STransH:一种改进的基于翻译模型的知识表示模型

陈晓军 向 阳
(同济大学电子与信息工程学院 上海 201804)

摘 要 最近,以深度学习为代表的表示学习技术受到广泛关注。表示学习旨在将研究对象的语义信息表示为低维稠密实值向量。因此,一系列知识表示模型被提出,其中基于翻译模型的经典方法TransE不仅模型复杂度低、计算效率高,而且具有良好的知识表达能力。但是,TransE方法在处理自反、一对多、多对一和多对多等复杂关系时存在局限性。鉴于此,文中提出一种改进的知识表示模型STransH,分别在实体空间和关系空间建模,并采用单层神经网络的非线性操作来加强实体和关系的语义联系。同时,受TransH模型的启发,引入投影到特定关系超平面的机制,使得实体在不同的关系中有不同的角色。在模型训练时,通过替换语义相似实体来提高生成负例的质量。最后,在公开的数据集FB15K和WN18上进行链接预测实验,分析和验证了所提方法的有效性。相比于TransE和TransH模型,STransH在各项性能指标上均取得了较大提升,其Hits@10和三元组分类准确率分别提高近10%。

关键词 知识图谱,表示学习,链接预测,三元组分类
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.09.026

STransH:A Revised Translation-based Model for Knowledge Representation

CHEN Xiao-jun XIANG Yang
(College of Electronic and Information Engineering,Tongji University,Shanghai 201804,China)

Abstract Recently,representation learning technology represented by deep learning has attracted many attentions in natural language processing,computer vision and speech recognition. Representation learning aims to project the interested objects into a low-dimensional,dense and real-valued semantic space. To this end,a number of models and methods were proposed for knowledge embedding. Among them,TransE is a classic translation-based method with low model complexity,high computational efficiency and favorable knowledge representation ability. However,it has limitations in dealing with complex relations including reflexive,one-to-many,many-to-one and many-to-many relations. In light of this,this paper proposed a revised translation-based method for knowledge graph representation,namely STransH. In this method,firstly,entity and relation embeddings are built in separate entity space and relation space,and then the non-linear operation of single-layer network layer is adopted to enhance the semantic connection between entity and relation. Inspired by TransH,this paper introduced the relation-oriented hyperspace model,thus projecting head and tail entities to the hyperspace of a given relation for distinction. Besides,it also proposed a simple trick to improve the quality of negative triplets. At last,it conducted extensive experiments on link prediction and triplet classification on benchmark datasets like WordNet and Freebase. Experimental results show that STransH performs significant improvements over TransE and TransH compared with TransE and TransH,and its Hits@10 and triplet classification accuracy are increased by nearly 10% respectively.

Keywords Knowledge graph,Representation learning,Link prediction,Triplet classification

1 引言

知识图谱是结构化的语义知识库,用于描述现实世界中的实体和关系,通常被表示为网络形式,网络中的每个节点代表实体(entity),而每条连边代表关系(relation)。因此,知识图谱中的连边及与之相连的两个实体通常以三元组的形式(头实体,关系,尾实体)来表示。

随着大数据技术的发展,大规模知识库的构建已经取得了很好的发展,人们构建了各种各样的知识库,如语言知识库WordNet[1]和世界知识库FreeBase[2]等。知识图谱是推动人工智能学科和智能信息服务发展的重要技术,已被广泛应用于问答系统、智能搜索和个性化推荐等领域。知识图谱(Knowledge Graph)的概念于2012年5月被Google正式提出,旨在提高搜索质量,增强用户的搜索体验。为了改善信息

到稿日期:2018-08-13 返修日期:2018-12-02 本文受国家自然科学基金(71571136),上海市科委基础研究项目(16JC1403000)资助。
陈晓军(1995-),男,博士生,CCF会员,主要研究方向为知识图谱、知识推理,E-mail:xiaojunchen@tongji.edu.cn;向 阳(1962-),男,博士,教授,CCF会员,主要研究方向为数据挖掘、机器学习等,E-mail:shxiangyang@tongji.edu.cn(通信作者)。

服务质量，国内外互联网公司纷纷推出自己的知识图谱，如百度的“知心”、搜狗的“知立方”等。正如谷歌在介绍 Knowledge Graph 时说，“The world is not made of strings, but is made of things”，大规模知识图谱研究和应用的热潮由此拉开。

知识图谱的研究目标是从无结构或半结构的信息中抽取有结构的知识，通过知识融合构建知识库，服务知识推理、自动问答等应用。其中，知识表示是知识获取与应用的基础。但随着知识库知识规模的不断扩大，知识形式更加复杂化。传统的知识表示形式表现出了局限性，三元组的表现形式无法度量和使用实体间的语义关系，还受到计算效率低和数据稀疏等问题的困扰，很难应用到大规模知识图谱上。

近年来，以深度学习为代表的表示学习[3]在自然语言处理领域受到广泛关注。词向量模型——word2vec[4]的提出，掀起了知识表示学习的热潮。顾名思义，知识表示学习就是指对知识库中的实体和关系进行学习，旨在将研究对象的语义信息表示为稠密低维实值向量。在低维向量空间中，两个对象之间的距离越近，说明它们的语义相似度越高。知识表示学习研究最近取得了重大进展，可以在低维向量空间中高效计算实体和关系之间的语义相似度，有效解决数据稀疏等问题，从而提升知识获取、知识推理的性能；此外，知识表示学习还被广泛应用于自动问答系统、关系抽取等任务中，并且展现出了巨大的应用潜力。

鉴于上述优点，研究者提出了若干知识表示模型，包括距离模型（Structured Embedding，SE）[5]、单层神经网络模型（Single Layer Model，SLM）[6]、语义能量匹配模型（Semantic Matching Energy，SME）[7]、张量分解模型、基于翻译的模型等。现有方法中，受 word2vec 模型中词向量在语义空间的平移不变现象启发而提出的 TransE 模型[8]最著名，其得分函数为 $f_r(\boldsymbol{h},\boldsymbol{t})=\|\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{t}\|_2^2$，其表示在向量空间中头实体经过关系转换之后与尾实体之间的欧氏距离。TransE 在取得较好预测表现的同时保持了简单、高效的特点，但它在处理复杂属性时的表现并不佳。TransE 方法在处理一对多、多对一、多对多和自反等复杂关系时存在局限性，不能很好地区分具有相同关系的实体。

为此，本文针对 TransE 模型存在的缺陷，在其基础上提出了一种改进的知识表示学习模型 STransH。具体地，将两个简单的知识表示模型——SE 模型与 TransE 模型进行结合[9]，借鉴 TransH 模型[10]的思想，引入投影到特定关系超平面的机制。首先，将三元组中的 h 和 t 映射到给定关系的超平面，以有效表示复杂关系；然后，用非线性张量层代替标准线性神经网络层；同时，在模型训练时，对负例三元组的抽样策略进行改进，利用一对多和多对一的映射关系选择替换实体，使尽可能多的实体得到训练。替换实体时，选择与其语义最相似的实体进行替换，以提高实体之间的区分度。在 WN18 和 FB15K 两个数据集上对链接预测和三元组分类两项任务展开评测。实验结果表明，STransH 在 MeanRank 和 Hist@10 两个指标上均有提高，从而验证了模型的有效性。

2　相关工作

目前，知识表示工作的主要思路是将知识库嵌入到一个连续的向量空间中，并保留原有知识库的某些特性。这些知识表示方法通过最小化全局损失函数来获得实体和关系的表示，这个损失函数涉及到知识图谱中所有的实体和关系，这就意味着实体和关系的表示是通过编码全局信息得到的[11]。

早期的知识表示模型主要关注提高模型表现力和普遍性，但也带来了模型复杂度增加和参数爆炸等问题。不仅如此，由于高复杂度的模型正则项较难设计，因此模型存在过拟合的可能性。当前，面向知识图谱的表示学习的研究主要集中在基于翻译的模型上，较有代表性的工作包括 TransE，TransH，TransR[12]，CTransR[12]，PTransE[13]，TransA[14] 和 TransD[15] 等方法。

基于翻译的模型认为，对于一个三元组 (h,r,t)，关系 r 可以看作是从头实体向量 $\boldsymbol{h}$ 到尾实体向量 $\boldsymbol{t}$ 的一个翻译（Translation）操作。受到向量空间平移不变现象的启发，Bordes 等[8]提出了 TransE 模型。对于每个三元组，TransE 希望 $\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{r}\approx 0$，并且定义了如下得分函数：

$$f_r(\boldsymbol{h},\boldsymbol{t})=\|\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{t}\|_2^2$$

TransE 模型参数少，计算复杂度低，在构建大规模知识图谱时具有简单、高效的特点。因此，自 TransE 被提出以来，大量学者在其基础上进行了扩展和应用。

TransH 模型将关系建模为一个超平面，较好地保留了关系的映射属性。对于关系 r，TransH 模型同时使用平移向量 $\boldsymbol{r}$ 和超平面的法向量 $\boldsymbol{w}_r$ 来表示。头实体和尾实体首先被映射到关系的超平面，得到向量 $\boldsymbol{h}_\perp=\boldsymbol{h}-\boldsymbol{w}_\perp^{\mathrm{T}}\boldsymbol{h}\boldsymbol{w}_\perp$ 和向量 $\boldsymbol{t}_\perp=\boldsymbol{t}-\boldsymbol{w}_\perp^{\mathrm{T}}\boldsymbol{t}\boldsymbol{w}_\perp$。因此，TransH 的得分函数变为 $f_r=\|\boldsymbol{h}_\perp+\boldsymbol{r}-\boldsymbol{t}_\perp\|$。

TransE 和 TransH 模型都假设实体和关系处于相同的语义空间中。然而，实体和关系本质上是不同的客观事物，将它们放置在同一个空间中进行刻画是不恰当的，这在一定程度上限制了模型的表示能力。TransR 模型认为，实体是多种属性的综合体，不同关系关注实体的不同属性，不同的关系应该拥有不同的语义空间。对于一个三元组 (h,r,t)，首先将实体向量向关系 $\boldsymbol{r}$ 进行空间投影，得到 $\boldsymbol{h}_r=\boldsymbol{h}\boldsymbol{M}_r$ 和 $\boldsymbol{t}_r=\boldsymbol{t}\boldsymbol{M}_r$，原来在实体空间中与头、尾实体相似的实体在关系空间中就被区分开了。CTransR 是 TransR 模型的拓展，通过将关系 r 对应的实体对的向量差值 $\boldsymbol{h}-\boldsymbol{t}$ 进行聚类，将关系细分为多个子关系，并且 CTransR 的参数数量比 TransE 和 TransR 多，导致复杂性过高，不如 Trans 模型和 TransH 模型简洁高效。为了改进 TransE 等模型孤立学习每个三元组的缺点，Lin 等[13]考虑关系路径的表示学习方法，提出 PTransE 模型。之前的 TransE，TransH 和 TransR 都认为每种关系只对应一种语义表示，而在实际情况中，关系 r 可能代表不同的含义。Ji 等[15]提出一种基于动态矩阵的 TransD 模型来解决这一问题。TransD 对于 (h,r,t) 来讲，有两种表示，一种是构建映射矩阵的表示，另一种是自身的语义表示。TransE 使用欧氏距离作为得分函数中的度量，每一个特征以相同的权重参与计算，会降低知识表示的准确性。针对这一缺陷，Xiao 等[14]提出 TransA 模型，利用玛氏距离（Mahalanobis Distance）来构造损失函数。

除了基于翻译的模型之外，距离模型是较早的知识表示

模型，每个实体用 d 维向量表示，所有实体被投影到同一个 d 维向量空间中。单层神经网络模型是距离模型的进一步改进，采用了单层神经网络的非线性操作来缓解 SE 无法精确刻画实体与关系的语义联系问题。虽然 SLM 是 SE 模型的改进版本，但是它的非线性操作仅提供了实体和关系之间比较微弱的联系。语义匹配能量模型定义若干投影矩阵，用于刻画实体与关系的内在联系。潜变量模型（Latent Factor Model，LFM）[16] 利用基于关系的双线性变换来刻画实体和关系之间的二阶联系。张量神经网络模型（Neural Tensor Network，NTN）[6] 的基本思想是，用双线性向量取代传统神经网络中的线性变换层，在不同的维度下将头、尾实体向量联系起来。RESCAL 模型[17] 是矩阵分解模型的代表，采用矩阵分解的方法进行知识的表示学习。

3 STransH 模型

本节详细介绍了改进的知识表示模型 STransH，并解释了其原理。STransH 也是属于基于翻译的模型，它分别采用单层神经网络和超平面模型来克服 TransE 模型的缺陷，并将两种想法同时集成到一个模型框架下，同时在模型训练时使用一种简单的技巧来生成高质量的负例三元组。

首先给出常用符号的说明：h 表示头实体，r 表示关系，t 表示尾实体。$\boldsymbol{h},\boldsymbol{r},\boldsymbol{t}$ 是相应的嵌入式表示，S 表示正例元组的集合，S' 表示负例元组的集合。

3.1 TransE 模型与距离模型结合

SE 为每个关系 r 定义了两个矩阵 $\boldsymbol{W}_{r,1},\boldsymbol{W}_{r,2}$，用于三元组中头实体和尾实体的投影操作；同时为每个三元组定义了如下得分函数：

$$f_r(h,t)=|\boldsymbol{W}_{r,1}\boldsymbol{h}-\boldsymbol{W}_{r,2}\boldsymbol{t}|_{L_1/L_2}$$

然后在向量空间中计算两投影间的距离，该距离反映了 2 个实体在关系 r 下的语义相似度。将 SE 模型和 TransE 模型进行结合，并置于统一的模型框架中。具体地，在 SE 模型的损失函数中添加关系 r 的表示，改进后的得分函数如下：

$$f_r(h,t)=\|\boldsymbol{W}_{r,1}\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{W}_{r,2}\boldsymbol{t}\|_{L_1/L_2}$$

接着，通过单层神经网络来减轻在距离模型中无法精确刻画实体和关系的语义联系的问题，非线性操作在一定程度上增强了实体之间的联系。添加单层神经网络后的评分函数为：

$$f_r(h,t)=g(\|\boldsymbol{W}_{r,1}\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{W}_{r,2}\boldsymbol{t}\|_{L_1/L_2})$$

其中，$g(\ )$ 是为 tanh 函数，$\boldsymbol{W}_{r,1}$ 和 $\boldsymbol{W}_{r,2}$ 为投影矩阵。

3.2 面向关系的超平面映射

结合 SE 模型后的 TransE 模型处理复杂关系的能力仍然较弱，为解决前文中提到的问题，首先对 TransE 模型进行分析。TransE 模型将实体和关系都表示在同一个空间中，造成无法区分复杂关系的问题，如图 1 所示。假设知识库中有两个三元组，分别为（姚明、出生于、上海）和（Angelababy、出生于、上海），TransE 不能正确区分这两个实体，如果用 TransE 从这两个三元组学习知识表示，将会使姚明和 Angelababy 的向量变得相同。这显然与事实不符，姚明和 Angelababy 除了出生地是上海这一属性比较相似外，其他方面有很大差异。正是这些复杂关系的存在，模型将两个实体收敛到过于接近，导致 TransE 学习得到的实体表示的区分度较低。

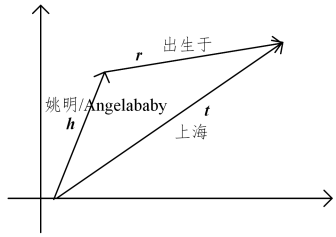

图 1 复杂关系的实例

Fig. 1 Instance of complex relation

受 TransH 模型的启发，引入投影到特定关系超平面的机制，使得实体在不同关系中有不同的角色。具体地，对于一个三元组，h 和 t 首先被映射到超平面 w_r 上，分别表示为 $\boldsymbol{h}_\perp$ 和 $\boldsymbol{t}_\perp$，然后在超平面上再将 $\boldsymbol{h}_\perp$ 和 $\boldsymbol{t}_\perp$ 与关系向量 $\boldsymbol{r}$ 联系起来。

$$\begin{cases}\boldsymbol{h}_\perp=\boldsymbol{h}-\boldsymbol{w}_\perp^{\mathrm{T}}\boldsymbol{h}\boldsymbol{w}_\perp\\ \boldsymbol{t}_\perp=\boldsymbol{t}-\boldsymbol{w}_\perp^{\mathrm{T}}\boldsymbol{t}\boldsymbol{w}_\perp\end{cases}$$

综合上式，可以得到应用超平面模型后的得分函数，即：

$$f_r(h,t)=g(\|\boldsymbol{W}_{r,1}\boldsymbol{h}_\perp+\boldsymbol{r}-\boldsymbol{W}_{r,2}\boldsymbol{t}_\perp\|_{L_1/L_2})$$

可以使用 L_1 或 L_2 距离，实验结果显示使用 L_1 距离的效果更好。

在应用了超平面模型后，如图 2 所示，姚明和 Angelababy 这两个实体通过不同的映射向量投影到“出生于”关系的超平面上，从而这两个实体得以区分，保证了知识表示的准确性。

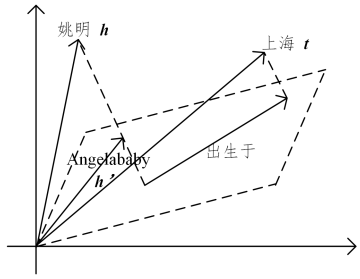

图 2 超平面模型的效果

Fig. 2 Effect of hyperspace model

3.3 模型训练

如前文所述，训练模型时，需要基于黄金三元组（Golden Triplets）构建负例三元组。已有方法只是通过随机打乱黄金三元组来获得负的三元组。例如，在 TransE 中，对于一个黄金三元组 (h,r,t)，一个负例三元组就是从实体集 E 中随机抽取一对实体 (h',t')。但是这种抽样方法在处理一对多、多对一和多对多的关系时，会出现将原来正确的三元组标记成错误三元组的情况。然而，对于一个尚未补全的真实知识图谱而言，这样简单的抽样方式可能会在训练中引入许多错的假阴性标签。本文以不同的概率替换头实体或尾实体，并且选择最相近的实体进行替换，以提高模型对实体的区分度。

（1）以不同的概率替换

在生成负例时，根据关系的类型设置不同的替换策略。对于一对多关系，我们以更高的概率替换头实体；对于多对一关系，我们以更高的概率替换尾实体。一个实体有多个属性，在处理一对多关系时，替换头实体能使头实体的每个属性都得到充分的训练；在处理多对一关系时，替换尾实体能使尾实

体的每个属性也得到充分的训练。

具体地,在一个关系的所有三元组中,首先可以获得以下两组数据:1)每个头实体对应的尾实体的平均数量 tph;2)每个尾实体对应的头实体的平均数量 hpt。然后定义参数为 $p=tph/(tph+hpt)$的伯努利分布来抽样。对于一个给定的黄金三元组,在构造负例三元组时,以概率 p 替换头实体,以概率1− p 替换尾实体,这样可以使总的概率为 1,且抽样方式符合伯努利分布,因此,我们采用伯努利分布。使用该分布既可以大大降低产生错误负例三元组(False Negative Triplets)的概率,又可以减小模型的计算复杂度。

我们规定,如果 $tph<1.5$ 且 $hpt<1.5$,则认为关系 r 是一对一的;如果 $tph>1.5$ 且 $hpt>1.5$,则认为关系 r 是多对多的;如果 $tph\geqslant 1.5$ 且 $hpt<1.5$,则认为关系 r 是一对多的;如果 $tph<1.5$ 且 $hpt\geqslant 1.5$,则认为关系 r 是多对一的。

(2)选择语义最相近的实体

相同类型的实体在向量空间中会聚集到比较相近的区域,如关系 president 对应的头部实体通常为人名,尾部实体通常为国家名,人名会较为集中地分布在一个区域,国家名会集中分布在另一个区域。如图 3 所示,聚集在一个区域的实体往往较难区分,容易导致预测错误。因此,我们在替换头尾实体时尽可能选择语义相近的实体,以提高模型的区分度。

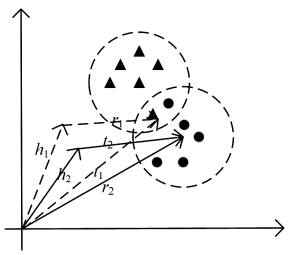

图 3　易混淆的三元组示例

Fig. 3　Instance of indistinguishable triplet

判断实体之间的相似度时,需要首先理解实体的语义信息,传统的字符级相似度计算方法的效果并不好。在知识图谱中,实体的属性可以作为相似度判断的主要依据。在分布式表示模型中,实体或者关系之间的语义相似度通常使用向量之间的相似度表示[18]。定义实体的相似度为:

$$dis(h,h')=\sqrt{\sum_{i=1}^{k}(h_i-h_i')}$$

因此,给定一个黄金三元组(h,r,t),替换头实体生成负例三元组(h',r,t)时,选择 h'使得 $dis(h,h')$最小;替换尾实体生成负例三元组(h,r,t')时,选择 t'使得 $dis(t,t')$最小。

在模型训练过程中,为了区分正确三元组和错误三元组,采用如下基于边际的排名损失函数作为训练模型的优化目标函数:

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

其中,S 是正确三元组所属集合,S^- 为错误三元组所属集合,$\max(x,y)$表示返回 x 和 y 之间较大的值,γ 为正确三元组损失函数得分与错误三元组损失函数得分之间的距离,4.2 节将详细介绍构造 S^- 的细节。因此,该目标函数的优化目标就是最大程度地将正确的三元组和错误的三元组分离。

在最小化目标函数 L 的过程中,我们考虑了下面的约束条件,主要包括:

$$\forall h,t\in E,\ \|\boldsymbol{h}\|_2\leqslant 1,\ \|\boldsymbol{t}\|_2\leqslant 1 \tag{1}$$

$$\forall r\in R,\ |\boldsymbol{w}_r^{\mathrm{T}}r|/\|\boldsymbol{r}\|_2\leqslant\varepsilon \tag{2}$$

$$\forall r\in R,\ \|\boldsymbol{r}\|_2=1 \tag{3}$$

限制条件(1)保证实体向量长度不超过 1,限制条件(2)保证翻译向量 $\boldsymbol{r}$ 在超平面上,限制条件(3)保证超平面为单位法向量。

具体训练模型时,采用随机梯度下降(Stochastic Gradient Descent)法来优化上述目标函数。对知识图谱中的三元组进行多次随机遍历,当一个正确三元组被访问到时,会同时生成负例三元组。一个 mini-batch 后,梯度及模型的参数会自动进行更新。算法 1 给出了 STransH 的学习算法。

算法 1　Learning STransH

input: Training set $S=\{(h,r,t)\}$, entities and rel. sets E and R, margin γ, embeddings dim. k.

1. initialize r← uniform$(-\frac{6}{\sqrt{k}},\frac{6}{\sqrt{k}})$ for each r∈R
2. r←r/‖r‖ for each r∈R
3. e←uniform$(-\frac{6}{\sqrt{k}},\frac{6}{\sqrt{k}})$ for each entity e∈E
4. loop
5. e←e / ‖e‖ for each entity e∈E
6. S_{batch}←sample(S,b) // sample a minibatch of size b
7. T_{batch}←∅ // initialize the set of pairs of triplets
8. for $(h,r,t)\in S_{batch}$ do
9. (h',r,t')←sample($S'_{(h,r,t)}$) // sample a corrupted triplet
10. $T_{batch}\leftarrow T_{batch}\cup\{((h,r,t),\frac{tph}{tph+hpt}(h',r,t),\frac{hpt}{tph+hpt}(h,r,t'))\}$
11. end for
12. Update embeddings w. r. t $\sum_{((h,l,t),(h',l,t'))\in T_{batch}}\nabla[\gamma+d(\boldsymbol{h}+\boldsymbol{r},\boldsymbol{t})-d(\boldsymbol{h}'+\boldsymbol{r},\boldsymbol{t}')]_+$
13. end loop

4　实验

本节介绍 STransH 模型的实验,分别在 WN18 和 FB15K 数据集上进行链接预测和三元组分类,这两项工作从不同角度评估了模型的性能。首先介绍这两项工作的评价指标和实验结果;然后分析 STransH 模型的实验效果,并与其他知识表示方法进行对比。

4.1　实验数据

实验数据是从 Wordnet 和 Freebase 中抽取的:WN18,WordNet 的子集;FB15K,FreeBase 中一个相对稠密的子图(其统计资料如表 1 所列)。

表 1　数据集的统计

Table 1　Statistics of datasets

Data Set	Entities	Relationships	Train	Valid	Test
WN18	40943	18	141442	5000	5000
FB15K	14951	1345	483142	50000	59071
WN11	38696	11	112581	2609	10544
FB13	75043	13	316232	5908	23733

4.2 链接预测

链接预测以预测一个关系事实元组(h,r,t)中缺失的h或者t为目标。在该任务中,对于缺失实体的每个位置,系统从知识图谱中对一组候选实体进行排序,而不是仅仅给出一个最好的结果。

(1)评价指标。为了更好地与 TransE 等模型进行对比,本文采用与 TransE 相同的评价准则。对于每一个三元组(h,r,t),将尾实体用知识图谱中的每一个实体e替换,同时用评分函数计算替换后的三元组(h,r,e)的得分,并据此对这些实体进行降序排列。同理,得到替换头实体的三元组的得分。

将所有的测试三元组进行综合,有 2 个评价指标:1)正确实体的平均排名,记为 *MeanRank*;2)正确实体排在前 10 名的概率,记为 *Hits*@10。事实上,如果一个损坏的三元组在知识图谱中存在,即该三元组实际上是正确的,将其排在原始三元组之前也是合理的。为了消除这种因素的影响,实验中在得到每一个测试三元组的排名得分之前,将上述产生"干扰"的损坏三元组从训练集、验证集和测试集中去除,从而保证了该损坏的三元组不属于任何数据集。该设置称为"Filt"。将未经上述处理的实验设置称为"Raw"。更低的 *MeanRank* 和更高的 *Hits*@10 意味着更好的实验结果。

(2)实验实现。将本文方法与几种现有的方法进行比较,包括 SE[5],SLM[6],TransE[8],TransH[10]。由于实现和参数调整的问题,我们没有得到相应文献中的最好结果。因为实验所用数据集相同,我们直接使用了各个模型在相应文献中的最优实验结果作为对比依据。为了减少参数随机初始化对结果造成的影响,对每一组参数都进行 10 次实验,并取其平均值作为最终的结果。训练 STransH 时,在随机梯度下降过程中使用了{0.001,0.005,0.01}中的学习率α、在{0.25,0.5,1,2}中的边际γ、在{50,100}中的嵌入维度k,以及在{20,75,1 200,4 800}中的 batch 的大小B。最佳的参数由验证集确定。

用"unif"表示传统的等概率替换头实体或者尾实体的方式,用"bern"表示使用伯努利抽样策略的方法,即用不同的概率分别来替换头实体和尾实体。在 unif 设置下,最佳配置为:在 WN18 上,$\alpha=0.01$,$\gamma=1$,$k=50$,$B=75$;在 FB15K 上,$\alpha=0.005$,$\gamma=0.5$,$k=100$,$B=4\,800$。在 bern 设置下:在 WN18 上,$\alpha=0.01$,$\gamma=1$,$k=50$,$B=1\,200$;在 FB15K 上,$\alpha=0.005$,$\gamma=0.25$,$k=100$,$B=4800$。对于这两个数据集,本实验将所有训练三元组迭代 1000 次。

(3)实验结果。通过观察发现,在 WN18 数据集上,TransE 和 STransH 的 *MeanRank* 比其他方法都要好。这可能是因为 WN18 中关系的数量比较少,所以忽视掉不同类型的关系也是合理的。在 FB15K 数据集上,STransH 的表现比其他方法好。在 *Hits*@10 这一指标上,与 TransE 和 TransH 相比,STransH 在 WN18 上分别提高了 0.5%和 7.3%,在 FB15K 上分别提高了 19.4%和 2.1%,性能提升明显。

表 2 链接预测实验结果

Table 2 Results of link prediction

Method	WN18				FB15K			
	MeanRank		*Hits*@10/%		*MeanRank*		*Hits*@10/%	
	Raw	Filt	Raw	Filt	Raw	Filt	Raw	Filt
Unstructured	315	304	35.3	38.2	1074	979	4.5	6.3
RESCAL	1180	1163	37.2	52.8	828	683	28.4	44.1
SE	1011	985	68.5	80.5	273	162	28.8	39.8
SME(Linear)	545	533	65.1	74.1	274	154	30.7	40.8
SME(Bilinear)	526	509	54.7	61.3	284	158	31.3	41.3
LFM	469	456	71.4	81.6	283	164	26.0	33.1
TransE	263	251	75.4	89.2	243	125	34.9	47.1
TransH	401	388	73.0	82.3	212	87	45.7	64.4
STransH(unif)	364	352	76.2	89.5	204	83	46.6	68.3
STransH(bern)	347	330	77.1	90.6	196	68	46.6	69.5

为了证实 STransH 能够更好地处理复杂关系,深入分析了不同关系类型的实验结果。在 1345 个关系中,24%的关系是 1-1 的,23%的关系是 1-*n* 的,29%的关系是 *n*-1 的,24%的关系是 *m*-*n* 的。实验结果如表 3 所列,相比于 TransE 模型,STransH 在复杂关系类型上确实有明显改善。

表 3 FB15K 各类关系的 *Hits*@10 值

Table 3 *Hits*@10 of each type of relations in FB15K

(单位:%)

Method	Predicting Left				Predicting Right (Hits@10)			
	1-1	1-*n*	*n*-1	*m*-*n*	1-1	1-*n*	*n*-1	*m*-*n*
Unstructured	34.5	2.5	6.1	6.6	34.3	4.2	1.9	6.6
SE	35.6	62.6	17.2	37.5	34.9	14.6	68.3	41.3
SME	30.9	69.6	19.9	38.6	28.2	13.1	76.0	41.8
TransE	43.7	65.7	18.2	47.2	43.7	19.7	66.7	50.0
TransH	66.8	87.6	28.7	64.5	65.5	39.8	83.3	67.2
STransH(unif)	76.8	88.1	35.5	68.4	73.6	42.1	85.3	70.2
STransH(bern)	76.7	88.2	35.8	68.1	73.6	42.4	85.2	70.6

4.3 三元组分类

三元组分类用于确定一个给定三元组(h,r,t)是否正确,其主要任务是对一个三元组进行"正确"或"错误"的二元分类。对于一个三元组(h,r,t),如果其得分大于给定的阈值σ_r,那么预测为错误,反之则正确。σ_r由验证集获得最大分类精度时的阈值决定。

实验时首先采用了 WordNet 的子集 WN11 和 FreeBase 的子集 FB13;由于 WN11 和 FB13 包含的关系非常少,因此我们也使用了包含更多关系的 FB15K。实验数据集的统计信息见表 1。

(1)评价准则

三元组分类任务使用准确率作为评价指标,计算方法如下所示:

$$ACC=\frac{T_p+T_n}{N_{\text{pos}}+N_{\text{neg}}}$$

其中,T_p表示预测正确的正例三元组个数;T_n表示预测正确的负例三元组个数;N_{pos}和N_{neg}分别表示训练集中的正例三元组和负例三元组的个数。*ACC* 越高,表示模型在三元组分类这一任务上的效果越好。

(2)实验实现

在 SGD 过程中,选择了{0.1,0.01,0.001}中的学习率、{1,2,4}中的边际γ,实体向量和关系向量的维度均从{20,50,100}中选取,*batch_size* 从{20,120,480,960,4 800}中选取。最佳配置的精度由验证集确定。WN11 上的最佳配置为:$\alpha=0.001$,$\gamma=2$,$k=100$,$B=4800$,并且使用 L_1 作为相似

性度量；FB13 上的最佳配置为：$\alpha=0.001$，$\gamma=1$，$k=100$，$B=4800$，并且使用 L_1 作为相似性度量；FB15K 上的最佳配置为：$\alpha=0.01$，$\gamma=1$，$k=100$，$B=4800$，并且使用 L_1 作为相似性度量。

(3)实验结果

表 4 列出了三元组分类的评估结果。可以看出，在 WN11 和 FB13 上，STransH 模型比 TransE 和 TransH 方法好；而在 FB13 上，NTN 模型同样表现出色。但是在 FB15K 数据集上，TransE 和 STransH 的表现更加出色，说明本文模型更适用于大规模知识图谱。这是因为，FB13 中只有 13 个关系，NTN 模型在 FB13 上通过张量分解建模具有一定优势。但是，FB15K 相对稀疏，NTN 不再适用。

表 4　不同模型的三元组分类精度

Table 4　Accuracy of triplet classification of different models

(单位：%)

Method	WN11	FB13	FB15K
Distant	53.0	75.2	—
SLM	69.9	85.3	—
SME	73.8	84.3	—
NTN	70.4	87.1	66.5
TransE	75.87	81.5	79.7
TransH	78.80	83.8	87.7
STransH(unif)	79.5	85.3	89.2
STransH(bern)	79.6	85.2	89.6

结束语　本文提出了一种新的知识图谱嵌入模型 STransH，所提方法主要有 3 方面贡献。

1)将 SE 模型和 TransE 模型结合，并且采用单层神经网络的非线性操作来精确刻画实体与关系的语义联系。

2)在 STransH 模型中应用了面向关系的超平面的投影思想，将头尾实体映射至给定关系的超平面加以区分。

3)在基于 WordNet 和 FreeBase 的大规模真实数据集上进行链接预测和三元组分类这两项任务。实验结果表明，STransH 取得了更优的效果，且未增加模型的复杂度和训练难度，可以应用到大规模知识图谱补全和推理等任务上。

未来我们将 STransH 模型进行进一步改进，将关系路径考虑在内；另外，除了将 STransH 模型用于链接预测和三元组分类等任务，还计划将其用于知识推理、关系抽取等任务中。

参 考 文 献

[1] MILLER G. Wordnet-a Lexical Database for English [J]. Communications of the Acm, 1995, 38(11): 39-41.

[2] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge[C]// Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data. New York: ACM, 2008: 1247-1250.

[3] BENGIO Y, COURVILLE A, VINCENT P. Representation learning: a review and new perspectives [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2013, 35(8): 1798-1828.

[4] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed Representations of Words and Phrases and their Compositionality [J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.

[5] 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. Palo Alto: AAAI Press, 2011: 301-306.

[6] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]// Neural Information Processing Systems 2013. Lake Tahoe: NIPS, 2013: 926-934.

[7] 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.

[8] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating Embeddings for Modeling Multi-relational Data[C]// Proceedings of Neural Information Processing Systems 2013. Massachusetts: MIT Press, 2013: 2787-2795.

[9] NGUYEN D Q, SIRTS K, QU L, et al. STransE: a novel embedding model of entities and relationships in knowledge bases[C]// Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics. San Diego: ACL, 2016: 460-466.

[10] WANG Z, ZHAN J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes[C]// Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2014: 1112-1119.

[11] JIANG T W, QIN B, LIU T. Open Domain Knowledge Reasoning for Chinese Based on Representation Learning[J]. Journal of Chinese Information Processing, 2018, 32(2): 34-41. (in Chinese)
姜天文，秦兵，刘挺. 基于表示学习的开放域中文知识推理 [J]. 中文信息学报，2018，32(2)：34-41.

[12] LIN Y, LIU Z, ZHU X, et al. Learning entity and relation embeddings for knowledge graph completion[C]// Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2015: 2181-2187.

[13] LIN Y, LIU Z, LUAN H, et al. Modeling Relation Paths for Representation Learning of Knowledge Bases[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon: ACL, 2015: 705-714.

[14] XIAO H, HUANG M, HAO Y, et al. TransA: An adaptive approach for knowledge graph embedding [J]. arXiv: 1509.05490, 2015.

[15] 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: ACL, 2015: 687-696.

[16] JENATTON R, ROUXN L, BORDES A, et al. A latent factor model for highly multi-relational data[C]// Proceedings of Neural Information Processing Systems 2012. Massachusetts: MIT Press, 2012: 3167-3175.

[17] NICKEL M, TRESP V, KRIEGELH P. A Three-Way Model for Collective Learning on Multi-Relational Data[C]// Proceedings of the 28th International Conference on Machine Learning. New York: ACM, 2011: 809-816.

[18] AN B, HAN X P, SUN L, et al. Triple Classification Based on Synthesized Features for Knowledge Based[J]. Journal of Chinese Information Processing, 2016, 30(6): 84-89. (in Chinese)
安波，韩先培，孙乐，等. 基于分布式表示和多特征融合的知识库三元组分类 [J]. 中文信息学报，2016，30(6)：84-89.