计算机科学
COMPUTER SCIENCE

# 融合实体描述的水泥制造领域知识图谱嵌入模型

周泓林, 宋华珠, 张娟

引用本文

周泓林, 宋华珠, 张娟. 融合实体描述的水泥制造领域知识图谱嵌入模型[J]. 计算机科学, 2024, 51(3): 251-256.

ZHOU Honglin, SONG Huazhu, ZHANG Juan. Knowledge Graph Embedding Model with Entity Description on Cement Manufacturing Domain [J]. Computer Science, 2024, 51(3): 251-256.

---

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221200080

# 融合实体描述的水泥制造领域知识图谱嵌入模型

周泓林　宋华珠　张　娟
武汉理工大学计算机科学与技术学院　武汉 430070
(zhouhonglin@whut.edu.cn)

**摘　要**　针对知识图谱嵌入模型在进行知识嵌入时大多缺乏对语义信息的考虑,不能很好地提取水泥制造领域专业性的实体语义信息问题,文中将实体描述文本加入到水泥制造领域知识图谱(CMFKG)的嵌入工作中,提出了融合实体描述的知识图谱嵌入模型(KGEED)。该模型采用 TransE 模型得到 CMFKG 结构信息的嵌入,采用基于 CNN 的实体描述嵌入模块获得 CMFKG 基于语义的嵌入,并用 CNN 对结构信息嵌入与语义信息嵌入的三元组进行融合,从而可以很好地考虑水泥制造领域知识图谱丰富的实体描述文本信息。经实验表明,该模型在水泥制造领域知识图谱的嵌入工作中取得了不错的效果。

**关键词**:水泥制造领域;知识图谱;实体描述嵌入;实体链接
**中图分类号**　TP181

## Knowledge Graph Embedding Model with Entity Description on Cement Manufacturing Domain

ZHOU Honglin,SONG Huazhu and ZHANG Juan
School of Computer Science and Technology,Wuhan University of Technology,Wuhan 430070,China

**Abstract**　To address the problem that many knowledge graph embedding models lack the consideration of semantic information when performing knowledge embedding and cannot extract the semantic information of entities specialized in cement manufacturing domain well. The entity description text is added to the embedding work of cement manufacturing domain knowledge graph (CMFKG),and the knowledge graph embedding with entity description model(KGEED) is proposed,which adopts the TransE model to get the embedding of structural information of CMFKG. The CNN-based entity description embedding module is used to obtain the semantic-based embedding of CMFKG,and the triples of structural information embedding and semantic information embedding are fused with CNN,so that the rich entity description text information of the knowledge graph in cement manufacturing domain can be well considered. Experiments show that the model achieves good results in the embedding work of the knowledge graph in cement manufacturing domain.

**Keywords**　Cement manufacturing domain,Knowledge graph,Entity description embedding,Entity link

## 1　引言

水泥是一种重要的建筑材料,对国家经济发展具有巨大的意义,而我国水泥产量居世界首位。问答系统与水泥制造领域的智能制造结合有利于推动产业链智能化,知识图谱问答在解析问句的基础上结合知识图谱信息得到问句的答案,但是,由于自然语言问句的复杂性和知识图谱的残缺性,问答准确度不高。而知识图谱推理技术可以从知识图谱中发现有价值的隐藏知识。知识图谱嵌入又称知识图谱表示学习,目的在于把知识图谱投射到连续向量空间中,同时保留知识图谱中蕴涵的一些重要的语义信息,合适的知识图谱嵌入方法能够有效解决水泥制造领域知识图谱数据稀疏问题,从而使推理的准确性得以提高。TransE[1]是最先提出来的一种简便可行的图谱嵌入方法,但它不能处理某些特定的关系,例如一对多、多对一、对称和传递关系,为了解决这些限制,TransH[2],TransR[3],TransD[4]相继被提出。此外,HAKE[5]建模语义层次结构,而不是基于关系模式建模在极坐标空间上。TransRHS[6]首先考虑关系层次结构(RHS)无缝嵌入,也有一些尝试更充分的相互作用的平移模型,如 TripleRE[7]和 InterHT[8]。但这些模型重点考虑三元组的结构信息,忽略其语义文本信息,导致其不能合理地解决在水泥制造领域中复杂关系的嵌入问题,嵌入效果尚有很大的改善余地。针对研究的问题,本文提出了融合实体描述的知识图谱嵌入模型(Knowledge Graph Embedding with Entity Description,KGEED)。该模型通过 TransE 模型来获得基于结构信息的嵌入,通过 CNN(Convolutional Neural Networks)编码器处理水泥制造领域知识图谱的实体描述文本,从而得到基于语义信息的嵌入,最后利用 CNN 的特性将实体语义向量和三元组结构向量进行融合,进而可以有效考虑图谱中实体描述的相关信息来辅助模型训练,最终完成融合实体描述文本

到稿日期:2022-12-12　返修日期:2023-06-08
通信作者:宋华珠(shuaz@whut.edu.cn)

的水泥制造领域知识图谱嵌入。

## 2 相关工作

具体地说，知识图谱推理技术分为3类：基于逻辑规则的推理、基于嵌入表示的推理和基于深度学习的推理[9]。

基于逻辑规则的方法是利用一些逻辑规则[10]与机器学习的统计特征来对其进行推理。Quecole等[11]提出了NELLs模型，该模型建立了一系列的学习概率规则来辅助进行推理。Jiang等[12]提出了一个NELL清理系统，允许知识库使用联合概率推理。Chen等[13]利用sql的一些思想来改进推理算法，将其知识进行补全，构建具有推理规则的知识库。

基于嵌入表示的知识图谱推理方法中使用最为广泛的就是TransE模型，鉴于它具有方便、简单以及成熟等诸多特点，大量的工作在它的基础上进行改进，该模型被普遍用在知识图谱问答中。而RotatE[14]提出将KG嵌入到复向量空间中来建模关系的对称或反对称。此外，KGEs中还利用了许多其他数学空间，如概率空间[15]、双曲空间[16]和球面空间[17]。

基于深度学习的知识图谱推理方法可以学习更深层次的特征，Dettmers等[18]提出了ConvE，卷积神经网络与KGR模型集成在一起，该模型可以扩展到大型知识图谱。ConvKB[19]通过删除重塑操作来扩展ConvE，并在事实中捕获全局和过渡特征，以实现更有信息的表达。此外，在InteractE[20]中设计了特征重塑、特征置换、循环卷积等新颖操作，用于处理更多KG内的交互。同时，ConEx[21]将仿射变换和复值嵌入上的厄米内积与卷积运算相结合，具有良好的表达性。

## 3 研究方法

### 3.1 融合实体描述的知识图谱嵌入模型框架

基于翻译的知识图谱嵌入方法可以很好地把握三元组的内在特性，而且计算效率很高，但是这种计算方式单纯依靠数据驱动，应用在水泥制造领域则不能很好地提取该领域专业性的实体语义信息。水泥制造领域知识图谱中的实体蕴含着大量的实体描述文本，这种信息对该领域知识图谱嵌入具有很大的帮助。

实体描述文本对传统知识图谱嵌入方法具有良好的补充功能，但目前的引入实体描述文本的模型(Description-Embodied Knowledge Representation Learning，DKRL)[22]存在描述文本信息嵌入不够充足的问题。针对这个问题，本章首先利用卷积神经网络编码器来进一步理解描述文本，并利用隐藏在词序中的内部文本信息，再使用TransE模型获取实体基于结构的表示。最后利用CNN对两种表示加以融合，进而更有效地使用描述文本，提升水泥制造领域知识图谱嵌入效果。因此，本文提出了融合实体描述的知识图谱嵌入模型(Knowledge Graph Embedding with Entity Description，KGEED)。融合实体描述文本的水泥制造领域知识图谱嵌入模型的框架如图1所示。

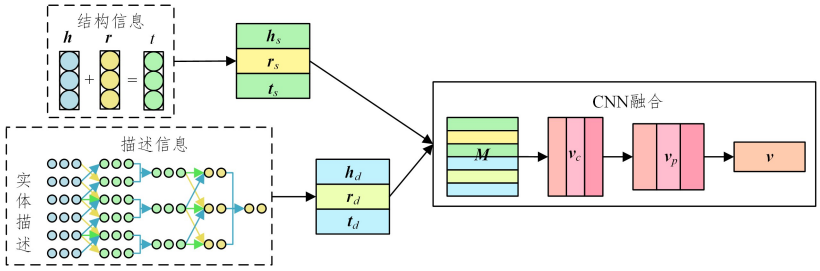


图1 融合实体描述文本的水泥制造领域知识图谱嵌入模型框架

Fig. 1 Model structure of KGEED on cement manufacturing domain

### 3.2 融合实体描述的知识图谱嵌入模型

#### 3.2.1 知识图谱嵌入算法

本章提出的KGEED算法的流程如算法1所示。

**算法1** 融合实体描述的知识图谱嵌入算法

输入：水泥制造领域知识图谱，水泥制造领域实体描述文本

输出：知识图谱的三元组向量表示

Begin

1. 初始化水泥制造领域知识图谱的三元组；
2. 通过TransE训练得到($\mathbf{h}_s$，$\mathbf{r}_s$，$\mathbf{t}_s$)；
3. 通过CNN编码器训练得到($\mathbf{h}_d$，$\mathbf{r}_d$，$\mathbf{t}_d$)；
4. 通过CNN融合($\mathbf{h}_s$，$\mathbf{r}_s$，$\mathbf{t}_s$)和($\mathbf{h}_d$，$\mathbf{r}_d$，$\mathbf{t}_d$)；
5. 将向量($\mathbf{h}_s$，$\mathbf{r}_s$，$\mathbf{t}_s$)和($\mathbf{h}_d$，$\mathbf{r}_d$，$\mathbf{t}_d$)进行组合操作，得到维度为k×3的矩阵$\mathbf{M}$；
6. 将矩阵$\mathbf{M}$通过卷积层得到三元组的$\tau$个特征图$v_c$；
7. 对矩阵$\mathbf{M}$使用ReLU激活函数；
8. 对$\mathbf{v}_c$进行最大池化运算得到$\tau$个特征图$\mathbf{v}_p$；
9. 将$\mathbf{v}_p$连接成一个向量$\mathbf{v}$；
10. 输出三元组向量表示；

End

#### 3.2.2 基于CNN的实体描述嵌入模块

为了在水泥制造领域实体的描述文本中尽可能完整地获取到重要的领域信息，充分利用实体描述文本，从而降低领域信息中产生的噪声，本章使用卷积神经网络作为水泥制造领域实体描述文本嵌入的编码器。卷积神经网络所形成的文本不但能够挖掘领域文本中的局部信息和字序信息，同时比较了循环神经网络等其他模型，还具备参数小和训练速度快的优点，因此可以更有效地满足水泥制造领域知识图谱的需要。

采用CNN编码器对水泥制造领域实体描述文本进行嵌入操作，获得三元组基于描述的向量。CNN编码器有5层，对领域实体的整个描述文本进行预处理后将结果作为输入，经过两次卷积与两次池化后获得文本的领域特征，最后输出该实体的基于描述的表示。基于CNN的实体描述文本编码的流程如图2所示。

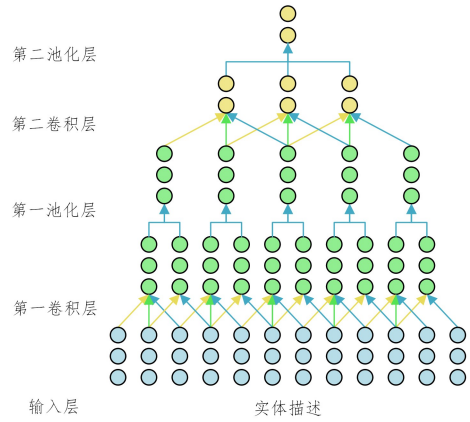


图 2 基于 CNN 的实体描述文本编码流程

Fig. 2 CNN-based text encoding process for entity description

1)输入层

首先对水泥制造领域实体描述文本进行初始化，具体来说，就是首先去除实体表述文本中的停用词，而此处的停用词表则选择了水泥制造领域知识图谱来构建。训练过程中使用的停用词表，之后将由 word2vec 训练的词嵌入向量作为 CNN 编码器的输入。

用 $\boldsymbol{D}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_n\}$ 代表水泥制造领域实体描述文本的输入，其中，$\boldsymbol{x}_i(i\in\{1,2,\cdots,n\})$ 代表描述文本中单词的次序，$n$ 代表描述文本中单词的数量。经过输入层的初始化，可以得到描述文本中的单词向量为 $\boldsymbol{x}_i$，将每个词向量按次序排列，即可得到实体描述的初始化形式 $\boldsymbol{D}=\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n\}$，$\boldsymbol{D}\in R^{d\times n}$，$d$ 代表 $\boldsymbol{x}_i$ 的维数。

2)卷积层

CNN 编码器有两个卷积层，将 $\boldsymbol{X}^{(m)}$ 设为第 $m$ 个卷积层的输入，将 $\boldsymbol{Z}^{(m)}$ 设为第 $m$ 个卷积层的输出。首先，采用一个尺寸为 $k$ 的窗口对 $\boldsymbol{X}^{(m)}$ 进行滑动处理，得到结果 $\boldsymbol{X}'^{(m)}$。特别是在第一层，$\boldsymbol{X}^{(1)}$ 是预处理后的描述，表示为一组向量 $\boldsymbol{D}=\{\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n\}$，$\boldsymbol{D}\in R^{d\times n}$ 就是第一层卷积层的输入。滑动窗口的过程如式(1)所示：

$$\boldsymbol{x}'^{(m)}_i=\boldsymbol{x}^{(m)}_{i:i+k-1}=[\boldsymbol{x}^{(m)\mathrm{T}}_i,\boldsymbol{x}^{(m)\mathrm{T}}_{i+1},\cdots,\boldsymbol{x}^{(m)\mathrm{T}}_{i+k-1}]^{\mathrm{T}} \tag{1}$$

其中，$\boldsymbol{X}^{(m)}$ 中的第 $i$ 个向量是在输入文本的对应窗口中连接 $k$ 个列向量得到的，因为描述文本的长度并不相同，在进行滑动窗口时输入的长度不是固定的，因此每个输入向量的末端添加零来填充向量，以确保窗口拥有相同的长度，从而方便模型训练。输出向量 $\boldsymbol{z}^m_i$ 的计算式如式(2)所示：

$$\boldsymbol{z}^m_i=\sigma(W^{(m)}\boldsymbol{x}'^{(m)}_i+\boldsymbol{b}^{(m)}_i) \tag{2}$$

其中，$W^{(m)}$ 为滑动窗口后第 $m$ 层卷积层的所有输入向量的卷积核；$\boldsymbol{b}^{(m)}_i$ 为偏置项；$\sigma$ 是激活函数。另外，补零操作不应对正向传播有任何影响，也不应在反向传播中有任何迭代。由此，实现统一输入文本的可变长度。

3)池化层

在每个卷积层后使用池化来缩小 CNN 的参数空间和噪声文本特征。由于描述文本被作为模型的输入，因此为不同的卷积层设置了不同的池化策略。

在第一个池化层中用尺寸为 $n$ 的非重叠窗口对卷积层的输出向量加以划分。在每个窗口中，选取特征映射的最大值生成新的向量，这个操作称为最大池化操作。使用最大池化是为了在一个大小为 $n$ 的窗口内、在输入向量的每个维度中确定最关键的特征值，其过程如式(3)所示：

$$\boldsymbol{X}_i=\max(\boldsymbol{z}^{(1)}_{n\times i},\cdots,\boldsymbol{z}^{(1)}_{n\times(i+1)-1}) \tag{3}$$

最大池化策略可以将特征图的尺寸减小为原来的 $1/n$，降低了模型的复杂性并减少了参数数量。

然而，水泥制造领域描述文本是复杂的，以至于描述中的不同句子可能有局部信息的不同含义，仅仅使用最大池化可能导致水泥制造领域信息的丢失。在这种情况下，对于最后一个池化层，在激活之前使用平均池化而不是最大池化来实现实体嵌入表示。平均池化操作如式(4)所示：

$$\boldsymbol{X}^{(3)}=\frac{1}{l}\sum_{i=1}^{l}\boldsymbol{z}^{(2)}_i \tag{4}$$

其中，$\boldsymbol{X}^{(3)}$ 是最终获得的基于描述的向量；$l$ 是平均池化层的输入矩阵长度；$\boldsymbol{Z}^{(2)}$ 是平均池化操作的输入矩阵。平均池化操作将输入矩阵按行相加后取平均值，考虑了所有的输入信息，最终得到三元组基于描述的嵌入。

所有包含不同局部信息的输入向量都可能对实体嵌入有影响，并可以在反向传播过程中进行更新。因为选择了不同的池化方式，模型能够应对可变长度的向量输入，并为每个实体获得固定长度的表示，而且不会造成太多的领域信息损失。

综上，水泥制造领域实体描述文本提供了丰富的领域信息，能发现领域实体之间的语义关联性，更好地实现水泥制造领域的知识图谱嵌入。

### 3.2.3 基于 CNN 的融合模块

卷积神经网络是一种广泛应用在图像处理上的有效模型，并且已经被证明在一些自然语言处理任务上是高效的，如关系提取和关系分类[23]。同时，CNN 也被用于知识图谱嵌入，能够捕获三元组结构中更深层次的特征[18]。由此可见，CNN 不仅可以学习语义特征，还可以学习结构特征。因此，利用 CNN 融合语义信息嵌入和结构信息嵌入，不同的语义和结构组合可以对应不同的卷积核，能够更好地融合结构信息和语义信息的嵌入。基于 CNN 的合模块实现过程如图 3 所示。

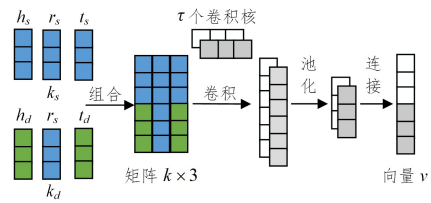


图 3 基于 CNN 的融合模块实现过程示意图

Fig. 3 Schematic of CNN-based fusion module implementation process

以图 3 为例，对基于 CNN 的融合模块实现过程进行描述，具体如下。

1)初始化

将水泥制造领域知识图谱中的一个三元组形式化地表示为 $(h,r,t)$，基于 CNN 的融合模块的任务是得到实体和关系的综合向量，并以此完成对水泥制造领域知识图谱的嵌入，因此首先对水泥制造领域知识图谱的三元组数据进行初始化

处理。KGEED模型分为两个部分的数据初始化，第一部分是通过TransE翻译模型得到基于结构的嵌入表示($\boldsymbol{h}_s$,$\boldsymbol{r}_s$,$\boldsymbol{t}_s$)，向量维数为$k_s$；第二部分则是通过基于CNN的实体描述嵌入模块得到基于描述的嵌入表示($\boldsymbol{h}_d$,$\boldsymbol{r}_d$,$\boldsymbol{t}_d$)，向量维数为$k_d$。对于三元组$(h,r,t)$来说，基于描述的嵌入表示$\boldsymbol{h}_d$与$\boldsymbol{t}_d$共享关系$r$，因此有$\boldsymbol{r}_d=\boldsymbol{r}_s$，$k_d=k_s$，相同维数的向量有利于三元组的语义信息嵌入和结构信息嵌入的互相学习。

2)组合

对于三元组$(h,r,t)$，初始化得到基于三元组结构与基于实体描述文本的向量后，通过组合得到$k\times3$的矩阵$\boldsymbol{M}=[\boldsymbol{v}_h,\boldsymbol{v}_r,\boldsymbol{v}_t]\in R^{k\times3}$，这里$k=2k_s$，$\boldsymbol{v}_h=\boldsymbol{h}_s\oplus\boldsymbol{h}_d$，$\boldsymbol{v}_r=\boldsymbol{r}_s\oplus\boldsymbol{r}_d$，$\boldsymbol{v}_t=\boldsymbol{t}_s\oplus\boldsymbol{t}_d$，$\oplus$代表将向量$\boldsymbol{a}\in R^{m\times l}$和向量$\boldsymbol{b}\in R^{m\times l}$连接成向量$\boldsymbol{c}\in R^{(m+n)\times l}$。$\boldsymbol{M}_i\in\boldsymbol{R}^{l\times3}$表示矩阵$M$的第$i$行。

3)卷积

由于组合后得到$k\times3$的矩阵$\boldsymbol{M}$，因此在卷积层上选择使用卷积核$\boldsymbol{w}\in R^{m\times l}$获得水泥制造领域知识图谱的三元组特征。对于具有维数相等的$\boldsymbol{v}_h$,$\boldsymbol{v}_r$和$\boldsymbol{v}_t$,$\boldsymbol{w}$能够得到三元组($\boldsymbol{v}_h$,$\boldsymbol{v}_r$,$\boldsymbol{v}_t$)在每一维度上的全局关系。$\boldsymbol{w}$在$\boldsymbol{M}$的每一行重复卷积从而得到特征图$\boldsymbol{v}_c=[v_1,v_2,\cdots,v_k]\in R^k$，其中$v_i=g(\boldsymbol{w}\times\boldsymbol{M}_{i,:}+b)$，$b\in R$代表偏置项，$g$是激活函数ReLU，激活函数能使矩阵得到三元组之间的非线性特征。

4)池化

对卷积获得的$\boldsymbol{v}_c=[v_1,v_2,\cdots,v_k]\in R^k$进行最大池化操作，从而获得($\boldsymbol{v}_h$,$\boldsymbol{v}_r$,$\boldsymbol{v}_t$)在相近维度上的最大特征，最终获得水泥制造领域三元组的特征图$\boldsymbol{v}_p=[v_1,v_2,\cdots,v_{k/2}]\in R^{k/2}$。

5)展平

$\boldsymbol{\Omega}$是卷积核的集合，$\tau$是卷积核的数量，即$\tau=|\boldsymbol{\Omega}|$，卷积运算后会得到$\tau$个特征图。在完成最大池化运算后，$\tau$个特征图将被连接成一个向量$\boldsymbol{v}=[\boldsymbol{v}_{p_1},\boldsymbol{v}_{p_2},\cdots,\boldsymbol{v}_{p_\tau}]\in R^{\tau k/2}$，与权重矩阵$\boldsymbol{w}\in R^{\tau k/2\times1}$做点积运算，就能获得三元组$(h,r,t)$的得分。展平运算的过程如式(5)所示：

$$f_r(h,t)=flatten(subsample(g([\boldsymbol{h}_s\oplus\boldsymbol{h}_d,\boldsymbol{r}_s\oplus\boldsymbol{r}_d,\boldsymbol{t}_s\oplus\boldsymbol{t}_d]*\boldsymbol{\Omega})))\times\boldsymbol{w} \tag{5}$$

其中，$\boldsymbol{w}$代表权值矩阵$\boldsymbol{w}\in R^{\tau k/2\times1}$；$*$代表卷积运算符；$g$代表激活函数ReLU；$subsample$代表最大池化运算；$flatten$代表将特征图展平成一个向量的操作。

## 4 水泥制造领域知识图谱嵌入实验

### 4.1 实验数据

通过对水泥制造领域知识图谱[24]中三元组的统计可知，实体共有4760个，关系共有287个，并且有4760个实体描述与每个实体对应。对水泥制造领域来说，因为实体描述文本数据，所以需要人工搜集，水泥制造领域实体描述的数据大多来源于维基百科以及百度百科，对于少数搜索不到的词条，则人工给出语义描述信息。水泥制造领域数据集(Cement Clinker Production Field，CCPF)的训练集包括7825个三元组，测试集有1052个三元组，验证集有972个三元组。

### 4.2 实验的设置和评价指标

1)实验参数设置

融合实体描述的知识图谱嵌入模型实验的硬件环境如下：CPU E5 2680 v2，内存16 GB，硬盘SSD 512 GB。实验软件环境如下：操作系统为Ubuntu 16.04，深度学习平台为Python3.6，pytorch1.10.0，cuda11.1。其他的主要参数如表1所列。

表1 KGEED模型的参数设置

Table 1 Parameter settings of KGEED model

| 参数 | 参数含义 |
|---|---|
| $ks=50$ | 基于结构信息的向量维数 |
| $kd=50$ | 基于描述信息的向量维数 |
| $\alpha=0.0005$ | Adam的学习率大小 |
| $\tau=200$ | 卷积核的个数 |
| $L2\_value=0.001$ | L2正则化系数 |
| $batch\_size=256$ | 批处理的大小 |
| $\gamma=3$ | TransE模型向量训练的间隔 |
| $num\_epochs=300$ | 模型迭代的总次数 |

2)实验评价指标

实验评价指标参考文献[1]，使用Mean Rank和Hits@$k$作为融合实体描述的知识图谱嵌入模型效果的评估指标。

### 4.3 实验结果与分析

根据4.2节对知识图谱嵌入模型效果的评价指标介绍，本节对实体预测的实验结果做出评估，下文将对评估指标上的实验结果展开讨论，并进行分析。

1)在水泥制造领域数据集(CCPF)上对比实验结果

在实体预测的评测中，评价指标Mean Rank值越低，同时Hits@10值越高，证明实验结果越好。在水泥制造领域数据集上，本节提出的融合实体描述的知识图谱嵌入模型与经典的知识图谱嵌入模型进行对比实验，每个模型进行了5次实验后取实验的平均值，实验结果如表2所列。

表2 实体预测实验结果

Table 2 Entity prediction experiment results

| Model | Mean Rank | | Hits@10/% | |
|---|---|---|---|---|
| | Raw | Filter | Raw | Filter |
| TransE[1] | 270 | 235 | 32.8 | 50.2 |
| ConvE[18] | 251 | 219 | 37.5 | 55.8 |
| DKRL(CBOW)[22] | 231 | 202 | 51.3 | 69.4 |
| DKRL(CNN)[22] | 219 | 188 | 46.5 | 64.4 |
| KGEED(本文) | 212 | 180 | 48.6 | 66.7 |

由实验结果可以看出，对于同一个模型，“Filter”下的实验结果总是比“Raw”下的实验结果好。这是因为“Filter”下的结果是对“Raw”的取值的再过滤，进一步表明水泥制造领域知识图谱中确实包含复杂的关系。

ConvE模型相比TransE模型在Mean Rank和Hits@10两种评价指标上的效果都更好，具体来看，Mean Rank指标下的最好情况的排序降低了19，Hits@10指标下最多提高了5.6%的占比。这是由于ConvE采用2d卷积神经网络实现知识图谱嵌入任务，表明了CNN比简单的平移距离模型TransE有更好的提取水泥制造领域知识图谱中三元组特征的能力。

DKRL方法取得了明显的提升，具体来看，Mean Rank指标下的最好情况排序降低了20，Hits@10指标下最多提高了13.8%的占比。这是因为ConvE和TransE模型都只考虑了三元组的结构特征，DKRL模型还在TransE模型实现结构

嵌入的基础上加入了实体描述信息。这证明在知识图谱嵌入过程中加入额外补充信息确实能够改善三元组的嵌入表示，从而能使事实三元组的实体平均排名更靠前。其中，DKRL(CNN)模型与 DKRL(CBOW)模型相比实验效果也有一定的改善，这表明了基于 CNN 的实体描述嵌入方法优于基于 CBOW 的实体描述嵌入方法，CNN 对实体描述信息的提取效果更好。

最后，本文提出的 KGEED 模型在 Mean Rank 和 Hits@10 两种评价指标上取得了最好的实验结果，与基线模型相比，KGEED 模型在评价指标上的效果都更好，在“Filter”情况下的 Mean Rank 指标取得最好的值 180，Hits@10 指标更是取得了最好的 66.7%。KGEED 模型比次优的 DKRL(CNN)模型在“Filter”情况下 Mean Rank 指标降低了 8 个名次的排序，在 Hits@10 指标下最多提高了 2.3%的占比。这表明基于 CNN 的结构和描述信息融合方法能更好地利用信息，从而证明了 KGEED 模型在水泥制造领域知识图谱嵌入任务中的优越性。由于其他较新的嵌入方法并未考虑语义信息，缺少对水泥制造领域专业性的语义信息的考虑，并不具有参考意义，因此不将其与本文进行对比。而如何将最新的结构嵌入方法与水泥熟料专业领域的描述信息融合，也是我们接下来要努力的方向。

2)KGEED 模型在不同 $k$ 取值的 Hits@$k$ 指标上的实验结果

本文提出的融合实体描述的知识图谱嵌入模型在 $k$ 取不同值时在水泥制造领域数据集上进行实体预测实验，实验结果的可视化如图 4 所示。

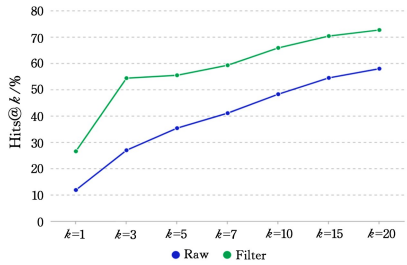


图 4 模型在不同 $k$ 取值的结果折线图

Fig. 4 Line graph of the model at different values of $k$

**结束语** 本文针对知识图谱嵌入模型大多未考虑语义描述信息的问题，提出了融合实体描述的知识图谱嵌入模型 KGEED，使用 TransE 获取水泥制造领域的知识基于结构信息的嵌入，使用基于 CNN 的实体描述嵌入模块获取实体基于描述的嵌入，最后利用 CNN 对两种嵌入表示加以融合，进而达到更有效使用描述文本，提升水泥制造领域知识图谱嵌入效果。最后，用链接预测实验来验证融合实体描述的知识图谱嵌入模型是否有效，且实验结果表明，KGEED 的确能改善水泥制造领域知识图谱三元组的嵌入效果。在未来的工作中，我们会考虑如何能够将水泥熟料领域的描述信息与最新的结构嵌入模型融合，使其具有更好的嵌入效果。

## 参 考 文 献

[1] 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：Curran Associates Inc，2013：2787-2795.

[2] WANG Z，ZHANG J，FENG J，et al. Knowledge graph embedding by translating on hyperplanes[C]// Proceedings of the AAAI Conference on Artificial Intelligence. New York：AAAI Press，2014：1112-1119.

[3] 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. New York：AAAI Press，2015，291-297.

[4] JI G，HE S，XU L，et al. Knowledge Graph Embedding via Dynamic Mapping Matrix[C]// Processing of the Association for Computational Linguistics & the International Joint Conference on Natural Language. New York：Association for Computational Linguistics，2015：687-696.

[5] ZHANG Z，CAI J，ZHANG Y，et al. Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction[C]// Proceedings of the AAAI Conference on Artificial Intelligence. New York：AAAI Press，2020：3065-3072.

[6] ZHANG F，WANG X，LI Z，et al. TransRHS：A Representation Learning Method for Knowledge Graphs with Relation Hierarchical Structure[C]// Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. New York：International Joint Conferences on Artificial Intelligence Organization，2020：2950-2956.

[7] YU L，LUO Z，LIU H，et al. Triplere：Knowledge graph embeddings via tripled relation vectors[J]. arXiv：2209.08271，2022.

[8] ZHU Y，LIU H，WU Z，et al. Representation Learning with Ordered Relation Paths for Knowledge Graph Completion[J]. arXiv：1909.11864，2019.

[9] MA R X，LI Z Y，CHEN C K，et al. A review of knowledge graph inference research[J]. Computer Science，2022，49(S1)：74-85.

[10] QU M，CHEN J，XHONNEUX L P，et al. RNNLogic：Learning Logic Rules for Reasoning on Knowledge Graphs[J]. arXiv：2010.04029，2020.

[11] QUECOLE F，DUARTE M C，HRUSCHKA E R. Coupling for Coreference Resolution in a Never-ending Learning System[J]. Journal of Information and Data Management，2018，9(2)：124-134.

[12] JIANG S，LOWD D，DOU D. Learning to Refine an Automatically Extracted Knowledge Base Using Markov Logic[C]// Proceedings of IEEE International Conference on Data Mining. New York：IEEE Press，2012：912-917.

[13] CHEN Y，WANG D Z. Knowledge expansion over probabilistic knowledge bases[C]// Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data. New York：Association for Computing Machinery，2014：649-660.

[14] CHEN W，XIONG W，YAN X，et al. Variational Knowledge Graph Reasoning[J]. arXiv：1803.06581，2018.

[15] NGUYEN D Q，VU T，NGUYEN T D，et al. Quatre：Relation-aware quaternions for knowledge graph embeddings[C]// Com-

panion Proceedings of the Web Conference. New York: Association for Computing Machinery, 2022: 189-192.
[16] NGUYEN D Q, VU T, NGUYEN T D, et al. Quatre: Relation-aware quaternions for knowledge graph embeddings[C]// Companion Proceedings of the Web Conference. New York: Association for Computing Machinery, 2022: 189-192.
[17] XU C, CHEN Y Y, NAYYERI M, et al. Temporal knowledge graph completion using a linear temporal regularizer and multivector embeddings[C]// Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Mexico: Association for Computational Linguistics, 2021: 2569-2578.
[18] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2d knowledge graph embeddings[C]// Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2018: 1-9.
[19] NGUYEN D Q, NGUYEN T D, NGUYEN D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network[J]. arXiv, 2017, 1712(02121): 1-7.
[20] VASHISHTH S, SANYAL S, NITIN V, et al. Interacte: Improving convolution-based knowledge graph embeddings by increasing feature interactions[C]// Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2020: 3009-3016.
[21] DEMIR C, NGOMO A C N. Convolutional complex knowledge graph embeddings [C]// Proceedings of European Semantic Web Conference. European: Springer Cham, 2021: 409-424.
[22] XIE R, LIU Z, JIA J, et al. Representation learning of knowledge graphs with entity descriptions[C]// Proceedings of the AAAI Conference on Artificial Intelligence. New York: AAAI Press, 2016: 2659-2665.
[23] NGUYEN T H, GRISHMAN R. Relation extraction: Perspective from convolutional neural networks[C]// Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing. New York: Association for Computational Linguistics, 2015: 39-48.
[24] CHEN D. Construction of Knowledge Map in Cement Clinker Production Field [D]. Wuhan: Wuhan University of Technology, 2020.

**ZHOU Honglin**, born in 1999, postgraduate. Her main research interests include knowledge graph and reinforcement learning.

**SONG Huazhu**, born in 1970, Ph.D, associate professor, master supervisor, is a senior member of CCF(No. 12201S). Her main research interests include artificial intelligent and data mining, semantic and knowledge abstraction.

(责任编辑:喻藜)