

融合实体类型信息的本体-实例联合学习方法

游乐圻, 裴忠民, 罗章凯

(航天工程大学 复杂电子系统仿真重点实验室, 北京 101416)

摘要: 对表示知识图谱的本体图和实例图进行联合学习能够提高嵌入学习效率,但不能区别表示实体在不同场景下的不同意义。在嵌入时考虑三元组中实体的关系类型特征,提出一种融合实体类型信息的本体-实例联合学习方法 JOIE-TKRL-CT,达到在联合学习中表示多义实体、提高知识图谱嵌入学习效率的目的。在视图内部关系表示上,利用实体分层类型模型融入实体类型信息,在两个独立的嵌入空间中分别表征学习;在视图间关系表示上,将表征在两个独立空间的本体和实例通过非线性映射的方法跨视图链接。基于 YAGO26K-906 和 DB111K-174 数据集的实验结果表明,JOIE-TKRL-CT 能够准确捕获知识图谱的实体类型信息,提高联合学习模型性能,与 TransE、HolE、DisMult 等基线模型相比,其在实例三元组补全和实体分类任务上均获得最优性能,具有较好的知识学习效果。

关键词: 联合学习; 实体多义性; 跨视图转换; 分层类型模型; 三元组补全; 实体分类

开放科学(资源服务)标志码(OSID):



中文引用格式: 游乐圻, 裴忠民, 罗章凯. 融合实体类型信息的本体-实例联合学习方法[J]. 计算机工程, 2022, 48(7): 82-88.

英文引用格式: YOU L Q, PEI Z M, LUO Z K. Ontology-instances joint learning method fusing entity type information[J]. Computer Engineering, 2022, 48(7): 82-88.

Ontology-Instances Joint Learning Method Fusing Entity Type Information

YOU Leqi, PEI Zhongmin, LUO Zhangkai

(Key Laboratory of Complex Electronic System Simulation, Space Engineering University, Beijing 101416, China)

[Abstract] The joint learning of the ontology graph and the instance graph that represent the knowledge graph can effectively improve the efficiency of embedded learning, but it cannot represent the ambiguity of entities in different scenarios. In this regard, this paper integrates the relationship type characteristics of the entities in the triples when embedding, and proposes an ontology-instances joint learning method, JOIE-TKRL-CT, that fuses entity type information to achieve the goals of representing polysemous entities in joint learning and improving the efficiency of knowledge map embedded learning. This method expresses the relationship within the view, integrates the entity type information based on the entity hierarchical type model, and characterizes learning in two independent embedding spaces. Meanwhile, it expresses the relationship between the views, passes the ontology and instance represented in the two independent spaces, and uses non-linear mapping methods for cross-view linking. Multi-group comparison experiments are performed on the YAGO26K-906 and DB111K-174 data sets, and the results show that JOIE-TKRL-CT can accurately capture the hierarchical type information of the knowledge graph and improve the performance of the joint learning model. Compared with baseline models such as TransE, HolE, and DisMult, this model achieves the best performance in instance triple completion and entity classification tasks, and has a better knowledge learning effect.

[Key words] joint learning; entity polysemy; cross-view transformation; hierarchical type model; triple completion; entity classification

DOI: 10.19678/j.issn.1000-3428.0062113

0 概述

知识图谱^[1]可以形式化展示事物之间的关系,其本质是一种语义网络。近年来,知识图谱已成为支撑对话代理^[2]、智能问答^[3]、新闻推荐^[4]等智能生活应用的技术基础。目前,已经有许多比较成熟的

知识图谱,例如 Freebase^[5]、YAGO^[6]、DBPedia^[7]、ConceptNet^[8]等,这些知识图谱都可以分为两种基本视图,即实例视图及其本体视图。本体视图和实例视图相辅相成,对两者进行联合学习可以掌握更全面的知识。通常,研究者将实例填充到本体视图中形成一个完整的知识图谱,再用知识表示学习的方法

基金项目:国防科技重点实验室基础研究项目(DXZT-JC-ZZ-2018-007)。

作者简介:游乐圻(1997—),男,硕士研究生,主研方向为人工智能、系统科学;裴忠民(通信作者),副研究员、博士;罗章凯,助理研究员、博士。

收稿日期:2021-07-19 修回日期:2021-09-07 E-mail: 1152549473@qq.com

法嵌入学习。文献[9]提出将知识图谱的实例和本体联合嵌入的学习方法 JOIE, 并在公开数据集 YAGO26K-906 和 DB111K-174 上验证了该模型的优越性。该方法避免了耗时耗力的本体实例填充工作, 并且与传统方法相比更完整地保存了实例之间的关系。同时, 作者在文末也指出联合学习模型的损失还有待进一步优化, 可以考虑复杂的嵌入模型、融入知识图谱全局结构、类型信息等方式来改进模型性能。其中, 同一实体在不同关系下具有多义性的问题有待进一步研究。

本文通过考虑三元组中实体的关系类型特征, 提出一种融合实体类型信息的本体-实例联合学习方法 JOIE-TKRL-CT。在本体和实例视图的内部关系表征上, 利用实体分层类型模型融入实体类型信息, 在两个独立的嵌入空间中分别表征学习。同时, 通过分层类型模型将每个三元组的头实体和尾实体转化成具有特定类型的投影矩阵嵌入, 以此来表征多义性实体。在视图间关系的表征上, 运用文献[9]提出的跨视图转换模型, 将表征在两个独立空间的本体和实例通过非线性映射的方法跨视图链接, 从而区别表示实体在不同场景下的不同意义。

1 相关研究

知识图谱可以形式化展示事物之间关系, 其本质是一种语义网络。近年来, 知识图谱已成为支撑对话代理、智能问答等智能生活应用的技术基础。知识图谱可以用(实体, 关系, 实体)三元组来表示, 现有的知识图谱大多可以分为两种基本视图, 即实例视图及其本体视图, 如图1所示。实例视图主要包含三元组特定实体及其之间的关系, 其本体视图主要描述的是抽象概念及其语义元关系。两个视图之间通过本体和实例的类属关系链接。

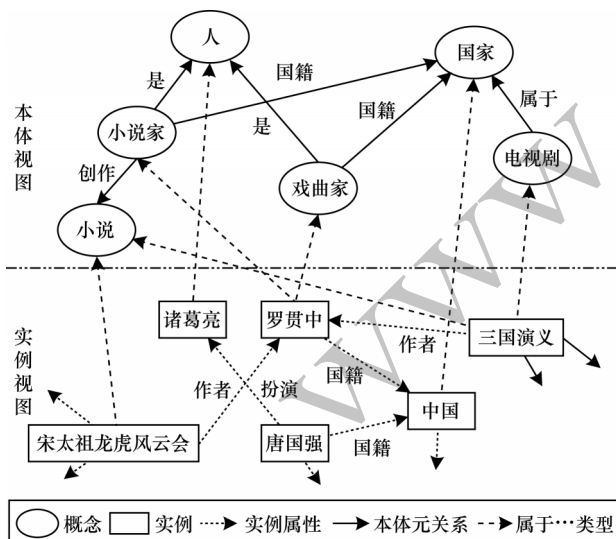


图1 知识图谱中本体和实例视图示例

Fig.1 Example of ontology and instance views in knowledge graph

通过对本体视图和实例视图进行联合学习, 可以掌握更全面的知识。一方面, 实例可以为其对应

的本体概念提供详细的信息; 另一方面, 本体提供了对应实例的高级抽象, 这在发现一些数量较少的实例时尤为重要。可以考虑复杂的嵌入模型、融入知识图谱全局结构、类型信息等方式来改进模型性能。其中, 同一实体在不同关系下多义性的问题就有待进一步研究。如在图2中, 有多个属性指向的有“罗贯中”“三国演义”两个实体, 其中两个属性指向的“三国演义”实体具有多义性, 一个指的是小说, 一个指的是电视剧。如果计算机像人一样, 能够分辨出不同场景对应的实体语义, 那么就可以在一定程度上提高模型性能。但是实体的多义性在计算机嵌入学习过程中无法分辨。因此, 使得拥有多种类型的实体在不同的场景中有不同的表示, 是提高联合学习模型识别性能的有效途径。

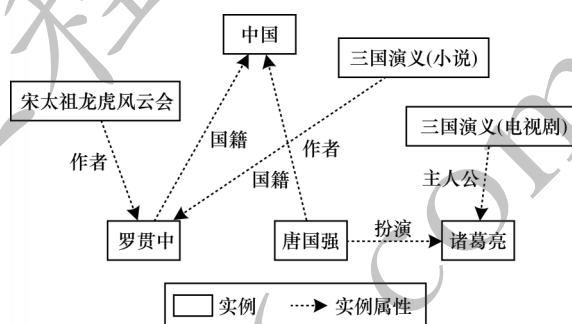


图2 知识图谱中的实体多义性示例

Fig.2 Example of entity polysemy in knowledge graph

知识图谱嵌入方法的不断优化推动了图谱构建和应用技术的快速发展。知识图谱嵌入就是对图谱中三元组的实体和关系表示学习。给定三元组 (h, r, t) , 其中 h, t 分别表示头实体和尾实体, r 表示实体之间的关系, 知识图谱嵌入即是将其转化为向量空间模型。在嵌入模型中, 基于平移距离模型的模型有 TransE^[10]、TransH^[11]、TransR^[12]、TransD^[13]、TranSparse^[14]、KG2E^[15] 等, 基于语义匹配的模型有 DisMult^[16]、HolE^[17] 等。其中, TransE 模型最为经典, 其设计思想是期望标准三元组 (h, r, t) 有 $h+r=t$, 定义距离 d 表示向量之间的距离, 因此机器学习到正确的三元组的距离越小越好, 而错误的三元组的距离越大越好, 其损失函数可以表示为:

$$E(h, r, t) = \|h + r - t\| \quad (1)$$

将表示知识图谱的本体图和实例图联合学习, 在性能上优于单个完整知识图谱的嵌入学习方法。自文献[18]提出本体嵌入方法 On2Vec 以来, 本体中的元信息嵌入表示得到广泛关注。文献[9]提出将知识图谱的实例库和本体库联合嵌入的 JOIE 模型, 采用跨视图和视图内建模的方法, 从知识图谱的本体元层次和实例层次分别表示学习。

在嵌入过程中融合实体类型信息可以有效提高模型性能。文献[19]提出一种能够区分概念和实例信息的 TransC 模型, 其将知识图谱中的概念编码为球体, 实例用向量表示, 在 YAGO 数据集上嵌入性能达到了相对最优。文献[20]将实体类型信息作为先验知识融入模型中, 模型在链接预测任务中性能提

升高达77%。文献[21]提出语义平滑嵌入的SSE模型,该模型的核心思想是让同一语义类别的实体在嵌入空间中位置相近。文献[22]提出TKRL模型,其将类型信息分层表示,并作为关系的特定约束。实验结果表明,利用此方法嵌入学习能捕获知识图谱中重要意义的分层类型信息。文献[23]建立一种基于知识表示学习的知识可信度评估模型PTCA,利用实体类型信息对知识可信度进行计算,从而有效应用知识图谱,对知识可信度进行准确评估。

结合现有研究,本文借鉴TKRL模型中对类型信息分层表示的思想,将类型信息融入本体-实例联合学习模型中,目的是表示实体在不同场景的不同意义,进一步提高知识图谱嵌入的学习效率。

2 JOIE-TKRL-CT方法

为弥补本体-实例联合学习模型不能区分表示多义性实体的不足,本文提出一种融合实体类型信息本体-实例联合学习方法JOIE-TKRL-CT。该方法将本体和实例知识库分别嵌入,再进行联合学习,其由视图内学习、视图间学习和联合学习3个模型组成。在本体和实例视图内部的嵌入学习中,为了在不同的三元组语境凸显相应的类型信息,本文将实体类型信息分层表示,在构建实体的投影矩阵时,类型信息由各类型层加权表示。视图间的学习采用跨视图转换方法来学习本体嵌入空间和实例嵌入空间之间的关联,在正样本中关联即为实体和本体间的类属关系。

2.1 视图内学习

2.1.1 本体和实例视图嵌入

在进行视图内部学习时,本文基于TransE模型,在表示实体时将类型信息作为实体的补充信息一同嵌入。具体而言,将每个类型 c 的层次信息构造成投影矩阵 M_c ,这样头实体 h 和尾实体 t 对应的特定类型就为 c_{rh} 、 c_{rt} ,相应的类型矩阵用 M_{rh} 、 M_{rt} 表示。再根据式(1),可得能量函数和损失函数分别为:

$$E(h, r, t) = \|M_{rh}h + r - M_{rt}t\| \quad (2)$$

$$L = \sum_{(h, r, t) \in T} \sum_{(h', r', t') \in T} \min(\gamma + E(h, r, t) - E(h', r', t')) \quad (3)$$

其中: $\gamma > 0$; $E(h, r, t)$ 是正确的三元组的能量函数; $E(h', r', t')$ 是错误的三元组的能量函数; (h', r', t') 通过 (h, r, t) 替换头实体和尾实体得到,并且替换后的三元组在视图中不存在。学习一个融合类型关系的知识图 T 中所有节点的嵌入,最小化所有的三元组损失,可得:

$$J_{\text{Intra}}^T = \frac{1}{|T|} \sum_{\substack{(h, r, t) \in T \\ (h', r', t') \notin T}} [\gamma^T + (h', r', t')^2 - (h, r, t)^2] \quad (4)$$

机器学习的过程就是运用梯度下降法更新上述函数,其中, $\gamma^T > 0$,是一个正余量。

在嵌入过程中需要考虑到实例内部视图和本体内部视图。本文用 $(h^{(i)}, r^{(i)}, t^{(i)}) \in T_i$ 表示实例视图,用 $(h^{(o)}, r^{(o)}, t^{(o)}) \in T_o$ 表示本体视图,分别用 $J_{\text{Intra}}^{T_i}$ 、 $J_{\text{Intra}}^{T_o}$ 表示计算得到的损失函数,再将两个损失函数通过组合权重 α_i 来调控内部视图的结构丢失。视图内嵌入的整体学习模型可表示为:

$$J_{\text{Intra}} = J_{\text{Intra}}^{T_i} + \alpha_i J_{\text{Intra}}^{T_o} \quad (5)$$

2.1.2 分层类型模型

本节用分层类型模型来表示实体在不同场景下的类型信息,主要是考虑判别三元组中实体类型的情境通过三元组的关系传达,例如在三元组(三国演义,作者,罗贯中)中,通过关系“作者”就可以辨别实体“三国演义”是属于小说类,而不是电视剧。每个类型分多层来表示,是因为可以通过不同的权重组合对应不同的类型。具体说明如下:

设某实体有 n 个类别,则其类别集合可表示为 $c = \{c_1, c_2, \dots, c_n\}$, c_i 表示实体的第 i 个类别。又设每个类别是由多个分层子类别集成控制,如图3所示。假设子类别个数为 m ,则 $c_i = \{c_i^{(1)}, c_i^{(2)}, \dots, c_i^{(m)}\}$,其中: $c_i^{(j)}$ 表示实体第 i 个类型的第 j 个子类型。

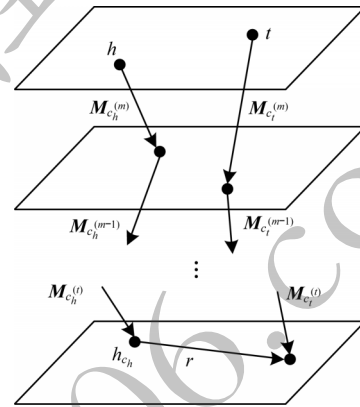


图3 分层类型模型

Fig.3 Hierarchical type model

在分层结构中,不同子类型的不同粒度在映射实体时,需要通过权重相加的方法达到区分实体不同类型信息的目的,如下所示:

$$M_c = \sum_{i=1}^m \beta_i M_{c^{(i)}} = \beta_1 M_{c^{(1)}} + \beta_2 M_{c^{(2)}} + \dots + \beta_m M_{c^{(m)}} \quad (6)$$

$$\beta_i : \beta_{i+1} = (1 - \eta) : \eta, \sum_{i=1}^m \beta_i = 1, \eta \in (0, 0.5) \quad (7)$$

其中: $M_{c^{(i)}}$ 是 $c^{(i)}$ 的投影矩阵; β_i 是 $c^{(i)}$ 的权重。由式(6)可知, $\beta_i > \beta_{i+1}$,以此控制在 M_c 上 $c^{(i)}$ 的影响比 $c^{(i+1)}$ 大。头实体的投影矩阵可表示为:

$$M_{rh} = \frac{n \sum_{i=1}^n \varepsilon_i M_{c_i}}{\sum_{i=1}^n \varepsilon_i}, \varepsilon_i = \begin{cases} 1, c_i \in C_{rh} \\ 0, c_i \notin C_{rh} \end{cases} \quad (8)$$

其中: C_{rh} 表示由关系特定类型信息给出的相对 r 的头类型集; M_{c_i} 为 c_i 的投影矩阵; ε_i 为 M_{c_i} 对应的权重,通过控制 ε_i 的大小分布达到区分不同类型下相同实体的目的。尾部实体的投影矩阵 M_{rt} 与头部实体的投影矩阵形式相同。

2.2 视图间学习

视图间学习采用文献[9]中提出的跨视图转换方法。如图4所示,本体图 O 和实例图 I 分别嵌入在不同的向量空间中,跨视图转换旨在通过学习本体和对应实例的类属关系,使得计算机在嵌入空间中认为本体和相应的实例通过一种非线性映射关系来链接。

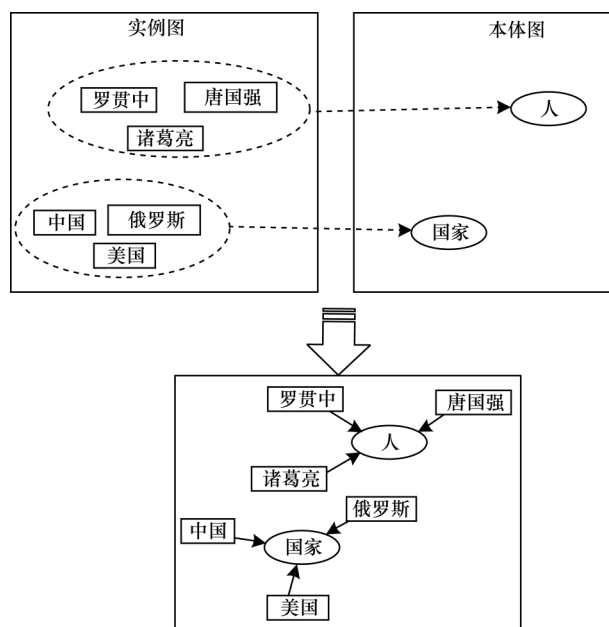


图4 跨视图转换示意图

Fig.4 Schematic diagram of cross-view transformation

假设属于本体 o 的实例 i 通过关系 f 会映射到对应的本体 o' , $o \in O, i \in I$, 即:

$$o \leftarrow f(i), \forall (o, c) \in S \quad (9)$$

其中: $f(i) = \sigma(\mathbf{w} \cdot i + b)$, 是一个非线性映射; \mathbf{w} 为权重矩阵; b 为偏置项; σ 为激活函数 \tanh 。因此, 视图间学习的损失函数可以表示为:

$$J_{\text{cross}} = \frac{1}{|S|} \sum_{\substack{(i, o) \in S_A \\ (i, o') \in S_B}} [\gamma^c + \|o - f(i)\|_2 - \|o' - f(i)\|_2]_+ \quad (10)$$

通过上述跨视图的转换, 可以让计算机学习到一个虚拟的空间中, 本体及其对应的实例会在空间中聚合为距离较近的状态。

2.3 联合学习

将视图内学习模型和视图间学习模型整合, 可得本体-实例联合学习模型:

$$J = J_{\text{Intra}} + \omega J_{\text{cross}}, \omega > 0 \quad (11)$$

其中: ω 是权重系数。在训练的过程中, 先更新底层的 $J_{\text{Intra}}^T, J_{\text{Intra}}^O, J_{\text{cross}}$, 再更新 J 。

3 实验与结果分析

通过三元组补全和实体分类两个任务, 验证本文模型的有效性。

3.1 实验前期准备

实验前期准备工作包括使用的数据集、实验评价指标、实验参数设置。

3.1.1 数据集

通常, 知识图谱中实例的数量会远大于本体的数量。为验证模型的有效性, 分别在 YAGO26K-906 和 DB111K-174 数据集上进行实验(数据来源: <https://github.com/JunhengH/joie-kdd19>), 这两个数据集是在 YAGO、DBpedia 基础上分离出本体、实例制作而成。数据集数据分布统计情况如表 1 所示。

表1 数据分布统计

Table 1 Data distribution statistics

数据集	实例图			本体图			类属
	实体数量	关系数量	三元组数量	实体数量	关系数量	三元组数量	关系数量
YAGO26K-906	26 078	34	390 738	906	30	8 962	9 962
DB111K-174	111 762	305	863 643	174	20	763	99 748

实验所需错误三元组的构建是将正确三元组的头实体、尾实体或者关系变成知识图谱中原来就存在的实体或者关系。在整个数据集中保持正确三元组与错误三元组之比为 1:1。

3.1.2 评价指标

实验选用以下评价指标:

1) MRR。将预测中排名第一的相关结果位置记作 k_i , 则 MRR 为 $\frac{1}{N} \sum_i \frac{1}{k_i}$ 。

2) Hit@10, 表示正确的实体或关系在预测中排名前十的个数占总数的比例。

3) Hit@1, 表示正确的实体或关系在预测中排名第一的个数占总数的比例。

3.1.3 参数设置

为保证公平性, 实验过程中保持以下参数一致: 本体图嵌入维度为 100, 实例图嵌入维度为 300, $\alpha_1 = 2.5$, $\gamma^T = \gamma^O = 1$, 实体类型维度 $n = 50$, epoch 为 100, 训练学习率 lr 为 0.000 5, 单视图数据 batch size 为 128, 联合学习中实例图数据 batch size 为 256, 本体图数据 batch size 为 64。训练的过程使用 AMSGrad 优化器优化联合损失。为了避免有的实体在训练过程中被忽略, 每次训练之前将实体和关系向量归一化。

3.2 三元组补全

三元组补全由文献[24]提出, 目标是在三元组 (h, r, t) 中 h, r, t 缺失时将其补全, 在知识图谱嵌入模型验证时经常被用作测试模型学习嵌入性能的检验方法。测试的方法是在模型内部将预测的实体或关系的似然函数进行计算和排名。在本次实验中, 分为本体三元组补全和实例三元组补全两个子任务。

设置以下 6 组对照模型:

1) TransE 模型^[10]。该模型参数少, 计算复杂度低, 并且具有较好的性能和可扩展性, 是经典的基于平移距离的知识图谱嵌入模型。

2) TransC 模型^[19]。该模型是融合了实体类型信息的知识图谱嵌入模型, 与同期模型相比具有较好的效果。

3) DisMult 模型^[16]。该模型是基于语义匹配的知识图谱嵌入模型, 其利用学到的关系来挖掘逻辑规则。

4) HoIE 模型^[17]。该模型结合了语义匹配模型的表达能力和 DisMult 的效率, 与同期模型相比具有较好的嵌入性能。

5) TKRL 模型^[22]。该模型在嵌入过程中融入了实体的层次类型信息, 得到了较好的嵌入效果。

6) JOIE-TransE-CT 模型^[9]。该模型是本体-实例联合学习模型, 在部分任务上能够达到较整个知识图谱嵌入更好的效果。

图 5 和图 6 显示了各模型在两个数据集上训练过程中损失变化(epoch: 30~100)。

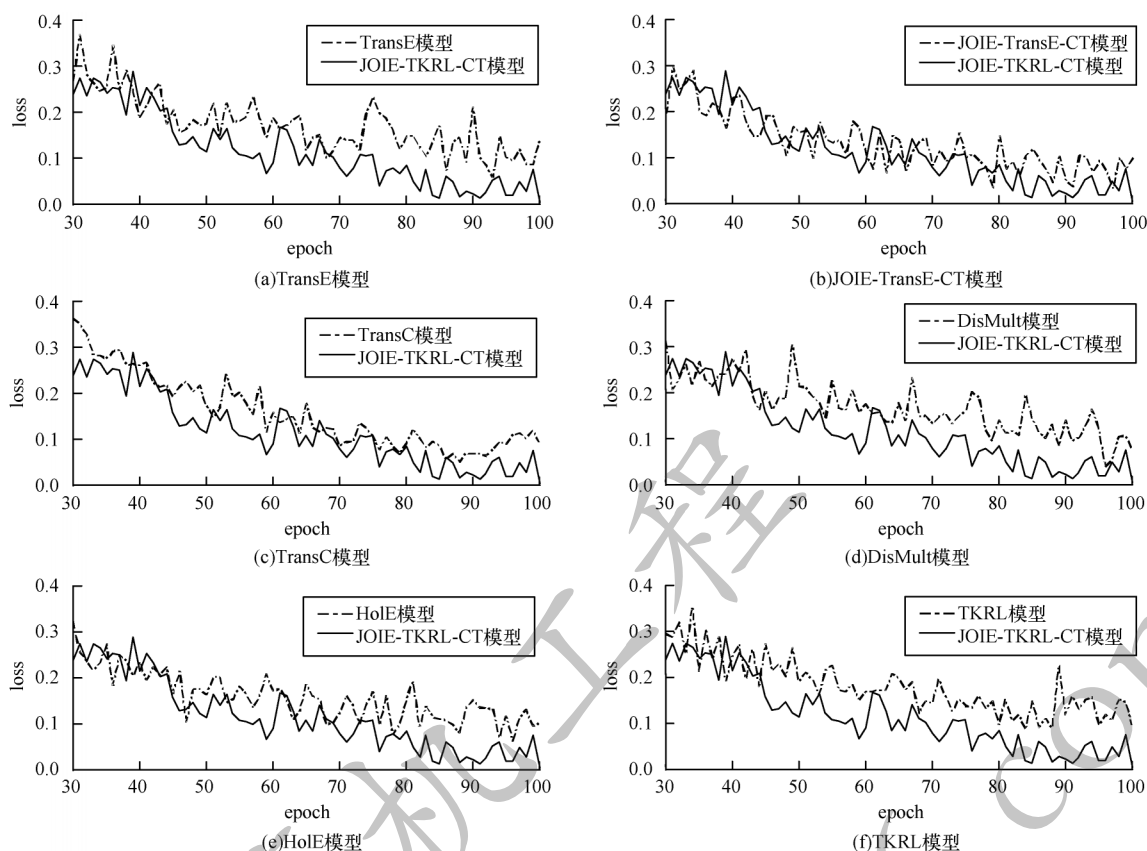


图5 YAGO26K-906数据集上各模型训练损失变化

Fig.5 Changes of the loss in each model training on YAGO26K-906 data set

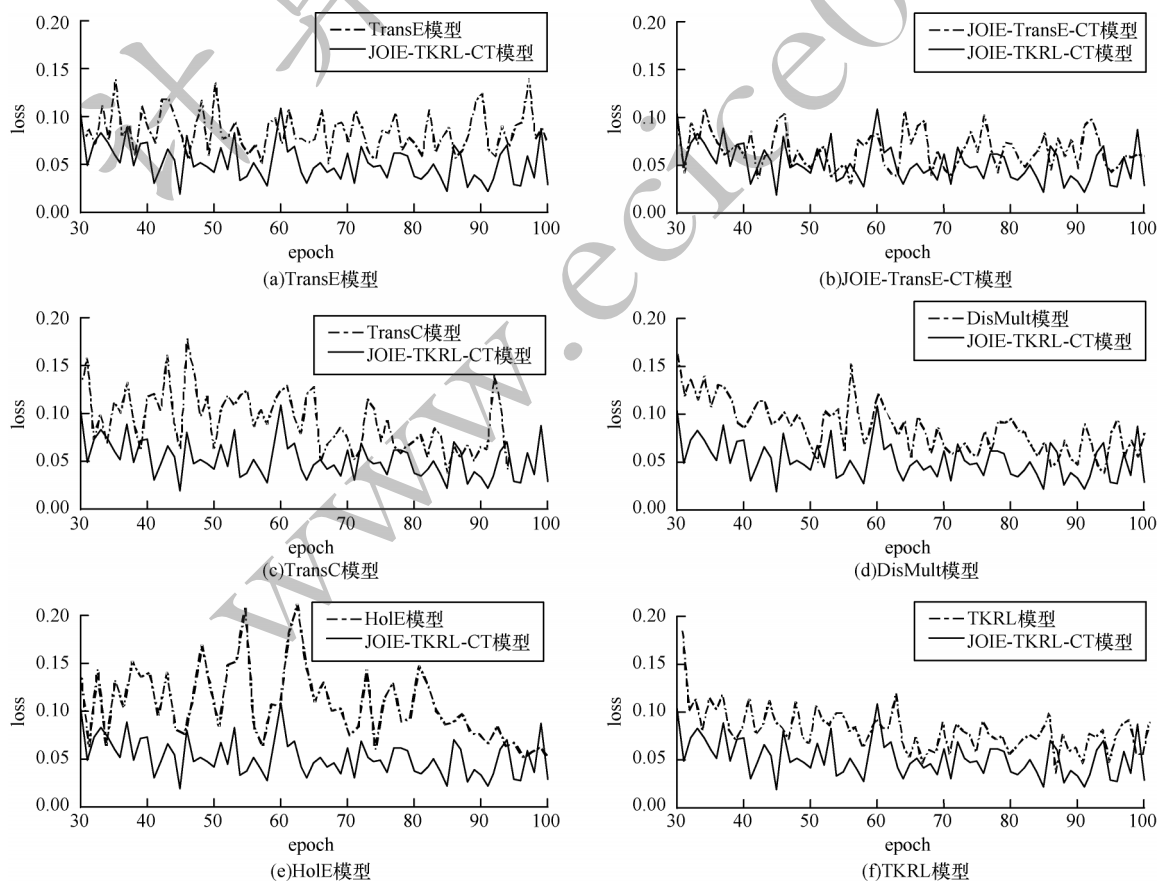


图6 DB11K-174数据集上各模型训练损失变化

Fig.6 Changes of the loss in each model training on DB11K-174 data set

损失(loss)在模型的训练中是一个关键角色。模型的训练就是最小化 loss 过程,减少预测的误差,寻找最优解的过程。一定程度上来说,模型的 loss 越小,模型的预测能力越好。由图 5 和图 6 整体来看,在训练达到一定的程度后(epoch>60),两种联合学习模型 loss 都会比其他模型降得更小。而进一步通过两个联合学习模型对比可知,JOIE-TKRL-CT 的 loss 更小。因此,JOIE-TKRL-CT 模型性能优于其他模型。

各模型具体实验结果如表 2 和表 3 所示,其中,加粗表示最优数据。值得注意的是,TransC 模型编码时区分了概念和实例,但是没有将概念的元关系编码,因此涉及对本体三元组的补全。

表 2 在 YAGO26K-906 数据集上的三元组补全实验结果

Table 2 Experimental results of triple completion on YAGO26K-906 data set

模型	本体三元组补全			实例三元组补全		
	MRR	Hit@1 /%	Hit@10 /%	MRR	Hit@1 /%	Hit@10 /%
TransE	0.139	12.18	20.36	0.187	13.45	32.98
TransC	—	—	—	0.248	15.59	37.54
DisMult	0.195	17.69	25.04	0.251	22.89	28.77
HolE	0.192	18.62	20.26	0.264	25.56	28.21
TKRL	0.168	15.62	21.54	0.213	16.21	34.28
JOIE-TransE-CT	0.231	14.13	33.21	0.286	18.63	44.02
JOIE-TKRL-CT	0.230	14.12	33.48	0.305	19.48	49.25

表 3 在 DB111K-174 数据集上的三元组补全实验结果

Table 3 Experimental results of triple completion on DB111K-174 data set

模型	本体三元组补全			实例三元组补全		
	MRR	Hit@1 /%	Hit@10 /%	MRR	Hit@1 /%	Hit@10 /%
TransE	0.301	23.14	45.93	0.309	22.01	48.62
TransC	—	—	—	0.348	24.67	49.17
DisMult	0.232	15.14	29.03	0.263	25.91	27.59
HolE	0.224	18.78	32.69	0.296	29.03	31.48
TKRL	0.325	23.89	46.32	0.334	24.87	49.05
JOIE-TransE-CT	0.603	58.08	72.94	0.441	32.05	67.84
JOIE-TKRL-CT	0.604	57.06	70.26	0.562	36.87	70.96

表 2 和表 3 的实验结果表明:

1) 在两个数据集上,联合学习模型都取得了较其他模型更好的性能,其中 JOIE-TKRL-CT 在实例三元组补全上具有优势,原因是融合的实体类型信息更多是针对实例,在本体三元组补全上反而弱于 JOIE-TransE-CT。

2) 对比 JOIE-TKRL-CT 和 JOIE-TransE-CT、TKRL 和 TransE 两组模型,可以发现在两种数据集上,融合实

体层次类型信息可以有效提高模型性能,在所有的指标中都有较大的提升。

3.3 实体分类

实体类型任务目标简单来说就是预测给定实体的对应本体概念。测试方法同三元组补全任务。

实验设置了 TransE、DisMult、HolE、JOIE-TransE-CT 这 4 个模型对照实验,实验结果如表 4 所示,其中,加粗表示最优数据。

表 4 实体分类实验结果

Table 4 Experimental results of entity classification

模型	YAGO26K-906			DB111K-174		
	MRR	Hit@1 /%	Hit@10 /%	MRR	Hit@1 /%	Hit@10 /%
TransE	0.139	7.21	35.03	0.501	43.62	60.75
DisMult	0.409	36.01	55.24	0.549	49.78	67.94
HolE	0.387	34.75	54.68	0.501	44.72	65.24
JOIE-TransE-CT	0.839	74.68	91.98	0.845	74.23	93.89
JOIE-TKRL-CT	0.841	75.01	92.34	0.862	76.36	93.98

就实体分类任务实验结果而言,在两种数据集上,联合学习模型的实验性能都远超过单纯基于距离和语义相似度的模型,其中 JOIE-TKRL-CT 较前者 MRR、准确率平均高了 1%,这也证明了融合实体了层次类型信息可以提高联合学习的性能。

综合以上实验结果可知,JOIE-TKRL-CT 模型在实例三元组补全和实体分类上都取得了比 JOIE-TransE-CT 模型更好的成绩。因此,融合了实体层次类型信息的本体-实例联合学习方法可以达到区分实体多义性的目的,具有更好的知识学习效果。

4 结束语

为区别表示实体在特定场景的独特意义,本文在嵌入时融合三元组中实体的关系类型特征,提出一种本体-实例联合学习方法 JOIE-TKRL-CT。在 YAGO26K-906 和 DB111K-174 数据集上针对三元组补全和实体分类两个任务进行多组实验,结果表明,JOIE-TKRL-CT 在实例三元组补全和实体分类任务上较 TransE、TransC、DisMult 等模型具有更好的性能,验证了本文方法的有效性。下一步将在模型中融入知识图谱的结构信息和深层的类型信息,获得更优的知识学习效果。

参考文献

[1] ADAMS T. Google and the future of search; amit singhal and the knowledge graph[EB/OL]. [2021-04-10]. <https://www.mendeley.com/catalogue/a2ceb8b0-2e58-31cb-ae6a-09f3e89a8b88/>.

- [2] HE H, BALAKRISHNAN A, ERIC M, et al. Learning symmetric collaborative dialogue agents with dynamic knowledge graph embeddings [EB/OL]. [2021-04-10]. <https://arxiv.org/abs/1704.07130>.
- [3] BORDES A, WESTON J, USUNIER N. Open question answering with weakly supervised embedding models [M]// CALDERS T, ESPOSITO F, HÜLLERMEIER E, et al. Machine learning and knowledge discovery in databases. Berlin, Germany: Springer, 2014: 165-180.
- [4] 陶天一, 王清钦, 付聿炜, 等. 基于知识图谱的金融新闻个性化推荐算法 [J]. 计算机工程, 2021, 47(6): 98-103, 114.
- TAO T Y, WANG Q Q, FU Y W, et al. Personalized recommendation algorithm for financial news based on knowledge graph [J]. Computer Engineering, 2021, 47(6): 98-103, 114. (in Chinese)
- [5] BOLLACKER K, EVANS C, PARITOSH P, et al. Freebase: a collaboratively created graph database for structuring human knowledge [C]//Proceedings of 2008 ACM SIGMOD International Conference on Management of Data. New York, USA: ACM Press, 2008: 1247-1250.
- [6] LUONG T, SOCHER R, MANNING C D. Better word representations with recursive neural networks for morphology [2021-04-10]. https://www.researchgate.net/publication/270878536_Better_Word_Representations_with_Recursive_Neural_Networks_for_Morphology.
- [7] REISINGER J, MOONEY R. Multi-prototype vector-space models of word meaning [C]//Proceedings of 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics. [S. l.]: Association for Computational Linguistics, 2010: 109-117.
- [8] SPEER R, CHIN J, HAVASI C. ConceptNet 5.5: an open multilingual graph of general knowledge [EB/OL]. [2021-04-10]. <https://arxiv.org/pdf/1612.03975.pdf>.
- [9] HAO J H, CHEN M H, YU W C, et al. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts [EB/OL]. [2021-04-10]. <https://arxiv.org/abs/2103.08115>.
- [10] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data [M]// GLEN D T, SUZANNA B, ZOUBIN G. Advances in neural information processing systems. Berlin, Germany: Springer, 2013: 2787-2795.
- [11] WANG Z, ZHANG J W, FENG J L, et al. Knowledge graph embedding by translating on hyperplanes [C]//Proceedings of the 28th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI Press, 2014: 1112-1119.
- [12] LIN Y K, LIU Z Y, SUN M S, et al. Learning entity and relation embeddings for knowledge graph completion [C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI Press, 2015: 2181-2187.
- [13] 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 (Volume 1: Long Papers). [S. l.]: Association for Computational Linguistics, 2015: 687-696.
- [14] JI G, LIU K, HE S, et al. Knowledge graph completion with adaptive sparse transfer matrix [C]//Proceedings of the 30th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI Press, 2016: 1-10.
- [15] HE S Z, LIU K, JI G L, et al. Learning to represent knowledge graphs with Gaussian embedding [C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York, USA: ACM Press, 2015: 623-632.
- [16] YANG B S, YIH W T, HE X D, et al. Embedding entities and relations for learning and inference in knowledge bases [EB/OL]. [2021-04-10]. <https://arxiv.org/abs/1412.6575>.
- [17] TROUILLON T, NICKEL M. Complex and holographic embeddings of knowledge graphs: a comparison [EB/OL]. [2021-04-10]. <https://arxiv.org/abs/1707.01475>.
- [18] CHEN M H, TIAN Y T, CHEN X L, et al. On2Vec: embedding-based relation prediction for ontology population [EB/OL]. [2021-04-10]. <https://arxiv.org/abs/1809.02382>.
- [19] LÜ X, HOU L, LI J Z, et al. Differentiating concepts and instances for knowledge graph embedding [EB/OL]. [2021-04-10]. <https://arxiv.org/abs/1811.04588>.
- [20] KROMPA D, BAIER S, TRESP V. Type-constrained representation learning in knowledge graphs [C]//Proceedings of SWC'15. Berlin, Germany: Springer, 2015: 640-655.
- [21] SHU G, QUAN W, WANG B, et al. Semantically Smooth knowledge graph embedding [C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). [S. l.]: Association for Computational Linguistics, 2015: 1-10.
- [22] FORSYTH D. Representation learning [J]. Computer, 2015, 48(4): 6.
- [23] 张晓明, 孙维雅, 王会勇. 基于知识表示学习的知识可信度评估 [J]. 计算机工程, 2021, 47(7): 44-54.
- ZHANG X M, SUN W Y, WANG H Y. Evaluation of knowledge credibility based on knowledge representation learning [J]. Computer Engineering, 2021, 47(7): 44-54. (in Chinese)
- [24] BORDES A, WESTON J, COLLOBERT R, et al. Learning structured embeddings of knowledge bases [C]//Proceedings of the 25th AAAI Conference on Artificial Intelligence. [S. l.]: AAAI Press, 2011: 301-306.