

一种基于改进向量投影距离的知识图谱表示方法



李鑫超 李培峰 朱巧明

苏州大学计算机科学与技术学院 江苏 苏州 215006

江苏省计算机信息处理技术重点实验室 江苏 苏州 215006

(20175227013@stu.suda.edu.cn)

摘要 表示学习在知识图谱推理中有着重要的研究价值,将知识库中的实体和关系用连续低维向量进行表示,可实现知识的可计算。基于向量投影距离的知识表示学习模型在面对复杂关系时有较好的知识表达能力,但在处理一对一简单关系时容易受到无关信息的干扰,并且在一对多、多对一和多对多等复杂关系上存在性能提升空间。为此,文中提出了一个基于改进向量投影距离的知识表示学习模型 SProjE,该模型引入自适应度量方法,降低了噪声信息的影响。在此基础上,通过进一步优化损失函数来提高复杂关系三元组的损失权重。该模型适用于大规模知识图谱的表示学习任务。最后,在标准知识图谱数据集 WN18 和 FB15K 上分析和验证了所提方法的有效性,基于链路预测任务的评测实验结果表明,相较于现有的模型和方法, SProjE 在各项性能指标上均取得了明显的进步。

关键词: 知识图谱;表示学习;自适应度量;链路预测

中图法分类号 TP391.1

Knowledge Graph Representation Based on Improved Vector Projection Distance

LI Xin-chao, LI Pei-feng and ZHU Qiao-ming

School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China

Provincial Key Laboratory for Computer Information Processing Technology, Suzhou, Jiangsu 215006, China

Abstract Representation learning is of great value in knowledge graph reasoning, which realizes the computability of knowledge by embedding entities and relationships into a low-dimensional space. The representation learning model based on vector projection distance has better ability of knowledge representation on complex relationships. However, the model is easily susceptible to irrelevant information, especially when dealing with one-to-one relationships, and it still has space to improve performance in representing one-to-many, many-to-one and many-to-many relationships. In this paper, we proposed an improved representation learning model SProjE, which introduces an adaptive metric method to reduce the weight of noise information and optimizes the loss function to improve the loss weight of complex relation triples. The proposed model is suitable for large scale knowledge graph representation learning. At last, the experimental results on the WN18 and FB15k data sets show that SProjE achieves significant and consistent improvements compared with the existing models and methods.

Keywords Knowledge graph, Representation learning, Adaptive metric, Entity link prediction

1 引言

知识图谱本质上是一种语义网络的知识库,其将实体 h 和 t 及实体间的关系 r 以三元组 (h, r, t) 的形式存储起来构成庞大的知识网络。它在知识问答^[1]、实体消歧^[2]、信息抽取^[3]、语义解析^[4]等人工智能领域发挥了巨大的作用。在构建知识图谱时,通常采用网络的形式组织知识,大部分的知识借助 RDF 和 OWL 等技术标准以三元组的形式表示,这种表示形式虽然在表示结构化数据时非常高效,但是存在数据稀疏、计算效率低等问题。为了解决基于网络的表示方法带来的问题,知识表示学习被提出并得到了越来越广泛的关注。

知识表示学习是面向知识图谱中的实体和关系进行表示学习的,通过将知识图谱中的隐含语义信息表示为稠密低维连续向量来实现知识的可计算性,显著降低了知识图谱的计算难度,缓解了数据稀疏的问题,并提高了知识推理的性能^[5]。知识表示学习根据打分函数的不同,可分为基于翻译模型、语义匹配模型和基于向量投影距离模型。在基于向量投影距离模型中,一个三元组 (h, r, t) 是否可信是由候选实体 t 的向量在目标向量 $C(h, r)$ 上的投影距离决定的,投影距离越长,三元组的可信度就越高。这类模型打分函数的特殊性,使其针对一对多、多对一和多对多这些复杂关系有着较强的表示能力。

收稿日期:2019-03-08 返修日期:2019-05-03 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(61836007,61772354,61773276)

This work was supported by the National Natural Science Foundation of China(61836007,61772354,61773276).

通信作者:李培峰(plli@suda.edu.cn)

但是,这类模型在计算投影距离时的方法相对简单,在计算损失时未考虑模型特性,因此在面对多样且复杂的实体关系时表示能力不足。其主要原因如下:1)在打分过程中,基于投影的模型的各个维度的特征信息在投影距离中所占的比重固定,这会引入无关维度信息带来的噪声影响并且忽略有关重要特征信息,从而降低了模型的表示能力和预测候选实体的精度;2)当输入实体关系对 (h, r) 有多个正确候选实体时,使用基于交叉熵的成列损失函数,简单地对每个正确候选实体的目标概率加上归一化权重用于训练,导致复杂关系三元组在训练时得不到充分更新,影响了整体性能。

为此,本文对基于投影模型 ProjE^[6]进行了改进:1)采用一种自适应度量的方法,对于不同的关系,赋予实体特征维度不同的权重,有效避免了过拟合现象,提升了泛化能力,同时减小了无关维度信息带来的噪声;2)优化损失函数,对于有多个候选实体的实体关系对,更改原先简单的归一化权重,解决了复杂关系三元组在使用基于交叉熵的成列损失函数时更新不足的问题。

2 相关工作

随着知识表示学习这一研究的提出,研究者们迅速给予了其关注,并做了大量的工作。根据衡量三元组成立的打分标准,可以将这些研究工作大致分为:基于翻译模型、语义匹配模型和基于向量投影距离的模型。

在词向量模型 word2vec^[7]中,具有相同关系的词对在隐式特征空间中有相同的相对位置分布。受此启发,基于翻译模型遵循几何原理 $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$,将三元组 (h, r, t) 中的关系 r 认为是从头实体向量 \mathbf{h} 到尾实体向量 \mathbf{t} 的一个翻译操作,这要求在向量空间中正确候选实体 t 应尽量分布在 $\mathbf{h} + \mathbf{r}$ 的附近,因此基于翻译的模型采用 $\mathbf{h} + \mathbf{r}$ 和 \mathbf{t} 之间的欧氏距离作为打分标准,定义如下:

$$f_r(h, t) = -\|\mathbf{h}_r + \mathbf{r} - \mathbf{t}_r\|_{1/2} \quad (1)$$

其中, \mathbf{h}_r 和 \mathbf{t}_r 是实体嵌入在不同关系空间中的表示。大多数基于翻译模型的区别在于实体在不同关系空间中的表示方法不同。

TransE^[8]是最具代表性的基于翻译模型,它将所有实体都固定表示在原始空间,与关系无关,即 $\mathbf{h}_r = \mathbf{h}, \mathbf{t}_r = \mathbf{t}$,因此虽然该模型非常简单高效,但是在面对复杂关系时其表示能力不足。为此,其他模型引入与关系相关的实体嵌入表示,例如 TransH^[9]模型将实体投影到与关系相关的超平面上,即 $\mathbf{h}_r = \mathbf{h} - \mathbf{w}_r^T \mathbf{h} \mathbf{w}_r, \mathbf{t}_r = \mathbf{t} - \mathbf{w}_r^T \mathbf{t} \mathbf{w}_r$; TransR^[10]模型使用投影矩阵将实体投影到与关系相关的超空间中,即 $\mathbf{h}_r = \mathbf{M}_r \mathbf{h}, \mathbf{t}_r = \mathbf{M}_r \mathbf{t}$,但其中引入了关于关系的投影矩阵,导致模型较为复杂。TransD^[11]和 TransSparse^[12]模型通过进一步简化 TransR 中的投影矩阵,获得了更优的性能。虽然将实体投影到不同的超平面和超空间提高了模型对复杂关系的表示能力,但是由于打分依据过于简单,模型的表示能力受限。为了进一步提升基于翻译模型的泛化能力,许多模型尝试在损失计算方面做出改进。TransM^[13]在三元组 (h, r, t) 原有打分的基础上给定一个与关系相关的权重,打分函数变为:

$$f_r(h, t) = -\theta_r \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|_{1/2} \quad (2)$$

该函数使模型在面对复杂关系时三元组的损失权重更

小,从而使 t 处于与 $\mathbf{h} + \mathbf{r}$ 更远的位置。TransA^[14]为每个关系引入一个对称非负矩阵,并且使用自适应的马氏距离替代原先的欧氏距离作为打分依据,在处理复杂关系时更加灵活。打分函数的定义为:

$$f_r(h, t) = -(\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|)^T \mathbf{M}_r (\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|) \quad (3)$$

语义匹配模型采用基于相似度的评分标准,通过匹配实体关系内的隐式语义信息来衡量三元组的可信性。常见的语义匹配模型有如下几种。

RESCAL^[15]是一种张量分解模型,将知识图谱中的每个三元组存在可能性分解为实体和关系的表示。打分函数定义为:

$$f_r(h, t) = \mathbf{h}^T \mathbf{M}_r \mathbf{t} \quad (4)$$

SME^[16]模型通过矩阵乘积和 Hadamard 乘积来获取实体和关系之间的语义相关性,其定义了两种打分函数,分别是线性形式和双线性形式。

$$f_r(h, t) = (\mathbf{M}_1 \mathbf{h} + \mathbf{M}_2 \mathbf{r} + \mathbf{b}_1)^T (\mathbf{M}_3 \mathbf{h} + \mathbf{M}_4 \mathbf{r} + \mathbf{b}_2) \quad (5)$$

$$f_r(h, t) = (\mathbf{M}_1 \mathbf{h} \otimes \mathbf{M}_2 \mathbf{r} + \mathbf{b}_1)^T (\mathbf{M}_3 \mathbf{h} \otimes \mathbf{M}_4 \mathbf{r} + \mathbf{b}_2) \quad (6)$$

其中, \otimes 表示 Hadamard 乘积。

NTN^[17]是一种张量神经网络架构模型,用双线性张量取代传统神经网络的线性变换层。该模型打分函数的定义为:

$$f_r(h, t) = \mathbf{u}_r^T f(\mathbf{h}^T \mathbf{W}_r \mathbf{t} + \mathbf{W}_n \mathbf{h} + \mathbf{W}_n \mathbf{t} + \mathbf{b}_r) \quad (7)$$

可以发现,NTN 模型中所有的参数都与关系相关,参数的数量非常巨大,在训练时所需的内存和时间远远大于其他模型,很难应用于较大的知识图谱。

基于向量投影距离模型将三元组 (h, r, t) 中尾实体的嵌入表示 \mathbf{t} 投影到由 \mathbf{h} 和 \mathbf{r} 组成的嵌入表示 \mathbf{Q} 上,并将投影距离作为衡量三元组可信的标准。由图1可以发现,与基于翻译模型相比,基于投影距离模型不再将正确实体向量表示束缚在目标向量周围,因此这类模型在表示复杂关系时有更好的灵活性。ProjE 模型在使用投影距离的基础上,将知识图谱补全任务视为多分类问题,在链接预测时,通常认为在组合向量 \mathbf{Q} 上投影距离长度排名靠前的候选实体正确,因此不同于基于翻译模型中常用的基于边缘的成对排序误差损失,该类模型采用基于交叉熵的成对排序误差损失函数,使其在负采样上更加灵活,并增强了处理大规模知识图谱的能力。

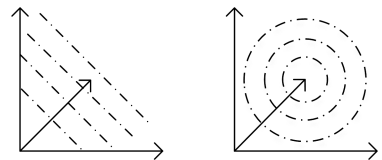


图1 基于向量投影距离模型(左)和基于翻译模型(右)打分等势面
Fig.1 Hyperplane of vector projection distance based model and transition distance based model

3 改进的向量投影距离模型

本节将介绍在基于向量投影距离模型 ProjE 的基础上改进的知识图谱表示方法 SProjE,并从打分函数和损失函数两个方面来解释其原理和算法。

3.1 打分函数

在 ProjE 模型中,打分函数可以分为两个部分,结合操作和投影距离计算。在计算三元组 (h, r, t) 的可信度时,先将 \mathbf{h} 和 \mathbf{r} 作为输入进行组合而得到目标向量。组合操作定义为:

$$\mathbf{e} \oplus \mathbf{r} = \mathbf{D}_e \mathbf{e} + \mathbf{D}_r \mathbf{r} + \mathbf{b}_e \quad (8)$$

其中, \mathbf{D}_e 和 \mathbf{D}_r 都为 $k \times k$ 维的对角矩阵。再将候选实体向量表示 \mathbf{t} 投影到该目标向量上, 求出投影距离作为三元组可信度, 因此 ProjE 中的打分函数定义为:

$$f_r(h, \mathbf{t}) = \sigma(\tanh(\mathbf{e} \oplus \mathbf{r}) \cdot \mathbf{t} + \mathbf{b}_p) \quad (9)$$

但是, 这种方法会带来两个问题: 1) 在组合操作时定义了全局的实体关系权重, 导致只会固定地对相应维度的特征信息放缩, 在面对不同关系三元组时容易带来无关维度的噪声信息; 2) 计算投影距离的方式相对简单, 每个特征维度信息在可信度计算中所占的权重比例相同, 无法突出有效信息, 在面对大规模复杂知识图谱时表现不够灵活。

本文提出的改进模型 SProjE 将原有的组合操作简化为直接相加, 以避免在模型输入阶段损失有效特征信息。借鉴 TransA 中的自适应度量方法, 在计算投影距离时, 引入一个与关系相关的权重矩阵 \mathbf{W}_r , 因此单个三元组的打分函数被重新定义为:

$$f_r(h, \mathbf{t}) = \sigma(\tanh(\mathbf{e} + \mathbf{r}) \mathbf{W}_r \mathbf{t} + \mathbf{b}_p) \quad (10)$$

遵循 ProjE 中不同维度之间不相互作用的约定, 设定 \mathbf{W}_r 为 $k \times k$ 维的对角矩阵。由于将整个知识图谱补全任务视为多分类任务, 输入的不是单个候选实体而是候选实体集合, 因此投影距离函数定义为:

$$h(\mathbf{e}, \mathbf{r}) = g(\mathbf{W}^e \mathbf{W}_r f(\mathbf{e} + \mathbf{r}) + \mathbf{b}_p) \quad (11)$$

其中, $\mathbf{W}^e \in \mathbb{R}^{n \times d}$ 是包含 n 个候选实体的矩阵; $f(x) = \tanh(x)$ 是非线性激活函数, 用来将输出范围限制在 $(-1, 1)$ 内; $g(\cdot)$ 采用 softmax 激活函数。

从打分等势面和特征权重两个角度来分析加入自适应权重的优势。1) 由图 2 可以发现, \mathbf{W}_r 可以视为对候选实体向量 \mathbf{t} 做缩放变换, 从而使打分等势面发生自适应转动, 以提升正确候选实体的排名; 也可以视为对组向向量 $\mathbf{h} + \mathbf{r}$ 做缩放变换, 以使打分等势面发生转动。2) 从特征权重角度来看, $\mathbf{W}_r = \text{diag}(\omega_1, \omega_2, \omega_3, \dots)$ 是对角矩阵, 以不同的特征维度乘上权重 ω_i , 实体间的关系由几个特定维度的信息来决定。在模型中, 采用同一种特征权重系数会带来其他无关维度产生的噪声信息, 从而影响模型性能, 加入自适应权重可以有效地规避这种现象。

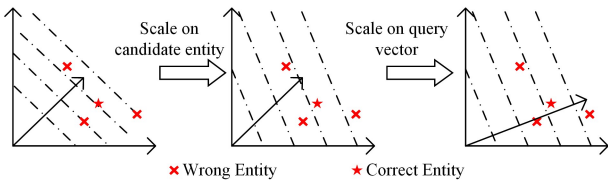


图 2 缩放变换效果图

Fig. 2 Effect of dimensional scaling

3.2 损失函数

在模型训练中, 采用基于交叉熵的成列排序误差损失函数作为训练模型的优化目标函数。ProjE 中采用的损失函数如下:

$$\mathcal{L}(\mathbf{e}, \mathbf{r}, \mathbf{y}) = - \sum_i^{|y|} \frac{1(y_i = 1)}{\sum_i \mathbf{1}(y_i = 1)} \log(h(\mathbf{e}, \mathbf{r})_i) \quad (12)$$

其中, $\mathbf{y} \in \mathbb{R}^n$ 是二元标签向量, $y_i = 1$ 表示第 i 个候选实体是正确的。对于有 n 个正确候选实体的情况, 在训练过程中输入实体关系对 (\mathbf{e}, \mathbf{r}) , 其损失值对应于训练集中的 n 个三元组实

例。由于每个三元组实例都会训练 n 次, 为了保证最后每个三元组的总损失值权重仍为 1, 在训练时会每个正确候选实体的选中概率乘上 $1/n$ 的权重用于归一化。然而, 在保证每个三元组总损失值不变的情况下, 无法保证损失多次分开训练与单次独立训练的效果一致。

在训练过程中, 模型采用 softmax 作为激活函数, 每个候选实体被选中的概率为:

$$P_j = \frac{e^{a_j}}{\sum_{k=1}^N e^{a_k}} \quad (13)$$

其中, N 表示候选的实体数, a_j 表示第 j 个候选实体在组向量上的投影长度。计算 P_j 对 a_i 的导数时, 分为两种情况。

1) 当 $i = j$ 时

$$\frac{\partial P_j}{\partial a_j} = P_j(1 - P_j) \quad (14)$$

2) 当 $i \neq j$ 时

$$\frac{\partial P_j}{\partial a_i} = -P_i P_j \quad (15)$$

这时就可以发现损失在反向传播时的问题: 对每次训练的损失值赋予 $1/N$ 的权重, 在反向传播中, 经过 softmax 层后的损失权重变为 $1/N^2$, 训练 N 次后, 反向传播中传递的总损失仅为单次独立训练的 $1/N$ 。复杂关系的三元组因为在训练中传递的损失过小, 导致梯度更新得不充分, 模型的代表效果受到了很大的影响。

因此需要重新定义归一化权重函数为 $f(n) = 1/\sqrt{n}$, 其中 n 为三元组中的正确候选实体数。在每一次的训练中, 每个正确三元组的损失值权重为 $1/\sqrt{n}$, 损失反向传播时经过 softmax 激活函数层后, 损失权重变为 $1/n$, 在完整的训练中每个三元组会参与训练 n 次, 因此复杂关系三元组多次分开训练的传递损失与单独训练的参数更新效果基本相同, 避免了将知识图谱补全任务视为多分类任务后带来的负面影响。损失函数重新定义为:

$$\mathcal{L}(\mathbf{e}, \mathbf{r}, \mathbf{y}) = - \sum_i^{|y|} \frac{1(y_i = 1)}{\sqrt{\sum_i \mathbf{1}(y_i = 1)}} \log(h(\mathbf{e}, \mathbf{r})_i) \quad (16)$$

4 实验

本节通过在通用知识图谱上进行链接预测实验来验证和评估模型的性能。

4.1 数据集和评测指标

本文采用 FB15k 和 WN18 作为实验语料。FB15k 是从 Freebase^[18] 知识图谱抽取出的部分数据构成的知识库, WN18 是从 wordnet 知识库中抽出的子集。WN18 包含的实体较多; 而 FB15k 中包含的关系类别较多, 实体间的联系也更加密集。通过两个数据集能较好地验证模型的实际性能。知识库的详细情况如表 1 所列。

表 1 数据集统计信息

Table 1 Statistics of datasets

Dataset	# Rel	# Ent	# Train	# Valid	# Test
FB15k	1345	14951	483142	50000	59071
WN18	18	40943	141400	5000	5000

实体链接预测任务是完善知识图谱的重要任务, 大多数模型利用实体链接预测的准确率来评估模型在知识推理方面

的能力。该任务主要是将三元组中的一对实体和关系作为已知信息来对候选实体进行排序,并根据排序位置的前后来预测三元组中缺失的实体。根据预测实体位置的不同,该任务分为头实体预测($?, r, t$)和尾实体预测($h, r, ?$)。在测试过程中,首先用知识图谱中所有的实体替换测试集中头实体或尾实体,然后根据模型的打分将替换的实体降序排列,通过观察正确实体在序列中的位置来评估模型的性能。

在实体链接预测中,通常将平均排序(mean rank)和 HITS@10 作为评测指标。平均排序指标用来衡量正确的实体在所有候选实体中的平均排名,该值越低,正确实体在候选实体队列中的排序越靠前,模型预测就越准确。HITS@10 指标用来衡量正确的实体在排名前 10 位的概率,该值越大,性能就越好。因为知识图谱存在 1-n, n-1 和 n-n 这些复杂关系,所以正确的候选实体往往有多个,只采用上述指标不能全面评估模型性能。因此,在对一个测试三元组排名打分之前,过滤掉其他不同但正确的候选实体,从而使原目标实体只与错误实体比较,更好地评估了模型的性能。这种设置被称为过滤(Filter)。基于此,本文在原有指标上增加了 Filter mean rank 和 Filter HITS@k 评测指标。

4.2 实验设置

在训练 SProjE 时,使用 Adam^[19]方法作为统计优化方法,并设置超参数: $\beta_1=0.9, \beta_2=0.999, \epsilon=1 \times 10^{-8}$ 。在训练过程中,对所有的参数采用了 L1 正则化,并在投影操作前对组合且经过缩放变换的向量做了 dropout 操作,用于防止过拟合并提升模型的泛化能力,dropout 率设为 0.5。在训练前,参照 TransE 中的设置,从均匀分布 $U[-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}}]$ 中随机初始化知识库中的实体和关系向量表示,并且最大训练迭代次数设为 100。对于给定的实体 e 、关系 r 和一个二元标签向量 y ,为了防止参与训练的候选实体数过大,只计算正确候选实体及一定比例的错误候选实体的投影长度,以提高训练效率。设定错误实体采样率 p_y ,即每个错误实体被采样训练的概率为 p_y ,则不被抽到的概率为 $1-p_y$ 。因此,对 y 中的每个错误候选实体,根据伯努利分布 $B[1, p_y]$ 来决定候选实体矩阵 W 中是否包含此实体。 p_y 为超参数,在实验中采用的采样率 $p_y = \{0.05, 0.1, 0.25, 0.5, 0.75, 0.95\}$ 。

训练 SProjE 模型时,在 FB15k 数据集上设置的实验参数如下:学习率 $lr=0.005$,维度 $d=200$,批次大小 $b=1024$,负候选实体采样率 $p_y=0.1$;在 WN18 数据集上设置的实验参数如下:学习率 $lr=0.01$,维度 $d=100$,批次大小 $b=1024$,负候选采样率 $p_y=0.005$ 。

4.3 链接预测

SProjE 模型的链接预测效果如表 2 所列,同时给出在 ProjE 基础上修改打分函数 ProjE+score 和修改损失函数 ProjE+loss 的性能,并与其他常见的知识表示模型进行对比。从表 2 中的实验结果发现:1) SProjE 相对于原模型 ProjE 在各个性能指标上都有明显的提升;2) 在两个公开数据集上, SProjE 在链接预测任务上都取得了更优的结果模型,与其他知识表示模型相比,其对知识图谱的表示学习能力更强;3) 对比 SProjE 在 WN18 和 FB15k 上的性能表现,可以发现

该模型在对关系种类多、信息稠密的知识图谱进行知识表示时更有优势;4) 虽然 SProjE 的平均排序和 HITS@10 性能弱于 TransA,但其在 Filter HITS@10 上的表现大大优于后者,表明 SProjE 在处理复杂关系时,多个正确候选实体的排名位置更加聚集且靠前,整体性能更加稳定。

表 2 链接预测的平均预测结果

Table 2 Average results of link prediction

Algorithm	WN18				FB15k			
	Mean Rank		HITS@10/%		Mean Rank		HITS@10/%	
	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter
RESCAL ^[15]	1180	1163	37.2	52.8	828	689	28.4	44.1
SE ^[20]	1011	985	68.5	80.5	273	162	28.8	39.8
SME (linear) ^[16]	545	533	65.1	74.1	274	154	30.7	40.8
SME (bilinear) ^[16]	526	509	54.7	61.3	284	158	31.3	41.3
TransE ^[7]	263	251	75.4	89.2	243	125	34.9	47.1
TransH ^[9]	401	388	73.0	82.3	212	87	45.7	64.4
TransR ^[10]	238	225	79.8	92.0	198	77	48.2	68.7
TransD ^[11]	224	212	79.6	92.2	194	91	53.4	77.3
TransA ^[14]	405	392	82.3	94.3	155	74	56.1	80.4
ProjEa	—	—	—	—	214	60	48.1	78.8
ProjE+score	339	329	84.0	93.6	192	42	52.5	83.1
ProjE+loss	322	312	83.0	94.8	173	37	54.5	85.5
SProjE	297	288	84.3	95.2	167	32	55.4	87.4

注: ProjEa 在 FB15k 数据集上的链接预测结果是由作者在修复原有错误代码后提供的最新结果。相关参数设置如下: 嵌入维度 $d=200$, 批次大小 $b=512$, 学习率 $lr=0.0005$, 负候选实体采样率 $p_y=0.1$, 最大迭代次数 $iter=50$

表 3 对 SProjE 模型在 FB15k 中各类关系上的表示性能进行了统计,并与其他模型进行了对比。可以发现,基于向量投影距离的模型本身针对复杂关系有较好的表示能力,这是基于翻译模型所不具有的。但这个特点也容易带来其他缺点,表示简单关系时容易带来脏信息,正确候选实体往往很难预测。在加入自适应权重后,干扰信息权重大幅下降,1-1 关系的链接预测性能得到了大幅提升,头实体预测提升了 25.8%,尾实体预测提升了 27.1%。这表明采用自适应度量方法后实体在面对不同关系时,向量表示发生了相应变化,模型因而能更好地区分正确与错误候选实体。同时,对损失函数进行优化,通过对多候选实体关系损失计算加上合适的权重,在 1-N 关系尾实体预测和 N-1 关系头实体预测这类其他模型较难处理的任务上有显著的性能改善。

表 3 SProjE 在 FB15k 各类关系上的 Filter HITS@10 值

Table 3 Filter HITS@10 of each type of relations in FB15k

(单位: %)

Algorithm	PredictHead(HITS@10)				PredictTail(HITS@10)			
	1-1	1-N	N-1	N-N	1-1	1-N	N-1	N-N
SE ^[20]	35.6	62.6	17.2	37.5	34.9	14.6	68.3	41.3
SME (bilinear) ^[16]	30.9	69.6	19.9	38.6	28.2	13.1	76.0	41.8
TransE ^[7]	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50.0
TransH ^[9]	66.8	87.6	28.7	64.5	65.5	39.8	83.3	67.2
TransD ^[11]	86.1	95.5	39.8	78.5	85.4	50.6	94.4	81.2
TransR ^[10]	78.8	89.2	34.1	69.2	79.2	37.4	90.4	72.1
TransA ^[14]	86.8	95.4	42.7	77.8	86.7	54.3	94.4	80.6
ProjE_ listwise	61.4	90.1	53.2	83.3	61.3	63.5	89.4	85.5
SProjE	87.2	96.3	71.4	86.2	88.4	80.8	95.8	89.7

为了更清楚地展示 SProjE 在不同关系类型三元组上的

效果提升,本文列出了4种关系类型的相应例子,性能对比如图3所示,从左到右3个关系类分别属于1-1,1-N,N-1和N-N。可以发现,1-1关系上的预测结果提升得最明显,简单关系的表示性能显著加强,噪声信息得到了有效控制。整体结果表明,SProjE模型相比ProjE有了全面的性能提升。

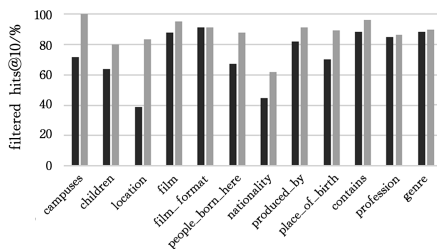


图3 ProjE和SProjE在不同关系类型三元组上的性能表现

Fig. 3 Performance comparison of ProjE and SProjE on different relational type triples

结束语 本文主要提出了一种改进向量投影距离的知识表示学习模型SProjE,在ProjE的基础上加入自适应度量方法,用于在面对不同关系时进一步区分正确与错误候选实体,提升模型的预测能力。另外,针对原有的损失函数进行权重优化,在归一化的基础上,给复杂关系损失赋予合适的权重,从而使损失分布更加一致,在加快模型收敛速度的同时,提高了知识表示的性能。实验数据表明,与ProjE和其他一些知识表示模型相比,SProjE取得了一致且显著的性能提升。

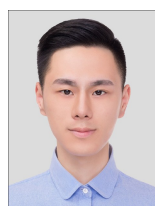
下一步将考虑以下研究方向:1)在SProjE的基础上加入额外的信息,如路径信息、实体类别信息等;2)考虑针对动态知识图谱,使实体关系的表示随着图谱的变化自动更新;3)如今大多数图谱都是自动构建的,存在很多错误的信息,考虑如何在自动构建的知识图谱中使用知识表示方法。

参考文献

- [1] BORDES A, WESTON J, USUNIER N. Open Question Answering with Weakly Supervised Embedding Models [J/OL]. [2014-04-16]. <https://arxiv.org/pdf/1404.4326.pdf>.
- [2] ZHENG Z, SI X, LI F, et al. Entity Disambiguation with Freebase [C] // 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology. IEEE Computer Society, 2012.
- [3] DAIBER J, JAKOB M. Improving efficiency and accuracy in multilingual entity extraction [C] // Proceedings of the 9th International Conference on Semantic Systems. ACM, 2013: 121-124.
- [4] BERANT J, CHOU A, FROSTIG R, et al. Semantic parsing on freebase from question-answer pairs [C] // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013: 1533-1544.
- [5] WANG Q. Knowledge graph embedding: A survey of approaches and applications [J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.
- [6] SHI B, WENINGER T. ProjE: Embedding Projection for Knowledge Graph Completion [C] // Proc of AAAI 2017, San Francisco; AAAI, 2017: 1236-1242.
- [7] MIKOLOV T, SUTSKEVER I, CHEN, et al. Distributed representations of words and phrases and their compositionality [C] //

Proc of NIPS 2013. Cambridge, MA; MIT Press, 2013: 3111-3119.

- [8] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data [C] // Proc of NIPS 2013. Cambridge, MA; MIT Press, 2013: 2787-2795.
- [9] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes [C] // Proc of AAAI 2014, Menlo Park, CA; AAAI, 2014: 1112-1119.
- [10] LIN Y K, LIU Z Y, SUN M S, et al. Learning Entity and Relation Embeddings for Knowledge Graph Completion [C] // Proc of AAAI 2015. Menlo Park, CA; AAAI, 2015: 2181-2187.
- [11] JI G, HE S, XU L, et al. Knowledge Graph Embedding via Dynamic Mapping Matrix [C] // Proc of ACL 2015, Beijing, China; ACL, 2015: 687-696.
- [12] JI G, LIU K, HE S, et al. Knowledge Graph Completion with Adaptive Sparse Transfer Matrix [C] // Thirtieth Aaai Conference on Artificial Intelligence. AAAI Press, 2016.
- [13] FAN M, QIANG Z, CHANG E, et al. Transition-based knowledge graph embedding with relational mapping properties [C] // Proc. of PACLIC. 2014.
- [14] XIAO H, HUANG M L, HAO Y, et al. TransA: An adaptive approach for knowledge graph embedding [J/OL]. [2015-09-28]. <https://arxiv.org/pdf/1509.05490.pdf>.
- [15] NICKEL M, TRESP V, KRIEGEL H P. A Three-Way Model for Collective Learning on Multi-Relational Data [C] // Proc of ICML 2011. New York, USA; ACM, 2011: 809-816.
- [16] BORDES A, GOROT X, WESTON J, et al. A semantic matching energy function for learning with multi-relational data [J]. Machine Learning, 2014, 94(2): 233-259.
- [17] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion [C] // Proc of NIPS 2013. Cambridge, MA; MIT Press, 2013: 926-934.
- [18] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge [C] // Sigmod Conference. 2008.
- [19] KINGMA D, BA J. Adam: A method for stochastic optimization [J/OL]. [2017-01-30]. <https://arxiv.org/pdf/1412.6980.pdf>.
- [20] BORDES A, WESTON J, COLLOBERT R, et al. Learning structured embeddings of knowledge bases [C] // Proc. of AAAI 2011. Menlo Park, CA; AAAI, 2011: 301-306.



LI Xin-chao, born in 1995, postgraduate. His main research interests include natural language processing and representation learning.



LI Pei-feng, born in 1971, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include natural language processing and machine learning.