

一种结合层次化类别信息知识图谱表示学习方法*

张金斗, 李京



(中国科学技术大学 计算机科学与技术学院, 安徽 合肥 230026)

通信作者: 李京, E-mail: lj@ustc.edu.cn

摘要: 知识图谱表示学习方法旨在将知识图谱中的实体和关系嵌入到低维连续的向量空间. 由于知识图谱本身具有数据稀疏性的问题导致学习出的向量表示性能欠缺. 实体的类别信息包含了丰富的语义, 引入它能够更好地指导向量表示的学习. 已有结合类别信息的表示学习方法要么不支持类别信息的层次化结构或者关系的类别约束, 要么对层次化结构的建模过于复杂. 提出一种结合层次化类别信息的表示学习方法. 我们将类别嵌入到不同的向量空间, 使用偏序关系建模类别的层次化结构. 同时, 将实体向量表示映射到类别向量空间中, 要求实体与其所属类别满足偏序关系, 且三元组的实体与其关系的类别约束也满足偏序关系. 最后, 在多个数据集上执行链接预测、三元组分类和实体分类任务的实验结果表明我们的方法相比其他基线方法学习出的向量表示性能更好.
关键词: 知识图谱表示学习; 类别信息; 层次化结构; 偏序关系

中图法分类号: TP18

中文引用格式: 张金斗, 李京. 一种结合层次化类别信息知识图谱表示学习方法. 软件学报, 2022, 33(9): 3331–3346. <http://www.jos.org.cn/1000-9825/6295.htm>

英文引用格式: Zhang JD, Li J. Knowledge Graph Embedding Combining with Hierarchical Type Information. Ruan Jian Xue Bao/Journal of Software, 2022, 33(9): 3331–3346 (in Chinese). <http://www.jos.org.cn/1000-9825/6295.htm>

Knowledge Graph Embedding Combining with Hierarchical Type Information

ZHANG Jin-Dou, LI Jing

(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China)

Abstract: Knowledge graph embedding aims to embed entities and relations into a low-dimensional continuous vector space. Due to the data sparsity of knowledge graphs, the performance of knowledge graph embedding is poor in vector representation. Since the type information of entities encompasses rich semantic information, it is introduced to improve the performance. However, the existing methods either do not support the hierarchical structure of type information or the type constraint of relations or complicate the model of the hierarchical structure. This study proposes a novel knowledge graph embedding method combining with hierarchical type information. Specifically, types are embedded into different vector spaces and the hierarchical structure of types is modeled by the partial order relation. Moreover, the vector representations of entities are mapped into the type vector space so that entities and their types can be required to satisfy the partial order relation. The entities and their type constraint of relations in triples are also made to satisfy the partial order relation. Finally, experimental results of link prediction, triple classification and entity typing task on four datasets show that the proposed method outperforms the state-of-the-art baseline methods in vector representation performance.

Key words: knowledge graph embedding; type information; hierarchical structure; partial order relation

知识图谱 (knowledge graph), 通常也被称为知识库, 是以一种有向图的结构描述客观世界的事物及其相互联系. 其中, 图的顶点由实体组成, 用来表示客观世界的事物; 边则由关系组成, 表示实体之间的相互联系. 一般地, 边使用事实三元组 (h, r, t) 的形式表示和存储, 其中, h , r 和 t 分别表示头实体、关系和尾实体. 例如, (北京, 首都, 中

* 基金项目: 中国科学院战略性先导科技专项 (A 类)(XDA19020102)

收稿时间: 2020-06-15; 修改时间: 2020-08-27, 2020-10-21, 2020-12-22; 采用时间: 2020-12-28; jos 在线出版时间: 2022-07-15

国)表示“北京是中国的首都”。作为人工智能领域中重要的基础技术,许多大规模知识图谱,比如 DBpedia^[1]、YAGO^[2]和 Freebase^[3],被构建出来为语义搜索^[4]、智能问答^[5,6]、信息抽取^[7]、推荐系统^[8]等应用提供服务并为它们知识赋能。

为了获取知识图谱中实体和关系潜在的语义信息,同时方便下游任务使用知识图谱,知识图谱的分布式表示(distributed representations)学习技术被提出并被广泛研究和应用。知识图谱表示学习,也称为知识图谱嵌入(embeddings),是将实体和关系分别映射到低维连续的向量空间中,对应的向量表示代表了它们的语义信息。虽然知识图谱嵌入可以快速有效地进行知识推理并方便其他任务使用,但是知识图谱本身具有数据稀疏性的问题导致学习出的知识图谱嵌入性能欠缺,无法很好地编码稀疏的实体。于是,研究人员提出通过引入一些外部信息或者先验知识的方式来提升表示学习的性能,这些外部信息包括实体的文本描述信息^[9,10]、类别信息^[11-13]和关系构成的一阶谓词逻辑规则^[14-16]等。其中,实体的类别信息作为知识图谱本体概念的一部分,包含了丰富的语义信息,能够很好地指导实体和关系向量表示的学习。

实体的类别信息放到知识图谱中可以用“属于”关系表示,它除了反映某个实体属于哪个类别,还具有另外两个重要的特性。第一,类别的层次化结构。类别与类别之间可能存在从属关系,这种从属关系构成层次化结构,它具有反对称性和传递性,对实体的向量表示学习具有指导意义。如图 1 所示,实体“艾伦·图灵”属于类别“计算机科学家”和“数学家”,同时这两个类别都属于类别“受过教育的人”,即使知识图谱中没有给出“艾伦·图灵”属于“受过教育的人”,通过层次结构我们同样可以得到这个重要的信息。第二,关系的类别约束。每个关系扮演着各自的角色,它规定了关系关联的头实体和尾实体分别应该属于什么类别。如图 1 所示,关系“毕业于”对应的头实体应该属于类别“受过教育的人”,而尾实体应该属于类别“教育机构”。在表示学习的过程中,该信息能够保证关系与满足其类别约束的实体之间语义相关,而与不满足约束的实体不相关,更加准确地学习出它们的潜在语义。

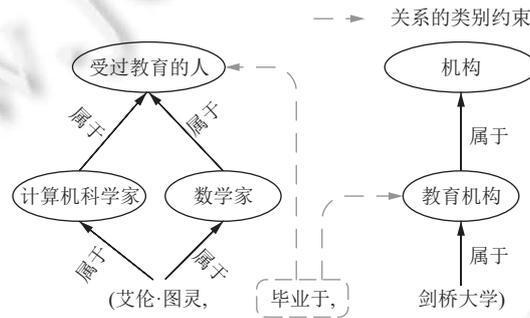


图 1 实体类别信息层次化结构和关系类别约束示例

如何把类别的层次化结构和关系的类别约束融合到已有的知识图谱表示学习中是引入实体类别信息的主要挑战。现有的相关工作可以分为 3 类：(1) 只考虑关系的类别约束。该类工作^[9]或者对每个关系只训练满足其类别约束的实体或者使用这些约束进行负样本采样。(2) 只考虑类别的层次化结构。该类工作可以再细分成两类：第 1 类工作^[17-19]把实体和类别映射到同一个向量空间,使用空间中的区域建模类别,区域之间的包含关系表示类别的层次化结构。然而,实体和类别的向量空间维度应该是不同的,类别通常比实体更稀疏且语义也不一致;另外,这种基于区域的方法通常会引入其他参数,增加额外的开销。第 2 类工作^[20,21]将类别映射到不同的低维向量空间,同时使用线性变换的方式建模类别的层次化结构,然而它们无法很好地表达层次结构的反对称性和传递性。(3) 同时考虑了类别的层次化结构信息和关系的类别约束。该类工作^[12,22]使用投影矩阵建模类别,并以矩阵的线性组合或者矩阵乘积的编码方式表达类别的层次化结构信息,然而这种方式同样无法表达其反对称性和传递性的特点。

通过对已有工作的分析,我们发现实体和类别应该在不同的向量空间中建模,同时在类别向量空间中要保证类别层次结构的反对称性和传递性。受到 Order Embedding 工作^[23]的启发,把它使用偏序关系建模词语、短语与图片之间层次结构的方法用来建模类别的层次化结构,Order Embedding 可以在不额外增加参数的情况下让有偏序关系的点在向量空间中仍然有序。另外,已有工作大多都假设关系的类别约束是已知的。然而,事实上并不是所有

的知识图谱都能够轻易地获取到这些信息. 虽然 Krompaß 等人的工作^[11]有考虑到关系的类别约束未知的情况, 并将关系对应的头实体(尾实体)集合近似作为它的类别约束, 但是他们独立地对待每个关系的类别约束, 使用这些约束进行负样本采样, 而忽略了关系之间类别约束的联系(他们对应的类别可能是相同的或者是具有层次结构的). 因此, 本文提出一种新颖的结合层次化类别信息的知识图谱表示学习方法 HiTE (hierarchy-aware type embodied knowledge graph embedding). 对于类别信息, 和 JOIE 模型^[20]类似, 我们将类别映射到不同的向量空间中学习它们的向量表示, 不同的是, 对于类别向量空间我们使用 Order Embedding 偏序关系建模类别的层次化结构, 保证学习出的类别向量表示满足偏序关系, 这样便可以保证具有层次结构的类别与类别之间具有反对称性和传递性. 同时, 我们将实体映射到类别向量空间, 保证它与所属类别满足偏序关系. 对于关系的类别约束, 我们将关系对应的头实体(尾实体)映射到类别的向量空间中, 并保证它与约束关系的类别满足偏序关系. 更重要地是, 即使关系的类别约束信息未知, 我们同样可以使用该约束. 具体地, 我们为每个关系分别指定不同的约束头实体和尾实体的类别, 根据知识图谱已有的三元组数据可以自动学习出这些类别的结构信息, 从而进一步指导实体向量表示的学习.

本文的主要贡献如下.

- 1) 首次将偏序关系建模层次结构的方法引入到知识图谱表示学习中, 使用它编码实体的类别层次化结构和关系的类别约束, 从而提高表示学习的能力;
- 2) 我们的方法在类别信息未知的情况仍然可以利用关系的类别约束, 帮助提高表示学习的能力;
- 3) 在 4 个基准数据集上的实验结果表明我们提出的方法能够很好地提高已有知识图谱嵌入模型的性能.

1 形式化定义

在本节中, 我们将给出与本文相关的概念的定义和需要用到的基本数学符号.

定义 1. 知识图谱. 给定一个有向图 $G = (E, R, T)$ 表示知识图谱. 其中, E 表示实体的集合, R 表示关系的集合, T 表示三元组的集合, 即 $T = \{(h, r, t) | h, t \in E, r \in R\} \subseteq E \times R \times E$.

定义 2. 实体的类别信息. 我们给定 C 表示所有类别的集合. 集合 $I_e = \{(e, c) | e \in E, c \in C\}$ 表示实体与其所属类别的集合; $I_c = \{(c_i, c_k) | c_i, c_k \in C \wedge c_i \neq c_k\}$ 集合表示类别与类别之间的从属关系的集合, 即类别的层次化结构; $DOMAIN = \{(c_h, r) | r \in R, c_h \in C\}$ 表示关系与其对应头实体类别的集合; $RANGE = \{(r, c_t) | r \in R, c_t \in C\}$ 表示关系与其对应尾实体类别的集合.

定义 3. 知识图谱表示学习. 知识图谱表示学习将实体和关系分别映射到低维连续的向量空间, 并学习实体和关系的向量表示. 给定实体向量 $h, t \in \mathbb{R}^{d_e}$ 和关系向量 $r \in \mathbb{R}^{d_r}$ (d_e, d_r 分别为实体和关系向量空间的维度大小), 知识图谱表示学习模型通过设计一个评分函数 $f: E \times R \times E \rightarrow \mathbb{R}$, 衡量输入的三元组的真实性, 评分函数的值越大表示三元组越有可能为真, 通过最大化所有已知三元组的评分函数值, 学习出实体和关系的向量表示.

定义 4. 负样本采样. 知识图谱的三元组数据在表示学习过程中被视为正样本, 为了模型的健壮性一般也需要负样本. 但是, 知识图谱的负样本无法直接给出, 因为不在知识图谱中的三元组也可能是真的. 构造负样本普遍采用的方法是随机地将已有三元组的头实体或者尾实体用其他实体替换进行采样. 定义负样本集合 $\tilde{T} = \{(h', r, t) | (h, r, t) \in T \wedge h \neq h' \wedge h' \in E\} \cup \{(h, r, t') | (h, r, t) \in T \wedge t \neq t' \wedge t' \in E\}$, 使用均匀分布或者伯努力分布从中采样负样本进行训练. 此外类别信息的负样本采样也使用类似上面的方法, 我们可以定义 $\tilde{I}_e = \{(e, c') | (e, c) \in I_e \wedge c \neq c' \wedge c' \in C\}$, $\tilde{I}_c = \{(c_i, c'_k) | (c_i, c_k) \in I_c \wedge c_k \neq c'_k \wedge c'_k \in C\}$.

2 相关工作

在本节中, 我们将概述已有的基础知识图谱表示学习模型, 结合类别信息的表示学习模型以及已有的层次结构建模的工作.

2.1 基础知识图谱表示学习模型

根据评分函数设计原理的不同, 我们可以将现有的表示学习模型分为以下 3 类, 有关这些模型更多的信息详

见综述文献 [24–26].

1) 基于翻译的模型. 该类模型把关系向量看作是头实体向量到尾实体向量的平移操作, 通过计算平移后的头实体与尾实体向量之间的欧式距离来衡量三元组的真实性. 其中, TransE^[27]是最具代表性的模型. 给定一个三元组 (h, r, t) , 它假设尾实体向量 t 等于头实体向量 h 和关系向量 r 的和, 即 $t \approx h + r$. 它的评分函数和损失函数如表 1 所示.

2) 线性模型. 该类模型把关系向量看作是线性变换操作, 通过计算线性变换后的头实体和尾实体向量的相似度来衡量三元组的真实性. 其中, DistMult^[28]是最简单且被广泛使用的模型. 它将关系向量 r 对角化变成对角矩阵, 作为线性变换矩阵, 使用该变换矩阵将头实体向量映射到尾实体向量. 它的评分函数和损失函数如表 1 所示.

3) 基于神经网络的模型. 该类模型利用多层神经网络非线性变换的特点, 计算三元组的概率来衡量三元组的真实性. 其中, ConvE^[29]使用卷积层和全连接层神经网络建模实体与关系之间的潜在语义, 它的评分函数和损失函数如表 1 所示.

表 1 3 种代表性知识图谱表示学习模型

方法	嵌入变量	评分函数 $f(h, r, t)$	损失函数
TransE	$h, r, t \in \mathbb{R}^d$	$-\ h + r - t\ _{1/2}$	$\sum_{(h,r,t) \in T} \sum_{(h',r',t') \in \bar{T}} \max\{0, \gamma - f(h, r, t) + f(h', r', t')\}$
DistMult	$h, r, t \in \mathbb{R}^d$	$h^\top \cdot \text{diag}(r) \cdot t$	$\sum_{(h,r,t) \in T \cup \bar{T}} \log(1 + \exp(-y \cdot f(h, r, t)))$
ConvE	$h, r, t \in \mathbb{R}^d$; $W \in \mathbb{R}^{d \times d}$	$g(\text{vec}(g(\text{concat}(\hat{h}, \hat{r}) * \omega)))^\top W t$	$\sum_{(h,r,t) \in T \cup \bar{T}} -(y \cdot \log(\sigma(f(h, r, t))) + (1 - y) \log(1 - \sigma(f(h, r, t))))$

注: $g(\cdot)$: 非线性函数; $\text{concat}(\cdot)$: 拼接函数; $\text{vec}(\cdot)$: 向量化函数; \hat{x} : 向量 ω : 一组过滤器; $*$: 卷积操作; $y = \pm 1$; $\sigma(\cdot)$: Sigmoid 函数; $\text{diag}()$: 向量转换为对角矩阵函数.

2.2 结合类别信息的知识图谱表示学习模型

实体的类别信息作为知识图谱本体概念的一部分, 能够用来补充实体和关系的语义信息. 因此, 有许多研究人员尝试将类别信息结合到知识图谱表示学习的模型中从而改进表示学习的性能. Guo 等人^[13]首先提出了语义平滑嵌入的方法, 要求属于同一类别的实体在向量空间中彼此邻近. 它使用两种学习算法衡量同一类别下实体间的距离, 并将其作为正则项约束实体嵌入的学习. 该方法的缺点是它假设一个实体只能属于一种类别, 且无法建模类别的层次结构和关系的类别约束. Krompaß 等人^[11]的工作考虑了关系的类别约束, 并把它引入到 3 种已有的知识图谱嵌入模型中, 包括 TransE、线性模型 RESCAL^[30]和一个多路神经网络模型. 对于 RESCAL 模型, 该工作通过约束每个关系只训练满足其 *DOMAIN* 和 *RANGE* 集合对应类别下的实体, 将关系的类别约束集成到该模型中; 对于 TransE 和神经网络模型则约束它们只从属于 *DOMAIN* 和 *RANGE* 集合里对应类别的实体中采样获取负样本. 另外, 作者指出关系的类别约束信息通常是缺失的, 因此他们提出本地闭合世界假设 LCWA (local closed-world assumptions), 利用知识图谱中已知的三元组, 将关系对应的头实体 (尾实体) 集合近似作为它的类别约束. TKRL^[12]同时考虑了类别的层次化结构和关系的类别约束, 它把类别看作是投影矩阵, 使用矩阵的线性组合或者矩阵乘积的编码方式表达类别的层次结构信息, 然后对每个关系根据其类别约束以及类别的层次化结构, 计算得到两个类别投影矩阵作为约束分别作用到头尾实体向量上. 此外, TKRL 使用 LCWA 假设进行负样本采样. 该方法的缺点是要求关系的类别约束信息都是已知的; 另外, 将每个类别建模成投影矩阵, 这增加了大量额外的参数和计算量. TEKRL 方法^[22]放宽了 TKRL 的要求, 允许关系的类别约束未知, 采用注意力机制, 分别以直接 (余弦相似性) 和间接 (缩放点积) 的方式来计算注意力分数, 以此得到实体类别和三元组关系之间的相关性, 用来表示关系的类别约束. 然而, 它们编码类别的层次结构信息的方式并没有表达其反对称性和传递性的特点, 无法准确地建模类别之间的层次结构.

最近提出的 JOIE 模型^[20], 将知识图谱和其本体概念信息融合在一起, 相互利用各自具备的语义信息, 学习出

更全面的知识图谱嵌入. 实体的类别信息作为本体概念的一部分, JOIE 使用了一种新的方法对其建模. 它将类别映射到不同的低维向量空间, 通过线性变换将实体向量映射到该空间上, 映射后的实体要求尽可能地与它所属类别的位置距离相近并远离不属于的类别; 同样地, 对于类别的层次结构, 类别与类别之间的从属关系也通过这种线性变换的方式实现. 但是这种用线性变换建模层次结构的方式无法很好地表达层次结构的反对称性和传递性, 并且 JOIE 没有利用到关系的类别约束. On2Vec^[21]方法使用类似 JOIE 的思路建模本体知识图谱中概念(类别)的层次化关系, 它也是让具有层次结构的概念在向量空间中尽量相互邻近, 不同的是它使用余弦距离衡量变换后的概念是否邻近, 而 JOIE 使用的是欧式距离. KEC 方法^[31]同样使用类别向量空间, 它以类别为桥梁, 建立一个实体与它所属类别下的其他实体之间的上下文联系, 使用 Skip-gram 模型^[32]学习实体在该空间上的向量表示, 然后将三元组向量投影到该空间的超平面, 保证其满足类别的约束.

2.3 层次结构建模

在实际应用中, 有许多数据具有层次化结构, 比如词语与词语之间的上下位关系, 商品的分类目录等等. 为了表示层次化结构数据, 研究人员尝试建模层次结构, 学习出具有层次特征的向量表示. 这方面的相关工作可以分为 3 类: 第 1 类是基于区域的模型. 该类工作利用空间中的区域建模类别, 区域之间的包含关系表示层次化结构, 该类的代表性工作包括 TransC^[17]、nball^[18]和 Box Embeddings^[19]. TransC 模型使用实体所在向量空间中的球体表示类别, 球体之间的包含关系表示类别的层次化结构. nball 模型使用相同的想法建模类别, 不同的是它使用后处理的方式增加类别的语义信息, 把实体也用球体来表示, 并对不同的关系三元组下的头尾实体构造不同的球体. Box Embedding 方法把类别看作是实体向量空间中的超矩形, 超矩形之间的包含关系则表示类别的层次化结构. 这类工作都把实体和类别放在同一个向量空间中, 没有区分它们之间的语义结构差异, 并且这类工作通常会引入其他参数(比如球体半径), 增加额外的空间和时间开销; 第 2 类是基于排序学习的模型. 该类工作主要集中在对细粒度的实体分类的研究, 它通过对层次化结构中不同层设定不同的间隔值, 学习出类别之间的排序关系, 这种排序关系反映了类别之间的层次结构, 该类的代表性工作包括 AFET^[33,34]和 Chen 等人^[35]的工作, 该类工作的缺点是无法学习出子类有多个父类别的情况; 第 3 类是基于向量的方向和模值的模型. 该类工作集中在词语蕴含关系的研究中, 它使用词语向量间的余弦相似度和向量的模值建模词语间的上下位层次结构. 向量间的余弦相似度越大则它们越有可能具有层次关系, 其中向量的模越大表示该向量处于层次结构的越上层位置, 该类的代表性工作包括 LEAR^[36]和它的扩展工作^[37], 该类工作的缺点和第 2 类工作类似, 当一个词语有多个不相关的上位词时, 这些上位词会被错误地学习出有上下位关系. 除此之外, 比较有代表性的工作还有 Order Embeddings^[23]和 Gaussian Embeddings^[38]. Order Embeddings 使用偏序关系建模层次结构, 要求有层次关系的点在向量空间中始终保持有序. 而 Gaussian Embeddings 使用概率密度作为向量表示, 把每个词语建模成一个高斯分布, 词语之间的 KL 散度(Kullback-Leibler divergence)代表了它们的层次关系. 另外, 最近越来越多的研究人员^[39-41]开始使用双曲空间建模层次结构. 在本文中, 我们使用 Order Embeddings 的想法, 将类别映射到不同的向量空间中, 要求它们满足偏序关系以此建模它们的层次结构.

3 我们的方法

在本节中, 我们详细介绍如何将具有层次化结构的类别信息和关系的类别约束结合到已有的知识图谱表示学习的模型中. 图 2 给出了我们的方法在二维向量空间上的示例, 我们将类别映射到不同的向量空间中学习它们的向量表示. 首先, 通过要求该空间里具有从属关系的类别与类别之间保持偏序关系, 建模类别的层次结构; 然后, 我们使用线性变换函数将在实体向量空间中的实体映射到类别向量空间, 同时保证变换后的实体与它所属类别也有偏序关系; 最后, 将三元组中的头实体和尾实体映射到类别向量空间, 分别与三元组中的关系对应的类别约束保持偏序关系. 具体地, 第 3.1 节给出如何使用偏序关系建模类别的层次结构. 第 3.2 节描述了如何实现关系的类别约束. 第 3.3 节介绍如何将类别信息结合到知识图谱嵌入模型中联合训练学习实体、关系以及类别的向量表示.

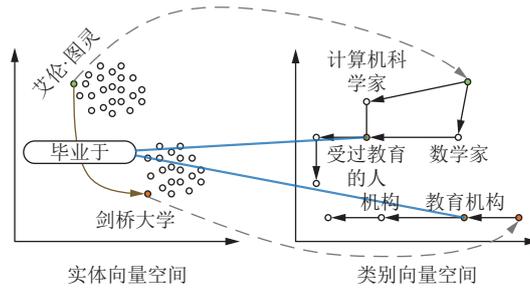


图 2 HiTE 建模类别层次化结构和关系类别约束示例

3.1 类别的层次结构建模

在类别的层次结构中,类别与类别之间的关系具有自反性、反对称性和传递性,可以把它视为偏序关系.我们需要解决的问题是如何将满足偏序关系的类别映射到向量空间以后仍然保持有序.这样的映射函数需要满足的性质如下定义所示:

定义 5. 类别向量空间映射函数. 我们给定映射函数 $F : (C, \leq_C) \rightarrow (Y, \leq_Y)$. 其中, C 是类别的集合, Y 是类别向量空间. 我们要求对于 $\forall u, v \in C$, 如果有 $u \leq_C v$ 当且仅当 $F(u) \leq_Y F(v)$.

我们选择这样一个满足偏序关系的向量空间 $Y = \mathbb{R}_+^d$, 即向量空间的每个坐标值都是正数, 并且对于 $\forall x, y \in Y$, 如果 $x \leq_Y y$ 当且仅当 $\wedge_{i=1}^d x_i \geq y_i$, 这表示坐标值越小类别的概念越概括, 值越大类别越具体. 该向量空间的偏序关系约束可以使用公式 (1) 所示的惩罚函数实现.

$$P(x, y) = \| [y - x]_+ \|_2 \tag{1}$$

其中, 函数 $[z]_+$ 只保留输入向量 z 的正数部分, 其余部分为 0, 即 $[z]_+ = \max\{0, z\}$. 可以看到, $x \leq_Y y$ 当且仅当 $P(x, y) = 0$. 当不满足偏序关系, $P(x, y) > 0$. 因此, 我们可以使用一个最大间隔损失函数, 去训练学习满足偏序关系的类别向量表示, 如公式 (2) 所示. 其中, I_c 是类别之间层次关系的集合, (u, v) 是满足偏序关系的正样本, (u', v') 是不满足的负样本, 它是通过第 1 节定义的正样本采样策略生成的. $u = F(u)$, $v = F(v)$ 分别是类别 u, v 的向量表示, 超参数 γ 是区分负样本的间隔.

$$\mathcal{L}_{HT} = \frac{1}{|I_c|} \sum_{(u,v) \in I_c} P(u, v) + \frac{1}{|\bar{I}_c|} \sum_{(u',v') \in \bar{I}_c} [\gamma - P(u', v')]_+ \tag{2}$$

3.2 实体和关系的类别约束建模

为了实现实体的类别约束, 我们将实体向量映射到类别向量空间中, 要求其与其所属类别向量满足偏序关系. 首先, 我们使用一层神经网络作为线性变换函数将实体向量映射到类别向量空间, 该变换函数定义如下所示:

$$\psi(e) = \delta(W \cdot e + b), \forall (e, c) \in I_e \tag{3}$$

其中, $W \in \mathbb{R}^{d \times d}$ 是权值矩阵, $b \in \mathbb{R}^d$ 是偏置向量, $\delta(\cdot)$ 是激活函数, 这里我们使用 ReLU 函数. 然后, 我们使用第 3.1 节中的惩罚函数 $P(x, y)$ 定义如下损失函数学习实体和类别在类别向量空间中的偏序关系.

$$\mathcal{L}_{ET} = \frac{1}{|I_e|} \sum_{(e,c) \in I_e} P(\psi(e), c) + \frac{1}{|\bar{I}_e|} \sum_{(e,c') \in \bar{I}_e} [\gamma - P(\psi(e), c')]_+ \tag{4}$$

和实体的类别约束类似, 我们将每个三元组的头实体和尾实体向量映射到类别向量空间中, 使其与三元组的关系对应的头实体和尾实体类别满足偏序关系. 我们使用公式 (5) 定义的损失函数, 学习关系的类别约束.

$$\mathcal{L}_{RT} = \frac{1}{|T|} \sum_{\substack{(h,r) \in T \wedge \\ (c_h,r) \in DOMAIN \wedge \\ (r,c_t) \in RANGE}} (P(\psi(h), c_h) + P(\psi(t), c_t)) + \frac{1}{|\bar{T}|} \sum_{\substack{(h',r') \in T \wedge \\ (c_h,r') \in DOMAIN \wedge \\ (r',c_t) \in RANGE}} ([\gamma - P(\psi(h'), c_h)]_+ + [\gamma - P(\psi(t'), c_t)]_+) \tag{5}$$

特别值得一提的是, 本文假设在关系的类别约束, 即关系的 $DOMAIN$ 和 $RANGE$ 集合, 中每个关系只对应一个头实体类别和一个尾实体类别. 而实际应用中人工给定或者自动设定的关系对应的头(尾)实体类别标签可能有

多个,比如“毕业于”关系设定的头实体类别标签可能有“数学家”“科学家”和“哲学家”.对于这种一对多的情况,我们的方法仍然可以支持,只需要增加一个虚拟类别,令给定的多个头(尾)实体类别都属于这个类别,然后使用该虚拟类别作为关系的类别约束即可.

3.3 结合层次化类别信息的知识图谱表示学习

由于我们将类别映射到不同于实体和关系的向量空间上,这使得把类别信息和大部分已有知识图谱嵌入模型结合在一起变得简单直接.假设已有知识图谱嵌入模型的评分函数和损失函数分别为 $f(h, r, t)$ 和 \mathcal{L}_{KG} . 本文使用 DistMult 模型作为基础的知识图谱嵌入模型,因此评分函数和损失函数定义为: $f(h, r, t) = h^T \cdot \text{diag}(r) \cdot t$, $\mathcal{L}_{\text{KG}} = \sum_{(h,r,t) \in T \cup \tilde{T}} \log(1 + \exp(-y \cdot f(h, r, t)))$, 如表 1 所示. 接下来,我们从 3 个方面讨论如何结合层次化类型信息.

1) 如果类别的层次化结构,实体所属类别以及关系的类别约束都是给定的,则将它们的损失函数线性组合在一起,联合训练学习实体、关系和类别的向量表示.这样得到的联合损失函数如公式 (6) 所示:

$$\mathcal{L}_{\text{ALL}} = \mathcal{L}_{\text{KG}} + \alpha \mathcal{L}_{\text{RT}} + (\mathcal{L}_{\text{HT}} + \mathcal{L}_{\text{ET}}) + \lambda \mathcal{L}_{\text{reg}} \quad (6)$$

其中, α 为关系的类别约束部分的权重, \mathcal{L}_{reg} 表示正则项,它是对实体和关系嵌入的 l_2 正则化, λ 为正则项的权重.同时,我们定义新的评分函数衡量三元组的真实性,如公式 (7) 所示:

$$f_{\text{all}}(h, r, t) = \sigma(f(h, r, t) + \alpha(P(\psi(h), c_h) + P(\psi(t), c_t))) \quad (7)$$

其中, c_h 和 c_t 分别是关系 r 对应的头实体和尾实体的类别向量, σ 是 Sigmoid 函数.算法 1 给出了损失函数 \mathcal{L}_{ALL} 训练过程.

2) 如果实体的类别信息、类别的层次化结构和关系的类别约束没有给定,我们使用无监督的方式去学习关系对应的类别的向量表示.我们假定每个关系对应的头实体和尾实体的类别都不同,并为每个类别设置一个类别标签,这样便可以得到 DOMAIN 和 RANGE 集合,然后使用 \mathcal{L}_{RT} 进行训练.类似 LCWA 假设,我们根据知识图谱自身的三元组数据可以近似地学习出关系的类别约束.最后,我们使用公式 (8) 定义的损失函数进行联合训练.

$$\mathcal{L}_{\text{TYPE_UNKNOWN}} = \mathcal{L}_{\text{KG}} + \alpha \mathcal{L}_{\text{RT}} + \lambda \mathcal{L}_{\text{reg}} \quad (8)$$

3) 如果关系的类别约束未知,而其他类别信息已知,我们同样可以使用 2) 中的方法加入关系的类别约束.

算法 1. 结合给定类别信息的知识图谱嵌入算法.

输入: 训练数据集 (三元组集合 T , 实体与类别集合 I_e , 类别与类别集合 I_c), 实体集 E , 关系集 R , 类别集 C , 关系头实体集合 DOMAIN , 关系尾实体集 RANGE , 实体关系嵌入维度 d , 类别嵌入维度 d' , 间隔 γ , 权重 α 和学习率 η ;
输出: 实体, 关系和类别的嵌入 Θ .

- ① 初始化: $e, r \leftarrow \text{uniform}(0, 1)$, $c \leftarrow \text{uniform}(0, 10)$ for $\forall e \in E, r \in R, c \in C$; $W \leftarrow \text{uniform}(0, 1)$, $b \leftarrow 0$
 - ② for $n \leftarrow 1$ to N /* N 为迭代次数*/
 - ③ Sample a mini-batch T^b, I_e^b, I_c^b from T, I_e, I_c ; /* 采样一个最小批量大小为 b 的训练数据集*/
 - ④ Generate a set of negative triples \tilde{T}^b , sets of negative pairs $\tilde{I}_e^b, \tilde{I}_c^b$ for T^b, I_e^b, I_c^b ; /* 负样本采样*/
 - ⑤ $\Theta^{(n)} \leftarrow \Theta^{(n-1)} - \eta \nabla \mathcal{L}_{\text{ALL}}$; /* 计算损失函数梯度,更新嵌入的值*/
 - ⑥ end for
-

算法 2. 类别信息未知的知识图谱嵌入算法.

输入: 三元组集合 T , 实体集 E , 关系集 R , 实体关系嵌入维度 d , 类别嵌入维度 d' , 间隔 γ , 权重 α 和学习率 η ;
输出: 实体, 关系和类别的嵌入 Θ .

- ① $C \leftarrow \emptyset, \text{DOMAIN} \leftarrow \emptyset, \text{RANGE} \leftarrow \emptyset$;
 - ② $\text{type_label} \leftarrow 0$; /* 类别标签生成器*/
 - ③ for $r \in R$ do
-

-
- ④ $c_h^r \leftarrow \text{type_label}; \text{type_label} \leftarrow \text{type_label} + 1;$
 - ⑤ $c_i^r \leftarrow \text{type_label}; \text{type_label} \leftarrow \text{type_label} + 1;$
 - ⑥ $C \leftarrow C \cup \{c_h^r\} \cup \{c_i^r\};$
 - ⑦ $\text{DOMAIN} \leftarrow \text{DOMAIN} \cup \{(c_h^r, r)\};$
 - ⑧ $\text{RANGE} \leftarrow \text{RANGE} \cup \{(r, c_i^r)\};$
 - ⑨ end for
 - ⑩ 初始化: $e, r \leftarrow \text{uniform}(0, 1), c \leftarrow \text{uniform}(0, 10)$ for $\forall e \in E, r \in R, c \in C; W \leftarrow \text{uniform}(0, 1), b \leftarrow 0;$
 - ⑪ for $n \leftarrow 1$ to N /* N 为迭代次数 */
 - ⑫ Sample a mini-batch T^b from T ; /* 采样一个最小批量大小为 b 的训练数据集 */
 - ⑬ Generate a set of negative triples \tilde{T}^b for T^b ; /* 负样本采样 */
 - ⑭ $\Theta^{(n)} \leftarrow \Theta^{(n-1)} - \eta \nabla \mathcal{L}_{\text{TYPE_UNKNOWN}};$ /* 计算损失函数梯度, 更新嵌入的值 */
 - ⑮ end for
-

3.4 算法复杂度分析

我们以 DistMult 模型为基础知识图谱嵌入模型为例, 对 HiTE 算法的时间和空间复杂度进行分析说明. 我们定义 $n_e = |E|, n_r = |R|, n_c = |C|$ 分别表示知识图谱中实体的数量, 关系的数量和类别的数量; d, d', b 分别表示实体和关系的向量空间维度大小, 类别的向量空间维度大小, 以及训练过程中每一次迭代训练的一个最小批量大小. 在一次迭代过程中, DistMult 模型的空间和时间复杂度为 $O(n_e d + n_r d)$ 和 $O(bd)$. 从算法 1 和公式 (6) 我们可以知道, 结合层次化类别信息以后, 我们的方法 HiTE 的空间复杂度则为 $O(n_e d + n_r d + n_c d' + dd')$, 一般情况下有 $d' < n_c < d \ll n_e$, 因此 HiTE 的空间复杂度可以近似为 $O(n_e d + n_r d)$; 由于需要计算类别信息, HiTE 的时间复杂度则为 $O(bd + bdd' + bdd') = O(bdd')$.

4 实验结果与分析

在本节中, 我们使用 DistMult 模型做为基础知识图谱嵌入模型来实现我们的方法 (用 $\text{HiTE}^{\text{DistMult}}$ 表示). 通过在 4 个数据集上执行链接预测, 三元组分类和实体分类任务对我们的方法进行评估. 另外, 我们选择 3 种不同的惩罚函数, 分析讨论它们对知识图谱嵌入性能的影响; 对于学习出的类别向量表示, 我们通过可视化的方式验证其正确性. 所有的实验都是在一台 Linux 服务器上完成的, 硬件配置为 Intel(R) Xeon(R) Gold 5118 CPU @ 2.30 GHz 处理器, 128 GB 内存和一块 NVIDIA GeForce GTX 1080 GPU.

4.1 数据集

我们选取 WN18、FB15K、YAGO26K 和 DB111K 作为实验的数据集. 其中, WN18 和 FB15K 分别来自于 WordNet^[42]和 Freebase, 由 Bordes 等人^[27]构建, 它本身不包含类别信息, 可以用来评估我们的方法在类别信息没有给定情况下的效果. YAGO26K 和 DB111K 分别来自于 YAGO 和 DBpedia, 由 Hao 等人^[20]构建, 它们自身都包含了一部分本体信息, 这里我们仅提取了实体对应的类别信息和类别层次结构信息. 然后, 我们将这些数据集中的三元组集合分别拆分成训练集, 验证集和测试集. 表 2 给出了这 4 个数据集的详细描述.

表 2 实验数据集的统计信息

数据集	#Entities	#Relations	#Train / Valid / Test	#Types	#Entity-Type Pairs	#Type-Type Pairs
FB15K	14 951	1 345	483142 / 50000 / 59071	—	—	—
WN18	40 943	18	141442 / 5000 / 5000	—	—	—
YAGO26K	26 078	34	321665 / 30000 / 39073	758	9 962	1 411
DB111K	111 762	305	559729 / 32925 / 65851	242	98 340	164

4.2 链接预测

链接预测是知识图谱表示学习方法的基准测试任务. 它是用来预测一个事实三元组 (h, r, t) 缺失的头实体 h 或尾实体 t , 即给定 (r, t) 预测 h 或者给定 (h, r) 预测 t .

4.2.1 实验部署和度量标准

我们使用 $\text{HiTE}^{\text{DistMult}}$ (见算法 1) 和 $\text{HiTE}^{\text{DistMult}}_{\text{UT}}$ (见算法 2) 分别表示我们的方法在类别信息已知和未知两种情况下的实现. 对于数据集 YAGO26K 和 DB111K, 除了运行 $\text{HiTE}^{\text{DistMult}}$ 之外, 我们同时会删除类别信息这部分数据, 然后运行 $\text{HiTE}^{\text{DistMult}}_{\text{UT}}$, 通过比较这两种方式的运行结果, 可以一定程度上反映出我们的方法在建模类别层次结构的效果.

对于所有数据集, 我们设置最小批量训练集大小为 100, 最大迭代次数为 800. 同时我们使用网格搜索的方法, 通过计算验证集的 MRR 值确定超参数的最优值. 在训练过程中我们使用 Adam 算法^[43]作为梯度优化算法, 并设置学习率的初始值为 0.001. 各个超参数的搜索范围设置如下: 实体和关系嵌入的维度 $d \in \{150, 200, 250, 300\}$, 类别嵌入的维度 $d' \in \{25, 50, 100\}$, 间隔 $\gamma \in \{0.5, 1.0, 1.5\}$, 类别约束权重 $\alpha \in \{0.5, 1.0, 1.5\}$, 正样本对应采样负样本的比例 $\beta \in \{2, 6, 10\}$, l_2 正则化权重 $\lambda \in \{0.003, 0.01, 0.03\}$. 通过搜索得到各个数据集的最优超参数设置如表 3.

表 3 超参数设定

数据集	d	d'	γ	α	β	λ
FB15K	250	100	1.0	0.5	6	0.003
WN18	150	50	0.5	0.5	2	0.003
YAGO26K	250	100	1.0	1.0	6	0.01
DB111K	300	50	1.5	1.0	6	0.03

为了衡量学习出的知识图谱嵌入的好坏, 我们使用两种常用的评估指标: 平均倒数排名 (MRR) 和 $Hits@n$ (有效测试三元组的头实体/尾实体在所有实体集评分排序的前 n 名中的比例, $n=1, 3, 10$). 另外, 我们使用过滤设置, 即计算指标值时不考虑任何知识图谱中已经出现的三元组. MRR 和 $Hits@n$ 的值越大表示学习出的嵌入性能越好, 它们的计算公式如下所示:

$$MRR = \frac{1}{2|T|} \sum_{(h,r,t) \in T} \left(\frac{1}{\text{rank}_{(h,r,t)}^h} + \frac{1}{\text{rank}_{(h,r,t)}^t} \right) \quad (9)$$

$$Hits@n = \frac{1}{2|T|} \sum_{(h,r,t) \in T} \left(\#(\text{rank}_{(h,r,t)}^h \leq n) + \#(\text{rank}_{(h,r,t)}^t \leq n) \right) \quad (10)$$

$$\text{rank}_{(h,r,t)}^h = 1 + \sum_{(h',r,t) \in \tilde{T}} I[f(h,r,t) < f(h',r,t)] \quad (11)$$

$$\text{rank}_{(h,r,t)}^t = 1 + \sum_{(h,r,t') \in \tilde{T}} I[f(h,r,t) < f(h,r,t')] \quad (12)$$

其中, $I[x]=1$ 当且仅当 x 为真, 否则 $I[x]=0$.

4.2.2 对比方法

我们将对比方法分为两类: (1) 基础的知识图谱嵌入方法, 即没有引入任何外部信息的嵌入方法. 我们选取 TransE、DistMult 和 HolE^[44]作为该类的对比方法; (2) 结合实体类别信息的知识图谱嵌入方法. 我们选取 TKRL、TEKRL (值得一提的是, TKRL 和 TEKRL 工作中使用的 FB15K 数据集加入了人工构造的实体类别信息和类别的层次结构信息) 和 JOIE、TransC、DistMult_LEAR、DistMult_L2R 作为该类的对比方法. 其中, DistMult_LEAR 和 DistMult_L2R 方法是在 DistMult 模型的基础上加入基于向量方向与模值^[36]和基于排序学习^[35]建模层次结构的方法以后得到的. 此外, 为了验证建模类别信息的有效性, 我们简单地将实体所属类别的信息使用新的关系“is_a”表示, 并将其转化为三元组添加到已有数据集中训练 DistMult 模型, 然后和该类的方法做比较, 我们使用 DistMult_all 表示这种设定. 除了 DistMult_LEAR 和 DistMult_L2R, 其他所有对比方法在链接预测上的实验结果

都是直接从已发表论文中获取。

4.2.3 实验结果及分析

表 4 和表 5 分别给出了不同方法在类别信息未知的数据集 FB15K、WN18 和类别信息已知的数据集 YAGO26K、DB111K 的测试集上链接预测任务的运行结果。从表中的数据我们可以看出, 对于类别信息未知的数据集 FB15K 和 WN18, 我们的方法与所结合的基础知识图谱嵌入模型 DistMult 相比, *MRR* 值和 *Hits@1* 的值分别提升了 6.7%、1.2% 和 15.2%、2.1%。同时, 对于 FB15K, 我们的方法在 *Hits@10* 指标上与引入类别信息的方法 TKRL 和 TEKRL 相比有 11.3% 和 6.7% 的提升。对于类别信息已知的数据集 YAGO26K, 我们的方法与 DistMult 相比, *MRR* 值、*Hits@1* 和 *Hits@10* 的值分别提升了 123.7%、110.5% 和 146.5%。对于数据集 DB111K, 我们的方法与 DistMult 相比, *MRR* 值、*Hits@1* 和 *Hits@10* 的值分别提升了 27.9%、5.0% 和 65.2%。虽然 DB111K 的实体与类别比例为 461.8, 比 YAGO26K 的 34.4 高, 但是后者的提升效果要明显高于前者。这是因为 YAGO26K 的关系数量比 DB111K 少接近 10 倍, 而且训练集中一个关系对应的不同实体个数的平均值为 2163.2, 比 DB111K 的 1383.7 多接近 1.6 倍。这也说明了我们的方法中使用的关系类别约束对知识图谱嵌入的学习起到了很大效果。同时, 从表 5 中我们可以看到, 对于 YAGO26K, 和其他 3 类建模层次结构的工作 TransC、DistMult_LEAR、DistMult_L2R 和类似的工作 JOIE 相比, 我们的方法在 3 个指标上都全面超越它们; 对于 DB111K, 我们的方法也在 3 个指标上都优于 DistMult_LEAR 和 DistMult_L2R。与 TransC 和 JOIE 相比, 我们的方法在 *Hits@1* 指标上都提升了 10.1%, 这说明了我们方法在建模类别信息上的优越性。

表 4 类别信息未知的两个数据集上链接预测任务的运行结果

方法	WN18			FB15K		
	<i>MRR</i>	<i>Hits@1</i>	<i>Hits@10</i>	<i>MRR</i>	<i>Hits@1</i>	<i>Hits@10</i>
TransE	0.495	0.113	<u>0.943</u>	0.380	0.231	0.641
DistMult	0.822	0.728	0.936	0.654	0.546	0.824
HolE	0.938	0.930	0.949	0.524	0.402	0.739
TKRL	—	—	—	—	—	0.734
TEKRL	—	—	—	—	—	0.766
HiTE ^{DistMult} _UT	<u>0.832</u>	<u>0.743</u>	0.941	0.698	0.629	<u>0.817</u>

注: 加粗表示最优的结果; 下划线表示次优的结果; “—”表示运行结果未知。由于 WN18 数据集的类别信息没有人工给定, TKRL 和 TEKRL 无法运行; TKRL 和 TEKRL 的论文中只给出 FB15K 的 *Hits@10* 结果, 因此没有给出其他指标的结果。

表 5 类别信息已知的两个数据集上链接预测任务的运行结果

方法	YAGO26K			DB111K		
	<i>MRR</i>	<i>Hits@1</i>	<i>Hits@10</i>	<i>MRR</i>	<i>Hits@1</i>	<i>Hits@10</i>
TransE	0.195	0.141	0.345	0.327	0.223	<u>0.490</u>
DistMult	0.253	0.229	0.288	0.265	0.260	0.276
HolE	0.265	0.259	0.283	0.301	0.292	0.315
DistMult_all	0.288	0.241	0.312	0.280	0.272	0.297
JOIE	0.296	0.194	0.455	0.369	0.248	0.559
TransC	0.252	0.157	0.378	<u>0.359</u>	0.248	0.493
DistMult_LEAR	0.479	0.393	0.634	0.166	0.051	0.361
DistMult_L2R	0.530	0.455	0.657	0.240	0.180	0.346
HiTE ^{DistMult} _UT	<u>0.547</u>	<u>0.458</u>	<u>0.702</u>	0.309	0.255	0.403
HiTE ^{DistMult}	0.566	0.482	0.710	0.339	<u>0.273</u>	0.456

另外, 和 DistMult_all 相比, 我们的方法 HiTE^{DistMult} 在各个指标上都要领先, 这也反映了我们的方法建模类别信息的有效性。同时, 在数据集 YAGO26K 和 DB111K 的类别信息未知和已知的两种情况下, 对比我们的方法的两种实现 HiTE^{DistMult}_UT 和 HiTE^{DistMult}, 从它们的运行结果可以看出结合类型的层次结构信息能够进一步增强知识图谱嵌入的性能。

4.2.4 消融实验

为了进一步说明引入类别的层次化结构和关系的类别约束对我们方法的影响,我们在数据集 YAGO26K 和 DB111K 上对它们分别做了消融实验. 实验结果如表 6 所示. 其中,“HiTE^{DistMult} w/o RT, w HT”方法表示我们的方法只使用实体的类别信息和类别的层次化结构,即在公式 (6) 中删除 \mathcal{L}_{RT} 这部分损失;而“HiTE^{DistMult} w/o HT, w RT”方法表示只使用实体的类别信息和关系的类别约束,即在公式 (6) 中删除 \mathcal{L}_{HT} ;“HiTE^{DistMult}”则表示使用类别的层次化结构和关系的类别约束. 从表 6 的结果我们可以看出,相较于类别的层次化结构信息,加入关系的类别约束能够学习出更好地知识图谱嵌入;而同时加入这两种类别信息会更进一步显著地提升知识图谱嵌入的性能. 这说明了类别信息的这两个特性对知识图谱嵌入的性能提升有累加的效果,同样也证明了建模类别的层次化结构和关系的类别约束的重要性.

表 6 数据集 YAGO26K 和 DB111K 上类别层次结构和关系的类别约束的消融实验结果

方法	YAGO26K			DB111K		
	MRR	Hits@1	Hits@10	MRR	Hits@1	Hits@10
HiTE ^{DistMult} w/o RT, w HT	0.531	0.442	0.688	0.284	0.211	0.416
HiTE ^{DistMult} w/o HT, w RT	0.552	0.468	0.708	0.327	0.260	0.450
HiTE ^{DistMult}	0.566	0.482	0.710	0.339	0.273	0.456

4.2.5 惩罚函数的选择

在第 3.1 节中,我们提到了使用惩罚函数来实现类别向量空间的偏序关系. 对于惩罚函数的选择可以是多种多样的,只要能满足 $\forall x, y \in Y, x \leq y$ 即可. 下面我们将选择 3 种惩罚函数来讨论它们对知识图谱嵌入的影响. 除了公式 (1) 给出的惩罚函数,另外两个的表示如下所示:

$$P'(x, y) = \text{maximum}([y - x]_+) \quad (13)$$

$$P''(x, y) = \|[y - x]\|_1 \quad (14)$$

其中, *maximum* 函数表示取一个向量表示中最大的坐标值. 我们分别将这 3 个函数在 HiTE^{DistMult} 代码中实现,并在数据集 FB15K 和 YAGO26K 上运行,得到结果如图 3 所示. 从图中我们可以看到,对于数据集 YAGO26K 的 3 个函数在各个指标上的结果相差不大,且趋势都是一致的. 平方和函数 $P(x, y)$ 要略好于最大值函数 $P'(x, y)$, 最大值函数略好于绝对值求和函数 $P''(x, y)$; 对于数据集 FB15K, 在 MRR 和 Hits@1 指标上,平方和函数 $P(x, y)$ 要明显好于其他两个函数. 因此,我们选取平方和函数作为我们方法的惩罚函数.

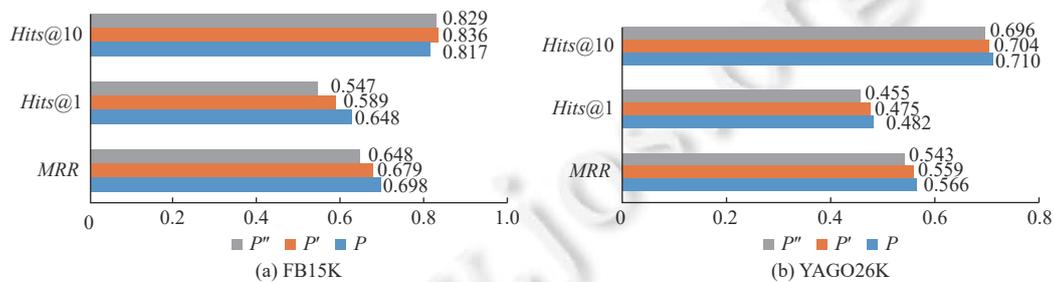


图 3 3 种惩罚函数在 FB15K 和 YAGO26K 数据集上运行链接预测的结果

4.2.6 类别向量表示的结果分析

为了验证我们的方法学习到的类别向量表示包含了类别的层次结构,我们使用最大值惩罚函数 $P'(x, y)$ 并设置类别的向量空间维度为 2, 然后分别在 FB15K 和 YAGO26K 数据集上运行我们的方法,最后得到所有类别的向量表示如图 4 和图 5 所示. YAGO26K 数据集的类别信息是已知的,图 5 只给出了这些已知类别的向量表示;而 FB15K 的类别信息是未知的,因此图 4 给出了通过无监督学习方式得到的所有关系的头实体和尾实体集合对应

的类别,同时给出了所有实体映射到类别向量空间中的位置.图中类别越是靠近原点表示该类别的概念越概括.对于 FB15K 数据集,大部分学习出的类别是靠近原点的,如图 4 中圈出的灰色框所示,它们与实体之间具有明显的偏序关系.另外,通过无监督学习的方式我们的方法还可以学到潜在的类别之间的层次关系.比如关系“/person/gender”的头实体集合对应的类别为 *person*,关系“/basketball_team/player”的尾实体集合对应的类别为 *athlete*,且 $athlete \leq person$,即使没有给出这些类别和它们之间的层次关系,我们的方法也同样学习出 *person* 的坐标值比 *athlete* 小,它们具有层次关系.这说明了我们的方法很好地学习出关系的类别约束.

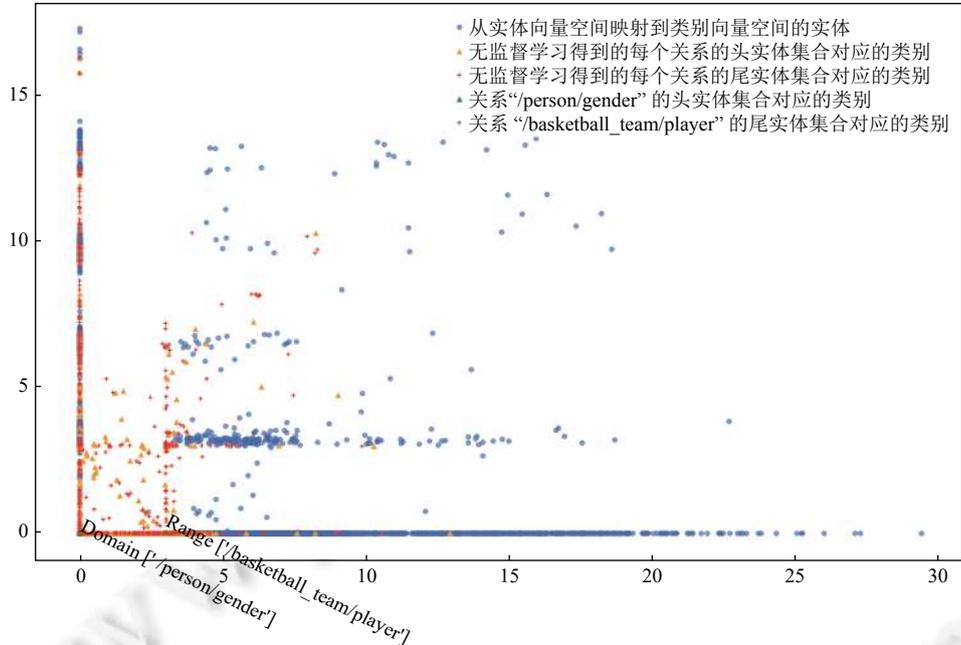


图 4 FB15K 数据集的类别和实体在二维空间上的向量表示

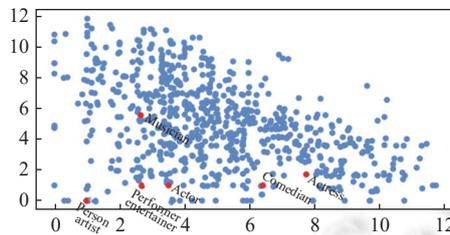


图 5 YAGO26K 数据集的类别在二维空间上的向量表示

对于 YAGO26K 数据集,我们从中选取了 8 个类别,如图 5 中红色的点所示.他们之间存在的层次结构为: $actress \leq actor$, $comedian \leq actor$, $actor \leq artist$, $actor \leq performer$, $musician \leq performer$, $performer \leq entertainer$, $entertainer \leq person$, $artist \leq person$.可以看到,图中它们在空间中的位置大部分都反映出它们的层次结构.由于采样的负样本数量较少,类别 *artist* 和 *person* 没有区别开来.尽管如此,我们的方法比较准确地将类别的层次结构学习出来.这也说明了我们的方法建模层次化类别的有效性.

4.3 三元组分类

三元组分类是知识图谱表示学习方法的另一个测试任务,它是用来判断一个给定的三元组是否为真.

4.3.1 实验设置和实验结果

三元组分类任务是一个二分类问题.我们使用和 TransC 一样的构造分类器的方法.我们为每个关系 r 设置一

个阈值 θ_r , 如果三元组的 $f_{\text{all}}(h, r, t)$ 得分大于 θ_r , 则该三元组为真, 否则为假. 该阈值是通过在验证集中最大化分类准确率得到的. 我们在数据集 YAGO26K 和 DB111K 上执行三元组分类任务, 程序运行时使用的超参数和链接预测任务一样, 对比方法我们分别选取了 TransE、DistMult、HolE、TransC、DistMult_LEAR 和 DistMult_L2R. 运行结果如表 7 所示. 从实验结果中我们可以看到, 相比较于 DistMult, 我们的方法在两个数据集的准确率上分别有 0.9% 和 1.6% 的提升; 和其他对比方法相比, 我们的方法在准确率上能够达到最优和次优的效果. 因此, 三元组分类任务的实验结果进一步验证了我们方法的优越性.

4.3.2 消融实验

为了更进一步地说明引入类别的层次化结构和关系的类别约束对我们方法的影响, 我们针对三元组分类任务在数据集 YAGO26K 和 DB111K 上分别对类别的层次化结构和关系的类别约束做消融实验. 实验结果如表 8 所示. 其中, “HiTE^{DistMult} w/o RT, w HT”和“HiTE^{DistMult} w/o HT, w RT”方法与第 4.2.4 节的一致. 从表 8 的准确率结果中我们也可以看出, 相较于类别的层次化结构信息, 加入关系的类别约束的方法分类准确率更高, 这是因为关系的类别约束能够很好地区分明显不满足约束的负样本.

表 7 数据集 YAGO26K 和 DB111K 上三元组分类任务的运行准确率结果 (%)

数据集	YAGO26K	DB111K
TransE	90.82	86.08
DistMult	92.73	85.99
HolE	89.17	75.50
TransC	93.56	90.74
DistMult_LEAR	90.33	81.34
DistMult_L2R	86.41	74.73
HiTE ^{DistMult} _UT	93.33	86.53
HiTE ^{DistMult}	93.59	87.39

表 8 三元组分类任务上对两个数据集进行类别层次结构和关系的类别约束的消融实验准确率结果 (%)

数据集	YAGO26K	DB111K
HiTE ^{DistMult} w/o RT, w HT	92.54	85.66
HiTE ^{DistMult} w/o HT, w RT	93.39	87.86
HiTE ^{DistMult}	93.59	87.39

4.4 实体分类

将实体的类别信息引入到知识图谱中, 用类别信息帮助知识图谱学习出更好的实体关系向量表示的同时, 我们也学习出语义丰富的类别向量表示和实体到类别向量空间的映射函数. 本节中, 我们使用实体分类任务验证我们的方法学习出的类别向量表示和实体到类别的映射函数的效果. 实体分类任务旨在预测给定的某个实体所属的类别.

4.4.1 实验设置

和 JOIE 工作一样, 我们选取数据集 YAGO26K 和 DB111K, 然后将数据集中实体与其所属类别的集合 I_e 以 6:4 的比例拆分成训练集和测试集, 在训练集和所有三元组数据集 T 、类别层次化结构数据集 I_c 上训练我们的方法, 训练过程中使用的超参数和第 4.2 节链接预测任务大致一样, 除了 YAGO26K 上的两个超参有变动, 即 $\alpha=\gamma=1.5$. 在测试阶段, 我们对测试集中的每一条数据 (e, c) , 分别计算实体和所有类别的得分值 $f_{\text{entity_typing}}(e, c') = [P(\psi(e), c') - \gamma]_+$, 并对这些得分进行排序, 然后计算测试集数据的 MRR 和 $Hits@n$ ($n=1, 3$) 值. 我们使用这些指标去评估各个方法的好坏.

和链接预测一样, 我们选取 TransE、DistMult、HolE 和 JOIE、TransC、DistMult_LEAR、DistMult_L2R 作为对比方法. 对于 TransE、DistMult 和 HolE 方法, 我们将集合 I_e 的数据 (e, c) 转化成三元组 $(e, r_T = \text{“is_a”}, c)$ 添加到已有三元组数据中, 再运行链接预测得到 MRR 和 $Hits@n$ 值, 这些方法和 JOIE 的实验结果是直接从 JOIE 论文中获得的; 而其他方法的结果是运行我们自己实现的代码以后得到的.

4.4.2 实验结果及分析

表 9 给出了各个方法在数据集 YAGO26K 和 DB111K 的测试集上运行实体分类任务的结果. 从表中的数据我们可以看出, 我们的方法在 MRR 和 $Hits@1$ 指标上分别达到次优和最优的结果. 对于数据集 YAGO26K, 我们的方法与类似的工作 TransC 相比, MRR 值和 $Hits@1$ 的值分别提升了 27.1% 和 57.4%; 与 JOIE 相比, 我们的方法在

表 9 数据集 YAGO26K 和 DB111K 上实体分类任务的运行结果

方法	YAGO26K			DB111K		
	MRR	Hits@1	Hits@3	MRR	Hits@1	Hits@3
TransE	0.144	0.073	0.353	0.503	0.437	0.608
DistMult	0.411	0.361	0.553	0.551	0.498	0.680
HolE	0.395	0.348	0.548	0.504	0.448	0.654
JOIE	0.865	0.816	0.918	0.778	0.694	0.857
TransC	0.639	0.507	0.752	0.798	0.721	0.859
DistMult_LEAR	0.549	0.406	0.633	0.645	0.479	0.792
DistMult_L2R	0.525	0.434	0.584	0.651	0.558	0.812
HiTE ^{DistMult}	0.812	0.798	0.815	0.830	0.787	0.851

5 总结

由于实体的类别信息包含了丰富的语义信息,能够很好地指导实体和关系向量表示的学习,本文提出一种新颖的结合具有层次化结构的类别信息的知识图谱表示学习方法.我们将类别映射到不同的向量空间中学习它们的向量表示,然后使用偏序关系建模类别的层次化结构,保证学习出的类别向量表示满足偏序关系,从而保证具有层次结构的类别与类别之间具有反对称性和传递性.同时,我们通过线性变换将实体向量映射到类别向量空间中表示实体与其所属类别之间的关系.另外,知识图谱中的每个关系潜在地扮演着不同的角色,它规定了关系关联的头实体和尾实体分别应该属于什么类别.我们同时也建模了这种关系的类别约束,即将关系对应的头实体(尾实体)映射到类别的向量空间中,并保证它与约束关系的类别满足偏序关系.然而,大部分知识图谱很难得到关系的类别约束信息,此时我们的方法仍然可以使用该约束.我们为每个关系分别指定不同的约束头实体和尾实体的类别,根据知识图谱已有的三元组数据可以自动学习出这些类别的结构信息,从而进一步指导实体向量表示的学习.

实验结果表明了我们的方法在 4 个数据集上的大部分评价指标结果比对比的基线方法效果好.同时,当类别未知时,我们仍可以使用关系的类别约束,实验结果也证实了该情况下我们的方法的有效性.在未来的工作中,我们将进一步考虑使用高斯分布建模类别信息,比较它与偏序关系之间的优劣.此外,还将进一步研究将类别信息和逻辑规则一起结合起来提升知识图谱表示学习能力.

References:

- [1] Lehmann J, Isele R, Jakob M, Jentzsch A, Kontokostas D, Mendes PN, Hellmann S, Morsey M, Van Kleef P, Auer S, Bizer C. DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 2015, 6(2): 167–195. [doi: 10.3233/SW-140134]
- [2] Rebele T, Suchanek F, Hoffart J, Biega J, Kuzey E, Weikum G. YAGO: A multilingual knowledge base from Wikipedia, WordNet, and GeoNames. In: Groth P, Simperl E, Gray A, Sabou M, Krötzsch M, Lécué F, Flöck F, Gil Y, eds. *Proc. of the 15th Int'l Semantic Web Conf.* Kobe: Springer, 2016. 177–185. [doi: 10.1007/978-3-319-46547-0_19]
- [3] Bollacker K, Evans C, Paritosh P, Sturge T, Taylor J. Freebase: A collaboratively created graph database for structuring human knowledge. In: *Proc. of the 2008 ACM SIGMOD Int'l Conf. on Management of Data.* Vancouver: ACM, 2008. 1247–1250. [doi: 10.1145/1376616.1376746]
- [4] Hua W, Wang ZY, Wang HX, Zheng K, Zhou XF. Understand short texts by harvesting and analyzing semantic knowledge. *IEEE Trans. on Knowledge and Data Engineering*, 2017, 29(3): 499–512. [doi: 10.1109/TKDE.2016.2571687]
- [5] Dong L, Wei FR, Zhou M, Xu K. Question answering over freebase with multi-column convolutional neural networks. In: *Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing (Volume 1: Long Papers).* Beijing: ACL, 2015. 260–269. [doi: 10.3115/v1/P15-1026]
- [6] Cui WY, Xiao YH, Wang HX, Song YQ, Hwang SW, Wang W. KBQA: Learning question answering over QA corpora and knowledge bases. *Proc. of the VLDB Endowment*, 2017, 10(5): 565–576. [doi: 10.14778/3055540.3055549]
- [7] Hoffmann R, Zhang CL, Ling X, Zettlemoyer L, Weld DS. Knowledge-based weak supervision for information extraction of overlapping

- relations. In Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Portland: ACL, 2011. 541–550.
- [8] Zhang FZ, Yuan NJ, Lian DF, Xie X, Ma WY. Collaborative knowledge base embedding for recommender systems. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 353–362. [doi: [10.1145/2939672.2939673](https://doi.org/10.1145/2939672.2939673)]
- [9] Wang Z, Zhang JW, Feng JL, Chen Z. Knowledge graph and text jointly embedding. In: Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing. Doha: ACL, 2014. 1591–1601. [doi: [10.3115/v1/D14-1167](https://doi.org/10.3115/v1/D14-1167)]
- [10] Xie RB, Liu ZY, Jia J, Luan HB, Sun MS. Representation learning of knowledge graphs with entity descriptions. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. Phoenix: AAAI Press, 2016. 2659–2665.
- [11] Krompaß D, Baier S, Tresp V. Type-constrained representation learning in knowledge graphs. In: Proc. of the 14th Int'l Semantic Web Conf. Bethlehem: Springer, 2015. 640–655. [doi: [10.1007/978-3-319-25007-6_37](https://doi.org/10.1007/978-3-319-25007-6_37)]
- [12] Xie RB, Liu ZY, Sun MS. Representation learning of knowledge graphs with hierarchical types. In: Proc. of the 25th Int'l Joint Conf. on Artificial Intelligence. New York: AAAI Press, 2016. 2965–2971.
- [13] Guo S, Wang Q, Wang B, Wang LH, Guo L. Semantically smooth knowledge graph embedding. In: Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int'l Joint Conf. on Natural Language Processing (Volume 1: Long Papers). Beijing: ACL, 2015. 84–94. [doi: [10.3115/v1/P15-1009](https://doi.org/10.3115/v1/P15-1009)]
- [14] Rocktäschel T, Singh S, Riedel S. Injecting logical background knowledge into embeddings for relation extraction. In: Proc. of the 2015 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Denver: ACL, 2015. 1119–1129. [doi: [10.3115/v1/N15-1118](https://doi.org/10.3115/v1/N15-1118)]
- [15] Guo S, Wang Q, Wang LH, Wang B, Guo L. Knowledge graph embedding with iterative guidance from soft rules. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. New Orleans: AAAI Press, 2018. 4816–4823.
- [16] Zhang JD, Li J. Enhanced knowledge graph embedding by jointly learning soft rules and facts. *Algorithms*, 2019, 12(12): 265. [doi: [10.3390/a12120265](https://doi.org/10.3390/a12120265)]
- [17] Lv X, Hou L, Li JZ, Liu ZY. Differentiating concepts and instances for knowledge graph embedding. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: ACL, 2018. 1971–1979.
- [18] Dong TS, Wang ZG, Li JZ, Bauckhage C, Cremers AB. Triple classification using regions and fine-grained entity typing. In: Proc. of the 33rd AAAI Conf. on Artificial Intelligence. Honolulu: AAAI Press, 2019: 77–85. [doi: [10.1609/aaai.v33i01.330177](https://doi.org/10.1609/aaai.v33i01.330177)]
- [19] Ren HY, Hu WH, Leskovec J. Query2box: Reasoning over knowledge graphs in vector space using box embeddings. In: Proc. of the 8th Int'l Conf. on Learning Representations. 2020.
- [20] Hao JH, Chen MH, Yu WC, Sun YZ, Wang W. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts. In: Proc. of the 25th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Anchorage: ACM, 2019. 1709–1719. [doi: [10.1145/3292500.3330838](https://doi.org/10.1145/3292500.3330838)]
- [21] Chen M, Tian Y, Chen X, Xue Z, Zaniolo C. On2Vec: Embedding-based relation prediction for ontology population. In: Proc. of the 2018 SIAM Int'l Conf. on Data Mining. San Diego: SIAM, 2018. 315–323.
- [22] Jin J, Wang HY, Lin YF. Knowledge graph representation learning fused with entity category information. *Computer Engineering*, 2021, 47(4): 77–83 (in Chinese with English abstract). [doi: [10.19678/j.issn.1000-3428.0057353](https://doi.org/10.19678/j.issn.1000-3428.0057353)]
- [23] Vendrov I, Kiros R, Fidler S, Urtasun R. Order-embeddings of images and language. arXiv: 1511.06361, 2016.
- [24] Wang Q, Mao ZD, Wang B, Guo L. Knowledge graph embedding: A survey of approaches and applications. *IEEE Trans. on Knowledge and Data Engineering*, 2017, 29(12): 2724–2743. [doi: [10.1109/TKDE.2017.2754499](https://doi.org/10.1109/TKDE.2017.2754499)]
- [25] Liu ZY, Sun MS, Lin YK, Xie RB. Knowledge representation learning: A review. *Journal of Computer Research and Development*, 2016, 53(2): 247–261 (in Chinese with English abstract). [doi: [10.7544/issn1000-1239.2016.20160020](https://doi.org/10.7544/issn1000-1239.2016.20160020)]
- [26] Guan SP, Jin XL, Jia YT, Wang YZ, Cheng XQ. Knowledge reasoning over knowledge graph: A survey. *Ruan Jian Xue Bao/Journal of Software*, 2018, 29(10): 2966–2994 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/5551.htm> [doi: [10.13328/j.cnki.jos.005551](https://doi.org/10.13328/j.cnki.jos.005551)]
- [27] Bordes A, Usunier N, García-Durán A, Weston J, Yakhnenko O. Translating embeddings for modeling multi-relational data. In: Proc. of the 26th Annual Conf. on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc., 2013. 2787–2795.
- [28] Yang BS, Yih WT, He XD, Gao JF, Deng L. Embedding entities and relations for learning and inference in knowledge bases. In: Bengio Y and LeCun Y eds., Proc. of the 3rd Int'l Conf. on Learning Representations. San Diego, 2015.
- [29] Dettmers T, Minervini P, Stenetorp P, Riedel S. Convolutional 2D knowledge graph embeddings. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. New Orleans: AAAI Press, 2018. 1811–1818.

- [30] Nickel M, Tresp V, Kriegel HP. A three-way model for collective learning on multi-relational data. In: Proc. of the 28th Int'l Conf. on Machine Learning. Bellevue: Omnipress, 2011. 809–816.
- [31] Guan NN, Song DD, Liao LJ. Knowledge graph embedding with concepts. Knowledge-Based Systems, 2019, 164: 38–44. [doi: 10.1016/j.knsys.2018.10.008]
- [32] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Su J, Duh K, Carreras X, eds., Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Austin: ACL Press, 2016. 1369–1378.
- [33] Ren X, He WQ, Qu M, Huang LF, Ji H, Han JW. AFET: Automatic fine-grained entity typing by hierarchical partial-label embedding. In: Proc. of the 2016 Conf. on Empirical Methods in Natural Language Processing. Austin: ACL, 2016. 1369–1378. [doi: 10.18653/v1/D16-1144]
- [34] Ren X, He WQ, Qu M, Voss CR, Ji H, Han JW. Label noise reduction in entity typing by heterogeneous partial-label embedding. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 1825–1834. [doi: 10.1145/2939672.2939822]
- [35] Chen TF, Chen YM, van Durme B. Hierarchical entity typing via multi-level learning to rank. In: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. Online: ACL, 2020. 8465–8475. [doi: 10.18653/v1/2020.acl-main.749]
- [36] Vulić I, Mrkšić N. Specialising word vectors for lexical entailment. In: Proc. of the 2018 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). New Orleans: ACL, 2018. 1134–1145. [doi: 10.18653/v1/N18-1103]
- [37] Kamath A, Pfeiffer J, Ponti EM, Glavaš G, Vulić I. Specializing distributional vectors of all words for lexical entailment. In: Proc. of the 4th Workshop on Representation Learning for NLP. Florence: ACL, 2019. 72–83. [doi: 10.18653/v1/W19-4310]
- [38] Athiwaratkun B, Wilson AG. Hierarchical density order embeddings. In: Proc. of the 6th Int'l Conf. on Learning Representations. Vancouver. 2018.
- [39] Nickel M, Kiela D. Poincaré embeddings for learning hierarchical representations. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 6341–6350.
- [40] Nickel M, Kiela D. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In: Proc. of the 35th Int'l Conf. on Machine Learning. Stockholm: PMLR Press, 2018. 3779–3788.
- [41] Ganea OE, Bécigneul G, Hofmann T. Hyperbolic entailment cones for learning hierarchical embeddings. In: Proc. of the 35th Int'l Conf. on Machine Learning. Stockholm: PMLR Press. 2018. 1646–1655.
- [42] Miller GA. WordNet: A lexical database for English. Communications of the ACM, 1995, 38(11): 39–41. [doi: 10.1145/219717.219748]
- [43] Kingma DP, Ba J. Adam: A method for stochastic optimization. In: Bengio Y and LeCun Y eds., Proc. of the 3rd Int'l Conf. on Learning Representations. San Diego. 2015.
- [44] Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs. In: Proc. of the 30th AAAI Conf. on Artificial Intelligence. Phoenix: AAAI Press, 2016. 1955–1961.

附中文参考文献:

- [22] 金婧, 万怀宇, 林友芳. 融合实体类别信息的知识图谱表示学习. 计算机工程, 2021, 47(4): 77–83. [doi: 10.19678/j.issn.1000-3428.0057353]
- [25] 刘知远, 孙茂松, 林衍凯, 谢若冰. 知识表示学习研究进展. 计算机研究与发展, 2016, 53(2): 247–261. [doi: 10.7544/issn1000-1239.2016.20160020]
- [26] 官赛萍, 靳小龙, 贾岩涛, 王元卓, 程学旗. 面向知识图谱的知识推理研究进展. 软件学报, 2018, 29(10): 2966–2994. <http://www.jos.org.cn/1000-9825/5551.htm> [doi: 10.13328/j.cnki.jos.005551]



张金斗(1988—), 男, 博士, 主要研究领域为知识图谱分布式表示, 分布式系统, 大数据处理.



李京(1966—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为分布式系统, 云计算, 大数据处理, 移动计算, 区块链.