



基于半边原理的知识图谱补全

程 涛¹, 陈 恒^{1,2}, 李冠宇¹

(1. 大连海事大学 信息科学技术学院, 辽宁 大连 116026; 2. 大连外国语大学 语言智能研究中心, 辽宁 大连 116044)

摘 要: 针对现有知识图谱补全算法耗时长和准确性差的问题, 构建一种基于半边的多层卷积模型。通过引入半边原理, 运用实体的描述信息和关系自身的特性, 结合两者的语义相似度对关系所连接的头尾实体进行约束, 组成半边, 在此基础上使用卷积神经网络进行知识图谱补全。该模型将只含有一个实体和关系的不完全 RDF 三元组以半边的形式保存, 便于补全扩充的知识图谱。实验结果表明, 与 TransE、DKRL 等模型相比, 该模型具有较优的实体和关系预测性能, 同时能有效缩短运行时间。

关键词: 半边原理; 卷积神经网络; 知识图谱补全; 实体预测; 关系预测

开放科学(资源服务)标志码(OSID):



中文引用格式: 程涛, 陈恒, 李冠宇. 基于半边原理的知识图谱补全[J]. 计算机工程, 2020, 46(11): 84-89.

英文引用格式: CHENG Tao, CHEN Heng, LI Guanyu. Knowledge graph completion based on half-edge principle[J]. Computer Engineering, 2020, 46(11): 84-89.

Knowledge Graph Completion Based on Half-Edge Principle

CHENG Tao¹, CHEN Heng^{1,2}, LI Guanyu¹

(1. Faculty of Information Science and Technology, Dalian Maritime University, Dalian, Liaoning 116026, China;
2. Research Center for Language Intelligence, Dalian University of Foreign Languages, Dalian, Liaoning 116044, China)

[Abstract] Existing knowledge graph completion algorithms are time-consuming and inaccurate. To address these problems, this paper proposes a multi-layer convolution model based on half-edge. The model introduces the half-edge principle, and uses the descriptive information of the entity and the characteristics of the relation itself to constrain the head and tail entities connected by the relation based on their semantic similarity, so as to form the half-edge. On this basis, the Convolutional Neural Network (CNN) is used to complete the knowledge graph. In this model, the incomplete RDF triples containing only one entity and relationship are saved in the form of half-edge, which facilitates the completion of extended knowledge graphs and provides a foundation for the dynamic completion of knowledge graphs. Experimental results show that the proposed model has better performance in entity prediction and relationship prediction than TransE, DKRL and other models, and can effectively reduce the running time.

[Key words] half-edge principle; Convolutional Neural Network (CNN); knowledge graph completion; entity prediction; relation prediction

DOI: 10.19678/j.issn.1000-3428.0055720

0 概述

在知识图谱中, 知识表示为三元组 <头实体, 关系, 尾实体> (即 RDF 三元组) 的形式, 使得知识能够得到结构化处理^[1]。人们必须为知识图谱应用设计各种基于图形的方法, 然而知识图谱中的知识不能作为完备的知识, 尤其是在大规模知识图谱上下文中, 需要不

断补充其中的实体关系进行知识图谱补全^[2]。同时, 现有的知识图谱均以三元组的形式表示知识, 这限制了知识必须包含头实体、关系和尾实体三部分, 但实际上三元组并不是适合所有知识的表示方式。例如管理员想查找某学生的导师 A 的博士论文, 但是已有知识库中并没有导师 A 的博士论文, 必须得到一个不完全的三元组 <A, 博士论文, ?>, 要补全这个三元

基金项目: 国家自然科学基金(61976032, 61371090, 61602076, 61702072); 国家社会科学基金(15BYY028); 辽宁省自然科学基金(20170540232, 20170540144, 20180540003); 大连外国语大学研究创新团队项目“计算语言学与人工智能创新团队”(2016CXTD06)。

作者简介: 程 涛 (1995—), 男, 硕士研究生, 主研方向为智能信息处理; 陈 恒, 副教授、博士研究生; 李冠宇, 教授、博士。

收稿日期: 2019-08-12 **修回日期:** 2019-11-08 **E-mail:** rabitlee@163.com

组,必须补充知识库。按照知识图谱的规则,这个不完全三元组无法加入现有的知识图谱,只能在以后通过扩充知识库的方式来补全成完整的三元组后加入。因此,研究保存不完全三元组的方法,实现知识图谱的动态补全,是一个具有研究意义的问题。

RDF三元组中关系本身的特性会对头尾实体进行限制。例如句子“Lennon was murdered by Champan outside the Dakota on Dec. 8. 1980”中蕴含的三元组为<Chapman, kill, Lennon>,其中关系“kill”决定了头尾实体都属于“人”这类实体,属于地理位置这类的实体“Dakota”则可以被排除。

在多数知识图谱中,关于实体有简明的描述^[3],这些丰富的语义信息对实体非常重要。以FreeBase中的三元组(William Shakespeare, book/author/works_written, Romeo and Juliet)为例,头实体“William Shakespeare”含有描述“William Shakespeare was an English poet, playwright, and actor...”,尾实体“Romeo and Juliet”含有描述“Romeo and Juliet is a tragedy written by William Shakespeare early in his career...”,这些描述对三元组的补全具有重要作用。

本文提出一个运用实体描述的知识图谱补全方法,即基于半边的多层卷积模型(Half-edged based Multilayer Convolution Model, HMCN)。该模型利用实体描述和关系自身可以对实体类别进行限制的特性组成半边,以保存缺失头实体或尾实体的不完全三元组。在此基础上,通过将对应类别的半边相结合,减少输入三元组的数量,再把候选的不完全三元组代入卷积神经网络(Convolutional Neural Network, CNN)进行知识图谱补全。

1 相关工作

1.1 知识图谱补全

针对知识图谱补全问题,研究者提出了较多的方法^[4-6],如基于结构与文本联合表示的知识图谱补全方法^[4]和基于双层随机游走的关系推理算法^[6]等。学习知识嵌入的代表模型有距离模型、单层神经网络模型、能量模型、双线性模型、矩阵分解模型和翻译模型等^[7]。张量神经网络模型^[8]和基于文档特征的向量空间模型^[9]在现有模型中加入多层神经网络来进行知识图谱的补全。文献[10-11]通过类比实验发现,词向量空间的平移不变现象普遍存在于词汇的语义关系和句法关系中。对此,文献[12]进一步提出翻译模型TransE。然而该模型解决一对一关系的问题效果较好,但无法应对一对多、多对一和多对多关系的问题。文献[13]提出TransH模型,通过将关系建模为超平面,并将 h 和 t 投影到关系特定超平面,从而让实体在不同的关系中扮演不同的角色,以解决TransE存在的问题。文献[14]提出的TransR模型在不同的语义空间中模拟实体和关系,

并在学习嵌入时将实体从实体空间映射到关系空间。文献[15]则提出一种基于多步关系路径的表示学习模型PTransE。文献[16]提出的DSKG模型利用循环神经网络实现了知识图谱的补全和实体预测。文献[17]提出的DKRL模型用CBOW和卷积神经网络把实体的描述运用到三元组的补全中。文献[18]提出的ProjE模型关注实体与关系之间的联系,不需要预训练,参数规模小,并且预测能力强。文献[19]提出的ConMask模型先以计算实体和关系语义相似度的方法对错误的三元组进行屏蔽,再进行补全工作,大幅减少了训练时间。

1.2 半边图

文献[20]介绍了半边原理及其在知识图谱中的应用。该文指出半边是一个基本概念,其将资源视为顶点和资源属性信息表示为半边,不同顶点的半边依据可结合性而结合成为边,表示资源属性之间的关联关系。此后,文献[21-22]介绍了半边在本体和因子图中的应用,文献[23]则证明了半边图在大规模图形数据库相似性搜索和高性能查询处理方面的优势。

2 基于半边的多层卷积模型

本文提出基于半边的多层卷积模型HMCN。该模型通过使用实体和关系约束组成两种半边,并根据结合规则使之相互结合,得到候选三元组集。对于需要补全的缺失三元组,在候选三元组集中查找对应的候选三元组代入CNN进行补全。本文模型的整体结构如图1所示,其利用两个两层的卷积神经网络,分别对头实体的描述和尾实体的描述进行训练。在此基础上,将训练得到的实体描述表示与实体结构表示相结合,使用TransE模型进行实体和关系的补全。

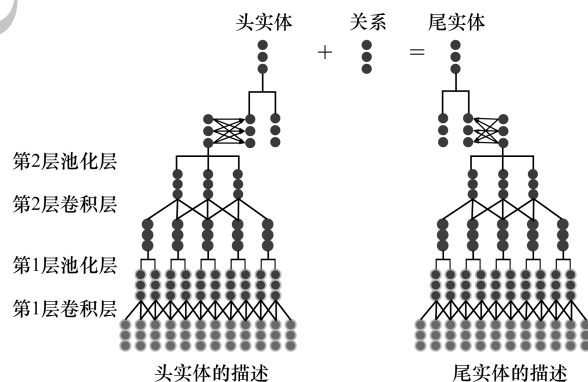


图1 HMCN模型结构

Fig.1 HMCN model structure

2.1 半边的定义

以图的顶点表示实体,以边表示关系,是知识表示的常用方法。这种方法将顶点和边作为不可分割的原子概念。半边即是针对这一问题提出的概念,其定义^[17]如下:

定义 1(半边) 半边是描述资源属性的概念,资源有若干属性,对应于一个顶点有若干个半边。半边具有以下特点:

1) 一个半边属于某个顶点,且分为不同的半边类型。半边类型由资源特征属性决定。

2) 半边之间相互独立,即属性之间相互独立,但半边与其他的半边可以相互结合。资源特征属性之间的关联关系,表示何种类型的两个半边可以结合,称为半边结合类型。

3) 每个半边有一个数值性度量值,称为半边权值,一般取大于 0 的实数,与资源属性的载荷情况成反比,即半边权值越大,资源属性的载荷越小。

图 2 展示了常见半边的形态。

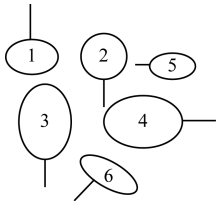


图 2 半边示意图

Fig.2 Schematic diagram of half-edge

参考半边的定义,结合知识图谱中三元组的特性,HMCM 所使用的三元组半边定义如下:

定义 2(三元组半边) 三元组半边为知识图谱补全所使用的特殊半边,分为头结点半边和尾结点半边两种,每个顶点均只有一个半边与其相连。三元组半边符号定义如下:

1) 头结点半边 $HE1 = \{Hh, Rh, csm\}$ 。其中: $Hh = \{h, W_h\}$ 代表顶点,由三元组中的头实体 h 和该实体的描述 W_h 组成; $Rh = \{r, rs\}$ 代表与顶点相连的半边,由三元组中的关系 r 和该关系所连接的头尾结点类别 rs 组成; csm 为实体描述与关系的相似度得分。

2) 尾结点半边 $HE2 = \{Tt, Rh, csm\}$ 。其中: $Tt = \{t, W_t\}$ 代表顶点,由三元组中的尾实体 t 和该实体的描述 W_t 组成; Rh 与 csm 定义与头结点半边中相同。

使用上述两种半边,即可得到 HMCM 所使用的半边集 $HE = \{HE1, HE2\}$ 。

2.2 半边的实现与结合

2.2.1 半边的实现

给定三元组 $(h, r, t) \in T$, 其中, $h, t \in E$ 代表实体; $r \in R$ 代表关系, E 是实体集, R 是关系集, T 是训练集。

定义实体类别 $S = \{s_1, s_2, \dots, s_n\}$, 其中每个实体 h (或 t) 都有自己的类别 $\{s_i, s_{i+1}, \dots, s_j\}$, 或是人, 或是位置等。对关系 r 追加候选实体类别 $rs = \{s_h, s_t\}$, 其中, s_h 是关系 r 所连接的头实体类别, s_t 是其尾

实体类别。例如关系 $father$ 的候选实体类 $rs = \{s_1, s_1\}$, s_1 为生物类的实体集, 这可以解释为生物的父亲一定是生物。因此, 那些逻辑上错误的三元组 (比如某人的父亲是某个地理位置) 就可以被过滤掉。

最后处理实体的描述。先从描述文本中删除所有停用词, 再标记所有短语并将这些短语视为单词。这些描述单词的集合定义为 $W_h = [w_1, w_2, \dots, w_i]$, 即词向量 w_i 是实体 h 的一个描述单词。

通过以上符号可以定义三元组的两种半边 HE1 和 HE2, 如式 (1) 和式 (2) 所示:

$$HE1 = \{h, r, rs, W_h, csm\} \quad (1)$$

$$HE2 = \{t, r, rs, W_t, csm\} \quad (2)$$

其中, 实体 h (或 t) 和该实体的描述 W_h (或 W_t) 为顶点, 关系 r 为半边, 候选实体类 rs 和 csm 为半边的权值。 csm 是实体的描述与关系相似度得分, 其定义如式 (3) 所示:

$$csm = \frac{1}{i} \sum_{x=1}^i \left(a \frac{w_x * r}{\|w_x\| * \|r\|} \right) \quad (3)$$

其中, a 是大于 1 的常数, 用于缓和求平均值所造成的相似度差距缩小的现象。

2.2.2 半边的结合

设实体集 E 中的两个实体 h 和 t , 若 $h \in HE1$, $t \in HE2$, $HE1.r = HE2.r$, 并满足式 (4):

$$|HE1.csm - HE2.csm| > c \quad (4)$$

其中, c 为设定阈值。三元组 $\langle h, r, t \rangle$ 是一个补全的候选三元组, 由此得到候选三元组的集合 HRT, 如式 (5) 所示:

$$HRT = \left\{ \begin{array}{l} \langle h_i, r_j, t_k \rangle \mid h_i \in HE1, t_k \in HE2 \\ HE1.r = HE2.r = r_j \\ |HE1.csm - HE2.csm| > c \end{array} \right. \quad (5)$$

HRT 将被用于进行实体和关系预测。

2.2.3 实体预测与关系预测

对于实体预测, 即缺失一个头实体或尾实体的三元组, 首先使用其中的实体和关系组成半边, 在 HRT 中选择满足该半边的候选三元组, 再将这些候选三元组代入 HMCM 进行补全。

同理, 对于关系预测, 可在 HRT 中得到同时含有已知头尾实体的三元组集合, 再使用 HMCM 进行补全。

2.3 卷积神经网络

卷积神经网络 (CNN) 是一种广泛应用于图像的高效模型, 其对某些自然语言处理任务和关系分类任务非常有效^[17]。同时, CNN 也可以用于关系分类任务^[24-25]。鉴于实体描述词汇量大、信息环境复杂、推理规则不明确等特点, 本文模型需要一个具有良好的容错能力且并行处理能力和自学习能力、运行速度快、自适应能力强的模型来进行训练, 而卷积神经网络正符合此需求。

参考 DKRL 模型, HMCM 使用两层的卷积神经网络, 第1层池化层为 max-pooling 层, 第2层池化层为 mean-pooling 层, 具体描述如下:

在卷积层中, 设 Z_i 为第 i 个卷积层的输出, X_i 为第 i 个卷积层的输入, 大小为 k 的卷积窗滑过 X_i 中的输入向量以得到 $X_i(k)$ 。在第1个卷积层中, X_1 是实体的描述向量 (x_0, x_1, \dots, x_n) 。因此, 卷积层的第 i 层输出向量如式(6)所示:

$$Z_i = \sigma(W_i X_i + b_i) \quad (6)$$

其中, $W_i \in \mathbb{R}^{n_2 \times n_1}$ 是第 i 个卷积层所有输入向量的卷积核, b_i 是偏差, $n_1 = k \times n$, n 为输入向量的维数, n_2 是输出向量的维数, σ 是激活函数, 此处选择 tanh。

模型的第1层输入为头实体和尾实体对应的描述单词集 W_h 和 W_t , 即 $X_1 = W_h$ (或 $X_1 = W_t$)。

卷积层之后使用池化层来缩小 CNN 的参数空间和滤波器噪声。卷积层的输出向量被分成多个大小为 n_L 的非重叠窗口。对于第1层池化层, 选取每个窗口的最大值以构成新的向量。 n_L -max-pooling 得到 n_L 大小的窗口内输入向量中最大的特征值, 如式(7)所示:

$$X_2 = \max(Z_{n_L \times i}, Z_{n_L \times (i+1)}, \dots, Z_{n_L \times (i+1)-1}) \quad (7)$$

n_L -max-pooling 可以缩小 n_L 次特征表示的大小, 降低 CNN 编码器的复杂性和参数学习的成本。然而, 一些描述非常复杂, 不同的句子可能表示信息的不同方面, 仅使用 max-pooling 会导致大量的信息丢失。对此, 第2个池化层使用 mean-pooling 以减少信息丢失, 使所有包含不同本地信息的输入向量都能对最终实体嵌入有贡献, 并可以在反向传播期间进行更新, 所得特征值如式(8)所示:

$$X_3 = \sum_{i=1}^m \frac{Z_{2i}}{m} \quad (8)$$

最终结果通过一层全连接神经网络与第2层池化层相连, 输出结果为与实体结构向量同维的向量, 记为实体的描述向量。

使用卷积神经网络训练得到的实体描述表示与实体结构表示组合形成最终的实体表示 h_c, t_c 。基于描述的实体表示记为 h_d, t_d , 基于结构的实体记为 h_s, t_s , 则 h_c 和 t_c 的计算公式如式(9)和式(10)所示:

$$h_c = \frac{h_s + h_d}{2} \quad (9)$$

$$t_c = \frac{t_s + t_d}{2} \quad (10)$$

最后使用 TransE 模型进行实体和关系预测, 如式(11)所示:

$$S = \|h_c + r - t_c\| \quad (11)$$

由于已经使用半边筛选了实体的类别, 此处实体的输入量仅为某个类中实体的数量, 因此可缩短运行时间。

2.4 损失函数

选择 TransE 的损失函数作为训练目标, 如式(12)所示:

$$L = \sum_{(h,r,t) \in T} \sum_{(h',r',t') \in T'} \max\{\gamma + \|h_c + r - t_c\| - \|h'_c + r' - t'_c\|, 0\} \quad (12)$$

在式(12)中, γ 是边际超参数, $\gamma > 0$, T' 是训练集 T 的负例的集合, 如式(13)所示:

$$T' = \{(h', r, t) \mid h' \in E \cup (h, r', t) \mid r' \in R \cup (h, r, t') \mid t' \in E\} \quad (13)$$

T' 为三元组中头实体、尾实体或关系被另一个实体或关系随机替换所组成的错误三元组集。如果三元组已经在 T 中, 则不会将其视为负样本。由于 h 和 t 有两种表示(基于结构的表示和基于描述的表示), 因此在基于边距的得分函数中, 实体的表示是基于结构表示和基于描述表示的平均值。

2.5 算法描述

HMCM 算法伪代码如下:

算法1 HMCM 算法

```

输入 三元组  $T = \{(h, r, t)\}$ , 每个实体对应的描述  $W_h$ 
for e in E: // 在实体和关系之间组合, 形成半边
  for r in R:
    if (e.s = rs, sh):
      csm = a * (Wh * r) / (||Wh|| * ||r||); HE1.append(h, r, csm)
    if (e.s = rs, st):
      csm = a * (Wt * r) / (||Wt|| * ||r||); HE2.append(t, r, csm)
    else: continue
  for i in HE1: // 按照结合规则进行半边的结合, 利用c
    // 进行筛选
    for j in HE2:
      if (|i.csm - j.csm| > c): HRT.append(<h, r, t>)
    for <h, r, t> in HRT: // 实体描述代入CNN得到最终
      // 的实体向量
      hd = Wh → CNN; hc = (hs + hd) / 2
      td = Wt → CNN; tc = (ts + td) / 2
      (hc, r, tc) → TransE // 最终结果代入 TransE 模型

```

3 实验与结果分析

本文实验的数据集使用 FB15K 和 WN18, 这两个数据集均是从典型的大型知识库中提取的数据集, 可用于评估知识图谱补全模型。为确认每个实体都能进行基于描述表示学习的训练, 从数据集中删除了部分实体和包含这些实体的所有三元组。处理后的 FB15K 和 WN18 数据集的数据分布如表1所示。其中: FB15K 训练集有 472 860 个三元组、14 904 个实体和 1 341 个关系, 测试集有 57 303 个三元组; WN18 中训练集有 140 229 个三元组、39 852 个实体和 18 个关系, 测试集有 49 276 个三元组。

表 1 FB15K 和 WN18 的数据分布
Table 1 Data distribution of FB15K and WN18

数据集	关系	实体	训练集	无效训练集	测试集
FB15K	1 341	14 904	472 860	48 991	57 803
WN18	18	39 852	140 229	47 568	49 276

训练模型采用多次实验来寻找最佳的参数设置,其中,实体和关系维度 n 分别为 50、80 和 100,学习率 λ 分别为 0.000 5、0.001 0 和 0.002 0,边缘参数 γ 分别为 0.5、1.0、1.5 和 2.0。第 1 个池化层使用 4-max-pooling。

使用正确实体的平均等级和排在前 10 名(实体)或第 1 名(关系)的正确率这两种指标对模型性能进行评价,同时参照文献[13],遵循 Raw 和 Filter 2 个评估标准并对比 CBOW 模型和 CNN 模型,得到实体预测和关系预测的比较结果,如表 2 和表 3 所示,表中数据均以 FB15K/WN18 的形式列出。

表 2 实体预测结果对比
Table 2 Comparison of entity prediction results

模型	平均等级		Hits@ 10/%	
	Raw	Fliter	Raw	Fliter
TransE	210/232	119/223	48.5/87.2	66.1/77.3
DKRL(CBOW)	236/253	151/249	38.3/76.1	51.8/59.8
DKRL(CNN)	200/219	113/220	44.3/84.6	57.6/65.2
DKRL(CNN) + TransE	181/205	91/197	49.6/87.9	67.4/78.5
HCMC	173/191	82/185	51.9/90.3	69.5/81.7

表 3 关系预测结果对比
Table 3 Comparison of relation prediction results

模型	平均等级		Hits@ 10/%	
	Raw	Fliter	Raw	Fliter
TransE	2.91/1.81	2.53/1.67	69.5/82.6	90.2/92.1
DKRL(CBOW)	2.85/1.73	2.51/1.63	65.3/78.3	82.7/84.9
DKRL(CNN)	2.91/1.86	2.55/1.62	69.8/83.9	89.0/90.7
DKRL(CNN) + TransE	2.41/1.45	2.03/1.41	69.8/84.4	90.8/93.6
HCMC	2.45/1.47	2.05/1.39	69.6/85.1	90.1/92.9

可以看出,HCMC 与 DKRL(CNN) + TransE 在平均等级和 Hits@ 10 上显著优于 TransE 和 CBOW。在实体预测方面,HCMC 略优于 DKRL,而在关系预测方面,DKRL 模型相对占优。这表明在基于结构补全方面,HCMC 与 DKRL 都没有优于 TransE。这是由于 TransE 在结构补全方面较为优秀,而 DKRL 和 HCMC 都是在 TransE 的基础上进行优化,并没有对其进行较大改进,因此,相比之下优势并没有非常大,但相较于 DKRL,HCMC 对实体和关系的组合进行了一定限制,提高了精确度,因此,在实体预测方面 HCMC 更占优。而对于关系预测,HCMC 对

结构与描述做了平均处理,削弱了关系描述中隐藏的语义对于关系预测的作用,导致预测关系的能力变弱。同时由于在半边的结合规则中,关系是非常重要的的一环,缺失关系对半边的结合精确度有所影响,因此 HCMC 的预测能力与 DKRL 相差无几。为此进行改进,后续将从实体的描述中发掘头尾实体之间的联系来加强关系预测的能力,弥补其不足。

从训练集中选取 100 000 个数据对 DKRL 和 HCMC 分别进行测试,记录模型运行时间,如图 3 所示。可以看出,当数据量较少时,HCMC 的效率略低于 DKRL,而随着数据量不断增长,HCMC 的运行效率逐渐比 DKRL 高。这是因为 HCMC 将三元组的范围扩展为半边,这虽然能发现潜在语义关系,但也一定程度上扩大了搜索范围,因此,当数据量较少时,其补全效率比 DKRL 略低,而随着数据量的增长,HCMC 中半边的作用逐渐体现,数据过滤效果逐渐增强,因此,运行时间也较 DKRL 越来越少。

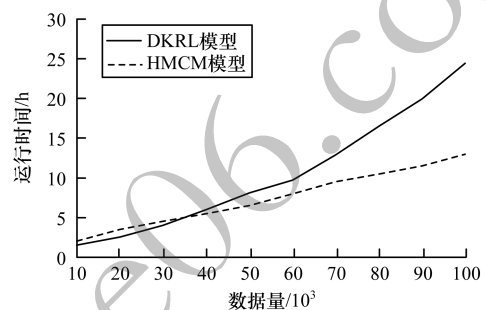


图 3 DKRL 模型与 HCMC 模型的运行时间对比
Fig. 3 Comparison of running time by DKRL model and HCMC model

此外,HCMC 把不完全的三元组以半边的形式加以暂存,这让只能通过补充知识库才能进行补全的三元组在后续补全工作中可以使用现有的半边来加快补全进程,避免了不必要且重复的遍历工作,从而实现知识图谱的动态补全。

4 结束语

本文提出 HCMC 模型,通过引入半边优化具有实体描述的知识图谱学习,使用深度卷积神经网络提取实体描述的语义,并将描述与结构相结合进行知识图谱补全。实验结果表明,该模型在实体预测和关系预测中均能达到预期目标,并且在知识图谱的动态补全方面具有优势。但 HCMC 模型仅考虑表示学习的实体描述,而未将其他类型的文本信息运用到模型中。此外,模型中虽然引入了半边来提高补全的精度,但是人工标记关系和实体分类依然耗时耗力。针对以上不足,下一步将对模型进行优化,完成实体的自动分类,同时对关系的约束进行定义,进一步探索时变半边图与知识图谱结合的方法。

参考文献

- [1] SINGHAL A. Introducing the knowledge graph: things, not strings [EB/OL]. [2019-07-08]. <http://googleblog.blogspot-spot.ie/2012/05/introducing-knowledge-graph-things-not.html>.
- [2] XU Zenglin, SHENG Yongpan, HE Lirong, et al. Review on knowledge graph techniques[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4): 589-606. (in Chinese)
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.
- [3] WANG Hui, YU Bo, HONG Yu, et al. Web information extraction system based on knowledge graph [J]. Computer Engineering, 2017, 43(6): 118-124. (in Chinese)
王辉, 郁波, 洪宇, 等. 基于知识图谱的 Web 信息抽取系统[J]. 计算机工程, 2017, 43(6): 118-124.
- [4] BAO Kaifang, GU Junzhong, YANG Jing. Knowledge graph completion method based on jointly representation of structure and text[J]. Computer Engineering, 2018, 44(7): 205-211. (in Chinese)
鲍开放, 顾君忠, 杨静. 基于结构与文本联合表示的知识图谱补全方法[J]. 计算机工程, 2018, 44(7): 205-211.
- [5] WANG Fei, LIU Jingping, LIU Bin, et al. Survey on construction of code knowledge graph and intelligent software development[J]. Journal of Software, 2020, 31(1): 47-66. (in Chinese)
王飞, 刘建平, 刘斌, 等. 代码知识图谱构建及智能化软件开发方法研究[J]. 软件学报, 2020, 31(1): 47-66.
- [6] LIU Qiao, HAN Minghao, JIANG Liuyi, et al. Two-tier random walk based relational inference algorithm [J]. Chinese Journal of Computers, 2017, 40(6): 1275-1290. (in Chinese)
刘峤, 韩明皓, 江浏伟, 等. 基于双层随机游走的关系推理算法[J]. 计算机学报, 2017, 40(6): 1275-1290.
- [7] LIU Zhiyuan, SUN Maosong, LIN Yankai, et al. Knowledge representation learning: a review [J]. Journal of Computer Research and Development, 2016, 53(2): 247-261. (in Chinese)
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.
- [8] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. New York, USA: ACM Press, 2013: 926-934.
- [9] NEELAKANTAN A, ROTH B, MCCALLUM A. Compositional vector space models for knowledge base completion [EB/OL]. [2019-07-08]. <https://arxiv.org/pdf/1504.06662.pdf>.
- [10] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[EB/OL]. [2019-07-08]. <https://arxiv.org/pdf/1301.3781.pdf>.
- [11] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality[C]//Proceedings of International Conference on Neural Information Processing Systems. [S. l.]: Curran Associates Inc., 2013: 3111-3119.
- [12] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data [C]//Proceedings of International Conference on Neural Information Processing Systems. [S. l.]: Curran Associates Inc., 2013: 2787-2795.
- [13] WANG Zhen, ZHANG Jianwen, FENG Jianlin, et al. Knowledge graph embedding by translating on hyperplanes[C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2014: 1-5.
- [14] LIN Yankai, LIU Zhiyuan, SUN Maosong, et al. Learning entity and relation embeddings for knowledge graph completion [C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2015: 1-5.
- [15] LIN Yankai, LIU Zhiyuan, LUAN Huanbo, et al. Modeling relation paths for representation learning of knowledge bases [EB/OL]. [2019-07-08]. <https://arxiv.org/pdf/1506.00379v2.pdf>.
- [16] GUO Lingbing, ZHANG Qingheng, GE Weiye, et al. DSKG: a deep sequential model for knowledge graph completion [C]//Proceedings of China Conference on Knowledge Graph and Semantic Computing. Berlin, Germany: Springer, 2018: 65-77.
- [17] XIE Ruobing, LIU Zhiyuan, JIA Jia, et al. Representation learning of knowledge graphs with entity descriptions[C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. New York, USA: ACM Press, 2016: 2659-2665.
- [18] SHI B, WENINGER T. ProjE: embedding projection for knowledge graph completion [C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2017: 1236-1242.
- [19] SHI B, WENINGER T. Open-world knowledge graph completion [C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. [S. l.]: AAAI, 2018: 1-5.
- [20] CHEN Tuanqiang, WANG Zhengming. Time-varying half-edge graph: a novel topology model for attributed relationships of Internet resources [J]. Journal of Computer Applications, 2007, 27(3): 512-513. (in Chinese)
陈团强, 王正明. Internet 资源属性关系的新拓扑模型——时变半边图 [J]. 计算机应用, 2007, 27(3): 512-513.
- [21] LASTRA-DÍAZ J J, GARCÍA-SERRANO A, BATET M, et al. HESML: a scalable ontology-based semantic similarity measures library with a set of reproducible experiments and a replication dataset [J]. Information Systems, 2017, 66: 97-118.
- [22] HOFFMANN C, ROSTALSKI P. Linear optimal control on factor graphs—a message passing perspective-[J]. IFAC-PapersOnLine, 2017, 50(1): 6314-6319.
- [23] ZHAO Peixiang. Similarity search in large-scale graph data-bases [M]//ZOMAYA A Y, SAKR S. Handbook of big data technologies. Berlin, Germany: Springer, 2017: 507-529.
- [24] ZENG Daojian, LIU Kang, LAI Siwei, et al. Relation classification via convolutional deep neural network [C]//Proceedings of the 25th International Conference on Computational Linguistic. Dublin, Ireland: [s. n.], 2015: 2335-2344.
- [25] ZHANG Dongxu, YUAN Bin, WANG Dong, et al. Joint semantic relevance learning with text data and graph knowledge [C]//Proceedings of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality. [S. l.]: ACL, 2015: 32-40.