文章编号:1671-4229(2021)03-0080-10

面向推理的知识图谱表示学习方法综述

翁金塔[1,2], 仇 晶[2], 张光华[3*]
(1. 中国科学院 信息工程研究所, 北京 100093; 2. 广州大学 网络空间先进技术研究院, 广东 广州 510006;
3. 河北科技大学 信息科学与工程学院, 河北 石家庄 050018)

摘 要: 知识图谱以其错综复杂、信息密集和语义关联的特征,推动了知识应用以及神经网络学科的发展。知识表示学习方法的发展以及日益增长的认知智能需求,使得知识表示与推理成为当前知识图谱和自然语言处理领域的研究热点。得益于良好的知识表示学习方法,基于表示学习的大规模知识推理已在相关数据集上得到验证。文章主要展开如下工作:①对已有表示学习与推理方法进行梳理,将知识图谱推理方法划分为:距离模型方法、双线性模型的语义相似度方法、神经网络学习方法和图神经网络学习方法;②对未来知识图谱表示学习与推理方法进行展望。知识图谱的表示学习与推理方法的发展,对推动整个知识社区和强人工智能发展具有重要意义。

关键词: 知识推理; 表示学习; 知识图谱; 网络学习; 链路预测
中图分类号: TP 18 **文献标志码**: A

The representation learning method of a knowledge graph for reasoning: A review

WENG Jin-ta[1,2], *QIU Jing*[2], *ZHANG Guang-hua*[3*]
(1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China;
2. Institute of Advanced Technology in Cyberspace, Guangzhou University, Guangzhou 510006, China;
3. School of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China)

Abstract: The knowledge graph has helped the development of knowledge application and neural network subject with its intricate, information-intensive and semantically related characteristics. The development of knowledge representation learning methods and the increasing demand for cognitive intelligence have made knowledge representation and reasoning a research focus in the current knowledge graph and natural language processing fields. Large-scale knowledge reasoning based on representation learning has also been verified on relevant data sets. This article mainly explores the following work: ① The existing representation learning and reasoning methods are sorted out, and the knowledge graph reasoning methods are divided into: Distance model method, semantic similarity method of bilinear model, neural network learning method, graph Neural network learning method. ② The article investigates the prospective future of the representation learning and reasoning methods of the knowledge graph. The development of representation learning and reasoning methods of the knowledge graph is of great significance to the entire knowledge community and to help the development of strong artificial intelligence.

Key words: knowledge reasoning; representation learning; knowledge graph; network learning; linking prediction

---

基金项目: 广东省高校创新团队资助项目(2020KCXTD007);广州市高校创新团队资助项目(202032854); 广州大学研究生创新能力培养资助计划(2019GDJC-M15);大学生创新创业训练资助项目(202011078028)

作者简介: 翁金塔(1997—),男,博士研究生. E-mail:552122632@qq.com
* 通信作者. E-mail: xian_software@163.com.

引文格式: 翁金塔, 仇晶, 张光华. 面向推理的知识图谱表示学习方法综述[J]. 广州大学学报(自然科学版),2021,20(3):80-89.

随着深度学习、大数据和云计算等新一代信息技术的发展,促使当前计算机发展从数据智能时代逐渐向认知智能跨越。以数据工程和算法分析设计为核心的计算方法层出不穷,促使数据的规模和知识外延不断扩大。研究者们不再局限于早期单一数据种类和关系的研究,而是加强了对多模态、多结构、多语义以及隐变量等图数据的探究[1]。

在数据种类明显和可获取的基础上,多种形式的数据都能通过以图或网状结构来表征。另一方面,自然语言处理和知识工程的发展,使得各种社会现象和数据系统通常也能被重新划分为具有一组节点和关系的网络。真实大数据的网络结构建模以及针对更复杂(动态、文本丰富和异构)的网络分析,也需要新算法和新模型进行更加有效的学习、推理和预测。同时,以网状数据为代表的各类网络任务的发展,如结构发现、链路预测和异常检测任务等,也开始引起了学者的关注。

知识图谱作为谷歌提出的一种用于搜素引擎优化的概念,已经发展成为良好的图形数据表征工具和知识存储库。知识图谱是由实体和链接标签组成的文本结构图数据[2],其标准形式为三元组,即 < 实体 1,链接,实体 2 > 的形式。相比标准的图数据,知识图谱中存储的是更为抽象化的文本知识。知识图谱以其简易的数据表征方式、多样化的知识表示和多层次的语义表达等优点,能够有效运用于挖掘关联知识和复杂关系。在大规模的开放知识图谱中,已经形成了诸如 Wikipedia、DBPedia 和 Yago 等千亿级别的知识图谱三元组数据。在垂直领域的知识图谱上,通过构建知识模型和制定信息抽取方式,已经形成了医疗、金融、安全和教育等面向领域应用的知识图谱。

知识图谱推理是指根据图谱中现有知识进行推理的过程。一般来说,推理过程分为简单推理和复杂推理。简单推理是指在已有的结点和关系上进行结点分类和链接预测的过程。复杂推理则是在融合外部知识的前提下,实现非原有关系和原有节点的推理。本文所指的推理,主要指的是简单推理,即实体分类和关系预测任务。目前,在文本表示学习方法和知识表示归纳出的各种方法,使得在知识图谱上开展推理任务成为可能。知识图谱上的推理工作主要分为 2 类:①局部任务,主要为结点分类、链路预测和知识补全与推理等;②全局任务,主要分为子图分类,图趋势预测等。相比局部任务,全局任务对在图的全局信息和局部信息上的学习能力要求更高。目前较多工作集中于结点分类和链路预测。在图上的知识推理算法上,早期工作主要集中于图的聚类算法、图的最短路径和图的联通性算法层面。对于更为深层次的挖掘,目前主要方式仍然是将图上的数据转化为以路径的方式进行学习或者采样后再学习的方式。知识推理的方法上,主要分为以下 4 类:①基于距离知识表示方法;②基于语义相似度的三元组表示方法;③基于神经网络的表示方法;④基于图神经网络的方法。本文将详细介绍和讨论目前知识图谱局部任务上各推理方法的发展近况。

1　基于距离的模型方法

1.1　TransE 模型[3]

最早提出基于距离模型的表示学习方法的为 TransE 模型。该模型是假设三元组 (h,r,t) 存在前提下,尾部实体 (t) 的嵌入表示应接近头部实体 (h) 的嵌入加上表示关系类型 (r) 的表示向量。通过 L1 或 L2 范数的距离度量使得所有三元组都尽可能保证 $h+l\approx t$(尽可能降低 $h+r-t$)。即

$$f_r(h,t)=\|l_h+l_r-l_t\|_{L1/L2},$$

其中,l 为实体 h 与关系 r 在向量空间上的表征。

为了实现带偏差的拟合约束和提高训练的样本约束,一般通过随机替换头部实体、关系和尾部实体的方式实现负样本生成,并将最终损失调整为正负样本的综合距离损失。损失函数为

$$L_{loss}=\sum_{(h,r,t)\in S(h',r',t')\in S^-}\max(0,f_r(h,t)+\gamma-f_{r'}(h',t')),$$

其中,S 是正确的已有三元组集合,S-为随机替换后生成的负样本三元组集合,max 函数为较大值,$fr(h,t)$ 为 r 关系下实体间的距离差距,γ 为三元组最大支持间隔。

TransE 相比其他以往的机器学习方法,其参数较少,模型复杂度不高,能通过距离的方式直接度量实体之间的语义相似度。相关实验表明,通过 TransE 训练生成的关系和实体向量,能够大大提升以往知识推理任务的效果。但其在距离表征的可解释性以及复杂语义层面的解释与支持度

上,仍然有待探究。

1.2 TransH 模型[4]

为了改进 TransE 对自反关系、多对一以及一对多等关系的支持,以及改进距离悖论实体 A + B = C,A + B = D→C = D,改进的 TransH 算法则假设实体间的距离,需要根据不同实体间的关系所在平面进行计算。其距离的度量上转化为如下函数:

$$l_{h,t}^{\perp} = l_{h,t} - w^{\perp} l_{h,t} \mid w \mid ,$$

$$f_r(h,t) = \left\| l^{\perp} + l_r - l^{\perp} \right\|_{L_1/L_2} ,$$

其中,lh,t 为原有的向量的实体嵌入表示,$W\perp$ 为关系 r 平面所在的法向量,$|w|$ 为 h 或者 t 实体向量在 r 关系屏幕 w 上的投影长度。相比 TransE,TransH 虽在一定程度上增加了计算量,但在关系密集的知识三元组数据库中,TransH 对实体和关系的计算仍然是处于同一语义空间的。

1.3 TransR/CTransR 模型[5]

从语义学的角度而言,对于关系 h 和实体类型,应当分别在相应的实体语义空间和关系语义空间进行转换操作。因此,TransR 在实体和关系向量的转换上通过矩阵转换形式进行了改进,即

$$\boldsymbol{l}_{h,t}^{r} = \boldsymbol{l}_{ht} \boldsymbol{M}_r ,$$

$$f_r(h,t) = \left\| l_h^r + l_r - l_t^r \right\|_{L_1/L_2} ,$$

主要采用特定的 $\boldsymbol{M}_r$ 关系矩阵进行实体 h 和 t 的向量转换以生成对应的 $\boldsymbol{l}_h^r$ 和 $\boldsymbol{l}_t^r$。进一步地,由于 TransR 在实体距离的度量上并未考虑实体间关系可能存在的多类别情况,因此,CTransR 通过实现对关系的再分类和对每个关系子空间 c 分配特定的投影矩阵 $\boldsymbol{M}_{cr}$,并对这些投影矩阵转化后所形成的关系向量嵌入进行聚类,以便获取最终的关系向量表示。

$$\boldsymbol{M}_{h,t}^{r} = r_p \boldsymbol{D}_{h,t}^{p} + \boldsymbol{I}^{m \times n} ,$$

$$\boldsymbol{l}_{h,t}^{r} = \boldsymbol{l}_{ht} \boldsymbol{M}_{h,t}^{r} ,$$

$$f_r(h,t) = \left\| l_h^r + l_r - l_t^r \right\|_{L_1/L_2} ,$$

其中,$\boldsymbol{D}$ 为实体 h 和 t 在各自用于调整相应实体上的投影矩阵,$\boldsymbol{M}$ 为实体最终的投影矩阵,$\boldsymbol{l}_r$ 为实体 h 和实体 t 在矩阵初始嵌入经过 $\boldsymbol{M}$ 投影后的矩阵。相比其他算法,TransD 实现了更加复杂的语义空间向量,实体也需要参与投影向量的计算,并在一定程度上减少了参数数量和矩阵运算量。

1.4 小结

图 1 为 Trans 距离模型的集合表示模型,可以看到随着对距离模型研究的深入,这类模型开始由稀疏的、单关系的和同语义空间下的计算向密集、多关系和复杂语义空间过渡,使得模型得以学习更多信息。关于距离模型的相关改进不断演化,出现了 TransSpare[6]、TransG[7]、TransA[8] 和 KG2E[9] 等多种模型,从不同角度尝试解决复杂关系建模问题。

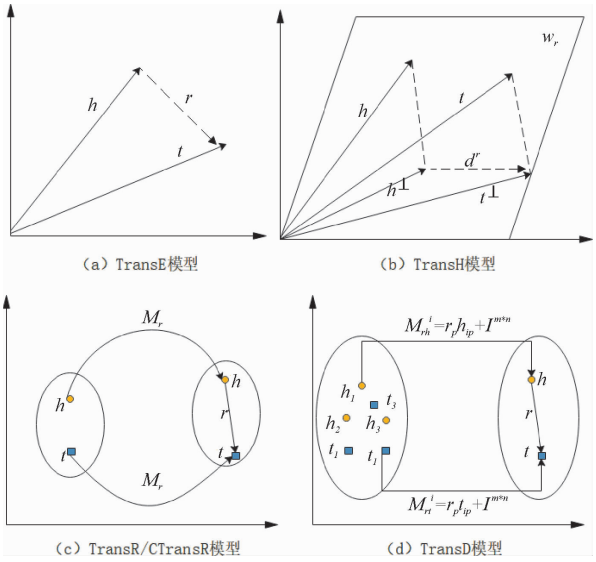

图 1 Trans 距离表示模型

Fig. 1 Trans distance representation model

基于距离的算法,如图 2 所示,在经过 200 代的正确例(正确的三元组)和错误例(错误的三元组或者人为修改的三元组)作为训练数据例的前提下,模型实现了最终的数据表征。通过该方法,电影《哈利波特》的女主角之一罗琳与科学、神秘和恐怖类的小说的实体结点在语义空间上比较接近,而与自传类小说距离较远,从语义空间的层度进一步验证了该类距离模型的有效性。

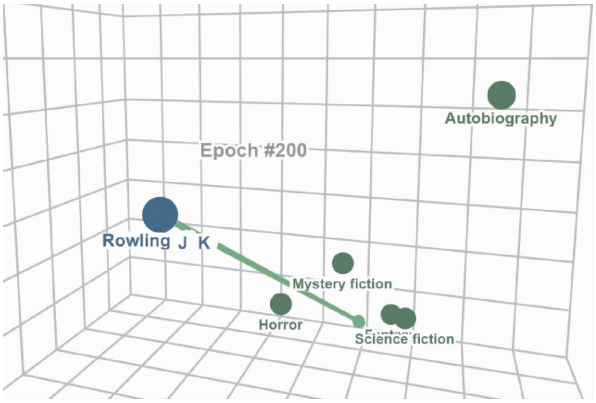

图 2 Trans 距离算法的几何验证

Fig. 2 Geometric verification of Trans distance algorithm

基于距离模型的算法能够在大规模稀疏知识图谱上取得较好的准确率和易用性,但是对于细粒度的、多跳知识关联和子关系信息的学习,模型对这类信息的关注仍然有所不足。此外,距离模型的语义解释性上相对较弱。

2　双线性模型及其变体

与基于距离的嵌入模型不同,双线性模型主要采用的假设依据为实体与关系之间是存在着潜在语义,通过匹配实体和关系的潜在语义空间相似度,能度量三元组事实的可信度。常见的模型可以分为 RESCAL 模型[10]及其变体 DistMult[11]、Hole[12]、ComplEx[13]和 ANALOGY 准则[14]。

2.1　RESCAL 模型

RESCAL 模型认为实体与关系之间的向量嵌入是存在潜在的语义交互作用,因此,在线性模型的假设上采用了如图 3 所示的复杂线性模型。与距离模型不同,RESCAL 模型采用的评价函数 f 为

$$f_r(h,t) = h^{\mathrm{T}}\boldsymbol{M}^r t = \sum_{i=0}^{d-1}\sum_{j=0}^{d-1}[M_r]_{ij} \cdot [h]_i \cdot [t]_j,$$

其中,d 为实体的编码层长度,$\boldsymbol{M}_r$ 为头部实体 h 到尾部实体 t 的关系矩阵。

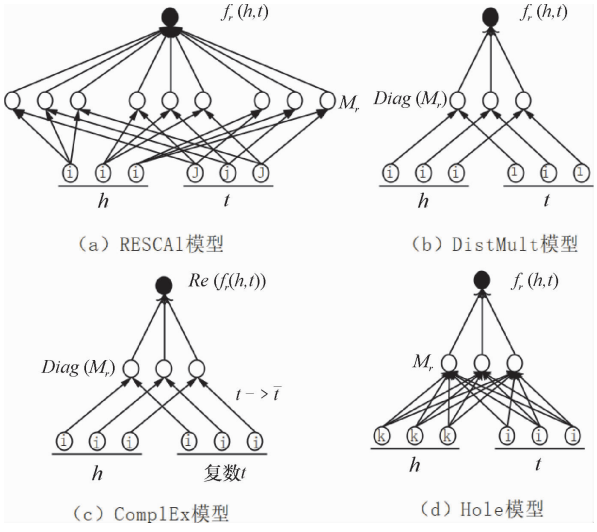

图 3　双线性模型及其变体

Fig. 3　Bilinear models and their variants

RESCAL 实体和关系的运算上采用的是矩阵运算,因此,每个实体 h 和 t 的深层编码信息能够实现交互计算,使得最终评价得分上包含了这部分语义相似信息。

2.2　DistMult 模型

与此同时,由高计算量带来的过拟合和复杂度问题 O(n2)是 RESCAL 模型在大规模知识图谱上较难开展的原因之一。因此,如图 3(b)所示,DistMult 通过简化的矩阵松弛方式对计算方式进行了更改,即:

$$f_r(h,t) = h^{\mathrm{T}}diag(M^r)t = \sum_{i=0}^{d-1}diag(M^r)_i \cdot [h]_i \cdot [t]_i,$$

其中,diag( )为对角矩阵上的主对角线值,i 为 h 和 t 的编码表示中的第 i 位。

DistMult 模型以较为简化的方式降低了计算量,其假设语义相似度的表示是同各自编码中的对应位置上的信息有关,并且这种信息相似度在关系矩阵中仅与矩阵对角线上的值有关。由计算公式可知,h 和 t 的位置变换并不会引起矩阵计算量值的变换。因此,对于具备非对称关系或者多重关系的三元组,DistMult 虽降低了计算量和复杂度 O(d),但仍然不足以表征这些复杂关系等信息。

2.3　ComplEx 模型

为了解决上述在非对称关系模型的信息表征不足问题,如图 3(c)所示,ComplEx 采用引入复数的方式对语义相似度度量方式进行了进一步改造。通过引入复数的操作,使得对 h 和 t 的非对称关系(互换位置)能够得到不同的信息表征,从而学习到非对称关系上的更多信息。具体计算方式如下:

$$f_r(h,t) = Re(h^{\mathrm{T}}diag(M^r)\bar{t}) = \mathrm{Re}(\sum_{i=0}^{d-1}diag(M^r)_i \cdot [h]_i \cdot [\bar{t}]_i),$$

其中,Re 为复数取实值操作,$\bar{t}$ 为尾部实体 t 的共轭复数。

通过复数表征和共轭函数操作,使得头部实体和尾部实体信息的互换所导致的计算得以不同,从而影响最终三元组的得分。

2.4　Hole 模型

本质上,RESCAL 的计算复杂度高和 DistMult 的计算信息不平衡问题并没有在 ComplEx 中得到融合。为此,Hole 模型中和了 RESCAL 的表达能力和 ComplEx 与 DistMult 的计算效率特点,将三元组计算方式度量为如图 3(d)所示的几何表示,其具体评价函数为

$$f_r(h,t) = r^{\mathrm{T}}(h \otimes t) = \sum_{i=0}^{d-1} [r]_i \sum_{k=0}^{d-1} [h]_k \cdot [t]_{(k+i)_{\mathrm{bmodd}}},$$

其中,h_k 为 k 位置上的实体编码,t 为与头部实体 h 对应的尾部实体的所有编码输入,r 为控制每个 i 关系的信息编码。

Hole 模型从一定程度上降低了计算量,另一方面通过降低矩阵关系 r 的编码并赋予语义空间的含义,使得多种实体和复杂关系的语义相似度计算得以区分。

2.5 ANALOGY

ANALOGY 模型将双线性模型进行了总结,通过是否支持实体互换操作制定了分类准则。此外,ANALOGY 还通过引入交换律和规则化的方式,对关系矩阵进行了总结,即双线性模型应当符合:

规则化:$\boldsymbol{M}_r\boldsymbol{M}_r^{\mathrm{T}} = \boldsymbol{M}_r^{\mathrm{T}}\boldsymbol{M}_r,\ \forall r \in \Re$

交互率:$\boldsymbol{M}_r\boldsymbol{M}_{r'} = \boldsymbol{M}_r\boldsymbol{M}_r,\ \forall r, r' \in \Re$

事实上,与基于距离的模型相比,双线性模型以尽可能符合三元组事实结构的方式进行了建模,具备一定的解释性。然而,其在多关系、子关系等复杂关系的表征上仍然不够全面,因此,往往不能在大规模知识图谱方面广泛应用。以(美国,总统,拜登)和(拜登,出生,美国)2 个事实三元组为例,这种级联关系并无法得到表征。相反,对于领域特定知识图谱或者关系较为简易的知识图谱,往往能够通过其较大的计算量和参数量取得不错的效果。

2.6 小结

线性模型是以 RESCAL 模型及其变体为代表的一类模型。这类模型认为三元组的信息是通过点积的方式直接对实体和关系进行复合后形成的。相比距离模型而言,这类模型能够抛开距离限制,赋予了一定的解释性和编码多样化表达的可能,但同时也增加了模型的计算量和复杂度。

3 神经网络模型的学习方式

基于神经网络的模型是另一种由深度神经网络结构所衍生的知识表示学习方法。通过神经网络模型在嵌入层获取实体和关系的表示层的向量表示,在隐藏层对这些向量表示进行运算,再在输出层进行相应的变化,以获取最终的知识三元组分数表征。

如图 4 所示,分别为基于能量聚合的 SME 模型[15]、NTN 模型[16]以及 MLP 模型[17]和 NAM 模型[18]的计算方式。

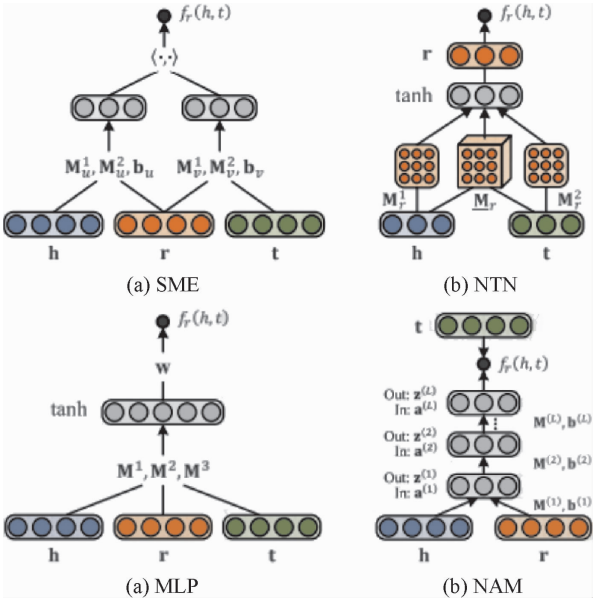

图 4 基于神经网络的推理模型(图片来源文献[19])

Fig. 4 Neural Network-based reasoning model (image from Ref. [19])

3.1 SME 模型

SME 模型首先将三元组的评价分为 2 个计算模块:头部实体 + 关系,尾部实体 + 关系的方式。独立地,项目还对每个独立模块下的实体和关系赋予独立的权值矩阵,具体的计算方式可采用如下 2 种方式之一:

$$f_r(h,t) = (\boldsymbol{M}_u^1 h + \boldsymbol{M}_r^2 r + b_u)^{\mathrm{T}}(\boldsymbol{M}_v^1 t + \boldsymbol{M}_v^2 r + b_v)$$

$$f_r(h,t) = ((\boldsymbol{M}_u^1 h) \cdot (\boldsymbol{M}_r^2 r) + b_u)^{\mathrm{T}}((\boldsymbol{M}_v^1 t) \cdot (\boldsymbol{M}_v^2 r) + b_v),$$

其中,$\boldsymbol{M}$ 为控制各部分信息的权值矩阵,b 为偏置值,h 与 t 为实体编码,r 为关系编码。

3.2 NTN 模型

不同于 SME 模型,NTN 神经网络计算模型将得分函数设计为 3 个部分:头部实体的信息、尾部实体的信息,以及实体关联的关系信息计算部分。通过对该 3 部分信息进行信息聚合后,使用 tanh 函数转化为影响关系变量的权值,再对权值与关系编码进行相乘得到最终三元组的得分表征:

$$f_r(h,t) = r^{\mathrm{T}} tanh(h^{\mathrm{T}} \underline{\boldsymbol{M}_r} t + \boldsymbol{M}_r^1 h + \boldsymbol{M}_r^2 t + b_r),$$

NTN 模型在设计上采用的是基于关系的模型计算方式,对于表征多关系或者复杂关系上的变化具备一定优势。

3.3　MLP 模型

相比 SME 模型和 NTN 模型的较大计算量和参数量,MLP 模型在设计上采用了简洁划分的神经网络思想,采用 3 个权重独立控制头部实体、尾部实体以及关系的信息。此外,在最终的激活后处理上,采用的是独立的权值矩阵代替关系编码的方式。其具体计算如下:

$$f_r(h,t)=w^{\mathrm{T}}tanh(\boldsymbol{M}^1h+\boldsymbol{M}^2t+\underline{\boldsymbol{M}_r^3}t+b_r),$$

其中,$\boldsymbol{M}^1$、$\boldsymbol{M}^2$ 和 $\boldsymbol{M}^3$ 是控制各部分信息的权值矩阵,w 为附加的用于控制最终得分矩阵。

3.4　NAM 模型

为了提高模型在解释性以及在深层关系上的细微变化,NAM 神经关联模型通过以头部实体和关系编码联结作为输入,然后将该输入放入 L 层神经网络中进行训练,通过激活层得到最终的权值得分。进一步地,将该权值得分与尾部实体编码相乘求得语义相似性最终得分,计算如下:

$$f_r(h,t)=t^{\mathrm{T}}Relu(\boldsymbol{M}^lz^{(l-1)}+b^{(l)}),\ l=1,\cdots,L,$$

$$Z^l=Relu(\boldsymbol{M}^lz^{(l-1)}+b^{(l)}),$$

其中,l 为层数,Z 为上一层输出,初始化为 $Z(0)=(h,r)\in\mathbf{R}2d$,$L$ 为层数,$\boldsymbol{M}$ 为第 l 层的参数矩阵,Zl 为每层的输出。

3.5　其他模型

在更为复杂的深度学习模型中,如图 5 所示的 Proje 模型结构[20]。

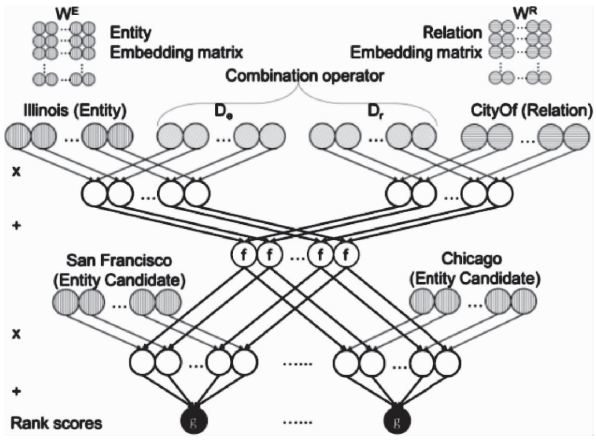

图 5　Proje 模型(来源文献[20])

Fig. 5　Proje Model (From Ref. [20])

Proje 模型首先将实体的候选集投影到输入数据的目标向量表示,使用可学习的组合算子将表示输入数据的嵌入向量组合成目标向量,优化了候选实体(关系)的集合。进一步的,模型通过相关函数 f 聚合了相应信息。模型为解决候选采样提供了大量的三元组数据集合,并使用聚合函数对这些信息进行了聚合,最后使用得分算子 g 的数值来实现三元组排名。

也有 Yao 等[21]从 BERT 模型的方式将每个标记的输入向量表示成其词向量、位置向量和段落向量的和,得到三元组在上下文相关的隐层表示后,传入一个 sigmoid 层用来计算三元组的得分。

3.6　小结

基于神经网络的模型与双线性模型的解释基础都是建立在信息聚合的前提下,但在计算思想上采用模块化计算和数据归一化的思想。在基于神经网络的模型当中,头部实体、尾部实体和关系不是通过 3 者直接点积操作实现,而是由独立权值控制各部分信息,并采用激活函数进行三元组信息平滑处理的思想。在网络结构的设计上,神经网络的方式更加灵活多变,能够根据具体任务所关注的需求进行相应改进,如侧重关系或者结点信息的提取。基于神经网络的模型能够以其更加庞大的计算量和参数量获得较强的学习能力和推理能力,但具备高复杂度和低解释性的特点。

4　基于图神经网络的方法

图神经网络是一种近年来提出的网络模型,其通过在图上开展相应的卷积学习方式并聚合图上的相应特征与信息,能够在图关系的学习中,获得良好的图谱推理任务效果。

目前图神经网络主要划分为 5 个类别:图卷积网络(Graph Convolution Networks, GCN)、图注意力网络(Graph Attention Networks)、图自编码器(Graph Autoencoders)、图生成网络(Graph Generative Networks)和图时空网络(Graph Spatial-temporal Networks)。本文主要以图卷积网络推理工作进行介绍。

图卷积神经网络的计算流程一般分为 2 个部分:①图结构模型的构建,在图结构模型构建中,需要根据已有的知识图谱结构,定义节点类型和链接类型,并根据下游具体任务进行适当的结构

改造;②图的信息传播和聚合过程,如图 6 所示,通过已有的知识图模型结构进行信息的传播和聚合,将某一结点的信息归结于所在结点的近邻结点和关系的复合表征。不同于以往方式,该方式将能以符合图谱结构的方式进行学习,并能使得最终模型获取到多层结点和关系上的图特征信息。

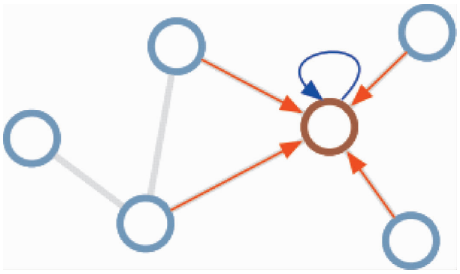

图 6 图神经网络的基本学习方式

Fig. 6 The basic learning mode of graph neural network

常见的信息传播和聚合包括:1-hop 邻接结点聚合,2-hop 邻接结点聚合,元路径学习等方式。1-hop 邻接结点聚合代表的是一个结点的信息,是由其结点相关联的关系的信息以及其邻结点的节点信息聚合而成。2-hop 邻接结点聚合是一个结点的信息,是由其 2 层结点相关联的关系的信息以及其 2 层邻结点的节点信息聚合而成。本节将主要面向结点分类和链路预测的 GCN,R-GCN 和 H-GCN 模型。

4.1 GCN[22]

GCN 是一种基于图结构的图卷积网络,也是图网络学习的雏形。一般而言,图卷积网络的基本公式如下。每一层的信息都由前一层的信息聚合而成。如下为每一层的信息更新:

$$H^{(l+1)} = \sigma(\tilde{\boldsymbol{D}} - \frac{1}{2}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}} - \frac{1}{2}H^{(l)}W^{(l)}),$$

其中,H_{l+1}代表更新之后的隐藏层信息,H_l 代表当前结点的特征,对于输入层而言,H 就是输入的图数据信息 X。$\boldsymbol{D}$ 代表图数据的邻接矩阵,$\boldsymbol{A}$ 代表结点之间的共享信息矩阵,一般为 $N \times N$ 维的矩阵,W 代表的是权重。

具体地,通过构造一个 2 层 GCN 的模式将能实现一定的推理任务。在第一层 GCN 中使用类似 Relu 函数进行激活隐层值。进一步,通过将隐层值作为新值,输入到外层 GCN 结构中,并在激活函数中选择 Softmax 归一化的方式进行标签权重的分配。整体的正向传播的公式为

$$Z = f(X,A) = \mathrm{softmax}(\tilde{A}\mathrm{ReLU}(\tilde{A}XW^{(0)})W^{(1)}).$$

对于实体分类任务而言,损失函数则被定义如下:

$$\mathcal{L} = -\sum_{l\in Y_L}\sum_{f=1}^{F} Y_{lt}\ln Z_{lf},$$

其中,F 代表节点的分类树,Y_L 则是有该标签的节点的集合,如 Y_{lf}为标签为 f 的所有点的集合。

由于基于 GCN 的推理模型往往具备较大的参数量和计算量,为防止过拟合效果,一般层数选择上大都小于 3 层。

4.2 R-GCN[23]

R-GCN(Relation-GCN)全名为基于关系的卷积网络构成。相比于 GCN,该模型(图 7)考虑了更多的图信息。首先,该模型将关系分为 3 类:①指向结点自身的关系;②指向外部结点的关系;③结点自身的关系。通过聚合这 3 类关系的信息向量,结合已有的 GCN 模型,最终经过 RELU 函数依据任务要求获取相应的概率分布。

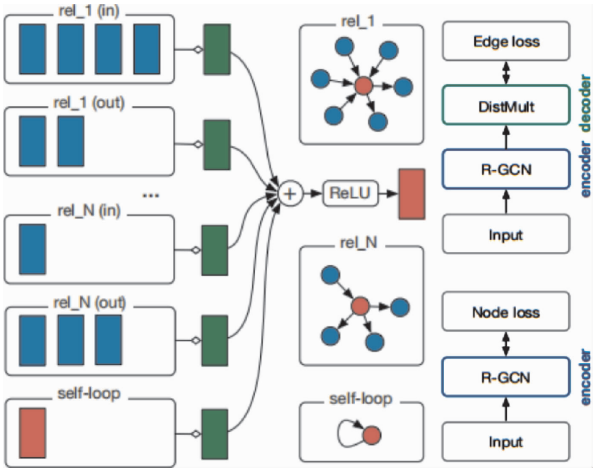

图 7 R-GCN 的模型示意图(图片来源于文献[23])

Fig. 7 Model schematic of R-GCN (image from Ref. [23])

相比 GCN,R-GCN 在隐藏层的信息单独分离了边的信息和结点自身的信息,使用了 Wr 来表示和调整其边上的临近结点的信息,弥补了 GCN 对于边相连结点信息的考虑不足特点。具体结点的信息计算公式如下:

$$h_i^{(1+1)} = \sigma\Big(\sum_{r\in R}\sum_{j\in N}\frac{1}{Ci,r}\boldsymbol{w}_r^{(1)}h_j^{(1)} + \boldsymbol{w}_0^{(1)}h_i^{(1)}\Big),$$

其中,h 代表隐藏层信息,i 代表结点编号,l 代表层数,c 代表边的数量,N 代表特点边类型的关系数量,j 代表其中一条边,j 代表与 i 相连的结点中

关系为 r 的结点的信息,$\boldsymbol{W}_0$ 代表原始结点信息的优化权重矩阵,$\boldsymbol{W}_r$ 代表临近边信息的优化权重矩阵。

在 R-GCN 中,高层结点的隐藏信息能够以低层结点的隐藏信息及其所连接的不同类型的边的信息聚合而成。相比结点的损失定义上,关系损失的定义使用了多一层的 DistMult 解码器。这使得能够在结点损失的基础上,进一步通过 DistMult 模型转化为边损失分数。最终,通过输入已经存在的知识图数据进行训练优化。

4.3　H-GCN[24]

H-GCN(Hierarchical GCN)中文名为层次 GCN,相比简单的 GCN 结构(通常只包含 1 ~ 2 层),其定义了多层次的 GCN。由于简单的进行多层次的 GCN 聚类会使得计算量增大,同时会使得模型过拟合。因此,层次 GCN 的任务进行了层级信息共享设计(图 8)。

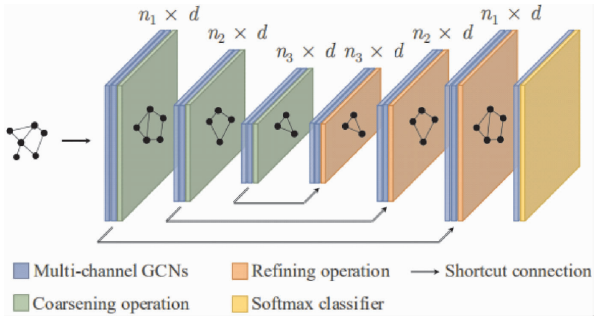

图 8　H-GCN 的模型架构图(来源于文献[24])

Fig. 8　H-gcn model architecture diagram (Image from Ref. [24])

在图 8 中,有 7 层 GCN 结构,其中,3 层为粗糙层,3 层为对称精炼层,1 层为输出层。H-GCN 是一个端到端的对称的网络结构。左侧部分中,图的表示矩阵在每次图卷积操作后,使用粗提取方式把结构相似的节点合并成超节点,实现逐层减小图的规模。对应地,右侧在每次 GCN 操作后将超节点进行还原,即可得到每个原始节点的表达。同时,在对称的神经网络层之间加了路径连接,从而可以更好地综合左右部分信息。具体公式为

$$H_i = \boldsymbol{M}_{l-i} \cdot \boldsymbol{G}_i + \boldsymbol{G}_{l-i},$$

其中,$\boldsymbol{M}$ 为每层赋予的权值矩阵,$\boldsymbol{G}$ 为图的表示矩阵,l 为层数,i 为层索引。

相比 R-GCN 而言,H-GCN 采用的策略是对粗细层次的图特征信息进行聚合的方式,而 R-GCN 则采用的是在原有图结构信息上获取更多的结点和边信息汇入方式。

4.4　小结

事实上,图卷积神经网络发展较为迅猛,用于结点分类和链路预测的推理模型远不止上述部分。本文仅以 GCN、R-GCN 和用于结点分类的 H-GCN 为例进行文献梳理。图神经网络符合了知识特性:知识三元组是动态调整与演进的,是路径触发的,非线性的。在网络结构上,能以符合知识图谱的网状神经网络方式进行学习,因此,能够取得较好的学习效果。此外,基于图神经网络的算法,采用更高的计算量和更加多样化的信息聚合和传播方式,因此,能够聚合图上更多特征信息。但是,由于参数计算庞大所带来的图学习过拟合特性,使得图神经网络模型往往无法适用于密集知识三元组以及知识图谱结构的变化。

5　总结与展望

得益于知识表示学习和神经网络科学的发展,知识图谱推理方法已经形成了包括本文所梳理的 4 类学习方式。本文将知识图谱推理方法分为以下 4 类:①基于距离模型的方法;②基于双线性模型的方法;③基于神经网络的方法;④基于图卷积神经网络的方法。此外,本文详细梳理了各类方法上的代表性模型及其计算方式,以快速掌握目前知识图谱表示学习与推理的研究近况。

事实上,目前关于知识表示学习方法也在随着知识表示和知识外延的扩充而在不断变化[25-28],如时空知识图谱、面向推理的事理图谱和元路径学习方式 Deepwalk 等。未来,面向知识推理的知识表示学习模型应该更加关注如下几点:

(1)图特征的知识表示学习

本文所指出的 4 类学习方法,都需要将完整的具备图语义的大图划分为系列知识三元组。事实上,这种由于训练前数据处理所带来的信息损失是可以避免的。通过融合图上学习到的相关特征信息到现有的向量嵌入或者模型设计中,能够在一定程度上缓解由于图特征缺失所带来的信息偏差。

(2)动态知识图谱的学习

知识图谱中的知识三元组是不断变化的,其所形成的三元组和图结构也是不断变化的。以金融知识图谱为例,一般具备三元组数据增长快、前后时刻知识关联的特征。因此,如何学习到知识图谱动态变化中的时序信息,并融入到现有的知识图谱推理工作中,是值得关注的。

(3)跨领域知识迁移的知识图谱学习

现有的知识图谱推理本质上是数据驱动的学习方法。因此,其所学习到的信息仅对当前状态下的知识图谱具备可靠性。对于其他类似领域的知识推理或者变换不大的知识图谱上,其知识推理能力一般不具备可迁移能力。如何复合跨领域的知识图谱,实现知识推理能力的迁移,是知识图谱推理工作的研究热点之一。

知识推理工作仍然处于弱知识推理工作[29-31],对于实现更加完备的强知识推理,还需要融合更多认知科学和脑科学机理,建立统一的知识联系接口以完善知识库,提高知识图谱构建的战略共识等。

参考文献:

[1] 熊回香, 杨滋荣, 蒋武轩. 跨媒体知识图谱构建中多模态数据语义相关性研究[J]. 情报理论与实践, 2019, 42(2): 13-18,24.

[2] Steiner T, Verborgh R, Troncy R, et al. Adding realtime coverage to the google knowledge graph[C]//11th International Semantic Web Conference (ISWC 2012). Aachen: Citeseer, 2012:65-68.

[3] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data[C]//Neural Information Processing Systems (NIPS). New York: Curran Associates, 2013:2787-2795.

[4] Wang Z, Zhang J, Feng J, et al. Knowledge graph embedding by translating on hyperplanes[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2014:1112-1119.

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

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

[7] Han X, 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. Berling: Association for Computational Linguistics, 2016: 2316-2325.

[8] Xiao H, Huang M, Hao Y, et al. TransA: An adaptive approach for knowledge graph embedding[EB/OL]. (2015-09-28)[2021-06-15]. https://arxiv.org/abs/1509.05490.

[9] 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: Association for Computing Machinery, 2015:623-632.

[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. Bellevue: Omnipress, 2011:809-816.

[11] Yang B, Yih W-t, He X, et al. Embedding entities and relations for learning and inference in knowledge bases[C]//Proceedings of the International Conference on Learning Representations (ICLR). San Diego: Computational and Biological Learning Society, 2015.

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

[13] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction[C]//Proceedings of the 33rd International Conference on Machine Learning (ICML). Cambridge: JMLR, 2016:2071-2080.

[14] Liu H X, Wu Y X, Yang Y M. Analogical inference for multi-relational embeddings[C]//Proceedings of the 34th International Conference on Machine Learning. Cambridge: JMLR, 2017:2168-2178.

[15] Bordes A, Glorot X, Weston J, et al. A semantic matching energy function for learning with multi-relational data[J]. Ma-

chine Learning, 2014, 94(2): 233-259.

[16] Socher R, Chen D, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]//Advances in neural information processing systems. Red Hook: Curran Associates Inc., 2013:926-934.

[17] Dong X, Gabrilovich E, Heitz G, et al. Knowledge vault: A web-scale approach to probabilistic knowledge fusion[C]// Proceedings of the 20th International Conference on Knowledge Discovery and Data Mining. NewYork: Association for Computing Machinery, 2014:601-610.

[18] Liu Q, Jiang H, Evdokimov A, et al. Probabilistic reasoning via deep learning: Neural association models[EB/OL]. (2016-08-03)[2021-06-10]. https://arxiv.org/abs/1603.07704v2.

[19] 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.

[20] Shi B, Weninger T. ProjE: Embedding projection for knowledge graph completion[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI Press, 2017:1236-1242.

[21] Yao L, Mao C, Luo Y. KG-BERT: BERT for knowledge graph completion[EB/OL]. (2019-09-11)[2021-06-10]. https://arxiv.org/abs/1909.03193.

[22] 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.

[23] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[C]//The Semantic Web: European Semantic Web Conference. Portoro: Springer International Publishing, 2018:593-607.

[24] Hu F, Zhu Y, Wu S, et al. Hierarchical graph convolutional networks for semi-supervised node classification[EB/OL]. (2019-06-10)[2021-06-05]. https:arxiv.org/abs/1902.06667.

[25] Qiu J, Du L, Zhang D, et al. Nei-TTE: Intelligent traffic time estimation based on fine-grained time derivation of road segments for smart city[J]. IEEE Transactions on Industrial Informatics, 2020, 16(4): 2659-2666.

[26] Qiu J, Chai Y, Tian Z, et al. Automatic concept extraction based on semantic graphs from big data in smart city[J]. IEEE Transactions on Computational Social Systems, 2020,7(1): 225-233.

[27] Xu D, Tian Z, Lai R, et al. Deep learning based emotional analysis of microblog texts[J]. Information Fusion, 2020, 64(12): 1-11.

[28] Weng J, Gan W, Ding G, et al. SESM: Emotional social semantic and time series analysis of learners´comments[C]// 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Piscataway: IEEE, 2020:4134-4139.

[29] Weng J, Luo J, Ding G, et al. Group-intelligence construction of linguistic-humanity kg in Chinese verses[C]//2019 15th International Conference on Semantics, Knowledge and Grids (SKG). Piscataway: IEEE, 2019: 119-122.

[30] Weng J, Gao Y, Qiu J, et al. Construction and application of teaching system based on crowdsourcing knowledge graph[C]// China Conference on Knowledge Graph and Semantic Computing. Berlin: Springer, 2019:25-37.

[31] Weng J, Qiu J, Gao Y. Ontology language xol used for cross-application communication[M]. London: Intech Open, 2020.

【责任编辑: 孙向荣】