

基于符号语义映射的知识图谱表示学习算法

杨晓慧^{1,2} 万睿¹ 张海滨¹ 曾义夫¹ 刘峤¹

¹(电子科技大学信息与软件工程学院 成都 610054)

²(中电科大数据研究院有限公司 贵阳 550008)

(yangxhui@std.uestc.edu.cn)

Semantical Symbol Mapping Embedding Learning Algorithm for Knowledge Graph

Yang Xiaohui^{1,2}, Wan Rui¹, Zhang Haibin¹, Zeng Yifu¹, and Liu Qiao¹

¹(School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054)

²(CETC Big Data Research Institute Co., Ltd., Guiyang 550008)

Abstract Learning graph embedding is a crucial research issue in the field of statistical relational learning and knowledge graph population, and it is important for the construction and application of knowledge graph in recent years. In this paper, we perform a comparative study of the prevalent knowledge representation based reasoning models, with detailed discussion of the general potential problems contained in their basic assumptions. A semantical symbol sensory projection based neural network model is proposed in order to learn graph embedding, whose basic idea is to utilize the recurrent neural network for encoding the compositional representation of symbol strings (composition of entity-relation) onto their target grounded symbol according to the existing relational data in knowledge. In addition, we introduce the inverse image of the relations into the system to deal with the symmetric/asymmetric properties of the relations, which makes the proposed model more adaptable to different types of reasoning tasks on a variety of homogeneous and heterogeneous networks than other solutions. The proposed model is suitable for large scale knowledge graph representation learning. Experimental results on benchmark datasets show that the proposed model achieves state-of-the-art performance on both of the knowledge based completion benchmark tests and the graph based multi-label classification tasks.

Key words representation learning; graph embedding learning; reasoning; link prediction; multi-label classification; knowledge graphs

摘要 图的分布式表示对于知识图谱的构建与应用任务至关重要。通过对当前流行的图表示学习模型进行比较,分析了现有模型存在的不合理之处,据此提出了一个基于符号语义映射的神经网络模型用于学习图的分布式表示,基本思想是依据知识图谱中已有的实体关系数据,采用循环神经网络对符号组合(实体-关系组合)进行语义编码,并将其映射到目标符号(实体)上。此外,通过为图中的每个关系类型

收稿日期:2018-04-02;修回日期:2018-06-01

基金项目:国家自然科学基金项目(61772117,61133016);装备预研领域基金项目(61403120102);四川省科技厅高新技术及产业化面上项目(2017GZ0308)

This work was supported by the National Natural Science Foundation of China (61772117, 61133016), the General Equipment Department Foundation (61403120102), and the Sichuan Hi-Tech Industrialization Program (2017GZ0308).

通信作者:刘峤 (qliu@uestc.edu.cn)

引入一个逆关系镜像,解决了关系的非对称性问题,使模型能够适应多种不同类型的(同构或异构)网络的关系推理任务.该模型适用于大规模知识图谱的表示学习任务.在公开数据集上的实验结果表明,该模型在知识图谱扩容任务和基于图的多标签分类任务上的性能表现优于相关工作.

关键词 表示学习;图嵌入学习;推理;链路预测;多标签分类;知识图谱

中图分类号 TP391

本文研究的是知识图谱上分布式表示学习的问题.知识图谱上表示学习的目标是将多关系数据中的实体和关系分别映射到低维的向量空间中,以促进知识图谱上的关系学习任务的发展^[1-2].机器学习领域的最新研究高度强调了符号数据的分布式表示(distributed representations)学习对于各类人工智能任务的重要性^[1-3],例如协同过滤^[4]、LBS服务^[5]、基于知识图谱的信息检索^[6]和文本概念建模^[7]等.然而由于知识图谱中本体与概念的多样性,知识图谱的表示学习仍然是一个具有挑战性的问题^[8].因此,能够为大规模知识图谱生成一个精确合理的表示具有重要的意义,因为它有助于从知识图谱中有限的已知概念推理得到未观测到的事实,揭示知识之间的区别与联系,催生新的应用^[9].

知识图谱是基于资源描述框架(resource description framework, RDF)建立的语义知识库,其中知识以(实体,关系,实体)三元组的形式表示与存储,实体间通过关系相互联结,构成网状知识结构^[10].本文使用三元组(head, rel, tail)来表示事实,head, rel和tail分别表示头实体、关系和尾实体.

当前的研究工作基于其基本假设的不同可以分为2类:矩阵分解模型(matrix factorization model, MFM)和随机游走模型(random walk model).矩阵分解模型试图将图的矩阵表示分解为节点和边的向量表示(称为embedding)^[11];而随机游走模型将网络建模为一系列有序节点序列组成的“文档”,然后用文本的词向量模型对其进行处理^[12].虽然基于这2种建模思路的算法在标准数据集的测试中获得了良好的效果(具体的细节将在第3节详细讨论),但是在多关系图数据上的表示学习研究中仍存在一些亟待解决的问题.例如,这些模型对于稀疏不平衡的数据的性能表现不稳定,泛化能力较差,模型理论基础不完备^[13].

为了解决上述问题,本文采用了基于符号语义映射(semantical symbol mapping)理论的建模框架^[14],该理论将实体和关系的唯一标识看作是基本的符号表示(symbol representation).为了让机器能够区

分实体或关系,需要为它们构建相应的标识表示(iconic representation),即实体和关系的embedding,而将实体关系组合映射后得到的组合表示视为分类表示(categorical representation).为了实现对知识图谱中知识的精准建模,将建立分类表示和对对应目标标识表示之间的映射关系.

本文的目的是研究一种能够灵活地对多种类型知识图谱进行建模的高效表示学习算法,并提出一些新的见解与证据以促进表示学习的进一步研究与发展.为了实现该目标,最重要的是定义一个合适的优化问题.针对上述问题,本文提出了一种基于符号语义映射的表示学习算法.受序列学习和机器翻译最新成果的启发^[15],本文在模型中引入了一个编码器-解码器(encoder-decoder)的框架来学习:1)实体和关系的联合表示空间;2)一个新颖的关系推理模型,将组合分类表示解码为标识表示的概率分布,从而将类别中的成员与非成员区分开.

本文的主要贡献有2个方面:

1)首次从符号语义映射的角度对图的表示学习问题进行建模,所提出的模型是首个基于循环神经网络(recurrent neural network, RNN)的表示学习和关系推理模型,实验结果表明所提算法的综合性能表现显著优于当前主流的相关工作;

2)将逆关系镜像的训练机制引入到模型中,通过区分实体在三元组中的不同位置,使模型能够处理关系的对称性/非对称性.

1 相关工作

近年来,分布式表示学习吸引了许多不同领域研究人员的关注,如推荐系统^[16]、知识图谱^[17]、复杂网络^[18]和生物信息学^[19],因为这是学习如何在现实世界网络中进行预测的第1步^[20-21].如引言所述,现有的方法可以分为2类:随机游走模型和矩阵分解模型,它们分别代表了对物理网络建模的不同选择.

随机游走模型将网络建模为一系列有序节点序列构成的“文档”,这些序列是通过随机游走在过程从

底层的网络中采样得来, Skip-gram 模型通常被用来学习序列中实体的 embedding^[12]. 随机游走模型被广泛应用于复杂网络研究领域, 它具有可在线学习的优势, 这意味着模型可以根据从图中变化区域采样生成的新路径来进行迭代的更新. 然而, 随机游走模型的采样策略可能非常棘手且计算复杂度较高, 实际上并没有明确的采样策略适用于所有的实际情况且不会产生冲突. 因此, 随机游走模型容易受到采样策略效率低的影响, 以及遭受过拟合与欠拟合的情况^[22]. 本文没有采用随机游走的建模思路, 但是 2 个最优的随机游走模型 DeepWalk^[12] 和 Node2vec^[18] 将被选做实验对比的基准实验.

矩阵分解模型(MFMs)在 Netflix 挑战赛使用非常广泛, 模型将网络视为邻接矩阵, 学习的策略则是将邻接矩阵分解为低维的实值向量^[16]. 知识图谱的研究人员对多关系数据推理学习的研究丰富了矩阵分解模型的研究领域^[23]. 本文的工作与 RESCAL^[24], NTN^[25], TransE^[26], ER-MLP^[17], DistMult^[27], HoIE^[11], ComplEx^[28] 和 ConvE^[29] 等模型相关. 这些工作将多关系数据看作一个 3 维的二元张量, 该张量的每一个切片对应一种关系的邻接矩阵. embedding 学习的策略则是用一些组合算子来分解该张量^[13]. 这些模型之间的区别在于它们的组合算子不同, 通常可以分为线性的组合算子和非线性的组合算子.

根据评分函数中的组合算子, RESCAL, TransE, DistMult, HoIE 和 ComplEx 可以划分为线性模型. TransE 和 RESCAL 的得分函数是实体向量相对于关系矩阵的双线性变换^[30]. 而 DistMult 是一种三线张量因子分解模型, ComplEx 是 DistMult 在复数空间中的扩展. HoIE 的组合运算被定义为实体向量的全息卷积, 然而理论证明, 在训练时对初始向量施加一定的约束, 则 HoIE 与 ComplEx 模型是相同的^[31]. 线性模型(如 TransE 和 DistMult)的一个普遍问题是不能很好地处理对称/非对称关系, 而 ComplEx 很好地解决了这个问题, 它考虑了实体的位置(作为头实体或尾实体)对建模的影响, 使用复共轭对不同位置的实体建立不同的表示^[28]. 同时, 为了解决关系对称性的问题, 本文提出的模型为图中每种关系引入了一个“逆关系镜像”, 如第 3 节所示. 使用逆关系镜像机制使模型能够从数据中捕获更多有用的模式, 在图表示学习任务中表现出更好的性能.

与本文提出的模型关联最密切的模型是 Dettmers 等人提出的 ConvE 模型^[29], 它在 embedding 上利用二维卷积和多层非线性特征提取来对知识图谱进行建模. ConvE 的评分函数定义为

$$\text{score}(\text{head}, \text{rel}, \text{tail}) = \sigma(f(\text{vec}([\bar{\mathbf{h}} \oplus \mathbf{r}] * \omega) \cdot \mathbf{W}) \cdot \mathbf{t}), \quad (1)$$

其中, f 定义了一个非线性方程, σ 是一个 sigmoid 函数, $*$ 表示向量对应位置元素相乘, $\bar{\mathbf{h}}$ 和 \mathbf{r} 是 \mathbf{h} 和 \mathbf{r} 的二维变化. $[\bar{\mathbf{h}} \oplus \mathbf{r}]$ 将 $\bar{\mathbf{h}}$ 和 \mathbf{r} 进行拼接作为卷积层的输入, 所得到的特征映射将转化为一个 d 维的向量. 评分函数的输出表示给定一对实体关系对(head, rel)观测到特定的尾实体 tail 的概率. 另一个相关的工作是 ProjE 模型^[32], 这是一个基于 $\text{head} = \text{rel} + \text{tail}$ 假设的 2 层感知机模型. 然而 ProjE 的计算复杂度较高, 性能也不能与 ConvE 和 ComplEx 相比.

1) 联系. ConvE 与本文提出的模型 SSME 有一些相似性, 它们都通过建立一个将概念从 $\{\text{head}, \text{rel}\}$ 组合表示到它的等价对象 tail 的神经映射机制来学习 embedding. 此外, SSME 与 ComplEx 的联系在于它们都明确区分了实体(头实体或尾实体)的语义功能, 以此解决关系的对称/非对称性质对建模带来的影响, 也因此它们比其他的相关工作更适用于广泛的本体定义^[13].

2) 差异. 然而 SSME 和 ConvE 有 2 处主要的区别: ① ConvE 使用了卷积神经网络来对实体关系对的拼接表示进行特征提取, 并通过张量变换操作来实现从特征映射建立语义组合表示. 而 SSME 是一个基于 RNN 的语言模型, 一个 2 跳的 RNN 模型用来对实体-关系对的符号组合进行语义编码, 然后通过一个多层感知机(multilayer perceptron, MLP)进行解码, 将其映射到目标符号上. ② 由于系统中引入了逆关系镜像, SSME 可以在一个统一的模型中完成双向推理, 这使得模型能够更好地利用图中的结构化信息, 特别是在有向图中(而 ConvE 的建模视角是单向的). 实验证明本文提出的模型能有效地捕捉到逆关系镜像与其对应的原始关系的语义关联, 如对称关系(如 antonymy 关系)和非对称关系(如 hyponymy 和 hypernym 关系), 使模型能适应多种不同类型(同构/异构)网络的关系推理任务.

2 符号语义映射模型

在本节中, 首先介绍本文的背景与使用的符号

系统,之后对所提出模型的直观性和算法框架进行讨论.

2.1 符号系统描述

多关系知识图谱可以视为一个3维的二元张量 \mathbf{G} ,其中每一个切片对应一个关系类型 rel 的邻接矩阵 \mathbf{G}_{rel} .如引言所述,用 $(head, rel, tail)$ 来表示一个事实单元,其中 $head, tail \in E$ 分别表示头实体和尾实体, E 表示实体集合, rel 是关系类型集合 R 中的一个关系类别.本文为图中的每一个实体和关系指定了唯一的实值向量.实体和关系的 embedding 位于相同的低维空间中,embedding 的维度为 d .为了区分符号表示和标识表示,用 r 来表示关系 rel 的 embedding 向量.

2.2 基础模型概述

本文的主要动机是设计一种具有适用性和灵活性的分布式表示学习方法,能够建模不同类型网络.为了实现这个目标,本文提出了一个基于符号语义映射的神经网络模型,它能够同时考虑到不同网络中具备的不同特性(例如有向/无向,同质/异质).这意味着,知识图谱将被建模为一个非符号/符号的混合系统,其中基本符号(elementary symbol)指命名实体和关系,被表示成2种非符号的 embedding,例如,给定一个事实(战狼,导演,吴京),其中的基本符号是“战狼”、“导演”和“吴京”.为了让机器能够区分这些输入,需要为它们构建相应的标识表示,即实体和关系的 embedding.然而如果想要学习出(战狼,导演,?)的标识表示,需要首先解决以下2个问题:

1) 如何基于“战狼”、“导演”的标识表示构建一个合适的分类表示?

2) 如何建立一个可靠的感知预测机制将属于(战狼,导演,?)答案集中的成员与其他易混淆的非成员区分开来?

我们从 Seq2Seq 模型中获得了一些灵感^[15],该模型已经被证明从符号序列中生成组合表示是可行的,同时该模型具有强大的解码功能,能够将组合表示直接映射到对应的符号系统中.由此,我们设计出 SSME 的基础模型,模型的目标是为知识图谱中的每一个实体和关系学习一个 d 维的实值向量表示 embedding.图1中给出了 SSME 基础模型的示意图. SSME 的建模思路是利用一个循环神经网络来读取输入的序列 $(head, rel)$,每一个时间步只读取序列中的一个元素,以此来生成一个关于符号组合(实体-关系组合)的语义表示 \mathbf{C} ,在 SSME 中 \mathbf{C} 是一个 d 维的向量.然后利用 MLP 网络建立从组合语

义表示 \mathbf{C} 到目标符号 $tail$ 的感知映射. MLP 本质上是一个以输入序列为条件的 embedding 翻译模型或解码模型.由于输入序列的长度固定为2,为了快速求解,SSME 使用了一个标准的 RNN 来进行编码.当然,也可以使用加门的 RNN(如 LSTM 或 GRU),或者深层(堆叠的 RNN)模型作为编码器.本文采用的编码器-解码器的框架与 Cho 等人和 Sutskever 等人的工作相近^[15],他们的工作清楚地证明了从富含语义的句子中生成信息组合表示的可行性,并且这一思路已成功应用于多种自然语言处理任务,如机器翻译.

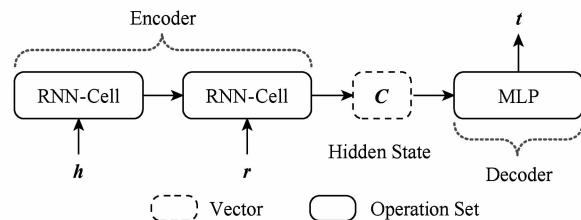


Fig. 1 Schematic illustration of the basic SSME model

图1 SSME 基础模型示意图

2.3 基本模型实现细节

如图1所示,一个2跳的 RNN 被用来编码输入序列 $(head, rel)$ 生成组合表示 \mathbf{C} ,其中神经网络中的单元结构定义如下:

$$s_{time} = f(\mathbf{W} \cdot \mathbf{x}_{time} + \mathbf{U} \cdot s_{time-1} + \mathbf{b}), \quad (2)$$

其中, $\mathbf{W} \in \mathbb{R}^{t \times d}$, $\mathbf{U} \in \mathbb{R}^{k \times k}$ 分别表示相对于当前输入 \mathbf{x}_{time} 和前一步的隐含状态 s_{time-1} 的权重矩阵;需要说明的是,下标 $time$ 表示时间步; $\mathbf{b} \in \mathbb{R}^k$ 定义了一个偏置向量; $f(\cdot)$ 表示一个非线性的激活函数,本文中所有的实验都是使用的 ReLU(rectified linear unit) 激活函数.

第2个时间步的输出状态中 s_{time} 将被用来作为组合表示 \mathbf{C} ,并输入到 MLP 网络中进行解码.解码用的 MLP 网络是一个只包含一层全连接层的前馈网络,如图2解码器部分所示.这个全连接层进行特征学习,负责学习从组合表示 \mathbf{C} 到目标符号(比如 $tail$) 的映射过程,它的输出状态 $\hat{\mathbf{y}}$ 定义如下:

$$score(head, rel, tail) = \hat{\mathbf{y}} = g[f(\mathbf{V} \cdot \mathbf{s}_{time} + \mathbf{b}_1)], \quad (3)$$

其中, $\mathbf{V} \in \mathbb{R}^{|E| \times k}$ 是一个权重矩阵, $\mathbf{b}_1 \in \mathbb{R}^{|E|}$ 是偏置向量. $g(\cdot)$ 定义了一个 softmax 函数,因此 $\hat{\mathbf{y}}$ 是一个长度为 $|E|$ 的 softmax 向量,其第 i 个元素表示 \mathbf{C} 相关的符号等价物是图中第 i 个实体的概率.

给定一个训练样本 $(head, rel, tail)$,关于 $\hat{\mathbf{y}}$ 的交叉熵损失函数被定义为

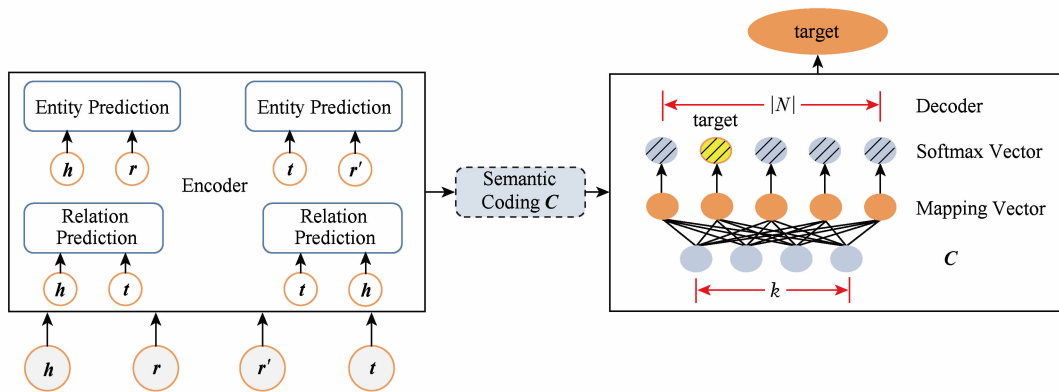


Fig. 2 Network structure illustration of the SSME model

图2 SSME模型网络结构示意图

$$L(\hat{\mathbf{y}}) = - \sum_{i=1}^{|\mathcal{E}|} \mathbf{y}[i] \log \hat{\mathbf{y}}[i] + (1 - \mathbf{y}[i]) \log(1 - \hat{\mathbf{y}}[i]), \quad (4)$$

其中 \mathbf{y} 表示 $tail$ 的独热 (one-hot) 向量. 为了加速随机优化过程, 本文使用小批量 (mini-batch) 训练策略进行训练. 当样本数量为 N 时, 平均交叉熵损失函数被重新定义如下:

$$L(\hat{\mathbf{y}}) = - \frac{1}{N} \sum_{i=1}^N \log(\hat{\mathbf{y}}[tail_i]), \quad (5)$$

其中, i 是当前训练批次中的第 i 个样本, $tail_i$ 则表示第 i 个样本的目标标签. 一个迭代随机梯度下降优化器将被用于优化上述的目标交叉熵损失函数.

从式(4)和(5)可以看出, 本文提出的模型可以用无监督的方式进行训练. 模型并不需要手动标记数据, 因为从知识图谱中获取的事实都可以看作是“自标记”的, 因此可以直接用于模型训练.

2.4 SSME 模型

本节我们将介绍 SSME 模型. 为了使如图 1 所示的基础模型适用于各种情况, 我们对其进行了如下的扩展:

1) 为了处理同一个映射框架下的实体预测和关系预测任务, 我们为 SSME 基础模型引入了一个附加的符号溯源任务模块, 它主要负责学习从组合表示 ($head, tail$) 到目标关系类型 rel 的映射模式. 为了与上述基础模型进行区分, 我们称这个模块为关系预测模块, 称基础模型为实体预测模块, 均作为模型的编码器部分, 基本结构如图 2 编码器部分所示.

这 2 个模块的结构非常相似, 但关系预测模块的输出向量大小为 $|R|$. 2 个模块的参数是独立于彼此的, 这使得模型可以从不同角度来构造映射方式. 模型的损失函数定义如下:

$$L(head, rel, tail) = L\{\hat{\mathbf{y}}(tail | head, rel)\} + L\{\hat{\mathbf{y}}(rel | head, tail)\}, \quad (6)$$

其中 $L\{\hat{\mathbf{y}}(tail | head, rel)\}$ 和 $L\{\hat{\mathbf{y}}(rel | head, tail)\}$ 是每个模块的平均交叉熵损失, 如式(5)定义. 它们是可以叠加的, 因为交叉熵损失可以解释为对于给定输入观察到目标的概率的负对数似然. 在模型训练时, 2 个模块的参数将根据式(6)同时进行反向传播更新.

2) 为了解决关系的非对称特性以及解决在同一 embedding 框架下进行头实体预测和尾实体预测的任务, 我们将关系的逆镜像引入系统. 这表明对于每一个 $rel \in R$, 它的逆关系镜像 rel' 将被加入到符号系统中. SSME 模型的优势是只需在训练时为每个 ($head, rel, tail$) 加入反向事实 ($tail, rel', head$), 就能将所有的 embedding (包括逆关系镜像 rel') 训练到, 而不需要再次训练调整. 通过这种方式, SSME 能够以“逆向思维”从事实中进行学习, 这与人类在学习新事物时很相似, 我们会在大脑中创建一个角度不断变化的场景来帮助理解事物. 更重要的是, SSME 模型能够处理关系的对称性/非对称性. 如 4.2 节所示, SSME 学习得到的分布式表示 embedding 能够有效地区分关系与其逆关系镜像的语义和语境的差别.

关系的对称特性在知识图谱中很常见. 在解决关系学习问题时, 关系的不同特性会导致不同类型的推理方式^[26]. 然而正如 Trouillon 等人^[13]指出的, 大多数流行的图表示学习算法缺乏对关系的对称性/非对称性精确建模的能力, 因此它们不能在不产生冲突的情况下适用于所有的实际情况. ComplEx 是一个例外, 它利用复共轭使实体在处于不同位置时拥有不同的表示来解决关系对称性的问题^[28].

在本工作中, 作者认为将实体表示限定在实数域比在复数域中更有利于下游应用的建模任务和计算分析, 例如实体链接和图挖掘. 操作实值向量比操作复数向量更方便, 也更符合 Mikolov 等人^[33]和 Pennington 等人^[6]对分布式表示的研究与定义, 即从文本语料库中学习单词和短语的实值分布式表示, 这已经被广泛地接受并应用于各种自然语言处理任务.

综上所述, 本文提出的 SSME 模型具有如下优点: 1) 本文提出的模型为实体预测和关系预测任务提供了一个统一的表示学习框架, 它可以综合 2 个不同任务提供的信息优势, 因此也可以更好地利用知识图谱的结构信息; 2) 通过为知识图谱中的每个关系引入逆关系, 模型可以解决实数域中关系对称性/非对称性建模的问题, 这个机制可以灵活地支持各种结构图的学习模型, 并且可以应用到广泛的下游应用中.

3 实验结果与分析

本节在 2 类不同的图表示学习任务上验证 SSME 模型的有效性, 任务的数据集取自不同的研究领域, 表 1 和表 2 给出了相应的统计信息. 本节还对 SSME 模型的优势和局限进行了详细的讨论与分析.

Table 1 Statistics of the Knowledge Graph Data Sets

表 1 知识图谱数据集的统计信息

Dataset	# Relation	# Entity	# Train	# Valid	# Test
WN18	18	40 943	141 442	5 000	5 000
WN18RR	11	40 943	86 835	3 034	3 134
FB15K	1 345	14 951	483 142	50 000	59 071
FB15K-237	237	14 541	272 115	17 535	20 466

Table 2 Statistics of the Complex Network Data Sets

表 2 复杂网络数据集的统计信息

Dataset	# Nodes	# Edges	# Categories	# Labels
BlogCatalog	10 312	333 983	39	14 476
PPI	3 890	38 292	50	6 640
Wikipedia	4 777	184 812	40	6 770

3.1 实验数据

首先在知识图谱扩容 (knowledge base completion, KBC) 任务上对 SSME 模型进行评估, 使用的数据集是基准数据集 FB15K 和 WN18^①, 以及它们的扩展数据集 FB15K-237 和 WN18RR^②. FB15K 是 Freebase 知识库的子集, 其知识来源是维基百科, FB15K 中的数据大部分与电影和体育主题相关. WN18 数据集是从 WordNet 知识库中采样得到的, 它主要包括语义概念和词汇关系, 并且实体以严格的等级方式进行组织. 这些数据集已经被用作比较评估的基准数据集. 然而 Toutanova 等人指出, FB15K 和 WN18 的测试集中包括大量与训练集中三元组相反的三元组, 即 (*head*, *rel*₁, *tail*) 属于训练集, 存在 (*tail*, *rel*₂, *head*) 属于测试集. 为了消除这种情况对实验结果的影响, Toutanova 等人引进了 FB15K-237, 去掉了 FB15K 中具有相同实体对的三元组^[34]. 同时本文也将在 Dettmers 等人提供的 WN18RR 数据集上进行测试^[29].

其次, 本文也通过在复杂网络研究领域的基于图的多标签分类 (multilabel classification, MLC) 任务评估 SSME 的有效性, 本文使用的基准实验数据集包括 BlogCatalog, protein-protein interaction (PPI) 和 Wikipedia^③. BlogCatalog 是一个从 BlogCatalog 网站上采样下来的社交网络, 是由博主之间的社交联系构成, 其中博主的标签代表了他们感兴趣的主体类别. 而 PPI 网络是一个生物网络, 采样于智人蛋白质交互作用网络, 其中节点 (蛋白质) 间的关系表示蛋白质间是否存在交互作用. Wikipedia 是维基百科中词共现网络, 其中节点的标签是 Grover 等人利用由斯坦福的 POS-Tagger 工具 (part-of-speech, POS) 推断的词性标注. 在 3 类网络数据的训练集中, 每个节点拥有一个或多个标签, 多标签分类任务则是预测测试集中节点的标签.

3.2 实验设置

第 1 个任务是为了验证 SSME 模型在知识图谱扩容任务上的有效性. 为了进行实验对比, 本文重现了相关工作的实验, 将其结果作为对比的基准, 包括现在性能最好的 2 个模型 ComplEx 和 ConvE, 以及在当前文献中引用最广泛的模型 TransE 和 DistMult.

① <https://everest.hds.utc.fr/doku.php?id=en:transe>

② <https://github.com/TimDettmers/ConvE>

③ <http://snap.stanford.edu/node2vec/>

在知识图谱扩容领域, $P@N$ 指标是性能分析中最常用的评估指标, 它表示被预测的事实出现在预测结果列表前 N 的概率. 本文报道了过滤后的 $P@N$ 指标, 即在统计命中率时过滤掉所有的已知事实.

本文所提出的 SSME 模型包含 4 个超参数, 需要在训练之前加以确定: SGD 算法的学习率 η 、RNN 隐含层的维度 k 、dropout rate p 以及向量表达空间的维度 d . 实验采用网络寻优法 (grid search) 对上述参数进行选择. 参考相关工作的实验参数, 设定 η 的搜索范围为 $\{0.0001, 0.001, 0.1, 1, 5, 10\}$, p 的搜索范围为 $\{0.0, 0.1, 0.2, 0.3\}$, k 的搜索范围为 $\{256, 512, 1024, 2048\}$, d 的范围为 $\{50, 100, 150, 200\}$. 在 4 个数据集上分别进行网络寻优, 取 $P@10$ 指标最优时的参数组合作为本文的实验参数: $\{d: 200, p: 0.2, \eta: 5.0, k: 512\}$. 训练的批量设置为 256, SGD 算法迭代次数统一设置为 50 轮.

第 2 个测评实验的目的是验证 SSME 模型在大型复杂网络中的 embedding 学习能力, 模型将在

多标签分类任务上与 DeepWalk 模型和 Node2vec 模型等最优的模型进行对比. 为了公平起见, 本文也将报道其他多关系表示学习模型的实验结果.

为了与相关工作保持一致, 我们抽取部分标记节点数据作为训练集, 其余部分用于测试. 这个过程重复了 9 次, 即训练比率为 $10\% \sim 90\%$. 测试时, 使用 Micro-F1 和 Macro-F1 指标作为评判标准. 采样过程具有随机性, 所以相同比率训练集的实验将重复进行 10 次, 实验结果取平均值. 对于多标签分类任务, 本文使用一个简单的感知机进行测试, 感知机的隐含层维度 $k=128$, 学习率为 0.1. 通过网格寻优获取模型的参数, 训练过程在 Macro-F1 达到最优时停止.

3.3 知识图谱扩容任务

第 1 组实验与知识图谱扩容任务相关, 我们在实体预测任务上测试了本文提出的模型和其他相关模型. 所有数据集上的实验结果由表 3 给出, 表 3 中每列指标中性能表现最好的一项用粗体标出.

Table 3 Entity Prediction Results on WN18, FB15K, WN18RR and FB15K-237 Data Sets

表 3 WN18, FB15K, WN18RR 和 FB15K-237 数据集上实体预测的实验结果

%

Dataset	TransE			DistMult			ComplEx			ConvE			SSME		
	$P@1$	$P@3$	$P@10$	$P@1$	$P@3$	$P@10$	$P@1$	$P@3$	$P@10$	$P@1$	$P@3$	$P@10$	$P@1$	$P@3$	$P@10$
FB15K	36.1	59.0	76.2	54.6	73.3	82.4	59.9	75.9	84.0	67.0	80.1	87.3	79.8	85.2	89.2
FB15K-237	17.6	29.6	44.6	10.7	21.2	37.7	11.9	22.5	38.4	22.0	33.0	45.8	21.7	33.7	48.6
WN18	44.5	85.9	93.8	72.8	91.4	93.6	93.6	94.5	94.7	93.5	94.7	95.5	94.1	94.6	94.9
WN18RR	2.7	33.1	42.7	35.6	36.4	37.1	37.7	41.4	44.0	30.6	36.0	41.1	39.3	41.6	44.3

Note: The best result of each column is highlighted in black boldface.

从表 3 可以看出, 在 4 个数据集上, 相比于其他模型, SSME 在 $P@N$ ($N=1, 3, 10$) 指标上都展现出了极强的优势. 如果考虑在 FB15K 上的 $P@1$ 预测准确度, SSME 的结果比 ConvE 高出 19.10%, 比 ComplEx 高出 33.22%, 优势更加明显. FB15K 采样自 FreeBase, 它是最大的协同知识图谱之一. 实验结果所表现出的积极信号表明 SSME 训练出的 embedding 将帮助提高基于 Freebase 知识库的其他下游应用程序的性能.

SSME 和 ConvE 所表现出的优异性能证明基于符号语义映射的神经网络表示学习模型是合理且有效的. 从实验可以发现, CNN 和 RNN 网络都可以从知识图谱中的事实有效地生成具有代表性的组合表示. 然而, 它们之间的联系仍不清楚, 也很难解释它们是否符合认知. 考虑到 RNN 模型在机器翻译上的成功应用, 我们推测基于 RNN 的解决方案

具有较好前景, 因为它可以进一步扩展以学习任意长度的序列 (由事实构成的链), 以便从链数据中学到更多的知识.

3.4 多标签分类任务

3.3 节评估了 SSME 模型在多关系知识图谱中的有效性, 接下来将考虑同质网络中的表示学习问题. 同质网络中实体间往往仅拥有一种特定的关系, 典型的例子是社交网络和生物网络. 本文使用了 BlogCatalog 网络和蛋白质交互作用网络 (PPI), 以便于与早期的相关工作进行对比, 实验的结果由图 3 展示, 横轴为训练数据占总数据量的比例, 纵轴为模型在相应评价指标上的得分. 从图 3 中可以看出, 与其他性能突出的模型相比, SSME 模型更具有优越性, 特别是在节点之间的联系是客观和科学地测量的 PPI 网络的情况下.

从图 3(a) BlogCatalog 中的 Micro-F1 指标可以

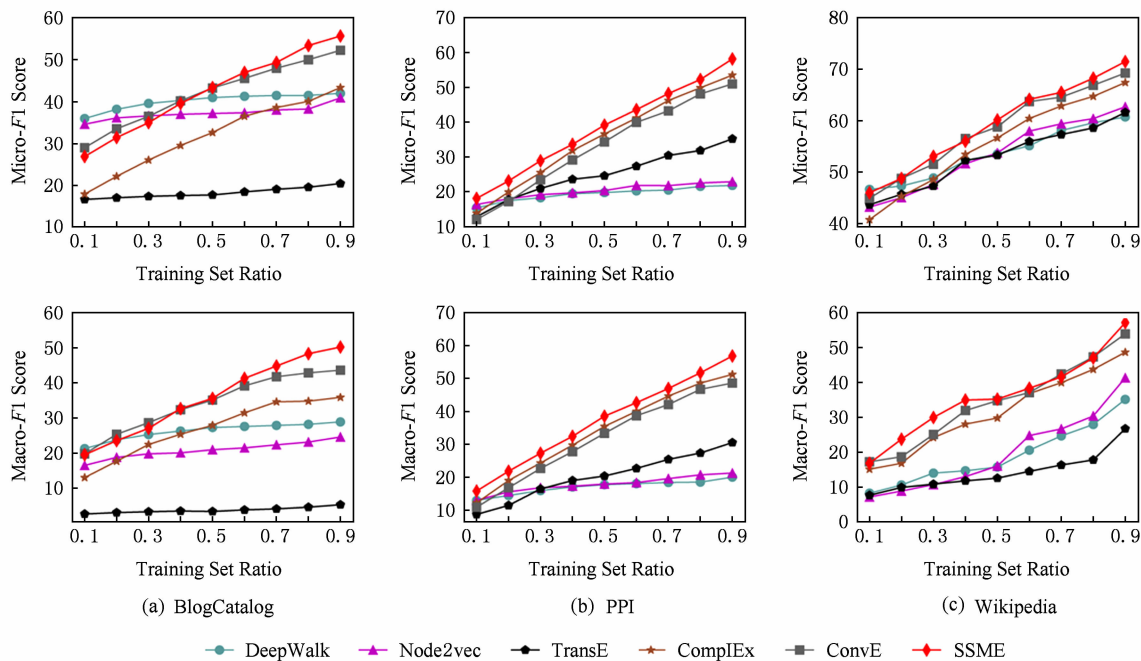


Fig. 3 Multi-label classification results on BlogCatalog, PPI and Wikipedia data set

图3 BlogCatalog, PPI 和 Wikipedia 数据集上的多标签分类结果

看出,标签数据越稀疏 DeepWalk 表现得越好.但是当给出 50% 的数据以后,SSME 模型在 Micro-F1 和 Macro-F1 上的结果都明显优于其他模型.实验结果证明当给定足够多的数据时,本文提出的表示学习模型都表现出更好的性能,在捕捉社交网络的结构信息更具有表现力,对社交网络进行建模时是有效的,因此可以预见在大数据集上 SSME 会有更好的结果.且由于其实现简单,在处理大规模网络时比随机游走模型更高效.值得一提的是,ComplEx 和 ConvE 模型也展现出其相对于其他模型在同质网络上学习表达的灵活性与优越性.其中 ConvE 模型与 SSME 性能相当,这与它们在多关系推理任务上的表现一致.

图 3(b)PPI 中的实验结果则显示,在 PPI 数据集上,SSME 模型始终显著优于其他模型,事实上,给定 30% 的标记蛋白质作为训练数据时,SSME, ConvE 和 ComplEx 的结果已经比给定 90% 数据时 DeepWalk 和 Node2vec 模型的结果要好,而后者代表了复杂网络分析领域中性能最好的图表示学习模型,这说明了本文提出的模型对生物网络具有强大的表示学习能力.同时,本文是相关领域中第 1 次将多关系表示学习模型应用到复杂网络中,作者认为,多关系表示学习模型的强大性能表明了其对生物网络的学习具有足够的灵活性,而这个结果也希望能够为其他的相关应用领域提供一些新的见解.

需要注意的是,PPI 网络比 BlogCatalog 网络更稀疏,图 3 中的结果说明了随机游走模型的性能退化,而本文提出的模型和其他知识图谱嵌入学习模型不受网络稀疏性的影响.这可能是因为这些模型比基于随机游走的方法可以更有效地利用现有的事实,避免通过随机游走过程生成节点序列的需要.因此,在这种 embedding 方案基础上建立的监督模型不易受到全局欠拟合和局部过拟合的影响.

4 案例分析

在本节中,我们设计了一组实验来分析 SSME 与其他相关工作生成的 embedding 的联系与区别,并分析了为每个关系引入逆关系镜像的合理性与有效性.

4.1 embedding 的联系与区别

由于很难解释 SSME 生成的 embedding 的意义,因此本文设计了一组实验分析 SSME 模型与其他当前性能最优的 3 个模型之间的联系与区别,包括 DistMult, ComplEx 和 ConvE. 4 个模型在 FB15K 数据集上生成的 embedding 将作为研究的对象.对于 4 种 embedding 模式,我们分别计算每个模型的 1345 个关系之间的余弦相似度,并对计算结果进行倒序排序,获得一个排序列表,称为 p -列表,列表长度为 $903840 = 1344 \times 1345/2$.为了公平起见,本文没有考虑计算逆关系.

取 p -列表中的前 N 个元素计算相似度比值,表 4 展示了根据 top-500, top-1 000 和 top-2 000 的 p -列表计算出的模型间的相对相似度. 从表 4 中可以看出, SSME 模型训练出的 embedding 与其他的

模型表现出了不同的模式. 以表 4 中的 top-1 000 为例, 可以得到在前 1 000 个最相似的关系对中, SSME 和 ConvE 模型之间重合了 200 对, 这可以看作是对不同 embedding 方案的相对相似度的测度.

Table 4 Relative Similarity of the Models Measured by Overlapping of the top- N Relation Pairs in p -List on FB15K

表 4 根据 FB15K 上的 p -list 中的前 N 关系对的重叠数量测量模型的相对相似实验结果

%

Model	Model											
	SSME			ConvE			ComplEx			DistMult		
	top-500	top-1 000	top-2 000	top-500	top-1 000	top-2 000	top-500	top-1 000	top-2 000	top-500	top-1 000	top-2 000
SSME				9.20	20.00	27.25	0.20	0.26	6.70	8.80	11.70	14.25
ConvE	9.20	20.00	27.25				45.00	36.10	29.75	51.80	50.40	45.40
ComplEx	0.20	0.26	6.70	45.00	36.10	29.75				67.60	60.60	57.15
DistMult	8.80	11.70	14.25	51.80	50.40	45.40	67.60	60.60	57.15			

从表 4 中可以看出, ComplEx 和 DistMult 生成的 embedding 相似度很高, 这也印证了二者的建模方式在本质上是相同的, 但处理实体的 embedding 的方式是不同的 (ComplEx 在 DistMult 的基础上引入了复共轭来解决关系对称性的问题). 从表 4 观察可知, ConvE 与 DistMult, ComplEx 相似度也很高, 而 SSME 与这 2 种模型均不相似.

4.2 逆关系镜像机制的有效性

如 2.4 节所述, 建模实际场景中关系的对称性和非对称性对于提高图学习模型的应用性能至关重要, 因为它可以帮助减少由关系多样性引发的错误. 在此, 本文分析了 WN18 数据集中关系与其逆关系镜像表示的相似度, 以分析关系的对称特性.

首先我们定义一个相似度矩阵 S , S 的第 1 个维度表示 WN18 数据集中的 18 个原始关系 R , 第 2 个维度表示 18 个逆关系镜像 R' . S 中包含所有 $rel_i \in R$ 和 $rel'_j \in R'$ 根据余弦相似度指标计算出的相似度分数, 其中 rel'_j 是关系 rel_j 的逆关系. S 的对角线元素就是关系与其自身逆关系的相似度, 相似度分数可以看作是关系对称性的度量. 给定一个三元组 $(head, rel, tail)$, rel' 的 embedding 是通过相反的事实 $(tail, rel', head)$ 学习出来的. 因此, 如果 rel 与 rel' 高度相似于彼此, 那么从 $(head, rel, tail)$ 推断 $(tail, rel', head)$ 是一个事实是合理的, 即关系 rel 可能是一个对称关系; 否则, $(tail, rel', head)$ 可能不成立, 即 rel 是一个非对称关系. 因此, 关系与其自身逆关系的 embedding 的余弦相似度可以用来衡量或反映关系的对称性/非对称性.

表 5 展示了 WN18 数据集中关系与其逆关系的相似度. 将相似度矩阵 S 对角线上的元素取出, 按

倒序进行排序, 排序的结果在表 5 的第 1 列中展示. 同时相似度分数对应关系的 id 及名称在表 5 的第 2~3 列给出. 例如, rel_2 与其逆关系 rel'_2 的余弦相似度是 0.999. S 矩阵中, 对于大多数列, 最大值不在对角线上, 即与关系最相似的逆关系不是其本身的

Table 5 The Cosine Similarity Between Relations and Their Inverse Images on WN18 Data Set

表 5 WN18 数据集上关系与其逆关系的余弦相似度

sim	rid	Relation Name	Type	rid'	sim'
0.999	rel_2	_derivationally_related_form	$n:n$	rel'_2	0.999
0.988	rel_{14}	_verb_group	$1:1$	rel'_{14}	0.988
0.954	rel_4	_similar_to	$1:1$	rel'_4	0.954
0.727	rel_{17}	_also_see	$n:n$	rel'_{17}	0.727
0.194	rel_{13}	_has_part	$1:n$	rel'_{15}	0.996
0.191	rel_{15}	_part_of	$n:1$	rel'_{13}	0.998
0.163	rel_{12}	_synset_domain_usage_of	$n:1$	rel'_8	0.985
0.160	rel_8	_member_of_domain_usage	$1:n$	rel'_{12}	0.957
0.137	rel_6	_member_holonym	$n:1$	rel'_1	0.998
0.136	rel_1	_member_meronym	$1:n$	rel'_6	0.997
0.031	rel_{10}	_hyponym	$1:n$	rel'_5	0.999
0.030	rel_5	_hypernym	$n:1$	rel'_{10}	0.999
-0.314	rel_0	_member_of_domain_topic	$1:n$	rel'_9	0.987
-0.315	rel_9	_synset_domain_topic_of	$n:1$	rel'_0	0.996
-0.342	rel_7	_instance_hypernym	$n:1$	rel'_{11}	0.993
-0.345	rel_{11}	_instance_hyponym	$1:n$	rel'_7	0.991
-0.440	rel_{16}	_synset_domain_region_of	$n:1$	rel'_3	0.991
-0.445	rel_3	_member_of_domain_region	$1:n$	rel'_{16}	0.976

Note: sim means similarity.

逆关系. 因此, 我们还找出了 S 中每列最大的值, 并记录该值以及其对应的逆关系的 id , 如表 5 第 5~6 列. 例如, 与 rel_{10} 关系(关系名为“_hyponym”)最相似的是 rel_5 (关系名为“_hyponym”)的逆关系是 rel'_5 , rel_{10} 与 rel_5 的余弦相似度是 0.999.

为了更好地理解关系的对称性/非对称性与它们的结构特征之间的联系, 本文计算了 WN18 中 18 个关系的头尾比(hpt)和尾头比(tph), 计算的方式与 Liu 等人提出的方式一样^[35-36]. 根据这个比例将关系分为 4 类: 一对一(1:1)关系、一对多(1:n)关系、多对一(n:1)关系和多对多(n:n)关系^[26], 关系所属类别分类的结果展示在表 5 的“Type”列.

从表 5 可以观察到, 前 4 个关系与它们相应的逆关系高度相关. 从它们的名字和类型(1:1 或 n:n)可以看出, 这些关系在本质上是对称的. 这意味着对于任意给定的事实 ($head, rel, tail$), ($tail, rel', head$) 也是一个事实. 而本文提出模型的有效性可以从“rid”列看出, 它恰好显示了与前 4 个关系最相似的关系是它们自身对应的逆关系, 这说明模型很好地捕捉到了关系对称的特性.

值得注意的是, 另外 14 个关系以成对的形式出现在表中, 例如“_has_part”关系最相似的逆关系是“_part_of”, 反之亦然, 这是因为 WordNet 中的大多数关系是语义双重对偶的^[37]. 计算结果再次清晰地表明所提出的模型能够有效地捕获关系间的非对称性.

5 结束语

本文研究了图上的分布式表示学习问题, 提出了一种基于符号语义映射的神经网络表示模型, 该模型使用递归组合机制和神经映射机制来对知识图谱和复杂网络进行建模. 本文有 2 个重要的发现: 1) 提出的模型是基于符号语义理论框架的, 其思路是通过从组合表示到标记表示的语义映射来学习图的表示, 因此可以推广到各类同构图和异构图中; 2) 本文证明了 RNN 网络可以用来有效地从图符号系统中的事实生成具有代表性的组合表示, 并且提出的模型在 2 类不同结构图上的学习任务中都表现出了优异的性能. 此外, 由于模型的递归结构(参数共享)和神经映射机制, 本文提出的模型在计算上效率很高, 因此可以扩展应用到大规模知识图谱中.

该工作为研究图的分布式表示学习算法提供了

新的建模思路和解决方案, 同时也留下了一些值得继续研究的问题. 比如, 与机器翻译中使用的 RNN 架构相比, 本文提出的模型较为浅薄, 后续工作将考虑使用深层的 RNN 模型以及更长的输入序列, 目标是提出更为合理且高效的表示学习算法. 同时, 我们也会继续考虑递归组合的解释因素, 进一步探索基于符号语义映射的图表示学习解决方案.

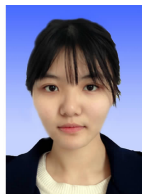
参 考 文 献

- [1] Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798-1828
- [2] 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)
- [3] Palangi H, Deng Li, Shen Yelong, et al. Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval [J]. IEEE/ACM Trans on Audio, Speech and Language Processing, 2016, 24(4): 694-707
- [4] Santos L D, Piwowarski B, Gallinari P. Gaussian embeddings for collaborative filtering [C] //Proc of the 40th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2017: 1065-1068
- [5] Zhao Zhou, Yang Qifan, Lu Hanqing, et al. Learning max-margin geosocial multimedia network representations for point-of-interest suggestion [C] //Proc of the 40th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2017: 833-836
- [6] Jameel S, Bouraoui Z, Schockaert S. MEMBER: Max-margin based embeddings for entity retrieval [C] //Proc of the 40th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2017: 783-792
- [7] Wang Chuanju, Wang Tinghsiang, Yang Hsiuwei, et al. ICE: Item concept embedding via textual information [C] / Proc of the 40th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2017: 85-94
- [8] Pennington J, Socher R, Manning C D. GloVe: Global vectors for word representation [C] //Proc of EMNLP 2014. Stroudsburg, PA: ACL, 2014: 1532-1543
- [9] Henaff M, Weston J, Szlam A, et al. Tracking the world state with recurrent entity networks [C] //Proc of the 5th Int Conf on Learning Representation. New York: CoRR, 2017: 1-15

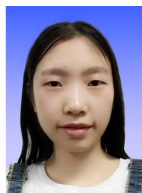
- [10] Liu Qiao, Li Yang, Duan Hong, et al. Knowledge graph construction techniques [J]. *Journal of Computer Research and Development*, 2016, 53(3): 582-600 (in Chinese)
(刘娇, 李杨, 段宏, 等. 知识图谱构建技术综述[J]. *计算机研究与发展*, 2016, 53(3): 582-600)
- [11] Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs [C] //Proc of the 30th AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2016: 1955-1961
- [12] Perozzi B, Alrfou R, Skiena S. DeepWalk: Online learning of social representations [C] //Proc of the 20th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701-710
- [13] Trouillon T, Gaussier E, Dance C R, et al. On inductive abilities of latent factor models for relational learning [J]. *CoRR*, arXiv preprint, arXiv:1709.05666, 2017
- [14] Harnad S. The symbol grounding problem [J]. *Physica D: Nonlinear Phenomena*, 1990, 42(1): 335-346
- [15] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks [C] //Proc of the 27th Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2014: 3104-3112
- [16] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems [J]. *Computer*, 2009, 42(8): 30-37
- [17] Dong Xin, Gabrilovich E, Heitz G, et al. Knowledge vault: A Web-scale approach to probabilistic knowledge fusion [C] //Proc of the 20th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2014: 601-610
- [18] Grover A, Leskovec J. Node2Vec: Scalable feature learning for networks [C] //Proc of the 22nd ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 855-864
- [19] Goyal P, Ferrara E. Graph embedding techniques, applications, and performance: A survey [J]. *Knowledge Based Systems*, 2017, 151(1): 78-94
- [20] Cai Hongyun, Zheng V W, Chang Chenchuan. A comprehensive survey of graph embedding: Problems, techniques and applications [J]. *arXiv preprint*, arXiv:1709.07604, 2017: 1-20
- [21] Wang Quan, Mao Zhendong, Wang Bin, et al. Knowledge graph embedding: A survey of approaches and applications [J]. *IEEE Trans on Knowledge and Data Engineering*, 2017, 29(12): 2724-2743
- [22] Tang Jian, Qu Meng, Wang Mingzhe, et al. LINE: Large-scale information network embedding [C] //Proc of the 24th Int Conf on World Wide Web. New York: ACM, 2015: 1067-1077
- [23] Nickel M, Murphy K, Tresp V, et al. A review of relational machine learning for knowledge graphs [J]. *Proceedings of the IEEE*, 2016, 104(1): 11-33
- [24] Nickel M, Tresp V, Kriegel H P. A three-way model for collective learning on multi-relational data [C] //Proc of the 28th Int Conf on Machine Learning. New York: ACM, 2011: 809-816
- [25] Socher R, Chen Danqi, Manning C D, et al. Reasoning with neural tensor networks for knowledge base completion [C] //Proc of the 26th Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2013: 926-934
- [26] Bordes A, Usunier N, Garcia-duran A, et al. Translating embeddings for modeling multi-relational data [C] //Proc of the 26th Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2013: 2787-2795
- [27] Yang Bishan, Yih W T, He Xiaodong, et al. Embedding entities and relations for learning and inference in knowledge bases [C] //Proc of the 5th Int Conf on Learning Representation. New York: CoRR, 2015: 1-12
- [28] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction [C] //Proc of the 33rd Int Conf on Machine Learning. New York: JMLR, 2016: 2071-2080
- [29] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2D knowledge graph embeddings [J/OL]. (2017-07-05) [2018-03-14]. <https://arxiv.org/abs/1707.01476>
- [30] Garcia-duran A, Bordes A, Usunier N. Effective blending of two and three-way interactions for modeling multi-relational data [C] //Proc of 2014 Joint European Conf on Machine Learning and Knowledge Discovery in Databases. Berlin: Springer, 2014: 434-449
- [31] Hayashi K, Shimbo M. On the equivalence of holographic and complex embeddings for link prediction [J/OL]. (2017-02-18) [2018-03-14]. <https://arxiv.org/abs/1702.05563>
- [32] Shi Baoxu, Weninger T. ProjE: Embedding projection for knowledge graph completion [C] //Proc of the 31st AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2017: 1236-1242
- [33] Mikolov T, Sutskever I, Chen Kai, et al. Distributed representations of words and phrases and their compositionality [C] //Proc of the 25th Advances in Neural Information Processing Systems. Cambridge, MA: MIT Press, 2012: 3111-3119
- [34] Toutanova K, Chen Danqi. Observed versus latent features for knowledge base and text inference [C] //Proc of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality. Stroudsburg, PA: ACL, 2015: 57-66
- [35] Liu Qiao, Jiang Liuyi, Han Minghao, et al. Hierarchical random walk inference in knowledge graphs [C] //Proc of the 39th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2016: 445-454
- [36] Liu Qiao, Han Minghao, Yang Xiaohui, et al. Representation learning based relational inference algorithm with semantical aspect awareness [J]. *Journal of Computer Research and Development*, 2017, 54(8): 1682-1692 (in Chinese)

(刘峤, 韩明皓, 杨晓慧, 等. 基于表示学习和语义要素感知的关系推理算法[J]. 计算机研究与发展, 2017, 54(8): 1682-1692)

- [37] Fellbaum C. Wordnet: An Electronic Lexical Database [M]. Cambridge, MA: MIT Press, 1998



Yang Xiaohui, born in 1993. Master candidate. Her main research interests include machine learning, natural language processing, social network analysis and representation learning.



Wan Rui, born in 1995. Master candidate. Her main research interests include representation learning, natural language processing, data mining and knowledge graph (rwan@std.uestc.edu.cn).



Zhang Haibin, born in 1993. Master candidate. Student member of CCF. His main research interests include natural language processing, machine learning and sentiment analysis (herb.zhang@std.uestc.edu.cn).



Zeng Yifu, born in 1995. Bachelor. His main research interests include nature language processing, sentiment analysis and recommendation system (ifz@std.uestc.edu.cn).



Liu Qiao, born in 1974. PhD and associate professor. Member of CCF. His main research interests include machine learning and data mining, natural language processing, and social network analysis.