计算机工程
Computer Engineering
ISSN 1000-3428,CN 31-1289/TP

《计算机工程》网络首发论文

题目：融合实体类别信息的知识图谱表示学习方法
作者：金婧，万怀宇，林友芳
DOI：10.19678/j.issn.1000-3428.0057353
网络首发日期：2020-04-02
引用格式：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方法．计算机工程. https://doi.org/10.19678/j.issn.1000-3428.0057353

网络首发时间：2020-04-02 11:24:13
网络首发地址：http://kns.cnki.net/kcms/detail/31.1289.TP.20200402.0849.001.html

融合实体类别信息的知识图谱表示学习方法

金婧，万怀宇，林友芳

(北京交通大学计算机与信息技术学院交通数据分析与挖掘北京市重点实验室，北京 100044)

摘 要：知识表示学习旨在将实体和关系嵌入到一个连续低维的语义空间中，以便于高效计算实体和关系的语义联系。本文提出了一种融合实体类别信息的知识表示学习模型（TEKRL），在引入多源信息的同时解决了其他模型在使用实体类别信息时需要引入额外规则的问题。该模型构建了基于结构和基于类别的两种实体表示，并通过引入注意力机制来捕获实体类别和三元组关系之间的潜在关联，自动地学习实体的不同类别对某种特定关系的不同重要程度，从而简化了人工制定规则这一繁琐的过程，更高效地利用实体类别信息进行知识表示学习。最后，通过知识图谱补全和三元组分类这两个任务对模型进行了评估，实验结果表明，TEKRL 模型在各项指标上取得了显著的提升，尤其是在在实体预测任务中，与其他方法相比 Hit@10 指标提升了约 7.2%，MeanRank 指标提升了约 23%，表明了该模型可以有效地利用类别信息来更好地进行知识表示。

关键词：知识图谱；知识表示学习；多源信息融合；注意力机制；实体消歧

开放科学（资源服务）标识码（OSID）：

Knowledge Graph Representation Learning via Incorporating Entity Type Information

Jin Jing, Wan Huaiyu, Lin Youfang

(Beijing Key Laboratory of Traffic Data Analysis and Mining, School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China)

【Abstract】Knowledge representation learning (KRL) aims to embed both entities and relations into a continuous low-dimensional semantic space, so as to efficiently calculate the semantic relationship between entities and relations. This paper proposes TEKRL model, a knowledge representation learning model, based on fusion of type information, which introduces multi-source information and solves the problem that other models often need to introduce additional rules when using type information. This model constructs two entity representations based on the structure and type information, and captures the potential association between entity types and triple relations by introducing attention mechanism. It could automatically learn the different importance of different types of entities for specific relations, thus simplifying the complicated process of manual rule-making and making more efficient use of entity type information for knowledge representation learning. We evaluate our model on two tasks including graph completion and triple classification. The experimental results show that the TEKRL model has made significant improvements in various indicators, especially in the entity prediction task, compared with other methods, the Hit@10 indicator has increased by about 7.2%, and the MeanRank indicator has increased by about 23%, indicating that our model can effectively use type information for better knowledge representation.

【Key words】knowledge graph; knowledge representation learning; multi-source information fusion; attention mechanism; entity disambiguation

DOI: 10.19678/j.issn.1000-3428.0057353

1 引言

知识图谱是推动人工智能发展和支撑智能信息服务应用的重要基础技术，其将人类知识构建成结构化的知识系统。在知识图谱中，知识通常以三元组 (h,r,t) 的形式来表示，其中 h 、t 分别代表头尾两个实体，r 代表头尾实体之间的关系。知识图谱以一种网络图的形式来构建整个知识系统，其中知识表示为知识图谱中知识获取和应用的基础，可以提升知识图谱在认知和推理方面的能力[1-4]。但是，随着越来越多的大型知识图谱被构建出来，如 Freebase、DBpedia 等，基于网络形式的知识表示在大规模的知识图谱下存在着计算效率低下和数据稀

基金项目：国家重点研发计划（2018YFC0830200）

作者简介：金婧（1994-），女，吉林吉林人，硕士研究生，主要研究方向为数据挖掘与知识表示学习；万怀宇（1981-），男（通讯作者），湖北宣恩人，副教授，主要研究领域为数据挖掘与信息抽取；林友芳（1971-），男，福建武平人，教授，主要研究领域为数据与知识工程。E-mail：hywan@bjtu.edu.cn

疏等问题。近年来，随着深度学习研究的不断深入，以深度学习为代表的表示学习技术获得了广泛的关注，其旨在将研究对象映射到一个连续低维的向量空间中，以便于高效计算实体和关系的语义联系，同时能有效地解决数据稀疏问题。

翻译模型是知识表示学习领域的一种主流方法，因为其简单性和有效性，自提出以来就成为了最受推崇的方法之一，并且随后还出现了许多在翻译模型的基础上进行改进的变体。近年来提出的很多方法不仅利用了知识图谱所固有的结构信息，还考虑了其他与实体相关的多源信息，例如实体描述信息、类别信息和图像信息等，大大提高了知识表示学习的性能。TKRL[5]就是一种利用了实体类别信息作为外部信息的知识表示学习方法，该模型认为不同类别的实体应该具有不同的表示，并且还关注到了实体类别的层次结构，利用两种类型的编码对层级结构进行建模，最终证实了实体类别可以在知识表示学习中发挥重要作用。但是，TKRL 模型存在着一定的使用限制，第一，需要依赖于具有层次结构的类别信息；第二，需要具有事先制定好的规则约束，这种规则约束具体是指：当给定一种关系时，约定了该种关系的头实体和尾实体的具体类别。这种约束对于现实世界的数据而言是不灵活的。而且，不仅 TKRL 这一模型利用了制定好的规则约束，还有很多其他融合实体类别信息的模型也都基于类似的规则。

为了有效地利用实体类别与三元组关系之间的潜在关联性，本文提出了一种新颖有效的融合实体类别信息的知识表示学习模型，即类别增强知识表示学习（Type-Enhanced Knowledge Representation Learning, TEKRL）。该模型引入实体基于类别的表示，采用注意力机制并通过定义不同的注意力计算方法，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。

此外，TEKRL 模型还为实体消歧任务提供了指导和借鉴意义。模型通过注意力机制可以学习到一个实体的各个类别在特定关系下的权重，基于这种特性，本文给出了一种解决实体消歧问题的新思路，并通过实验验证了提出的模型在处理这类问题上的有效性，为语义分析等自然语言处理中的关键性问题提供了一种可行的参考方法。

2 相关工作

知识表示是对知识进行描述的一种途径，旨在研究如何更准确地表示知识的语义信息以更好地利用知识图谱，从而使得计算机能够接受并运用知识，最终达到智能的目标。知识表示学习则是通过机器学习的方式将知识（知识图谱中的实体和关系）表示为稠密低维的实值向量，有效地解决了数据稀疏的问题；并且学习到的知识表示能够保留知识图谱中的结构和语义关系，从而可以高效地计算实体和关系之间的语义联系，便于广泛应用到知识图谱补全、自动问答和实体链接等下游任务中。

近年来，随着深度学习的发展，知识表示学习的方法也随之不断改进，取得了很大进展。以 TransE[2]为代表的翻译模型是知识表示学习中的热门方法，这类模型将关系向量作为头实体向量到尾实体向量之间的平移，即假设尾实体向量 $\boldsymbol{t}$ 近似于头实体向量和关系向量的和 $\boldsymbol{h}+\boldsymbol{r}$，并定义能量函数为：

$$E(h,r,t)=\left\|\boldsymbol{h}+\boldsymbol{r}-\boldsymbol{t}\right\| \tag{1}$$

TransE 模型因其参数少，计算复杂度低，具有简单高效的特点，在 1-1 这种简单的关系中表现较好。但对于 1-N、N-1 和 N-N 等更复杂的关系，由于其建模方式过于简单，存在着一定的局限性。为了解决这一问题，后续出现了许多以 TransE 为基础改进的模型，如 TransH、TransAH、TransA、TransG、TransR 和 TransD 等。TransH 通过将头、尾实体向量投影到对应关系的超平面上，从而令一个实体在不同的关系下拥有不同的表示[6]。TransAH 在 TransH 的基础上引入了一种自适应的度量方法，通过加入对角权重矩阵将得分函数中的度量由欧氏距离转换为加权欧氏距离[7]。TransA 同样提出了一种自适应的度量方法，为每个关系定义了一个非负的对称矩阵，从而为表示向量中的每一个维度添加了权重，有效地增加了模型的表示能力[8]。TransG 模型使用高斯混合来刻画实体间的多种语义关系，利用最大相似度原理训练数据，该模型能够有效地解决多语义问题[9]。TransR 假设不同的关系拥有不同的语义空间，将每个实体投影到对应的关系空间中[10]。TransD 通过设置两个关系-实体投影矩阵，考虑了头尾实体位置的属性，同时也解决了 TransR 参数过多的问题[11]。

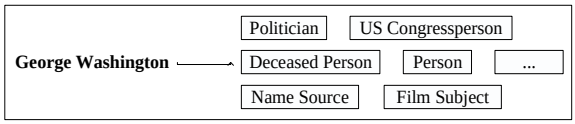
在知识表示学习的方法中，还存在着一些其他类型的方法，包括：1）距离模型，如 SE 模型[12]，

将头尾实体向量通过投影矩阵投影到对应空间中，并通过计算投影向量的距离来反应实体间的语义相似度；2）能量模型，如 SME 模型[13,14]，通过定义若干个投影矩阵，利用双线性函数刻画实体与关系的内在联系；3）矩阵分解模型，通过矩阵分解的方式得到低维向量表示，其中 RESCAL 模型[15,16]是矩阵分解模型的代表；4）双线性模型如 LFM 模型[17]提出利用基于关系的双线性变换，刻画实体和关系之间的二阶联系。以上模型仅利用了知识图谱自身所包括的三元组结构信息，但除了结构信息以外，还有大量与知识有关的其他信息没有得到有效利用，例如知识库中所包含的实体和关系的描述信息、类别信息以及知识库以外的海量互联网文本信息等。这些多源信息提供了知识图谱中三元组事实以外的额外信息，有助于更准确地学习知识表示。其中，NTN 模型[18]通过使用实体中的单词嵌入的平均值来表示实体，从而捕捉实体之间潜在的文本关系。DKRL 模型[19]考虑了实体的描述信息文本，利用两种模型来编码实体描述信息的语义。IKRL 模型[20]引入实体的图像信息，并利用神经网络构造实体图像的表示。TKRL 模型[5]通过引入具有层次结构的类别信息以及实体类别与关系之间的约束信息来提高知识表示的能力。但并非所有实体类别都具有层次结构，并且对实体类别和关系之间做约束的方式不具有普适性和灵活性。为了解决上述问题，本文提出了一种融合实体类别的知识表示学习模型 TEKRL，该模型直接利用数据集中最底层的这一层实体类别，通过引入注意力机制学习到实体的类别和关系之间的相关性大小，并利用得到的注意力分数对类别表示进行加权，以这种方式取代制定实体类别与关系之间的约束的操作，同时达到更准确地表示知识的目的。

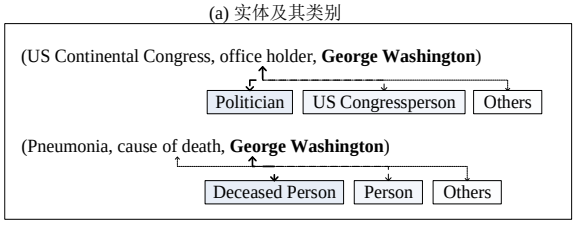
3 TEKRL 模型

在知识图谱中，常包含实体的类别信息，而类别信息作为实体属性的一部分，能够起到补充实体语义信息的作用。为了能够有效融合知识图谱中的实体类别信息，同时还能够兼备翻译模型的高效性能，本文提出了 TEKRL 模型，该模型基于 TransE 且在其基础上引入了实体的类别表示，旨在学习到三元组知识的同时，能够通过类别信息得到更加准确的知识表示。并且，该模型不需要依赖于实体类别与关系之间的固定映射，以一种高效的方式解决了大多数方法都需要固定这种映射关系的问题，便于将模型灵活地迁移到其他更加复杂、难以得到这种映射关系的场景中。同时，本文提出的模型对实体类别的组织形式没有要求，通过将类别的组织结构进行扁平化处理，可以适配各种应用场景对类别信息格式的要求，无论是像 FB15K 中具有层次结构的类别信息还是其他形式的类别数据，我们的模型均可以使用。

为了更清楚地表述模型的基本思想，图 1 列举了一个具体的例子说明了实体类别与三元组关系之间存在一定的语义关联，而 TEKRL 模型正是基于这种关联性提出的。图 1 中的子图（a）列举了实体乔治·华盛顿及其所包含的类别属性（图中仅列出了其中的一部分），包括：政治家、美国国会议员、死者、人、名称来源、电影主题。图 1 中的子图（b）以知识图谱中与乔治·华盛顿这一实体相关的两个具体的三元组为例，华盛顿的“政治家”和“美国国会议员”这两个类别在（美国大陆会议，官员，华盛顿）这个三元组中，比其他类别更具有相关性；而在（肺炎，死因，华盛顿）这个三元组中，“死者”则能表达出更多相关的信息。这说明了同一个实体的不同类别在不同的三元组中，更具体地说，在不同的关系中，是可以起到提供语义信息的作用的，并且不同的类别所起到的作用大小也与三元组的关系存在着一定的关联。进一步来说，在这种关联的指导下，实体的类别信息可以用来丰富实体的表示，使知识表示具有更多的语义信息。



(a) 实体及其类别



(b) 实体的类别与不同关系之间的潜在关联

图 1 实体类别与三元组关系之间存在一定的语义关联示例

Fig.1 Example of semantic relationships between the entity types and the relation of triple.

基于以上想法，本文提出了 TEKRL 模型，下面将对模型进行详细介绍。首先，为了更清晰地描述 TEKRL 模型，这里给出相关的定义和符号表示。将知识图谱定义为 $G=(E,R,S)$，其中 E 为实体集，

R 为关系集；$S \subseteq E \times R \times E$ 表示三元组的集合，每一个三元组用 (h,r,t) 来表示，h、r 和 t 分别代表头实体、关系和尾实体。除此之外，本文引入了类别这一概念，并用 C 来表示类别集合。另外，本文定义了两种类型的实体表示，即基于结构的表示和基于类别的表示，分别代表从知识图谱的三元组中学习到的实体表示，以及引入类别表示所得到的实体表示。

TEKRL 模型的整体框架如图 2 所示，斜线状的圆圈组成的椭圆形代表基于结构的向量表示，网格状的圆圈组成的椭圆形代表基于类别的向量表示，实心圆圈组成的椭圆形代表关系的向量表示，空心的圆圈组成的椭圆形代表实体类别的向量表示，a 表示的是注意力分数。为了将两种类型的表示进行融合，本文定义了如下的能量函数：

$$E = E_{SS} + \beta E_{CC} \tag{2}$$

其中，$E_{SS} = \left\| \boldsymbol{h}_s + \boldsymbol{r} - \boldsymbol{t}_s \right\|$，$E_{CC} = \left\| \boldsymbol{h}_c + \boldsymbol{r} - \boldsymbol{t}_c \right\|$，$\boldsymbol{h}_s$ 和 $\boldsymbol{t}_s$ 分别是基于结构的头、尾实体表示，$\boldsymbol{h}_c$ 和 $\boldsymbol{t}_c$ 分别是基于类别的头、尾实体表示，超参 β 用于调整基于类别的表示在模型中起到的作用大小。值得说明的是，实体基于结构的表示和基于类别的表示在训练过程中都使用统一的关系表示 $\boldsymbol{r}$，这样做保证了两种类型的表示空间可以通过相同的关系表示达到统一。

在训练过程中，首先通过注意力机制得到实体类

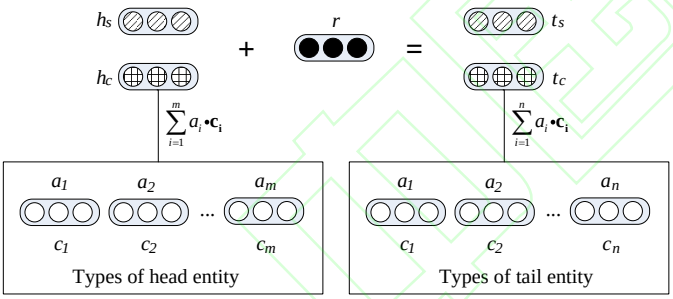


图 2 TEKRL 模型的整体架构

Fig. 2 Overall architecture of the TEKRL model.

别表示与三元组中关系的相关性，即相关的注意力分数，并利用该分数对类别表示加权求和作为实体的基于类别的表示。最后，通过相同的关系表示作为两种表示空间的联系，将基于结构的表示和基于类别的表示进行联合训练。

3.1 注意力机制

实体的不同类别信息可以从多个角度刻画实体，而同一个实体在不同的关系下会侧重关注其不同的含义，也可以表现为同一实体的各个类别与不同的关系之间的语义关联大小不同。为了能够利用三元组中的关系和实体的类别之间存在的潜在关联，本文通过注意力机制计算得到二者之间的相关性大小。我们采用两种注意力的计算方式：

1）基于相似度的注意力机制(Similarity-based Attention, SA)：这种方法受 STKRL 模型[21]中的注意力机制的启发，将实体的类别与三元组中关系之间的相关性定义为二者向量表示的相似性，并采用余弦相似度来计算这种相似性，如下：

$$att(\boldsymbol{c},\boldsymbol{r}) = \frac{\boldsymbol{c} \cdot \boldsymbol{r}}{\left\|\boldsymbol{c}\right\| \cdot \left\|\boldsymbol{r}\right\|} \tag{3}$$

其中 $att()$ 是求注意力分数 a 的函数，$\boldsymbol{c}$ 是类别的向量表示，$\boldsymbol{r}$ 是与类别 $\boldsymbol{c}$ 所对应的实体出现在同一个三元组中的关系。

2）缩放点积注意力机制(Scaled Dot-Product Attention, SDPA)：基于文献[22]中提出的注意力计算方法，结合本文提出的模型，将关系 $\boldsymbol{r}$ 作为 query 向量，类别 $\boldsymbol{c}$ 同时作为 key 向量和 value 向量。在实际实现过程中，为了加快处理效率，注意力是通过矩阵的形式来计算的，因此将多个关系的表示向量及其对应的类别表示向量分别拼接为关系矩阵 $\boldsymbol{R}$ 和类别矩阵 $\boldsymbol{C}$。然后，引入待训练的权重矩阵 $\boldsymbol{W}^{Q}$、$\boldsymbol{W}^{K}$ 和 $\boldsymbol{W}^{V}$，通过权重矩阵与关系矩阵和类别矩阵分别做矩阵相乘操作，然后得到 query、key 和 value 对应的矩阵 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$，计算过程如下：

$$\boldsymbol{Q} = \boldsymbol{R} \times \boldsymbol{W}^{Q} \tag{4}$$

$$\boldsymbol{K} = \boldsymbol{C} \times \boldsymbol{W}^{K} \tag{5}$$

$$\boldsymbol{V} = \boldsymbol{C} \times \boldsymbol{W}^{V} \tag{6}$$

其中，$\boldsymbol{C} \in \mathbb{R}^{|C| \times k}$，$\boldsymbol{R} \in \mathbb{R}^{|R| \times k}$，$\boldsymbol{W}^{Q} \in \mathbb{R}^{k \times d_k}$，$\boldsymbol{W}^{K} \in \mathbb{R}^{k \times d_k}$，$\boldsymbol{W}^{V} \in \mathbb{R}^{k \times k}$。

最后，注意力分数可以利用如下公式计算：

$$att(\boldsymbol{C},\boldsymbol{R}) = Softmax(\frac{\boldsymbol{Q}\boldsymbol{K}^{T}}{\sqrt{d_k}}) \tag{7}$$

通过以上两种方式得到的注意力分数越高，说明类别 $\boldsymbol{c}$ 与关系 $\boldsymbol{r}$ 的相关性越强，因此，本文考虑利用计算出的注意力分数对相关的各个类别表示赋予权重，再对加权后的所有表示求和，以得到对应的实体表示，即实体基于类别的表示。在矩阵形式下的计算公式如下：

$$\boldsymbol{E}_c = Softmax(\frac{\boldsymbol{Q}\boldsymbol{K}^{T}}{\sqrt{d_k}})\boldsymbol{V} \tag{8}$$

3.2 模型训练

与 TransE 的训练目标相同，为了增强知识表示的区分能力，本文采用最大间隔方法，定义如下目

标函数：

$$L=\sum_{(h,r,t)\in T}\sum_{(h',r',t')\in T'}\max(\gamma+E(h,r,t)-E(h',r',t'),0) \tag{9}$$

其中，$E(h,r,t)$ 是正例三元组的能量函数，$E(h',r',t')$ 是负例三元组的能量函数，γ 是表示间隔的超参数且 $\gamma>0$，T 代表训练集，T' 代表利用 T 进行负采样得到的集合，定义为：

$$T'=\{(h',r,t)\,|\,h'\in E\}\bigcup\{(h,r,t')\,|\,t'\in E\}\bigcup\{(h,r',t)\,|\,r'\in R\} \tag{10}$$

其中，头实体、尾实体或者关系被随机替换为其他的实体或者关系，另外，如果替换后的新三元组仍在 T 中，则不会被用作负样本。

在模型训练过程中，实体、关系和类别的表示均可以随机初始化，实体和关系的表示也可以采用简单的翻译模型预训练得到。在具体的模型实现过程中，为了使初始的类别表示具有一定的语义信息，我们借助预训练得到的实体表示对其进行初始化，使得相较于随机初始化能够缩短模型达到收敛的时间。初始化类别的具体方法为，利用所有包括类别 c_i 的实体表示的平均值作为这一类别的表示，形式化为：

$$\boldsymbol{c}_i=\frac{1}{\left|e_{c_i}\right|}\sum_{j=1}^{\left|e_{c_i}\right|}\boldsymbol{e}_j \tag{11}$$

其中，$\left|e_{c_i}\right|$ 代表具有 c_i 类别的实体的数量且 i 满足 $i\in\left[1,\left|C\right|\right]$。

最后，模型通过最小化目标函数同时学习基于结构和基于类别的两种表示。具体训练模型时，采用了 Adam[23]优化算法，这是一种计算每个参数的自适应学习率的方法。

4 实验及结果分析

本节将详细介绍模型验证实验，包括实验数据集、实验设置以及在不同任务中得到的实验结果和结果分析。

4.1 数据集

实验使用 FB15K 数据集，通过知识图谱补全和三元组分类任务对模型进行评估。FB15K 是从 Freebase 中抽取出的数据集，在实验中将其划分为训练集、验证集和测试集，具体的统计信息如表 1 所示，第一行中的#Rel 表示关系，#Ent 表示实体，#Train 表示训练集，#Valid 表示验证集，#Test 表示测试集合，第二行的数字为所对应内容的数目。所有的事实三元组，即实验中的训练集、验证集和测试集的并集，在下文中被称作黄金三元组。

表 1 FB15K 数据集统计信息

Table 1 Statistics of FB15K dataset.

Database	#Rel	#Ent	#Train	#Valid	#Test
FB15K	1,345	14,951	483,142	50,000	59,071

关于类别相关的数据，采用文献[5]公开的数据集，该数据集包含了 Freebase 知识库所涉及到的 type/instance 字段，即包含了文中提到的类别信息。这一数据集是通过匹配 Freebase 中 FB15K 所包含的实体，并为这些实体添加知识库中实体对应的类别信息得到的。在数据处理过程中，发现有 10 个实体出现在原始的 FB15K 数据集中，但没有与之对应的实体类别信息。在处理这一数据缺失问题上，为了保证原始的 FB15K 数据的完整性，就需要保留这 10 个实体及其所涉及到的所有三元组，使这 10 个实体也具有类别信息。经过数据统计发现 99%的实体都拥有 common/topic 这一类别，因此在实验中采用众数规则对上述 10 个缺失类别的实体人为地添加了 common/topic 这一类别。处理后的数据集具有 3852 个类别，一个实体平均约有 12 个类别。

4.2 实验设置

为了验证 TEKRL 模型的效果，我们将其与多个已有模型进行对比，包括 TransE、TransR 和 TKRL 等。为了提升 TransE 模型在关系预测方面的性能，在训练阶段对其增加了关系负采样的操作。对于 TransR，本文采用文献[10]的开源代码进行实验，并且与 TransE 在负采样过程的做法相同，在生成负样本时也对关系进行了替换。对于其他方法的实验结果，包括 RESCAL、SE、SME 以及 LFM，本文直接引用了文献[10]中所发表的结果。对于 TKRL 模型，本文引用了发表该模型的文献[5]中的实验结果。

关于模型的参数选择问题，实验设置初始学习率 α 为 $\{0.0005,0.001,0.002\}$，批量大小 B 为 $\{20,240,1200,4800\}$，实体和关系的向量维度 k 为 $\{50,100,200\}$，阈值 γ 为 $\{0.5,1.0,1.5,2.0\}$。对于缩放点积注意力机制，设置权重矩阵中的 d_k 为 $\{49,64,100\}$。并且考虑到基于结构的表示和基于类别的表示起到的作用大小是不同的，因此本文为基于类别的表示设置了权重，用超参数 β 来表示，用以调整其在模型中的重要程度。在实验中得到模

型的最优的参数设置为：$\alpha=0.001$，$B=4800$，$k=200$，$\gamma=1.0$，$d_k=100$，$\beta=0.5$。

4.3 知识图谱补全实验结果

知识图谱补全任务是在给定一个完整的事实三元组(h,r,t)中的两项，预测缺失的剩余的一项，即给定(h,r)预测t，给定(r,t)预测h，或给定(h,t)预测r，因此知识图谱补全包括实体预测和关系预测这两个子任务。

本文采用两种评估指标，分别是 MeanRank 和 Hit@n，分别代表正确的实体和关系在预测结果的平均排名以及正确的实体和关系排在预测结果前 n 名的比例。针对每种指标还有两种不同的设置，分别是 Raw 和 Filter。在 Raw 设置下，只要预测结果不是当前三元组所期待的结果，就视作为错误的预测结果，即使该预测结果属于黄金三元组。Filter 代表将预测结果中的属于黄金三元组的预测剔除后所得到的结果。

对于这两种设置，Raw 指标会忽略黄金三元组的存在，具体来说，如果预测出的结果属于黄金三元组，但非当前所关注的特定三元组，则认为预测结果错误，从而会导致预测指标变差，但这部分由于黄金三元组而造成预测“错误”的结果，实际上预测的结果是正确的，不应视为模型性能的问题，因此，本文认为 Filter 设置下的结果更具有说服力。

4.3.1 实体预测

模型在实体预测任务中结果如表 2 所示。结果表明 TEKRL 除了 MeanRank 的 Raw 指标较 TKRL 和 TransR 略低，其他的指标均取得了提升。在 Filter 设置下，TEKRL 模型与 TKRL 模型相比，Hit@10 指标提升了约 7.2%，MeanRank 指标提升了约 23%。

表 2 实体预测的评估结果

Table 2 Evaluation results on entity prediction.

Metric	MeanRank		Hit@10(%)	
	Raw	Filter	Raw	Filter
RESCAL	828	683	28.4	44.1
SE	273	162	28.8	39.8
SME(linear)	274	154	30.7	40.8
SME(bilinear)	284	158	31.3	41.3
LFM	283	164	26.0	33.1
TransE	250	102	46.1	69.6
TransR	199	77	47.2	67.2
TKRL	**184**	68	49.2	69.4
TEKRL(SDPA)	205	53	49.2	76.1
TEKRL(SA)	205	**52**	**49.3**	**76.6**

4.3.2 关系预测

关系预测任务的结果如表 3 所示，从中可以看出 TEKRL(SA)在 MeanRank 指标上效果超过了其他模型，这说明本文提出的模型在关系预测中具有较好的效果。同时也可以看到，在 Hit@1 指标中，TKRL 模型的效果要略好于 TEKRL。然而，TKRL 模型利用了关系与类别之间的约束关系来编码层次结构信息，也就是说，相较于本文的模型，TKRL 模型还额外引入了约束关系的信息来提升模型的性能。但想要得到这种约束关系，需要对数据集有一定的要求，才便于构造出约束，或者需要对一些不容易提取出约束关系的数据集进行人工构造，这样的要求使得 TKRL 模型失去了一定的通用性和灵活性。本文提出的 TEKRL 模型通过学习的方式获取上述约束信息，更为普适和灵活，更适合于基于多源信息融合的知识表示学习。

表 3 关系预测的评估结果

Table 3 Evaluation results on relation prediction.

Metric	MeanRank		Hit@1(%)	
	Raw	Filter	Raw	Filter
TransE	2.79	2.43	68.4	87.2
TransR	2.49	2.09	70.2	91.6
TKRL	2.12	1.73	**71.1**	**92.8**
TEKRL(SDPA)	2.36	1.95	70.3	91.8
TEKRL(SA)	**2.11**	**1.69**	69.4	92.2

4.4 三元组分类实验结果

三元组分类是一个二分类任务，用于判断给定的三元组是否正确。在生成负样本三元组时，本文采取了与文献[18]相同的策略，对生成负样本时所需替换的实体或者关系进行一定的限制，使得负样本更加难以区分，从而更能展现出模型在三元组分类任务中的能力。在分类过程中，对于给定的一个三元组(h,r,t)，如果其得分低于给定的阈值γ，则预测其为正确的三元组，反之则为错误的。每种关系的阈值设置是不同的，具体值通过最大化验证集中对应关系下的分类准确率来设置。三元组分类的评估结果如表 4 所示，表明 TEKRL 模型具有较优的分类性能。

表 4 三元组分类的评估结果

Table 4 Evaluation results on triple classification.

Methods	Accuracy(%)
TransE	83.1

TransR	83.7
TEKRL(SDPA)	**84.9**
TEKRL(SA)	83.4

4.5 案例分析

为了进一步验证 TEKRL 模型可以学习到特定关系下不同类别的相关程度，并且更加清晰地展示模型的作用效果，我们列举一个具体的案例进行说明。

表 5 展示了在三元组(Gangs of New York, film festivals, 2010 Berlin Film Festival)中，2010 Berlin Film Festival 作为尾实体所拥有的类别在实验中注意力分数的排名，其中“Head: Gangs of New York”表示头实体为 Gangs of New York（一部美国电影），“Relation: film_festivals”表示关系为 film festivals，“Tail(interest): 2010 Berlin Film Festival”表示尾实体为 2010 Berlin Film Festival，并且是我们所关注类别排名的实体。从这个根据类别注意力分数得到的排名结果可以看出，排在最靠前的类别与三元组中的关系相关性最强，排在靠后位置的类别一般覆盖范围更广。

表 5 实体 2010 Berlin Film Festival 的类别根据注意力分数的排名情况

Table 5 The types of 2010 Berlin Film Festival ranked by attention score.

Head: Gangs of New York
Relation: film_festivals
Tail(interest): 2010 Berlin Film Festival

Rank	Types of tail
1	/film/film_festival_event
2	/film/film_screening_venue
3	/time/event
4	/common/topic

由于本文的模型具有上述特性，即可以分辨出不同关系中实体类别所起作用的程度不同。因此，当具有一词多义的实体处于不同语境中，可以通过最相关的类别来判断其具体含义，用于辅助实体消歧任务。实体消歧是由于同一实体指称在不同上下文可以指代不同实体，为了能够明确实体指称所指代的实体而提出的任务，在语义分析、搜索和问答等诸多自然语言处理相关的应用中都是需要解决的关键性问题。本文利用实体类别中注意力分数最高的作为在不同语义环境下分辨实体的参考依据，并基于这一想法设计了以下实验。

在实验中，从测试集随机选取 100 个三元组，通过模型得到头尾实体类别的注意力分数，并检验最高分数的类别是否能够直接体现出对应实体在该三元组中的语义。结果显示，其中有 61 个三元组都符合上述实验假设。因此证明，TEKRL 模型在实体消歧任务中也可以提供一定的帮助。限于篇幅，此处不对这一实验进行详细阐述。

5 结束语

本文针对融合实体类别信息的知识表示学习方法中常会受到的类别与关系之间需要设置约束条件的问题，提出了一种采用注意力机制学习二者之间相关性强度的方法 TEKRL。为了验证模型的有效性，在具有实体类别信息的数据集 FB15K 上，通过知识图谱补全和三元组分类这两项任务对模型进行了评估。通过与现有模型的实验结果进行对比，TEKRL 模型取得了一定的性能提升，并且提供了案例分析说明注意力机制的有效性，还进一步提出与验证了利用注意力分数可以辅助实体消歧任务的想法。本文提出的模型只是利用了类别信息，对于知识图谱来说，还有很多具有丰富语义的多源信息并没有得到充分利用，可以将模型融入更加多元化的信息联合训练，这是我们未来研究的一个方向。并且也可以多加关注实体消歧领域，通过合适的数据对我们的模型进行优化，期望能够使其应用到更多的任务中。

参考文献

[1] DONG Xin, GABRILOVICH E, HEITZ G, et al. Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion[C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York, USA: ACM, 2014: 601-610.

[2] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]//Annual Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, USA: MIT Press, 2013: 2787-2795.

[3] YANG Bishan, YIH W, HE Xiaodong, et al. Embedding entities and relations for learning and inference in knowledge bases[C].//International Conference on Learning Representations. San Diego, CA, USA: 2015.

[4] NEELAKANTAN A, ROTH B, MCCALLUM A. Com-

positional vector space models for knowledge base completion[C]//Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Beijing, China: ACL, 2015: 156-166.

[5] XIE Ruobing, LIU Zhiyuan, SUN Maosong. Representation learning of knowledge graphs with hierarchical types[C]// International Joint Conference on Artificial Intelligence. New York, NY, USA: AAAI Press, 2016: 2965-2971.

[6] WANG Zhen, ZHANG Jianwen, FENG Jianlin, et al. Knowledge Graph Embedding by Translating on Hyperplanes[C]//Twenty-eighth AAAI Conference on Artificial Intelligence. Québec City, Québec, Canada: AAAI Press, 2014: 1112-1119.

[7] FANG Yang, ZHAO Xiang, TAN Zhen, et al. A Revised Translation-Based Method for Knowledge Graph Representation[J]. Journal of Computer Research and Development, 2018, 55(1): 139-150. (in Chinese)
方阳，赵翔，谭真等. 一种改进的基于翻译的知识图谱表示方法[J]. 计算机研究与发展，2018，55(1): 139-150.

[8] XIAO Han, HUANG Minlie, HAO Yu, et al. TransA: An Adaptive Approach for Knowledge Graph Embedding [EB/OL]. [2020-02-09]. http://arxiv.org/abs/1509.05490

[9] XIAO Han, HUANG Minlie, HAO Yu, et al. TransG: A generative mixture model for knowledge graph embedding[EB/OL].[2020-02-09].https://arxiv.org/pdf/1509.05488v2.pdf.

[10] LIN Yankai, LIU Zhiyuan, SUN Maosong, et al. Learning entity and relation embeddings for knowledge graph completion[C]// Twenty-Ninth AAAI Conference on Artificial Intelligence. Austin, Texas, USA: AAAI Press, 2015:2181-2187.

[11] JI Guoliang, HE Shizhu, XU Liheng, 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, China: ACL, 2015: 687-696.

[12] BORDES A, WESTON J, COLLOBERT R, et al. Learning structured embeddings of knowledge bases[C]// Twenty-Fifth AAAI Conference on Artificial Intelligence. San Francisco, California , USA: 2011.

[13] BORDES A, GLOROT X, WESTON J, et al. Joint learning of words and meaning representations for open-text semantic parsing[C]//Artificial Intelligence and Statistics. La Palma, Canary Islands: MIT Press, 2012: 127-135.

[14] 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.

[15] NICKEL M, TRESP V, KRIEGEL H P. A Three-Way Model for Collective Learning on Multi-Relational Data[C]// Proceedings of the 28th International Conference on Machine Learning. Bellevue, Washington, USA: Omnipress, 2011, 11: 809-816.

[16] NICKEL M, TRESP V, KRIEGEL H P. Factorizing yago: scalable machine learning for linked data[C]//Proceedings of the 21st International Conference on World Wide Web. New York, USA: Association for Computing Machinery. 2012: 271-280.

[17] JENATTON R, ROUX N L, BORDES A, et al. A latent factor model for highly multi-relational data[C]//Advances in Neural Information Processing Systems. Lake Tahoe, Nevada, USA: MIT Press, 2012: 3167- 3175.

[18] SOCHER R, CHEN D, MANNING C D, et al. Reasoning with neural tensor networks for knowledge base completion[C]//Advances in Neural Information Processing Systems. Lake Tahoe, Nevada, USA: MIT Press, 2013: 926-934.

[19] XIE Ruobing, LIU Zhiyuan, JIA Jia, et al. Representation learning of knowledge graphs with entity descriptions[C]//Thirtieth AAAI Conference on Artificial Intelligence. Phoenix, Arizona, USA: AAAI Press, 2016: 2659-2665.

[20] XIE Ruobing, LIU Zhiyuan, LUAN Huanbo, et al. Image-embodied Knowledge Representation Learning[C]//International Joint Conference on Artificial Intelligence. Melbourne, Australia: AAAI Press, 2017: 3140-3146.

[21] WU Jiawei, XIE Ruobing, LIU Zhiyuan, et al. Knowledge Representation via Joint Learning of Sequential Text and Knowledge Graphs [EB/OL]. [2020-02-09]. https://arxiv.org/pdf/1609.07075.pdf.

[22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. Long Beach, CA, USA: MIT Press, 2017: 5998-6008.

[23] KINGMA D, BA J. Adam: A Method for Stochastic

Optimization [EB/OL]. [2020-02-09]. http://arxiv.org/abs/1412.6980.