人工智能时代知识图谱表示学习方法体系

张翔[1],杨伟杰[2,3*],刘文文[1],张珣[1,3],段大高[3,4],韩忠明[3,4]

1. 北京工商大学计算机学院,北京 100048
2. 北京工商大学人工智能学院,北京 100048
3. 食品安全大数据技术北京市重点实验室,北京 100048
4. 北京工商大学国际经管学院,北京 100048

摘要 总结了不含辅助信息的知识图谱表示学习方法,主要是基于距离和基于语义匹配2类主流方法;研究了包含文本辅助信息和类别辅助信息的知识图谱表示学习方法;通过对比各类表示学习方法的优缺点,发现引入辅助信息能有效表达知识图谱中新实体,但时空开支大幅上升,因而在现阶段,不含辅助信息的方法更易应用于实际场景中。分析了知识图谱嵌入如何应用于三元组分类、链路预测、推荐系统等下游任务,整理归纳了应用于不同任务的数据集和开源库的集合,并展望了大规模、动态知识图谱等具有广泛应用前景的研究方向。

关键词 知识图谱;知识图谱嵌入;表示学习;深度学习

在人工智能时代,知识图谱(knowledge graph,KG)的建设和应用发展迅速。从语义解析(semantic parsing)[1]和命名实体识别(named entity disambiguation)[2],到信息提取(information extraction)[3]和问题回答(question answering)[4],大量的KGs,如Freebase、DBpedia、YAGO和NELL被创建并成功应用于现实世界。知识图谱中存储着大量的知识,知识能让机器具备认知能力,而认知能力就是人工智能的基础,知识图谱能帮助人工智能更好地认识人类世界。强大的人工智能可以帮助从客观世界中挖掘、获取信息,将碎片信息关联起来最终提炼成知识,这些知识和人工智能系统互相促进,共同进步。KG是一个由实体(节点)和关系(不同类型的边)组成的多关系图。每条边被表示为三元组

收稿日期:2021-03-13;修回日期:2021-07-29

基金项目:国家重点研发计划项目(2019YFC0507800);北京市自然科学基金项目(4172016);北京市教委科研计划一般项目(KM201710011006)

作者简介:张翔,硕士研究生,研究方向为互联网数据挖掘,电子信箱:1930401027@st.btbu.edu.cn;杨伟杰(通信作者),讲师,研究方向为大数据分析与信息检索,电子信箱:yangwj@btbu.edu.cn

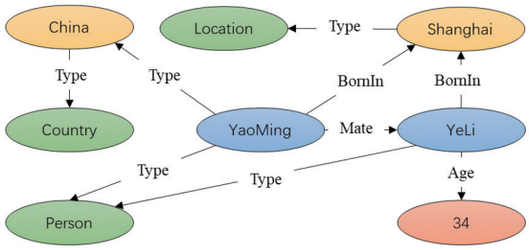
引用格式:张翔, 杨伟杰, 刘文文, 等. 人工智能时代知识图谱表示学习方法体系[J]. 科技导报, 2021, 39(22): 94-110; doi: 10.3981/j.issn.1000-7857.2021.22.011

(head entity, relation, tail entity),也被称为fact,简写为(h,r,t),表示2个实体是通过特定关系连接起来的。例如,(阿尔弗雷德·希区柯克, 导演, 惊魂记),在这个三元组中,"阿尔弗雷德·希区柯克"为头实体(head entity),"导演"为关系(relation),"惊魂记"为尾实体(tail entity)。知识图谱和知识库(knowledge base)在一定程度上是同义的。当考虑到知识图谱的图结构特性时,可以将其看作一个有向图[5];当知识图谱涉及到语义时,可以看作一个知识库用于对事实的解释和推理[6](图1)。但这种表示形式存在严重的计算效率问题,当利用知识图谱进行知识推理一类的任务时,三元组表示法往往需要设计专门的图算法来实现,而基于图的算法计算复杂度高、可扩展性差,一旦知识图谱达到一定规模,就很难满足实时计算要求。

为解决这个问题,研究者提出一个新的研究方向,称为知识图谱表示学习,也叫做知识图谱嵌入。其核心思想是将知识图谱的组件(包括实体和关系)嵌入到连续向量空间中,这些实体和关系嵌入可以进一步用于各种任务,如知识图谱完成[7]、关系提取[8]、实体分类[9-10]和实体解析[10-11]。既能简化操作,同时又可以保留知识图谱的固有结构,因而,一经提出迅速获得了巨大的关注。

(YaoMing, **Origin**, China)
(YaoMing, **Mate**, YeLi)
(YeLi, **Age**, 34)
(YaoMing, **BornIn**, Shanghai)
(YeLi, **BornIn**, Shanghai)
(YeLi, **Type**, Person)
(China, **Type**, Country)
(Shanghai, **Type**, Location)

(a) 知识库中的三元组



(b) 知识图谱中的实体与关系

图1　知识库与知识图谱示意

1　相关定义

知识图谱一词最早由Google在2012年[12]提出,初衷是为了提高搜索引擎的性能,改善用户的搜索质量和搜索体验。随着人工智能技术的发展和应用,知识图谱逐渐成为人工智能关键技术之一,现已被广泛应用于智能搜索、智能问答、个性化推荐、内容分发等领域。但是,到目前为止,知识图谱还没有标准的统一定义,现有的知识图谱定义都是通过描述一般语义表示或基本特征给出。例如,Ehrlinger等[13]分析了已有的知识图谱定义,提出了定义1,该定义主要强调了知识图谱的推理引擎,认为知识图谱可以获取信息并将信息整合到本体中,进而应用推理器来推导新的知识。而Wang等[14]则是将知识图谱定义为定义2中的多关系图,认为知识图谱是由实体和关系组成的多关系图,实体和关系在图中分别被表示为节点和不同类型的边。

为了能更形式化地界定知识图谱,将知识图谱定义为:

$$G=\{E,R,F\}$$

式中,E、R和F分别表示实体(entities)、关系(relations)和事实(facts)的集合。一个fact被表示为一个三元组$(h,r,t)\in F$。

网络表示学习起源于降维技术,其目标是通过分析原始网络数据,学习得到网络中每个节点的低维稠密实数向量表示。即,学习一个映射$f:V\rightarrow\mathbb{R}^d$,$x\in V$,$f(X)\in\mathbb{R}^d$。其中,$x$是网络中的一个节点,$f(X)$是$X$的特征表示。这些低维向量既保存了网络的拓扑结构和性质,又能更加方便地应用于网络重建和网络推断任务。其中,网络拓扑结构包括节点邻居结构,高阶节点相似性以及网络社区结构,网络性质包括网络传递性和结构平衡性。

知识图谱的表示学习是知识图谱应用的基础,其目的是构建一个模型,将知识图谱中的实体和关系映射到一个低维稠密实数向量。知识图谱的表示学习为许多知识获取任务和知识图谱的下游应用铺平了道路。表示学习得到的低维向量表示是一种分布式表示[15]。之所以如此命名,是因为孤立地看向量中的每一维,都没有明确对应的含义,只有综合各维形成一个向量,才能够表示对象的语义信息。

典型的知识图谱表示学习方法通常包括3个步骤:表示实体和关系;定义评分函数;学习实体和关系表示。表示实体和关系,是指在连续向量空间中确定实体和关系的表示形式,实体通常表示为向量,即向量空间[7,10,16-18]中的确定性点。但在一些研究中[19]也会进一步考虑实体的不确定性,并通过多变量高斯分布建立模型。关系通常被视为向量空间中的操作,可以表示为向量[7,16]、矩阵[11,17]、张量[18]、多变量高斯分布[19],甚至混合高斯分布[20]。实体和关系被表示为向量之后,需要定义评分函数$f_r(h,t)$,衡量每个三元组(h,r,t)的合理性。在知识图谱中观察到的三元组往往比未观察到的三元组得分更高。最后,为了学习这些实体和关系的表示(即嵌入),需要解决一个优化问题,即最大限度地提高观察到的三元组(h,r,t)的可信度。

2 研究现状

依据表示学习模型融合信息的情况,可以将知识图谱表示学习模型分为融合事实信息的模型和融合辅助信息的模型2大类(图2)。其中融合事实信息的模型仅利用知识图谱中的三元组信息,使用评分函数衡量事实的合理性,因而又被称为不含辅助信息的知识图谱表示学习方法;融合辅助信息的模型使用多模态的外部知识提供额外信息作为知识图谱嵌入的补充,如实体的本文描述、类型等。

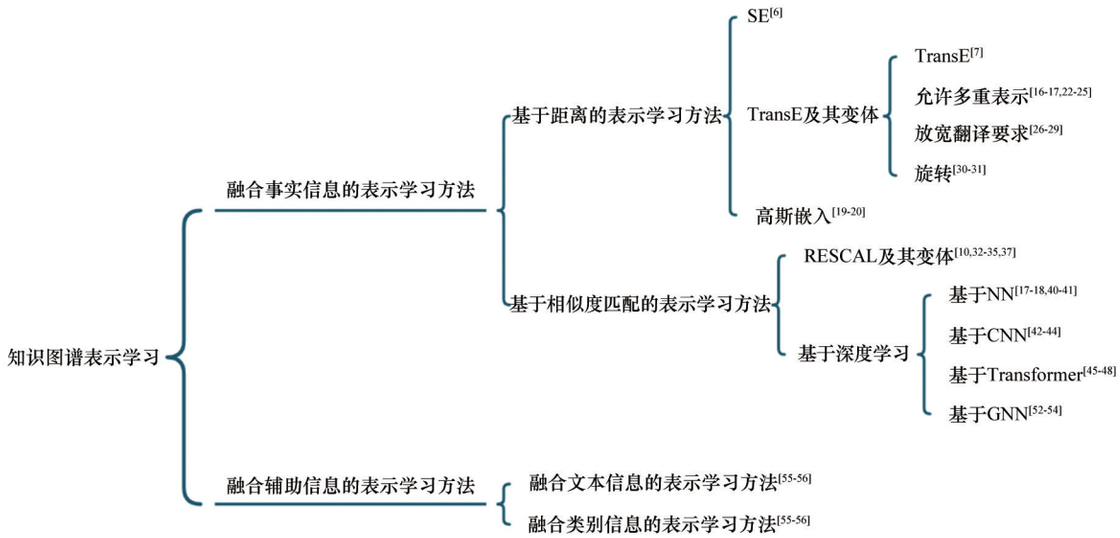


图2 知识图谱表示学习分类图示

按照评分函数,将融合事实信息的表示学习模型分为翻译距离模型和语义匹配模型。前者使用基于距离的评分函数,通过计算实体之间的距离来衡量事实的可信度,后者使用基于相似性的计分函数,通过语义匹配来衡量事实的可信度。基于距离的方法又可继续分为计算投影距离的结构嵌入(structural embedding,SE)、基于翻译思想的TransE及其变体和高斯嵌入,基于相似度匹配的则分为RESCAL及其变体和深度学习方法。融合辅助信息模型按照融合的信息类型分为融合文本信息和融合类别信息2类模型。

2.1　融合事实信息模型

2.1.1　基于距离的模型

基于距离的模型通过2个实体之间距离大小衡量三元组的真实性,其中一个直观的方法是计算实体关系投影之间的距离。SE[6]使用2个投影矩阵和L1距离(曼哈顿距离)学习结构嵌入。

另一个更常用的方法则将实体之间的关系表示为从头实体到尾实体的转化,以下分成2类介绍,一类是TransE及其变体,另一类是高斯嵌入。

1) TransE及其变体。

TransE[7]是最具代表性的转化距离模型,TransE从词向量学习方法获得启发,Mikolov[21]提出分布式词表征来捕捉语言规律性。在TransE中,将实体和关系都表示为同一空间中的向量,给定一个三元组(h,r,t),关系被解释为一个翻译向量$\boldsymbol{r}$,这样嵌入的头实体$\boldsymbol{h}$和尾实体$\boldsymbol{t}$可以通过$\boldsymbol{r}$以较低的误差连接起来,如图3所示。

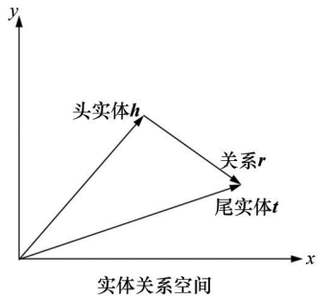


图3　TransE示意

即当(h,r,t)成立时,$\boldsymbol{h}+\boldsymbol{r}\approx\boldsymbol{t}$。得分函数设置如下:

$$f_r(\boldsymbol{h},\boldsymbol{t})=\left\|\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{t}\right\|_{1/2}$$

TransE的优点在于简单高效,但它的缺点也比较明显,在处理1对N、N对1和N对N关系时存在缺陷[16-17]。以1对N关系为例,由于它对所有(h,r,t_i)都执行规则$\boldsymbol{h}+\boldsymbol{r}\approx\boldsymbol{t}_i$,这就导致最后学到的$\boldsymbol{t}$向量会非常相似。例如,对于"导演"关系,尽管"惊魂记"、"蝴蝶梦"和"后窗"是完全不同的实体,TransE学习得出的向量表示会非常相似,因为他们都是"阿尔弗雷德·希区柯克"导演的电影。

为了解决这个问题,TransH[16]在TransE的基础上引入特定于关系超平面,TransH将实体建模为向量$\boldsymbol{h}$和$\boldsymbol{t}$,关系r表示为以$\boldsymbol{w}_r$为法向量的超平面上的向量$\boldsymbol{r}$。具体操作为:给定一个三元组(h,r,t),先将实体$\boldsymbol{h}$和$\boldsymbol{t}$投影到超平面上$\boldsymbol{h}_\perp=\boldsymbol{h}-\boldsymbol{w}_r^{\top}\boldsymbol{h}\boldsymbol{w}_r$,$\boldsymbol{t}_\perp=\boldsymbol{t}-\boldsymbol{w}_r^{\top}\boldsymbol{h}\boldsymbol{w}_r$,假设三元组存在,则在超平面上由$\boldsymbol{r}$以低误差连接起来。

TransR[17]与TransH有着很相似的想法,但它引入了特定于关系的空间,而不是超平面。在TransR中,实体表示为实体空间$\mathbb{R}^d$中的向量,而关系则是建模在关系空间$\mathbb{R}^k$中的平移向量。对于给定的三元组(h,r,t),TransR会先将实体表示$\boldsymbol{h}$和$\boldsymbol{t}$投影到特定于关系$\boldsymbol{r}$的空间中。这种方法为每种关系引入了一个投影矩阵,这使它失去了TransE和TransH的简单性和高效率。在后续研究中,有学者提出更为复杂的方法[22-23],每个关系都与2个矩阵相关联,一个矩阵与头实体相关,一个与尾实体相关。这些都使得TransR方法在效率上远不如原本的TransE方法。

TransD[24]则通过将投影矩阵分解成2个向量的乘积来简化TransR。对每个三元组,TransD引入了额外的映射向量$\boldsymbol{w}_h$、$\boldsymbol{w}_t$和$\boldsymbol{w}_r$,这样投影矩阵就能够简化为向量乘积。由于参数数量下降,TransD要比TransR效率更高,速度更快。另一个优化TransR的方向是增强投影矩阵稀疏性来简化TransR,Ji[25]提出2种版本,TranSparce(share)和TranSparce(separate)。Share版本对每个关系$\boldsymbol{r}$使用同一个稀疏投影矩阵,Separate版本中使用2个单独的投影矩阵,一个用于投影$\boldsymbol{h}$实体,一个用于投影$\boldsymbol{t}$实体。

另一个优化TransE的方向是放宽对$\boldsymbol{h}+\boldsymbol{r}\approx\boldsymbol{t}$过分严格的要求来进行改进。TransM[26]为每个三元组分配特定于关系r的权重θ_r,通过把1对N、N对1、N对N关系的权重调低,放宽这些关系中对于$\boldsymbol{h}+\boldsymbol{r}$和$\boldsymbol{t}$之间距离的限制。ManifoldE[27]将$\boldsymbol{h}+\boldsymbol{r}\approx\boldsymbol{t}$放宽到$\left\|\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{t}\right\|_2^2\approx\theta^2$,这样$t$可以近似位于一个超球体上。TransF[28]采用同样的思想,不严格要求$\boldsymbol{h}+\boldsymbol{r}=\boldsymbol{t}$,而只要求$\boldsymbol{h}$在$\boldsymbol{t}-\boldsymbol{r}$的同一方向上,同时$\boldsymbol{t}$在$\boldsymbol{h}+\boldsymbol{r}$的同一方向上。而在进行评分时将$\boldsymbol{t}$与$\boldsymbol{h}+\boldsymbol{r}$匹配,同时将$\boldsymbol{h}$与$\boldsymbol{t}-\boldsymbol{r}$匹配。TransA[29]引入非负对称矩阵

$\boldsymbol{M}_r$,评分函数变为$f_r(\boldsymbol{h},\boldsymbol{t}) = -(|\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{t}|)^{\mathrm{T}} M_r(|\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{t}|)$。

近年,有研究者将Trans系列模型中的"转化"视为从头实体到尾实体的旋转,代表性的工作有RotatE[30]和HAKE[31]。RotatE模型能建模和推断包括对称、反对称、反转和组合在内的多种关系模式;HAKE将实体映射到极坐标中,以极坐标中的同心圆模拟知识图谱中包含的语义层次结构。

2)高斯嵌入。

上述方法都是把实体和关系建模为向量空间中的确定点,而已有研究也开始考虑不确定性,将他们建模为随机变量[19-20]。KG2E[19]将实体和关系表示为从多维高斯分布中抽取的随机向量,受翻译思想假设$\boldsymbol{h}+\boldsymbol{r}\approx\boldsymbol{t}$的启发,KG2E通过测量$\boldsymbol{t}-\boldsymbol{h}$和$\boldsymbol{r}$(即分布$N(\mu_t-\mu_h,\sigma_t+\sigma_h)$和$N(\mu_r,\sigma_r)$)之间的距离作为打分函数。利用高斯嵌入,KG2E可以有效模拟知识图谱中实体和关系的不确定性。同样也是利用高斯分布建模实体和关系,但是TransG[20]认为一个关系可以有多个语义,因此在TransG中用混合高斯来表示关系$\boldsymbol{r} = \sum_i \pi_r^i \mu_r^i, \mu_r^i \sim N\left[\mu_t-\mu_h,(\sigma_h^2+\sigma_t^2)I\right]$。

2.1.2 基于语义的模型

语义匹配模型利用基于相似度的评分函数,通过匹配实体与关系之间的隐藏语义和他们在向量空间中的表示来衡量三元组的合理性。现有的语义匹配模型可分为两类,一类是RESCAL及其变体,另一类是深度学习方法。

1) RESCAL及其变体。

RESCAL[10](也被称为双线性模型[32])中实体表示为向量,每种关系表示为一个矩阵,用来代表实体之间所有潜在成分的相互作用,具体结构如图4所示。

Jenatton等[32]在RESCAL的基础上进一步假设所有的矩阵在对角矩阵上进行矩阵分解,来减少参数数量。TATEC[33]不仅将三元组建模为$\boldsymbol{h}^{\mathrm{T}}\boldsymbol{M}_r\boldsymbol{t}$,而且也建模了双向交互,即1个实体与1个关系之间的交互。

Yang等[34]提出的DistMult通过将$\boldsymbol{M}_r$限制为对角矩阵来简化RESCAL[10],对于每个关系,它引入了

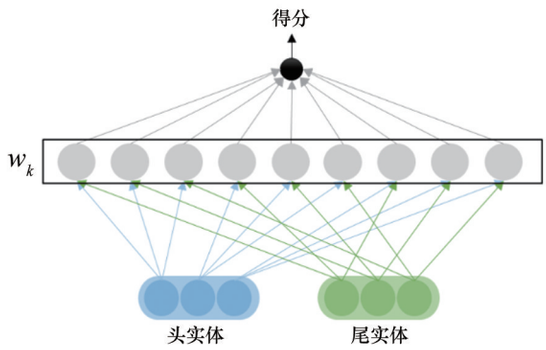


图4 RESCAL简单图示

一个向量嵌入$\boldsymbol{r}$,使$\boldsymbol{M}_r$=diag($\boldsymbol{r}$),因此只需要比较嵌入同一维的交互作用,每个关系的参数量就能降到d。但这种方法只能处理对称关系,不足以对一般的知识图谱进行建模。

Nickel[35]提出的HolE结合了RESCAL的表现力和DistMult的效率和简单性。它把实体和关系都表示为向量,给定一个三元组(h,r,t),首先用循环相关操作[36]将实体表示组成$\boldsymbol{h}\star\boldsymbol{t}\in\mathbb{R}^d$,然后将合成向量$\boldsymbol{h}\star\boldsymbol{t}$与关系表示进行匹配作为评分函数。循环相关操作可以对成对相互作用进行压缩,所以相较于RESCAL,HolE参数更少更有效率,同时能保留模拟非对称关系的能力。

Trouillon[37]提出的ComplEx引入了复值嵌入,能更好地对非对称关系进行建模。这样,实体和关系不再建模在实空间中,而是建模在复数空间$\mathbb{C}^d$中,评分函数定义为$f_r t(\boldsymbol{h},\boldsymbol{t}) = \mathrm{Re}\left(\boldsymbol{h}^{\mathrm{T}}\mathrm{diag}(\boldsymbol{r})\boldsymbol{t}\right) = \mathrm{Re}\left(\sum_{i=0}^{d-1}[\boldsymbol{r}]_i\cdot[\boldsymbol{h}]_i\cdot[\boldsymbol{t}]_i\right)$,因此评分函数也不再对称,来自非对称关系的三元组可以根据所涉及实体的顺序得到不同的分数。而Hayashi[38]的研究中表明HolE被ComplEx归为将结合对称性强加于嵌入的特殊情况。

而Liu等[39]提出的ANALOGY则扩展了RESCAL,以便进一步模拟实体和关系的类比性质。如"阿尔弗雷德·希区柯克"对于"惊魂记"的含义可以等同于"詹姆斯·卡梅隆"对于"阿凡达"的含义。虽然ANALOGY把关系表示为矩阵,但它要求可以分块对角化为一组稀疏的对角矩阵,大大减少了参数

的数量。之前介绍的DistMult、HolE和ComplEx都属于ANALOGY的特殊情况。

2）深度学习方法。

SME[11]使用神经网络架构进行语义匹配,结构如图5所示。给定一个三元组,它首先在输入层中投射实体及其关系的向量嵌入。然后在隐藏层将关系$\boldsymbol{r}$与头实体$\boldsymbol{h}$组合得到$g_u(\boldsymbol{h},\boldsymbol{r})$,与尾实体$\boldsymbol{t}$组合得到$g_v(\boldsymbol{t},\boldsymbol{r})$,它的评分函数通过点积匹配$g_u(\boldsymbol{h},\boldsymbol{r})$和$g_v(\boldsymbol{t},\boldsymbol{r})$:

$$f_r(\boldsymbol{h},\boldsymbol{t}) = g_u(\boldsymbol{h},\boldsymbol{r})^{\mathrm{T}} g_v(\boldsymbol{t},\boldsymbol{r})$$

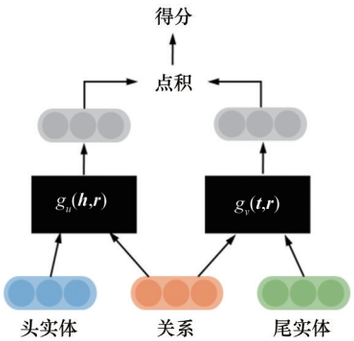


图5　SME示意图

神经张量网络(NTN)[18]是另一种神经网络结构。给定一个三元组,首先将实体投影成输入层的向量嵌入,然后通过特定于关系的向量$\boldsymbol{M}_r$组合实体$\boldsymbol{h}$和$\boldsymbol{t}$,映射到非线性隐藏层中,最后由特定于关系的线性输出层给出分数。NTN模型学习到的表示很准确,但是每个关系都需要设置d^2k个参数,若用来学习大规模的知识图谱,耗时太久,效率相对较低。而多层感知机MLP[40]是一种更简单的方法,每个实体和关系都表示为1个单独的向量,头尾实体$\boldsymbol{h},\boldsymbol{t}$与关系$\boldsymbol{r}$在输入层串联,映射到1个非线性层,最后由线性的输出层生成分数。参数$\boldsymbol{M}^1$、$\boldsymbol{M}^2$、$\boldsymbol{M}^3$和$\boldsymbol{w}^{\mathrm{T}}$都是所有关系共享的。神经关联模型NAM[41]使用"deep"架构进行语义匹配,给定一个三元组,它先在输入层将头实体和关系的嵌入$\boldsymbol{h},\boldsymbol{r}$串联,然后输入到一个由$L$层线性隐藏层组成的深层神经网络中。在前向处理后,将最后一层的输出和尾实体嵌入$\boldsymbol{t}$匹配给出分数。

自卷积神经网络(CNN)提出后,知识图谱的表示学习中开始引入卷积神经网络的方法,部分学者[42-44]开始使用CNN来学习知识图谱中的深层表示特征。CNN在图像中具有很强的特征提取的能力,为了能够在知识图谱表示学习中使用CNN,ConvE[42]模型先使用2D卷积将头实体和关系重塑为2维矩阵,然后建模实体和关系之间的交互,之后就可以通过多层非线性层来学习语义信息,具体结构如图6所示。

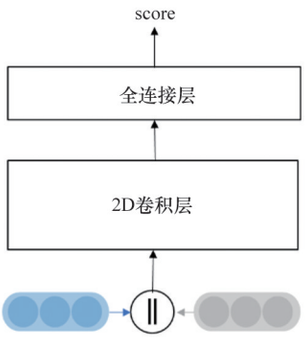


图6　ConvE的神经网络架构

ConvE注重于捕捉局部信息,为了优化这一点,提出了ConvKB[43],采用CNNs对实体和关系的连接进行编码,不需要进行ConvE中的重构。在ConvKB中,每个三元组都表示为一个3列矩阵,然后将这个3列矩阵送入1个卷积层,在该层中,对矩阵进行多个滤波器运算,以生成不同的特征映射。然后将这些特征映射连接到表示输入三元组的单个特征向量。评分函数为$f_r(h,t)$=concat(tanh([$\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}$]×$\boldsymbol{w}$))·$\boldsymbol{W}$。卷积生成的特征映射集合的串联提高了潜在特征的学习能力。与ConvE相比,ConvKB能更好地保留过渡特性。HypER[44]利用超网络H进行1D特定关系卷积滤波生成,实现多任务知识共享,简化了二维ConvE。其中,滤波器的大小限制了嵌入的哪些维度可以相互作用,同时通过使用较少的参数限制过拟合。

在最新的研究中,研究者开始尝试将基于转换器的模型引入知识图谱表示学习中,来促进知识图谱中的上下文语境信息的获取[45-48],提升表示学习的效果。之前提到的方法,都是对整个KG学习到的每个实体/关系分配了1个静态嵌入表示,

CoKE[45]借鉴了最近学习上下文单词嵌入的技术[49],对于给定的上下文输入,采用转换器模块对输入进行编码,并获得其中实体和关系的上下文表示。将边或路径作为1个输入序列,用1个特殊的标记[MASK]替换1个实体。然后将替换后的内容输入到转换器编码器块的堆栈中。与[MASK]标记相对应的最终隐藏状态被用来预测目标实体,使用交叉熵作为loss函数。

KG-BERT[46]借鉴了NLP中语言预训练的思想,将BERT中的双向编码器表示作为实体和关系的编码器。相比CoKE,KG-BERT的预训练任务共2个,分别为完型填空任务(masked language modeling)和下文预测任务(next sentence prediction)。完型填空任务用于三元组分类,而下文预测任务用来预测关系,即用2个实体h和t的嵌入预测它们之间的关系r。KEPLER[47]将预训练PLM[50]和传统的知识图谱嵌入模型结合,不仅能更好地将事实知识融入到PLMs中,而且能通过文本中丰富的信息有效学习知识图谱的表示。JAKET[48]从联合训练角度入手,关注如何同时对知识图谱和语言模型进行预训练。使用RoBERTa作为语言模型对文本进行编码,使用GAT模型来对知识图谱进行编码;由于文本和知识图谱的交集在于其中共有的若干实体,文中采用一种交替训练的方式来帮助融合两部分的知识。

知识图谱中还隐藏着丰富的拓扑结构信息,为了学习到这些结构信息,部分研究者开始将GNN[51]引入知识图谱表示学习中。Schlichtkrull[52]提出一个假设,即在进行知识图谱的补全任务时许多缺失的信息都可能存在于图的邻居编码中。根据这个假设,将GCN框架应用于知识图谱建模,提出了R-GCN。引入特定于关系的变换,即取决于边的类型和方向,收集来自相邻节点的信息,分别针对每种关系类型进行转换。同时为了防止对稀有关系的过拟合和模型过大,R-GCN引入两种单独的方法规范R-GCN的权重,基础分解法可看作是不同关系类型之间共享权重,对角块分解法是关系类型对权重矩阵的稀疏约束。这2种方法都减少了学习高度多元数据需要的参数量,能有效缓解过拟合现象。SACN[53]在ConvE的基础上引入了GCN,它由加权卷积网络WGCN编码器和CovE-TransE解码器组成。它定义了具有相同关系类型的2个相邻节点的强度,利用节点结构、节点属性和关系类型来捕捉知识图谱中的结构信息。解码器采用ConvE模型作为语义匹配度量,并保留了平移特性。CompGCN[54]利用了知识图嵌入技术中的实体-关系组合操作,CompGCN学习一个d维的关系特征嵌入和1个节点特征嵌入,将关系特征表示为向量,缓解了在关系图上应用GCNs时的过参数化问题。为了将关系嵌入合并到GCN公式中,CompGCN利用了在知识图嵌入方法中使用的实体-关系组合操作,$\boldsymbol{t}=\phi(\boldsymbol{h},\boldsymbol{t})$。组合操作的选择是决定嵌入学习质量的重要因素,Vashishth[54]提出可以采用未来开发的更好的知识图谱组合操作来进一步提高CompGCN的性能。

2.2 使用辅助信息的嵌入模型

知识图谱不同于普通的图结构,其富含结构信息以外的信息,如实体的文本描述,关系的语法含义,甚至在一些特殊的知识图谱中还会存在图片和视频。为了有效地利用这些额外信息,Xie、Xiao、Guo[55-59]等提出引入辅助信息的知识图谱嵌入模型。按照引入的信息类型可以分为两类,一类为引入文本信息的模型,一类为引入类别信息的模型。这些额外融入的信息能有效提升知识图谱表示的质量,改善数据稀疏问题。

1)引入文本信息的模型。

引入文本信息的知识图谱表示学习模型,旨在引入各知识库提供的文本描述,用以补充实体的语义。试图用文本信息表示知识图谱的"文本感知"嵌入通常可以追溯到NTN[18],NTN将实体名称的词嵌入作为实体的嵌入。DKLR[55]模型提供了两种文本表示的方法,其中CBOW将文本的词向量相加作为实体的文本表示,另一种使用卷积网络获取文本中的词序信息。DKLR能有效将文本与实体嵌入融合,但无法描述文本与三元组之间存在的强关联,语义空间投影模型SSP[56]提出,通过在语义子空间中执行嵌入过程,对三元组和文本描述之间的强相关性进行建模,建立2个信息源之间的交互,可

以为实体提供精确的语义嵌入。

2）引入类别信息的模型。

知识图谱中的实体存在类别和层级的语义，如“惊魂记”可划分为“电影”类别中的“惊悚片”，实体间的关系也表示出不同的语义类型。SSE[57]提出将同一类别的实体嵌入到同一语义空间中；TKRL[58]认为实体应在不同的类型上有不同的多种表示，使用2种类型的编码器对分层结构进行建模；KR-EAR[59]则将关系类型分为属性和关系，对实体描述之间的相关性进行建模。

2.3　对比分析

对目前主流的融合事实信息与融合辅助信息的知识图谱表示学习方法进行系统性的对比结果见表1。

表1　知识图谱表示学习方法对比

类别		模型	优点	缺点	参考文献
融合事实信息的模型	TransE及其扩展	TransE	简单，可解释性强	过于简化损失度量，表示关系的多样性不够灵活	[7]
		TransH			[16]
		TransR			[17,22-23]
		TransD			[24]
		TransSparce			[25]
		TransM	体现多样性和复杂性，灵活	建模未确定性点，不能体现实体和关系的不确定性	[26]
		TransA			[29]
		TransF			[28]
		ManifoldE			[27]
		RotatE	采用旋转思想，可建模多种关系		[30]
		HAKE			[31]
	Gaussia	KG2E	采用高斯分布可以表达语义的不确定性	参数更多，效率较低	[19]
		TransG			[20]
	RESCAL及其扩展	RESCAL	性能高，可扩展		[10,32]
		TATEC			[33]
		DistMult	参数少，效率高	无法模拟非对称关系	[34]
		HolE	保留非对称关系	训练时间长	[35]
		ComplEx	可扩展，适用情况广		[37]
		ANALOGY	参数少，适用于二元关系	非二元关系的提升不明显	[39]
	Neural Network	SME	可以学习隐藏语义信息，得到更丰富的表示	参数较多，训练时耗时长，内存消耗大	[11]
		NTN			[18]
		MLP			[40]
		NAM			[41]
	Convolutional Neural Network	ConvE	强有效学习网络局部特征	捕获不到长期依赖关系	[42]
		ConvKB			[43]
		HypER			[44]
	Transformer	CoKE	有效利用语义信息提升效果	预训练模型耗时长	[45]
		KG-BERT			[46]
		KEPLER			[47]
		JAKET			[48]
	Graph Neural Network	R-GCN	有效利用图的拓扑信息	忽视知识图谱大量文本信息	[52]
		SACN			[53]
		CompGCN			[54]
融合辅助信息的模型	融合文本信息	DKRL	引入外界信息增强知识图谱嵌入	需要在空间内引入非结构化信息，时间空间复杂度大幅上升	[55]
		SSP			[56]
	融合类别信息	SSE	引入外界信息增强知识图谱嵌入		[57]
		TKRL			[58]
		KR-EAR			[59]

不透明，缺乏可解释性

融合事实信息的模型中,基于翻译距离的方法,尤其是开创性的TransE[7],借鉴了分布式词表征学习的思想,并启发了许多后续方法,如指定复杂关系(1对N、N对1和N对N)的TransH[14]和TransR[17]。在语义匹配方面,许多方法利用数学运算或组成运算,包括SME[11]中的线性匹配、DistMult[34]中的双线性映射、NTN[18]中的张量乘积、HolE[35]和ANALOGY[39]中的循环相关等。而近几年出现的模型一般可分为双线性和神经网络2个系列。双线性模型主要基于乘法运算,参数比基于神经网络的方法少,但同时造成了性能上的局限性,如典型的双线性模型DistMult[34]只能建立对称关系的模型,其扩展的ComplEx[37]虽设法保留了反对称关系,但涉及大量的冗余计算。基于神经网络的编码模型从实体和关系的分布式表示出发,利用复杂的神经结构,如张量网络[18]、卷积网络[42-44]和变换器[45-48]学习更丰富的表示方法。这些深度模型取得了非常有竞争力的结果,但它们缺乏可解释性。

融合辅助信息的模型引入外界信息增强图谱嵌入,能有效提升表示学习的性能,对于新加入的实体,可以通过各种知识图谱外的信息生成实体的表示。但如果引入非结构化信息数据,模型需要将外部知识与内部实体进行匹配对齐,时间空间上的支出比仅融合事实信息的模型更大。

3 下游任务和评价指标

知识图谱的下游任务主要分为知识图谱内应用和知识图谱外应用。

3.1 知识图谱内应用

知识图谱内应用是指在知识图谱范围内进行的实体和关系嵌入的应用,有4种应用,即链路预测、三元组分类、实体分类和实体解析。

1)链路预测。

链路预测通常指的是预判一个实体与另一个给定实体有特定关系的任务,即给定(r,t)预测h或给定(h,r)预测t,前者表示为$(?,r,t)$,后者表示为$(h,r,?)$,例如,(?, 导演, 惊魂记)是预测电影"惊魂记"的导演,而(阿尔弗雷德·希区柯克, 导演, ?)则为预测"阿尔弗雷德·希区柯克"导演的电影。本质上是一个知识图谱完成任务,即把缺失的知识添加到图中。这种链路预测任务有时也被称为实体预测或实体排名。类似的思想也可以用来预测2个给定实体之间的关系,即,$(h,?,t)$,通常被称为关系预测[58-59]。

在评价时,通常的做法是将正确答案的等级记录在有序列表中,以便观察是否可以将正确答案排在不正确答案之前。在(?, 导演, 惊魂记)的例子中,正确答案"阿尔弗雷德·希区柯克"的排名越靠前表示表现越好。基于这样的排名,人们设计了各种评价指标,例如,MR(mean rank,排名的平均值)、MRR(mean reciprocal rank,平均倒数排名)、Hits@n(排名不大于n的比例),均为通用的排序任务的评价机制,其数学形式为:

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

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

$$Hits@n = \frac{Numberofrank < n}{Num}$$

式中,$rank$代表测试集中三元组正确答案得分的排名,Num为测试集三元组数量。

2)三元组分类。

三元组分类主要是指判断一个未出现过的三元组(h,t,r)是否为真。例如,(阿尔弗雷德·希区柯克, 导演, 惊魂记)应归为真事实,而(詹姆斯·卡梅隆, 导演, 惊魂记)为假事实。这个任务也可以看作是对输入KG的某种完成,在文献[16-18]中都有将设计的模型用于这个任务。

评估三元组分类任务可以使用传统的分类指标,如准确度指标。准确度的计算公式为$P = \frac{TP}{TP + FP}$,式中分母为所有测试集数量,分子为预测正确的数量;由于每个三元组的分类器,在输出二进制标签[0,1]的同时,还将输出一个分数,因此三元组分类任务也可以使用排名指标进行评估,如MR和MRR。

3)实体分类。

实体分类的目的是将实体分为不同的语义类

别,例如,“阿尔弗雷德·希区柯克”是一个人,而“惊魂记”是一个作品。在大多数情况下,编码实体类型的关系——“是一个”(IsA),包含在知识图谱中,并且可能在嵌入过程中被学习,实体分类可以被视为特定的链路预测任务,即(x,IsA, ?)。

评估方法可以直接参照链路预测的评估方法。

4) 实体解析。

实体解析包括验证2个实体是否指向同一个对象。在某些知识图谱中,许多节点实际上表示同一个对象,例如,在Cora数据集中,它包含了作者、标题和地点等字段的引文,作者的名字或地点可以用不同的方式书写,但本质上指的都是同一个对象。实体解析相当于是去重任务[10-11]。

Bordes等[11]考虑知识图谱中包含一个能说明2个实体是否指向一个对象的关系——“等价于”(EqualTo),并且在学习过程中能学到。在这种情况下,实体解析退化为三元组分类问题,即判断三元组(x,EqualTo,y)是否存在或有多大的可能性存在。这样就能使用三元组分类中的评价指标。但是,这种直观的策略并不总是有效,因为并非所有KG都编码“等价于”关系。Nickel等[10]提出了在实体表示的基础上执行实体解析任务。更具体地说,给定2个实体x,y和它们的向量表示$\boldsymbol{x}$,$\boldsymbol{y}$,计算$\boldsymbol{x}$,$\boldsymbol{y}$之间的相似度,用这个相似度分数来判断x,y是否为同一个对象。这时更常用的评价指标为ROC-AUC。ROC-AUC是ROC曲线下的面积,ROC曲线是通过在[0, 1]范围内选取阈值来计算对应的*TPR*和*FPR*,最终将所有点连起来构成的。其中*TPR*为所有正例中被预测为正例的比例,*FPR*为所有负例中被错误地预测为正例的比例,ROC-AUC指标能够有效衡量模型识别出唯一正确答案并对其排序的能力。

3.2 知识图谱外应用

知识图谱外应用是突破知识图谱边界并扩展到更广泛领域的应用,如关系提取、问题解答和推荐系统。

关系提取旨在从已检测到实体的纯文本中提取关系事实。例如,给定一句话“阿尔弗雷德·希区柯克导演了惊魂记”与实体h=“阿尔弗雷德·希区柯克”和t=“惊魂记”到关系提取器应提取出这2个实体之间的关系“导演”。可将TransE[7]与基于文本的提取器结合,来提升关系提取的效果。

知识图谱中的问题解答是给定一个自然语言表达的问题,任务是从一个知识图谱中检索出由一个三元组或一组三元组支持的正确答案[4]。Bordes等[4]为这项任务引入了一个基于嵌入的框架。他们方法的关键思想是学习单词和知识图谱成分的低维矢量嵌入,以便问题及其相应答案在嵌入空间中彼此接近。具体来说,将问题表示为$\boldsymbol{q}$,答案表示为$\boldsymbol{a}$,函数$S(\boldsymbol{q},\boldsymbol{a})$来评价问题与答案之间的相似性。

推荐系统向用户提供有关他们可能希望购买或查看物品的建议。由于用户与项目的交互可能非常稀疏,普通的推荐系统中采用的协同过滤技术并不总是能获得很好的效果,于是研究者开始探索利用知识图谱中的异构信息提高协同过滤的质量。

4 数据集

在知识图谱表示学习研究中,真实网络数据资源尤为重要。本文将常用的数据集分为通识数据、特定领域数据、特定任务数据3类进行介绍。

4.1 通识数据集

WordNet于1995年首次发布,是一个包含约11.7万个语法集的词汇数据库。DBpedia是一个从维基百科中提取的社区数据集,包含1.03亿个关系三元组,与其他开放的数据集互连后可以继续扩大。为了解决单源实体知识覆盖率低、质量不高的问题,YAGO利用维基百科分类页中的概念信息和WordNet中的概念层次信息,建立了一个高覆盖率和高质量的多源数据集。此外,它还可以通过其他知识源进行扩展,提供的实体超过1000万个。Freebase是一个2008年出现的可扩展的知识库,用于存储世界上的知识,截至目前它共拥有19亿个关系三元组。NELL是一个通过Never-Ending Language Learner智能代理在网络上建立起来的数据集,到目前为止有2810379个实体,置信度很高。Wikidata是一个为方便管理维基百科数据而人工编辑创建和维护的免费结构化知识库,它是多语言

的,包含358种不同的语言。具体统计数量如表2所示。

上述数据集是公开发表的,并且都是由社区或研究机构进行维护。还有一些商业数据集。Cyc知识库来自Cycor,包含约150万个概念和超200万个关系三元组。OpenCyc在2017年时被废弃,不再提供。谷歌知识图谱拥有超过5亿个实体和35亿个关系三元组。微软建立的Probase IsA知识图谱包含超1250万实体与超8500万IsA关系。

4.2 特定领域数据集

一些特定领域的知识图谱数据是为了评估特定领域的任务而设计和收集的。一些常见的领域包括生命科学、卫生保健和科学研究。还有一些数据集涵盖复杂的领域和关系,如化合物、疾病和组织。特定领域知识图谱的例子有中文医学知识图谱CMKG、统一的医学语言系统知识图谱UMLS和商业化临床术语知识图谱SNOMED CT,具体统计如表3所示。

表2 通识数据集统计

数据集	实体数	三元组数	网址
WordNet	117597	207016	https://wordnet.princeton.edu
OpenCyc	47000	30600	https://www.cyc.com/opencyc/
Cyc	250000	2200000	https://www.cyc.com
YAGO	1056638	5000000	http://www.mpii.mpg.de/~suchanek/yago
DBpedia	1950000	103000000	https://wiki.dbpedia.org/develop/datasets
Freebase	—	125000000	https://developers.google.com/freebase/
NELL	2810379	242453	http://rtw.ml.cmu.edu/rtw/ Wikidata
Wikidata	14449300	30264656	https://www.wikidata.org/wiki
Probase IsA	12501527	85101174	https://concept.research.microsoft.com/Home/Download
Google	>500 million	>3.5 billion	https://developers.google.com/knowledge-graph

表3 特定领域数据集统计表

数据集	实体数	三元组数	网址
CMKG	30000	300000	https://github.com/liuhuanyong/QASystemOnMedicalKG
UMLS	2000000	12000000	https://uts.nlm.nih.gov/uts/
SNOMED CT	321900	7000000	https://www.snomed.org/

CMKG由中国科学院软件研究所建设完成,内容包括寻医问药网上规范的半结构化医学知识,包括8807种疾病、3828种药物、5998种症状、3300种检查方法等7类实体、10种疾病相关属性作为关系。UMLS是美国国立医学图书馆于1986年开始建设的一体化医学知识语言,具有集成性、跨领域和工具化特点,包含超级叙词表、语义网络和专家词典。SNOMED CT是一部医学术语集,涵盖大多数方面的临床信息,如疾病、所见、操作、微生物、药物等,目前包括约321900个实体和超700万条关系。

4.3 特定任务数据集

产生特定任务的数据集的一种常用方式是从大型通用数据集中抽取子集,如表4所示。

FB前缀数据集是关系数据库Freebase的子集,WN前缀代表WordNet的子集,表4中WN18和FB15k存在测试集泄露问题,即存在一些逆向链接。

5 未来研究趋势

当前人工智能的发展仍然处于弱人工智能的阶段,只能在某些方面辅助人类工作而不能完全替代人类工作。2018年以来,美国人工智能协会收录关于认知智能层面的论文逐年增多,认知智能的发展主要依靠外部知识和逻辑推理。而知识图谱

表4　特定任务数据集统计

数据集	关系数	实体数	训练集	验证集	测试集
WN18	18	40943	141442	5000	5000
FB15K	1345	14951	483142	50000	59071
WN11	11	38696	112581	2609	10544
FB13	13	750463	316232	5908	23733
WN18RR	11	40943	86835	3034	3134
FB15k-237	237	14541	272115	17535	20466
FB5M	1192	5385322	19193556	50000	59071
FB40K	1336	39528	370648	67946	96678

是一种用图模型来描述知识和建模世界万物之间关联关系的大规模语义网络,能够有效帮助机器实现理解、解释和推理的能力,是认知智能的底层支撑。知识图谱的表示学习能有效降低各类图计算的时间,协助人工智能对知识图谱中蕴含的大量信息进行有效利用。因此,知识图谱表示学习近年来获得了愈来愈多国内外研究者的关注,也涌现了很多有价值的解决方案,但仍有一些问题需要完善。

1) 结合领域知识的知识图谱。

现在的知识图谱表示学习大多面向的是通识数据集,少有专门为某个特定领域设计的表示学习方法。特定领域的知识图谱数据来源于特定行业的语料,基于行业数据构建且具有一定的行业深度,旨在解决行业人员的问题,使用者也是这个行业内的从业人员或是这个领域的专业人员。这就使专业领域的知识图谱应用往往涉及决策分析支持,在进行表示学习时需要提供较强的可解释性。

如金融领域,其基础设施好、数据标准化程度高且信息化成熟,在信贷领域、理财领域、保险领域等应用场景下知识图谱都有广泛的应用前景。在金融知识图谱中,某个行业的任何变化可能起源于供需端的一系列因素,甚至包括气候变化、地缘政治等,且金融知识具有高度动态性,需要关注知识的时效性。在医疗领域,随着人工智能的飞速发展,精准医疗、智慧医疗的提出,医学知识图谱应用关注度日益上升。知识图谱在医疗决策支持系统和医疗问答系统中占据核心地位,问答模型更是建立在准确快速的知识图谱表示学习方法上,如何在进行特定领域知识图谱表示学习时结合领域特点,提升表示学习方法的性能,可能成为知识图谱表示学习的重要研究方向。

2) 大规模知识图谱的表示学习。

在处理大规模知识图谱时,需要模型在计算效率和模型表现力之间进行权衡,已知的能在超过100万个实体的图谱上运行的模型很少。如HolE[35]通过控制参数矩阵的稀疏性来降低计算成本,然而这些方法仍然难以扩展到包含数百万个实体和关系的知识图谱中。最新的ExpressGNN[56]尝试使用NeuralLP进行高效的规则归纳。尽管如此,在应对繁琐的深度架构和日益增长的知识图谱时,仍有很长的路要走。工业界的知识图谱往往具有数十亿节点和数百亿边,有些场景甚至可以到数百亿节点和数千亿边,如典型的电子商务领域的推荐系统。在这样规模的图上进行表示学习对于算力和算法设计的要求很高。比较常见的大规模图网络解决方法有采样子图、数据并行和模型并行。通过采样子图,模型可以不必在每次迭代时使用整个图结构进行运算;而数据并行和模型并行通过将数据与模型分配给不同的资源,进行并行计算,有效降低了单个GPU的内存需求,能有效降低模型运行时间。

3) 知识图谱表示学习预训练模型。

预训练模型可以从大规模无标签数据中学习到数据中的通用规律,在自然语言处理、视觉、语音等领域取得了广泛的成功,知识图谱作为一个重要的外部知识来源,可以提供很丰富的知识信息,将预训练模型移植到图上,可以提前将信息提取出来,而后将其运用到下游任务时,只需进行微调即

可获得良好的性能。在图上的预训练模型可以分为2类,一类使用一种固定的预训练GNN结构去处理同一类的图,类似于NLP中对于transformer的改进;另一类使用一些无监督的学习任务预先对GNN进行训练,之后再进行有监督的训练,主要进行预训练任务的创新。从方法论的角度,目前主流的自监督学习方法可以分为基于生成式的和基于对比学习的两大主要类别。基于生成式的自监督学习方法通过让模型对输入数据进行生成,重建学习到数据的潜在特征;基于对比学习的方法则主要是从输入数据中构造出正负样本,让模型在隐式表示空间对正负样本进行判别。

从现阶段的进展看,图上的自监督预训练模型仍处于起步阶段,大部分研究着眼于图结构与图属性的预训练,而知识图谱作为一种蕴含丰富的实体关系和文本信息图结构,如何捕捉到知识图谱中的实体关系和相关知识,也是未来非常值得研究的课题。同时知识图谱上的预训练模型能够有效降低知识图谱实际运用中花费的时间,若能将知识图谱的预训练模型运用到工业界的知识图谱中,对于常用的下游应用如智能检索、问题回答等,能有效降低时间消耗、提升用户体验。

4) 动态知识图谱的表示学习。

信息技术发展的脚步已逐渐加快,单一的技术及静态的数据已无法满足业务的发展需求,所以人们更加注重融合技术和动态数据的处理和应用。现有的神经网络模型通常聚焦于静态图谱中已经存在或出现过的实体与关系,但随时间变化的图谱在现实图谱实际应用中占绝大部分。例如,社交网络的用户会随着时间的推移添加关注或者移除好友,这时就需要更新用户的嵌入表示来反映其社会关系的演变;论文的引文网络由于频繁的引用和新技术的出现而不断丰富,文章的影响力甚至是学科分类都会随着时间产生变化,这时就需要更新节点嵌入以反映这种变化;在金融网络中,所有的金融交易会附带时间戳,而用户帐户的性质可能会因所涉及交易的特性而发生变化(例如,帐户参与洗钱或用户成为信用卡欺诈的受害者),尽早发现这种变化对于提高执法效率和减少金融机构的损失至关重要。这些都在促使研究者将研究重心转向动态知识图谱,通过编码关系数据的时间演变来建模只在特定时期成立的事实。

动态图的表示学习方法多为静态图的扩展,现阶段对于动态知识图谱的表示学习方法多采用递归神经网络(RNN)调节嵌入和学习图谱的动态信息,这些方法需要节点在整个时间跨度内的信息(包括训练集和测试集),这对节点集频繁变化的情况不太适用。而通常动态知识图谱除了已知的实体与关系更注重没有在知识图谱中出现过的关系和实体,这就造成了不同时间段的节点集可能完全不同。对于动态知识图谱,通过捕捉每个时间步(或切片)之间的动态关系来解决动态知识图谱中表示学习与知识推理是未来重要的研究方向。

6 结论

随着互联网的发展,网络数据内容呈现爆炸式增长的态势,作为人类知识集合的知识图谱已经引起了越来越多的关注。知识图谱以其强大的语义处理能力和开放组织能力,为人工智能时代的知识组织和智能应用奠定了基础。本文对知识图谱表示学习的方法进行了系统性的归纳,对现有的知识图谱表示学习方法进行了对比总结,并介绍了知识图谱的嵌入如何应用三元组分类、链路预测、推荐系统等下游应用中。为了方便未来对知识图谱的研究,整理并提供了通识数据集、特定领域数据集和特定任务数据集的开源库集合。最后,对人工智能时代知识图谱表示学习面临的挑战进行了总结,提出了4个未来可能的研究方向。知识图谱表示学习的应用能有效促进认知智能的发展,但距离真正实现知识图谱在人工智能领域的应用还有很长的路要走。

参考文献(References)

[1] Petroni F, Rocktäschel T, Lewis P, et al. Language models as knowledge bases[C]//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Pro-

cessing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong: Association for Computational Linguistics, 2019: 2463-2473.

[2] 刘峤, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. 计算机研究与发展, 2016, 53(3): 582-600.

[3] 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. Seattle, Washington, USA: Association for Computational Linguistics, 2011: 541-550.

[4] Bordes A, Weston J, Usunier N. Open question answering with weakly supervised embedding models[C]//Joint European conference on machine learning and knowledge discovery in databases. Berlin, Heidelberg: Springer, 2014: 165-180.

[5] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606.

[6] Bordes A, Weston J, Collobert R, et al. Learning structured embeddings of knowledge bases[C]//Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence. San Francisco, California USA: The AAAI Press, 2011: 301-306.

[7] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2013: 2787-2795.

[8] Weston J, Bordes A, Yakhnenko O, et al. Connecting language and knowledge bases with embedding models for relation extraction[C]//Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, Washington, USA: Association for Computational Linguistics, 2013: 1366-1371.

[9] Nickel M, Tresp V, Kriegel H P. Factorizing yago: Scalable machine learning for linked data[C]//Proceedings of the 21st international conference on World Wide Web. Lyon, France: Association for Computing Machinery, 2012: 271-280.

[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 International Conference on Machine Learning. Madison, WI, USA: Omnipress, 2011: 809-816.

[11] Bordes A, Glorot X, Weston J, et al. A semantic matching energy function for learning with multi-relational data[J]. Machine Learning, 2014, 94(2): 233-259.

[12] Blog G O. Introducing the knowledge graph: Thing, not strings[J]. Official Google Blog, 2012: 1-8.

[13] Ehrlinger L, Wöß W. Towards a definition of knowledge graphs[J]. Association for Computing Machinery, 2016, 48: 1-4.

[14] Wang Q, Mao Z, Wang B, et al. Knowledge graph embedding: A survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.

[15] Turian J, Ratinov L, Bengio Y. Word representations: A simple and general method for semi-supervised learning [J]. Association for Computational Linguistics, 2010(7): 384-394.

[16] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes[C]//Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. Québec City, Québec, Canada: AAAI Press, 2014: 1112-1119.

[17] Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion[C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin, Texas: AAAI Press, 2015: 2181-2187.

[18] Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion [C]//Twenty-seventh Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, United States: Curran Associates, Inc., 2013: 926-934.

[19] He S, Liu K, Ji G, et al. Learning to represent knowledge graphs with gaussian embedding[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York, NY, USA: Association for Computing Machinery, 2015: 623-632.

[20] Xiao H, Huang M, Zhu X. TransG: A generative model for knowledge graph embedding[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. Berlin, Germany: Association for Computational Linguistics, 2016: 2316-2325.

[21] Mikolov T, Yih W, Zweig G. Linguistic regularities in continuous space word representations[C]//Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies. Atlanta, Georgia: Association for Computational Linguistics, 2013: 746-751.

[22] Yoon H G, Song H J, Park S B, et al. A translation-based knowledge graph embedding preserving logical property of relations[J]. Association for Computational Linguistics, 2016: 907-916.

[23] Nguyen D Q, Sirts K, Qu L, et al. Stranse: A novel embedding model of entities and relationships in knowledge bases[J]. Association for Computational Linguistics, 2016: 460-466.

[24] 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. Beijing, China: Association for Computational Linguistics, 2015: 687-696.

[25] Ji G, Liu K, He S, et al. Knowledge graph completion with adaptive sparse transfer matrix[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Phoenix, Arizona: AAAI Press, 2016: 985-991.

[26] Fan M, Zhou Q, Chang E, et al. Transition-based knowledge graph embedding with relational mapping properties[C]//Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing. Chulalongkorn University, Phuket, Thailand: Department of Linguistics, 2014: 328-337.

[27] Xiao H, Huang M, Zhu X. From one point to a manifold: Knowledge graph embedding for precise link prediction [C]//Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York, USA: AAAI Press, 2016: 1315-1321.

[28] Feng J, Huang M, Wang M, et al. Knowledge graph embedding by flexible translation[C]//Proceedings of the Fifteenth International Conference on Principles of Knowledge Representation and Reasoning. Cape Town, South Africa: AAAI Press, 2016: 557-560.

[29] Xiao H, Huang M, Hao Y, et al. TransA: An adaptive approach for knowledge graph embedding[J]. arXiv preprint arXiv:1509.05490, 2015.

[30] Zhi S, Zhi D, Jian N, et al. RotatE: Knowledge graph embedding by relational rotation in complex space[C]//Seventh International Conference on Learning Representations. ICLR. Ernest N. Morial Convention Center, New Orleans. 2019: 1-18.

[31] Zhan Z, Jian C, Yong Z, et al. Learning hierarchy-aware knowledge graph embeddings for link prediction[C]//The Thirty-Fourth AAAI Conference on Artificial Intelligence. New York, USA: AAAI Press. 2020: 3065-3072.

[32] Jenatton R, Roux N, Bordes A, et al. A latent factor model for highly multi-relational data[C]//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc., 2012: 3167-3175.

[33] García-Durán A, Bordes A, Usunier N. Effective blending of two and three-way interactions for modeling multi-relational data[C]//Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Berlin, Heidelberg: Springer, 2014: 434-449.

[34] Yang B, Yih W, He X, et al. Embedding entities and relations for learning and inference in knowledge bases [C]//International Conference on Learning Representations 2015. San Diego, CA, USA: Conference Track Proceedings, 2015: 141-153.

[35] Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs[C]//Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. Phoenix, Arizona: AAAI Press, 2016: 1955-1961.

[36] Plate T A. Holographic reduced representations[J]. IEEE Transactions on Neural networks, 1995, 6(3): 623-641.

[37] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48, JMLR.org, New York, NY, USA: ACM, 2016: 2071-2080.

[38] Hayashi K, Shimbo M. On the equivalence of holographic and complex embeddings for link prediction[J]. Association for Computational Linguistics, 2017: 554-559.

[39] Liu H, Wu Y, Yang Y. Analogical inference for multi-relational embeddings[C]//Proceedings of the 34th International Conference on Machine Learning - Volume 70, JMLR.org. Sydney, NSW, Australia: ACM, 2017: 2168-2178.

[40] Dong X, Gabrilovich E, Heitz G, et al. Knowledge vault: A web-scale approach to probabilistic knowledge fusion [C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, USA: Association for Computing Machinery, 2014: 601-610.

[41] Liu Q, Jiang H, Evdokimov A, et al. Probabilistic reasoning via deep learning: Neural association models[C]// 25th International Joint Conference on Artificial Intelligence. New York City, NY, USA: Deep Learning for Artificial Intelligence, 2016: 271-278.

[42] Dettmers T, Minervini P, Stenetorp P, et al. Convolution-

al 2D knowledge graph embeddings[C]//32nd AAAI Conference on Artificial Intelligence, AAAI 2018. New Orleans, Louisiana USA: AAAI Publications, 2018, 32: 1811-1818.

[43] Dai Quoc Nguyen T D N, Nguyen D Q, Phung D. A novel embedding model for knowledge base completion based on convolutional neural network[C]//Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). New Orleans, Louisiana: Association for Computational Linguistics, 2018: 327-333.

[44] Balažević I, Allen C, Hospedales T M. Hypernetwork knowledge graph embeddings[C]//28th International Conference on Artificial Neural Networks. Munich, Germany: Springer, 2019: 553-565.

[45] Wang Q, Huang P, Wang H, et al. CoKE: Contextualized knowledge graph embedding[J]. arXiv preprint arXiv: 1911.02168, 2019.

[46] Yao L, Mao C, Luo Y. KG-BERT: BERT for knowledge graph completion[J]. arXiv preprint arXiv: 1909.03193, 2019.

[47] Wang X, Gao T, Zhu Z, et al. KEPLER: A unified model for knowledge embedding and pre-trained language representation[J]. Transactions of the Association for Computational Linguistics, 2021(9): 176-194.

[48] Yu D, Zhu C, Yang Y, et al. JAKET: Joint Pre-training of Knowledge Graph and Language Understanding[C]//Eighth International Conference on Learning Representations. Formerly Addis Ababa ETHIOPIA: ICLR, 2020: 1-11.

[49] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Minneapolis, Minnesota: Association for Computational Linguistics, 2018: 4171-4186.

[50] Zaremoodi P, Buntine W, Haffari G. Adaptive knowledge sharing in multi-task learning: Improving low-resource neural machine translation[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Melbourne, Australia: Association for Computational Linguistics, 2018: 656-661.

[51] 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.

[52] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[C]//5th SemWebEval Challenge at ESWC 2018. Heraklion, Greece: Springer, 2018: 593-607.

[53] Shang C, Tang Y, Huang J, et al. End-to-end structure-aware convolutional networks for knowledge base completion[C]//The Thirty-Third AAAI Conference on Artificial Intelligence. Honolulu, Hawaii, USA: AAAI Press, 2019, 33: 3060-3067.

[54] Vashishth S, Sanyal S, Nitin V, et al. Composition-based multi-relational graph convolutional networks[C]//Eighth International Conference on Learning Representations. Formerly Addis Ababa ETHIOPIA: ICLR, 2020: 1-15.

[55] Xie R, Liu Z, Jia J, et al. Representation learning of knowledge graphs with entity descriptions[C]//The Thirtieth AAAI Conference on Artificial Intelligence. Phoenix, Arizona, USA: AAAI Press, 2016: 2659-2665.

[56] Xiao H, Huang M, Meng L, et al. SSP: Semantic space projection for knowledge graph embedding with text descriptions[C]//The Thirty-First AAAI Conference on Artificial Intelligence. San Francisco, California, USA: AAAI Press, 2017: 3104-3110.

[57] Guo S, Wang Q, Wang B, et al. Semantically smooth knowledge graph embedding[C]//The 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Beijing, China: Association for Computational Linguistics, 2015(1): 84-94.

[58] Xie R, Liu Z, Sun M. Representation learning of knowledge graphs with hierarchical types[C]//The Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI). New York, USA: AAAI Press, 2016: 2965-2971.

[59] Lin Y, Liu Z, Sun M. Knowledge representation learning with entities, attributes and relations[C]//The Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI). New York, USA: AAAI Press, 2016: 2866-2872.

Knowledge graph representation learning method system in the era of artificial intelligence

ZHANG Hui[1], YANG Weijie[2,3*], LIU Wenwen[1], ZHANG Xun[1,3], DUAN Dagao[3,4], HAN Zhongming[3,4]

1. School of Computer Science and Engineering, Beijing Technology and Business University, Beijing 100048, China
2. School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China
3. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing 100048, China
4. School of Economics and Management, Beijing Technology and Business University, Beijing 100048, China

Abstract In recent years, the knowledge graph representation learning has been used to represent the components of the knowledge graphs in a low-dimensional vector embedding , as a mainstream way to combine the artificial intelligence with the knowledge graphs. This paper reviews the mainstream knowledge graph representation learning methods without auxiliary information, mainly, the distance-based and the semantic matching-based methods, and the knowledge graph representation learning methods containing textual auxiliary information and category auxiliary information, along with the advantages and the disadvantages of various representation learning methods. It is found that the introduction of auxiliary information can effectively represent new entities and relationships in the knowledge graph, but the time and space costs are significantly increased, and thus the methods without auxiliary information are more easily applied in practical scenarios at this stage. Finally, we show how the knowledge graph embedding can be applied to downstream tasks such as the triad classification, the link prediction and the recommender systems. A collection of datasets and open source libraries for different tasks is compiled and, and a comprehensive outlook on promising research directions such as large-scale, dynamic knowledge graphs is given.

Keywords knowledge graph; knowledge graph embedding; representation learning; deep learning

（责任编辑　傅雪）