次(参考文献[27]研究了 m 的数值, 并断定 $m=1$ 时效果最佳)。

4.3 图神经网络模型

图神经网络是将深度学习技术应用用于图结构数据的方法, 是最近人工智能中非常热门的研究领域之一。图神经网络天然地适用于知识图谱推理。关系图卷积网络 (relational graph convolutional network, R-GCN)^[29]是较早对关系进行建模实现知识图谱推理的图自编码器模型。R-GCN模型的每一层都使用了图8和式(17)所示的传播模型来编码知识图谱:

$$\mathbf{e}^{(l+1)} = \text{ReLU} \left(\mathbf{W}_{\text{self}}^{(l)} \mathbf{e}^{(l)} + \sum_{r \in R} \sum_{i \in E^r} \frac{1}{c_r} \mathbf{W}_r^{(l)} \mathbf{e}_i^{(l)} \right) \quad (17)$$

其中, $\mathbf{e}^{(l)}$ 和 $\mathbf{e}^{(l+1)}$ 分别表示知识图谱的实体 e 在R-GCN中的 l 层和 $l+1$ 层的向量。 E^r 表示与实体 e 具有关系 r 的所有邻接实体的集

合, 区分入边和出边。 R 表示知识图谱中所有关系的集合。 $\mathbf{W}_{\text{self}}^{(l)}$ 表示 l 层中实体自身的相关参数(自连接), 其目的是给 $l+1$ 层传播实体自身的信息。 $\mathbf{W}_r^{(l)}$ 则是与关系 r 相关的参数。 c_r 是归一化参数, 可以随模型学习, 或者提前设定。R-GCN传播模型可被直观地理解为依据每一个关系对邻接节点的信息进行学习, 进而表示实体。

式(17)的参数 $\mathbf{W}_r^{(l)}$ 会因知识图谱中关系的增加而急剧变大, 导致模型巨大, 并且非常容易过拟合。基分解 (basis decomposition) 和块对角分解 (block diagonal decomposition) 是解决这个问题的两种方法, 在使用R-GCN时可以根据实际情况选择其中一种。基分解的出发点是在关系间共享参数, 即将 $\mathbf{W}_r^{(l)}$ 分解为关系无关的 B 个参数矩阵 $\mathbf{V}_b^{(l)} \in \mathbb{R}^{d^{(l+1)} \times d^{(l)}}$ 的线性组合, 组合系数为关系相关的标量参数 $a_{rb}^{(l)}$:

$$\mathbf{W}_r^{(l)} = \sum_{b=1}^B a_{rb}^{(l)} \mathbf{V}_b^{(l)} \quad (18)$$

而块对角分解的定义为:

$$\mathbf{W}_r^{(l)} = \text{diag}(\mathbf{Q}_r^{(l)}, \dots, \mathbf{Q}_{Br}^{(l)}) \quad (19)$$

其中, $\mathbf{Q}_{br}^{(l)} \in \mathbb{R}^{\frac{d^{(l+1)}}{B} \times \frac{d^{(l)}}{B}}$ 表示块对角分解的参数矩阵。

R-GCN模型堆叠了 L 层式(17)作为编码器, 并使用DistMult^[30]作为解码器, 从而其打分函数为:

$$f_r(\mathbf{h}, \mathbf{t}) = \text{rgcn}_L(\mathbf{h})^\top \mathbf{R}_r \text{rgcn}_L(\mathbf{t}) \quad (20)$$

其中, $\mathbf{R}_r \in \mathbb{R}^{d \times d}$ 是关系特定的对角矩阵, 和 r 一起用来表示关系; rgcn_L 表示堆叠了 L 层式(17)的R-GCN模型, 输出编码了 \mathbf{h} 的 d 维向量。

RGHAT^[31]是一种通过加入两层注意力来进一步改善关系图神经网络的效果的知识图谱推理方法。ATTH^[24]使用注意力机制自动归纳出知识图谱的逻辑结构, 而RGHAT使用两层注意力分别归纳实

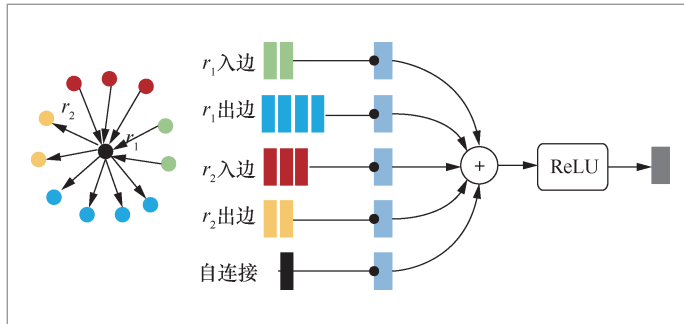


图8 R-GCN的传播模型, 也是R-GCN的“层”

体和同一关系下邻接实体的特征,其模型如图9所示,其中关系层注意力表达了不同关系对实体的影响权重:

$$\begin{aligned} \mathbf{a}_{h,r} &= \mathbf{W}_1 \text{concat}(\mathbf{h}, \mathbf{r}) \\ \alpha_{h,r} &= \text{softmax}_r(\mathbf{a}_{h,r}) = \frac{\exp(\sigma(\mathbf{p} \cdot \mathbf{a}_{h,r}))}{\sum_{r' \in R_h} \exp(\sigma(\mathbf{p} \cdot \mathbf{a}_{h,r'}))} \end{aligned} \quad (21)$$

实体层注意力则表达了同一关系的邻接实体的影响权重:

$$\begin{aligned} \mathbf{b}_{h,r,t} &= \mathbf{W}_2 \text{concat}(\mathbf{a}_{h,r}, \mathbf{t}) \\ \beta_{r,t} &= \text{softmax}_t(\mathbf{b}_{h,r,t}) = \frac{\exp(\sigma(\mathbf{q} \cdot \mathbf{b}_{h,r,t}))}{\sum_{t' \in E_{h,r}} \exp(\sigma(\mathbf{q} \cdot \mathbf{b}_{h,r,t'}))} \end{aligned} \quad (22)$$

其中, $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{d \times 2d}$ 和 $\mathbf{p}, \mathbf{q} \in \mathbb{R}^d$ 是注意力参数, R_h 表示实体 h 的所有邻接关系的集合, $E_{h,r}$ 是实体 h 在关系 r 下的尾实体集合, σ 表示激活函数 LeakyReLU。 $\alpha_{h,r}$ 是关系层注意力分数, $\beta_{r,t}$ 是实体层注意力分数, 三元组 $\langle h, r, t \rangle$ 的注意力分数定义为二者的乘积 $u_{h,r,t} = \alpha_{h,r} \times \beta_{r,t}$ 。RGHAT模型使用信息聚合器来聚合所有邻接节点的信息, 并使用与R-GCN模型一样的自连接^[29]来表达当前实体自身的信息, 即:

$$\mathbf{h}'_+ = \sigma \left(\mathbf{W}_3 \left(\mathbf{h} + \sum_{r \in R_h} \sum_{t \in E_{h,r}} u_{h,r,t} \mathbf{b}_{h,r,t} \right) \right) \quad (23)$$

其中, $\mathbf{W}_3 \in \mathbb{R}^{d \times d}$ 是训练参数。RGHAT使用式(14)的ConvE^[25]作为解码器, 打分

函数为:

$$\begin{aligned} f_r(\mathbf{h}, \mathbf{t}) &= \\ &\text{ReLU}(\text{flatten}(\text{ReLU}(\text{concat}(\mathbf{h}', \mathbf{r}) * \boldsymbol{\omega})) \mathbf{W}_{\text{dec}}) \mathbf{t} \end{aligned} \quad (24)$$

5 知识推理应用

5.1 知识图谱补全

知识图谱普遍存在知识缺失的事实, 也就是知识图谱的不完备性。如图10所示, 实线部分是知识图谱存在的关系, 比如 $\langle \text{张三}, \text{出生于}, \text{厦门市} \rangle$ 和 $\langle \text{张三}, \text{任职于}, \text{甲公司} \rangle$ 等, 而虚线部分是知识图谱中缺失的, 比如 $\langle \text{张三}, \text{工作于}, \text{上海市} \rangle$ 和 $\langle \text{张三}, \text{出生于}, \text{中国} \rangle$ 等。知识图谱补全 (knowledge graph completion) 的目标是发现这些缺失的知识, 并将其补充到知识图谱中, 使得知识图谱趋向完备。

知识图谱补全是知识推理应用最多的领域, 大量的知识图谱推理算法被提出来的初衷就是应用于知识图谱补全, 如MLN^[7]、TransR^[15]、CapsE^[27]、RGHAT^[31]等。前面提到的所有方法都可以通过在向量空间的推理来断定任意实体间是否存在某种关系, 进而实现知识图谱的补全。

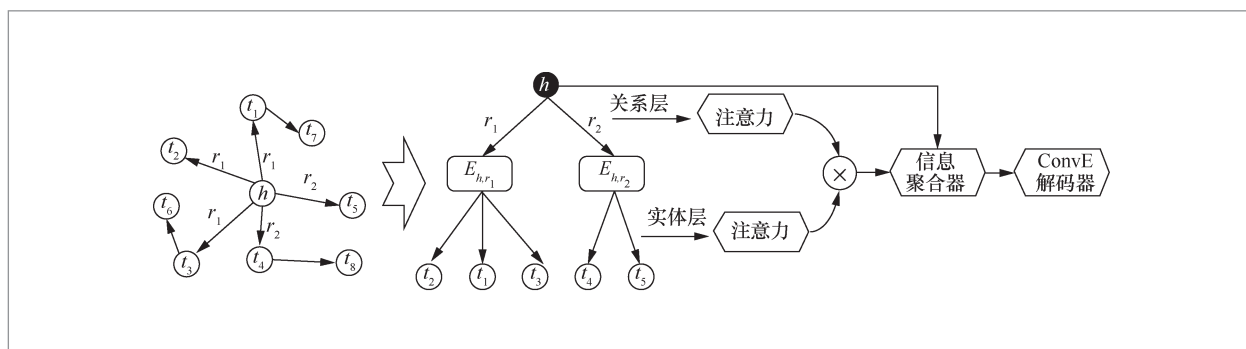


图9 RGHAT 模型的网络结构

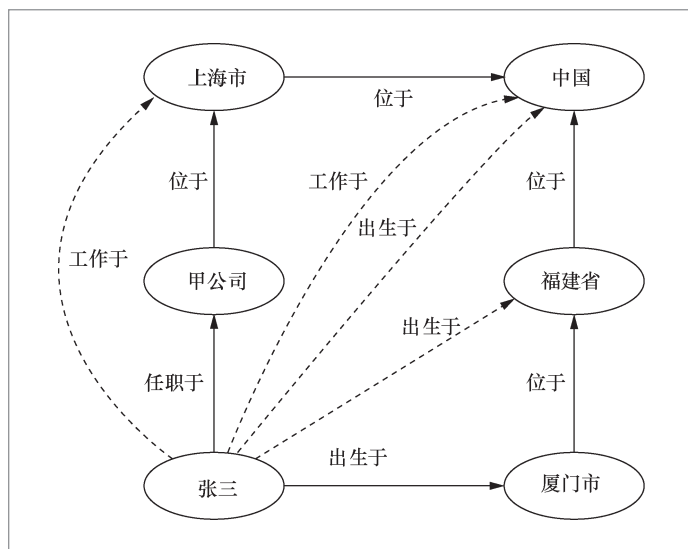


图 10 知识图谱的不完备性

5.2 知识问答

问答 (question answering, QA) 系统是自然语言处理领域一个重要的发展方向,其目标是通过自然语言的方式来获取知识。在很多真实应用场景下,基于知识图谱的问答系统能够带来极大的知识获取的便利性。但除了简单地从知识图谱中检索实体,更多的问题要求问答系统具有知识推理的能力^[32-33]。本文介绍的知识推理方法都可以用于基于知识图谱的知识问答系统中,用来回答复杂的问题,提升结果的准确性、完备性等^[34-35]。

除了将知识图谱嵌入通用的知识问答系统,也可将知识推理等技术直接应用于问答系统来解决特定的问题。比如用R-GCN来建模多轮对话问答系统的对话结构和背景知识^[36]。基于知识图谱嵌入的问答系统正在兴起,比如将TransE向量空间与搜索技术结合,实现了基于知识图谱嵌入的问答系统^[37]。

随着现代知识图谱推理方法的研究逐渐深入,知识问答的效果也持续改善,同时

将上述这些知识推理技术和问答技术结合用于解决特定问题的方法也在兴起。

5.3 推荐系统

推荐系统是随着互联网的发展、信息的极大丰富而兴起的技术,旨在理解用户并主动给用户推荐信息。随着知识图谱技术的发展,将知识图谱和推荐系统全方位地结合,从而更深入地理解用户,更好地匹配用户需求,同时提供更强的解释性。

在理解用户方面,知识图谱能够对用户之间的网状关系进行建模,通过购买同一件商品、看过同一个视频、对同一条信息点赞等用户行为构建行为图谱,并利用知识图谱推理技术推断用户间的关系,挖掘潜在的深度需求,配合协同过滤等推荐技术来改善推荐的效果。参考文献[38]利用知识图谱来丰富用户信息,并使用R-GCN对关系和结构建模推理,从而更好地理解用户。推荐的内容(如商品、信息、知识、人物等)同样可以通过各种关系来构建知识图谱,使用推理技术来挖掘潜在特征,推断潜在关系,增强对推荐内容的理解,改善推荐效果。参考文献[39]尝试使用TransR^[15]来建模结构化的推荐内容,以提升推荐效果。在此之上,推荐本身可以看成用户-推荐内容之间的复杂的网状关系,由此可应用知识图谱推理技术来直接实现推荐。KGAT^[40]结合TransR^[15]和深度学习实现了知识图谱注意力网络的推荐方法。

随着知识图谱和知识推理技术的进一步发展和成熟,各种现代的知识推理方法在推荐领域的应用也会越来越多,并逐渐成为主流。

5.4 个性化搜索

个性化搜索是充分利用搜索和点击等

历史行为记录和用户本身的信息来实现个性化的搜索,返回更加匹配用户的结果。比如同样搜索“苹果”,果农和电子产品爱好者对搜索结果的期望是不同的;而搜索“天气预报”,在上海的人和在北京的人对搜索结果的期望也是不一样的。个性化搜索是致力于解决这类问题的技术。推荐系统中的使用知识图谱及推理技术对用户建模的方法可以用于个性化搜索。

直接使用知识推理技术来实现个性化搜索的方法也在被尝试,参考文献[41]将用户、文档以及用户与文档产生的搜索、点击等交互关系构建成知识图谱,并使用TransE^[11]推理方法实现个性化搜索。参考文献[42]则将ConvE^[25]的改进版本ConvKB应用于个性化搜索。

6 行业应用

随着知识图谱的流行,知识推理在各行各业被广泛应用。在金融、投资、保险、地产、电力电网、能源、制造、生物医药、医疗、智慧城市^[43]等行业都能见到知识图谱推理技术的应用,本文主要介绍金融、生物医药和智能制造行业的应用情况。

6.1 金融行业

知识图谱被广泛应用于金融行业,在风险监测与控制、事件、舆情、对话机器人、金融市场事件监测等方面都有知识图谱的身影,大量的知识推理方法也被应用在具体的场景下。在风险监测与控制方面,在基于企业工商信息、诉讼信息、招聘信息以及风险事件等数据构建的大规模知识图谱中,可以利用知识推理技术挖掘隐藏的关联关系、风险传导链、担保圈链等。在金融舆情方面,可构建事件及其影响关

系的图谱,使用知识推理技术来完善图谱中的关联关系,识别事件的真伪和影响后果等。在对话机器人方面,针对规章制度、财务制度、财务知识、业务知识等建立知识图谱,并使用基于知识图谱的问答技术实现对话机器人,随时随地满足客户获取知识的需要。在投资研究方面,基于事件构建图谱,并使用知识图谱推理技术来研究对金融市场的影响^[44]。针对小微企业信贷业务的特点,构建全方位企业画像与企业关联图谱的贷前反欺诈模型,定量评估小微企业客户的欺诈风险,能够有效地帮助银行机构更准确地对企业申贷欺诈行为进行评估^[45]。

6.2 生物医药行业

生物医药行业的多个子领域深入应用了知识图谱和推理技术,包括药物图谱、疾病图谱、蛋白质图谱、基因图谱、药物分子信息图谱等,涉及生物医药行业的方方面面^[46-47]。这些图谱同样存在不完备的情况,可使用知识图谱推理技术来完善这些图谱。

生物医药的研究对象通常是复杂的微观结构以及这些微观结构间的关系,深度学习中很多与图结构相关的算法是生物医药领域的研究人员提出的^[48]。同样地,知识图谱推理技术也被广泛地直接应用于生物医药的研究,参考文献[49]利用R-GCN^[29]模型作为编码器,并使用张量分解作为解码器,使用知识图谱推理的方法来推断药物-蛋白质、蛋白质-蛋白质、药物-药物、药物-副作用等方面的相互作用。GrEDeL^[50]使用两个TransE^[11]结合LSTM的模型,从公开文献中挖掘潜在的治疗疾病药物,该模型把文献中构建的生物医药知识图谱拆分成语义图谱(semantic graph)和类型图谱(type

graph),并用两个TransE分别对这两个图谱进行建模,然后使用LSTM模型来挖掘适用于疾病的药物。

6.3 智能制造行业

智能制造行业极大地依赖于工程师和工人的知识与经验,为了适应越来越复杂的产品设计、生产制造等过程,知识图谱也被用来积淀、传承、管理和应用智能制造中理化、电子电气、工艺、失效、故障等的知识和经验,避免知识的流失,辅助设计工程师、质量工程师和制造工程师分析问题,帮助现场工人快速解决问题,以及激发创新创造等。知识图谱推理技术可以用来完善图谱数据,改善图谱质量,提升知识和经验的完备性^[51]。

智能制造通常涉及多个跨领域的知识、规范、标准等,可以使用知识图谱技术来跟踪、跟进这些规范标准,并保持更新。参考文献[52]是致力于将工业4.0的标准进行图谱化的一个尝试。中国电子技术标准化研究院也在尝试将国内各种标准图谱化。辅以知识问答技术和推荐技术,能够更加方便地推进制造业应用最新的知识、经验、规范、标准等。

进一步地,设备传感器采集的大量信息能够有效地监测制造过程。知识图谱推理技术同样能够用在这些数据所构建的设备和制造过程的图谱中,优化生产制造过程,优化保养周期、预测可能发生的故障,保证生产过程处于良好的状态^[53]。参考文献[54]利用TransE^[11]融合来自传感器的操作数据源和制造图谱的实体,为潜在的未知实体提供自动化融合的建议,从而保持数字孪生的同步,对于依赖于数据的实时对齐的监视和管理应用具有巨大价值。在更多专业的领域,知识图谱推理技术也在发挥其价值,参考文献[55]探

索了电子自动化设计(electronic design automation,EDA)中应用R-GCN的场景。在更广泛的智能制造场景中(如供应链管理、客户和供应商管理、BOM的自动创建、转换与管理、工厂员工管理等方面),知识图谱推理技术也被用来提升智能化管理水平。

7 挑战与未来展望

近些年来,现代的知识图谱推理技术正在快速发展,许多模型在各自的领域有非常好的效果,并且被广泛应用,但远未达到完美的状态,挑战和机遇并存。

- 知识图谱嵌入通常是嵌入欧几里得空间,近些年,MuRP、ATTH等模型探索了嵌入双曲空间的情形,获得了非常好的效果。但总体上,把知识图谱嵌入双曲空间的研究还是少数。而已有的模型表明双曲空间等非欧几里得空间能够更好地表达知识图谱。在非欧几里得空间进行知识图谱表示和推理是非常值得继续探索的。

- 图神经网络天然匹配知识图谱,本文介绍的R-GCN和RGHAT等模型还是早期的尝试,远未完美。设计出更加精巧的图网络结构来实现知识图谱推理会是热门且机会巨大的方向。

- 变换器(transformer)^[1,56]网络因其在自然语言处理领域强大的表达能力和高效的并行训练能力而大放异彩,并迅速地迁移到计算机视觉、图像处理和语音识别等领域,效果同样出众。相信变换器网络在知识图谱和知识图谱推理方面同样能表现出众。

- 基于预训练模型的迁移学习在自然语言处理、图像处理和计算机视觉等领域应用广泛,但在知识图谱推理方面还比较少见。探索知识图谱和知识图谱推理方面

的预训练模型是值得尝试的。

● 现代的知识图谱推理技术在数据集以及相应的竞赛、评测等方面同样存在巨大的机会,特别是中文知识图谱的数据集。

8 结束语

本文首先系统地研究了知识图谱推理的现代的方法和应用,提出了一种统一的框架来详细介绍现代主流的知识图谱推理模型,方便读者基于自身研究领域来比较其中的异同和优劣。现代的知识图谱推理方法通过几何的或深度学习的方法将知识图谱嵌入低维空间,包括欧几里得空间和双曲空间等。紧接着详细介绍了嵌入欧几里得空间的Trans*和RotatE等方法,同时介绍了能够更好地表达知识图谱结构和逻辑的双曲空间嵌入,并介绍了MuRP和ATTH等方法。在深度学习发展如火如荼的今天,基于深度学习的知识图谱推理方法不断被提出,本文使用相同的框架介绍了主流的深度卷积网络、胶囊网络、图神经网络等方法。随着知识图谱推理技术的日趋成熟,其应用也日趋广泛。随后详细介绍了知识图谱推理技术在知识图谱补全、知识问答、推荐系统和个性化搜索等方面的应用,同时还针对金融、生物医药和智能制造3个行业介绍了知识图谱及其推理技术的应用。最后梳理了知识图谱推理方面的一些挑战和值得关注的研究方向,供读者参考。

参考文献:

[1] 吴运兵, 杨帆, 赖国华, 等. 知识图谱学习和推理研究进展[J]. 小型微型计算机系统, 2016, 37(9): 2007–2013.
WU Y B, YANG F, LAI G H, et al. Research progress of knowledge graph

learning and reasoning[J]. Journal of Chinese Computer Systems, 2016, 37(9): 2007–2013.

- [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] MITCHELL T, COHEN W, HRUSCHKA E, et al. Never-ending learning[C]// Proceedings of the 29th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2015: 2302–2310.
- [4] SCHOENMACKERS S, DAVIS J, ETZIONI O, et al. Learning first-order horn clauses from web text[C]// Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. [S.l.]: Association for Computational Linguistics, 2010: 1088–1098.
- [5] KOK S, DOMINGOS P. Learning the structure of Markov logic networks[C]// Proceedings of the 22nd International Conference on Machine Learning (ICML 2005). New York: ACM Press, 2005: 441–448.
- [6] RICHARDSON M, DOMINGOS P. Markov logic networks[J]. Machine Learning, 2006, 62(1–2): 107–136.
- [7] ONDŘEJ K, JESSE D. Markov logic networks for knowledge base completion: a theoretical analysis under the MCAR assumption[C]// Proceedings of the 35th Uncertainty in Artificial Intelligence Conference. [S.l.:s.n.], 2020: 1138–1148.
- [8] CHEN X, CHEN H, ZHANG N, et al. OWL reasoning over big biomedical data[C]// Proceedings of the 2013 IEEE International Conference on Big Data. Piscataway: IEEE Press, 2013: 29–36.
- [9] LAO N, COHEN W. Relational retrieval using a combination of path-constrained random walks[J]. Machine Learning, 2010,

- 81(1): 53–67.
- [10] XIONG W H, HOANG T, WANG W Y. DeepPath: a reinforcement learning method for knowledge graph reasoning[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. [S.l.:s.n.], 2017.
- [11] BORDES A, USUNIER N, GARCÍA-DURÁN A, et al. Translating embeddings for modeling multi-relational data[C]// Proceedings of the 26th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2013: 2787–2795.
- [12] MIKOLOV T, CORRADO G, CHEN K, et al. Efficient estimation of word representations in vector space[C]// Proceedings of the International Conference on Learning Representations (ICLR 2013). [S.l.:s.n.], 2013.
- [13] WANG Z, ZHANG J W, FENG J L, et al. Knowledge graph embedding by translating on hyperplanes[J]. Proceedings of the 28th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2014: 1112–1119.
- [14] SUN Z, DENG Z H, NIE J Y, et al. RotatE: knowledge graph embedding by relational rotation in complex space[C]// Proceedings of the 7th International Conference on Learning Representations. [S.l.:s.n.], 2019.
- [15] LIN Y K, LIU Z Y, SUN M S, et al. Learning entity and relation embeddings for knowledge graph completion[C]// Proceedings of the 29th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2015: 2181–2187.
- [16] JI G, HE S, XU L, et al. Knowledge graph embedding via dynamic mapping matrix[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. [S.l.]: Association for Computational Linguistics, 2015: 687–696.
- [17] XIAO H, HUANG M L, ZHU X Y. TransG: a generative model for knowledge graph embedding[C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. [S.l.:s.n.], 2016: 2316–2325.
- [18] GRIFFITHS T L, GHAHRAMANI Z B. The Indian buffet process: an introduction and review[J]. Journal of Machine Learning Research, 2011, 12(2): 1185–1224.
- [19] SARKAR R. Low distortion delaunay embedding of trees in hyperbolic plane[C]// Proceedings of the 19th International Symposium on Graph Drawing. [S.l.:s.n.], 2011: 355–366.
- [20] OCTAVIAN G, GARY B, THOMAS H. Hyperbolic neural networks[C]// Advances in Neural Information Processing Systems. [S.l.:s.n.], 2018.
- [21] CHAMI I, YING R, RÉ C, et al. Hyperbolic graph convolutional neural networks[C]// Advances in Neural Information Processing Systems. [S.l.:s.n.], 2019: 4869–4880.
- [22] LIU Q, NICKEL M, KIELA D. Hyperbolic graph neural networks[C]// Advances in Neural Information Processing Systems. [S.l.:s.n.], 2019.
- [23] BALAEVI I, ALLEN C, HOSPEDALES T. Multi-relational Poincaré graph embeddings[C]// Advances in Neural Information Processing Systems. [S.l.:s.n.], 2019.
- [24] CHAMI I, WOLF A, JUAN D C, et al. Low-dimensional hyperbolic knowledge graph embeddings[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [S.l.:s.n.], 2020: 6901–6914.
- [25] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2D knowledge graph embeddings[C]// Proceedings of the 32nd AAAI Conference

- on Artificial Intelligence. Palo Alto: AAAI Press, 2018.
- [26] JIANG X T, WANG Q, WANG B. Adaptive convolution for multi-relational learning[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. [S.l.]: Association for Computational Linguistics, 2019: 978–987.
- [27] NGUYEN Q, VU T, NGUYEN D, et al. A capsule network-based embedding model for knowledge graph completion and search personalization[C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. [S.l.]: Association for Computational Linguistics, 2019: 2180–2189.
- [28] SABOUR S, FROSST N, HINTON G E. Dynamic routing between capsules[C]// Advances in Neural Information Processing Systems. [S.l:s.n.], 2017.
- [29] SCHLICHTKRULL M, KIPF T, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]// Proceedings of the 2018 European Semantic Web Conference. Heidelberg: Springer, 2018: 593–607.
- [30] YANG B, YIH S, 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 (ICLR 2015). [S.l:s.n.], 2015.
- [31] ZHANG Z, ZHUANG F, ZHU H, et al. Relational graph neural network with hierarchical attention for knowledge graph completion[C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 9612–9619.
- [32] 王文广, 徐永林, 贺梦洁, 等. 基于知识图谱的通用知识问答系统: 体系与方法[J]. 新一代信息技术, 2020, 3(7): 38–47.
- WANG W G, XU Y L, HE M J, et al. Knowledge graph based universal question answering system: framework and methods[J]. New Generation of Information Technology, 2020, 3(7): 38–47.
- [33] 邹艳珍, 王敏, 谢冰, 等. 基于大数据的软件项目知识图谱构造及问答方法[J]. 大数据, 2021, 7(1): 22–36.
- ZOU Y Z, WANG M, XIE B, et al. Software knowledge graph construction and Q&A technology based on big data[J]. Big Data Research, 2021, 7(1): 22–36.
- [34] SAXENA A, TRIPATHI A, TALUKDAR P. Improving multi-hop question answering over knowledge graphs using knowledge base embeddings[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. [S.l:s.n.], 2020: 4498–4507.
- [35] WANG X, ZHAO S, CHENG B, et al. HGMAN: multi-hop and multi-answer question answering based on heterogeneous knowledge graph (student abstract)[C]// Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 13953–13954.
- [36] LIU J, SUI D, LIU K, et al. Graph-based knowledge integration for question answering over dialogue[C]// Proceedings of the 28th International Conference on Computational Linguistics. [S.l:s.n.], 2020: 2425–2435.
- [37] HUANG X, ZHANG J, LI D, et al. Knowledge graph embedding based question answering[C]// Proceedings of the 12th ACM International Conference on Web Search and Data. [S.l:s.n.], 2019: 205–113.
- [38] CHEN Q, LIN J, ZHANG Y, et al. Towards knowledge-based recommender dialog system[C]// Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th

- International Joint Conference on Natural Language Processing. [S.l.]: Association for Computational Linguistics, 2019.
- [39] ZHANG F, YUAN J, LIAN D, et al. Collaborative knowledge base embedding for recommender systems[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2016: 353–362.
- [40] WANG X, HE X, CAO Y, et al. KGAT: knowledge graph attention network for recommendation[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM Press, 2019: 950–958.
- [41] VU T, NGUYEN D, JOHNSON M, et al. Search personalization with embeddings[C]// Proceedings of the 39th European Conference on Information Retrieval. Heidelberg: Springer, 2017.
- [42] NGUYEN D Q, NGUYEN T D, NGUYEN D Q, et al. A convolutional neural network-based model for knowledge base completion and its application to search personalization[J]. Semantic Web, 2018, 10(4): 1–14.
- [43] 臧根林, 王亚强, 吴庆蓉, 等. 智慧城市知识图谱模型与本体构建方法[J]. 大数据, 2020, 6(2): 96–106.
- ZANG G L, WANG Y Q, WU Q R, et al. Model and construction method of the ontology of knowledge graph of smart city[J]. Big Data Research, 2020, 6(2): 96–106.
- [44] DING X, ZHANG Y, LIU T, et al. Knowledge-driven event embedding for stock prediction[C]// Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers. [S.l.:s.n.], 2016: 2133–2142.
- [45] 金磐石, 万光明, 沈丽忠. 基于知识图谱的小微企业贷款申请反欺诈方案[J]. 大数据, 2019, 5(4): 100–112.
- JIN P S, WAN G M, SHEN L Z. Knowledge graph-based fraud detection for small and micro enterprise loans[J]. Big Data Research, 2019, 5(4): 100–112.
- [46] ZHENG S, RAO J, SONG Y, et al. PharmKG: a dedicated knowledge graph benchmark for biomedical data mining[J]. Briefings in Bioinformatics, 2020.
- [47] WISHART D, FEUNANG Y, GUO A, et al. DrugBank 5.0: a major update to the DrugBank database for 2018[J]. Nucleic Acids Research, 2018, 46(D1): 1074–1082.
- [48] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph attention networks[C]// Proceedings of the 6th International Conference on Learning Representations. [S.l.:s.n.], 2018.
- [49] MARINKA Z, MONICA A, JURE L. Modeling polypharmacy side effects with graph convolutional networks[J]. Bioinformatics, 2018, 34(13): 457–466.
- [50] SANG S, YANG Z, LIU X, et al. GrEDeL: a knowledge graph embedding based method for drug discovery from biomedical literatures[J]. IEEE Access, 2018, 7: 8404–8415.
- [51] HE L, JIANG P. Manufacturing knowledge graph: a connectivism to answer production problems query with knowledge reuse[J]. IEEE Access, 2019, 7: 101231–101244.
- [52] BADER S, GRANGEL-GONZALEZ, NANJAPPA P, et al. A knowledge graph for industry 4.0[C]// Proceedings of the 2020 European Semantic Web Conference. Heidelberg: Springer, 2020: 465–480.
- [53] GAROFALO M, PELLEGRINO M, ALTABBA A, et al. Leveraging knowledge graph embedding techniques for industry 4.0 use cases[J]. arXiv preprint, 2018, arXiv:1808.00434.
- [54] RINGSQUANDL M, LAMPARTER S, LEPRATTI R, et al. Knowledge fusion of manufacturing operations data using representation learning[C]// Proceedings of

- the 2017 IFIP International Conference on Advances in Production Management Systems. Heidelberg: Springer, 2017: 302–310.
- [55] MA Y, HE Z, LI W, et al. Understanding graphs in EDA: from shallow to deep learning[C]// Proceedings of the 2020 International Symposium on Physical Design. [S.l.:s.n.], 2020: 119–126.
- [56] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2017: 6000–6010.

作者简介



王文广 (1984–), 男, 达而观信息科技(上海)有限公司高级工程师、副总裁, 中国计算机学会会员、中国中文信息学会语言与知识计算专业委员会委员、中国人工智能学会深度学习专业委员会委员, 主要研究方向为知识图谱、自然语言处理、计算机视觉、深度学习、深度强化学习等。

收稿日期: 2021-03-02

通信作者: 王文广, wangwenguang@datagrand.com