

基于贝叶斯规则的具有层次注意力的知识补全

单晓欢, 赵雪, 陈廷伟

引用本文

单晓欢, 赵雪, 陈廷伟. [基于贝叶斯规则的具有层次注意力的知识补全](#)[J]. 计算机科学, 2023, 50(11): 234-240.

SHAN Xiaohuan, ZHAO Xue, CHEN Tingwei. [Bayesian Rule-based Knowledge Completion with Hierarchical Attention](#) [J]. Computer Science, 2023, 50(11): 234-240.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[QubitE:用于知识图谱补全的量子嵌入模型](#)

QubitE:Qubit Embedding for Knowledge Graph Completion

计算机科学, 2023, 50(11): 201-209. <https://doi.org/10.11896/jsjcx.221100217>

[多因素特征融合的EBSN活动推荐方法](#)

Event Recommendation Method with Multi-factor Feature Fusion in EBSN

计算机科学, 2023, 50(7): 60-65. <https://doi.org/10.11896/jsjcx.220900036>

[基于熵权-AHP与云模型的国产BIM建模软件多维度评价研究](#)

Multidimensional Evaluation Method for Domestic Building Information Modeling Software Based on Entropy-Weight-AHP and Cloud Model

计算机科学, 2023, 50(6A): 220400216-9. <https://doi.org/10.11896/jsjcx.220400216>

[基于自适应注意力机制的知识图谱补全算法](#)

Adaptive Attention-based Knowledge Graph Completion

计算机科学, 2022, 49(7): 204-211. <https://doi.org/10.11896/jsjcx.210400129>

[基于光照叠加的颜色恒常计算方法](#)

Light Superposition-based Color Constancy Computational Method

计算机科学, 2021, 48(11A): 386-390. <https://doi.org/10.11896/jsjcx.210200053>

基于贝叶斯规则的具有层次注意力的知识补全

单晓欢 赵雪 陈廷伟

辽宁大学信息学院 沈阳 110036

(shanxiaohuan@lnu.edu.cn)

摘要 知识图谱作为大数据时代的人工智能,被广泛应用于诸多领域,然而知识图谱普遍存在不完备性及稀疏性等问题。知识补全作为知识获取的子任务,旨在通过知识库中已知三元组来预测缺失的链接。然而现有方法普遍忽略了实体类型信息联合邻域信息对提高知识补全准确性的辅助作用,同时还存在特征信息被紧密编码到目标函数,导致集成操作高度依赖训练过程等问题。为此,提出了一种基于贝叶斯规则的具有层次注意力的知识补全方法。首先将实体类型和邻域信息视为层次结构,按关系进行分组,并独立计算组内各类信息的注意力权重。然后将实体类型和邻域信息编码为先验概率,将实例信息编码为似然概率,且按照贝叶斯规则将二者进行组合。实验结果表明,所提方法在 FB15k 数据集上的 MRR(Mean Reciprocal Rank)指标比 ConvE 提高 14.4%,比 TKRL 提高 10.7%;在 FB15k-237 数据集上的 MRR 指标比 TACT 提高了 2.1%。在 FB15k,FB15k-237 和 YAGO26K-906 数据集上,其 Hits@1 达到了 77.5%,73.8%和 95.1%,证明了引入具有层次结构的类型信息和邻域信息能够为实体嵌入更丰富、准确的描述信息,进而提升知识补全的精度。

关键词: 贝叶斯规则;实体类型;多层注意力;知识图谱补全

中图法分类号 TP311

Bayesian Rule-based Knowledge Completion with Hierarchical Attention

SHAN Xiaohuan,ZHAO Xue and CHEN Tingwei

College of Information,Liaoning University,Shenyang 110036,China

Abstract As artificial intelligence in the big data era,knowledge graphs are widely used in many fields. Knowledge graphs generally suffer from incompleteness and sparsity. As a sub-task of knowledge acquisition,knowledge completion aims to predict missing links from known triples in the knowledge base. However,existing knowledge completion methods generally ignore the auxiliary role of entity type jointly with neighborhood information,which can improve the knowledge completion accuracy. There are other problems such as feature information closely encodes into the objective function,and integration operations depend on the training process highly. To this end,a Bayesian rule-based knowledge completion method with hierarchical attention is proposed. Firstly,it regards entity type and neighborhood information as hierarchical structures,groups by relationship. It calculates each type information's attention weights independently. Then the entity types and neighborhood information encoding are regarded as the prior probability. The instance information encoding as likelihood probability. The two are combined according to the Bayesian rule. Experimental results show that the mean reciprocal rank(MRR) metric in the FB15k dataset improves 14.4% over ConvE and 10.7% over TKRL. The MRR metric in the FB15k-237 dataset improves 2.1% over TACT. In the FB15k,FB15k-237 and YAGO26K-906 datasets,its Hits@1 reaches 77.5%,73.8% and 95.1% respectively,which demonstrates the introduction of type information and neighborhood information with hierarchical structure can embed richer and more accurate descriptive information for entities,and thus improve the accuracy of knowledge completion.

Keywords Bayesian rule,Entity type,Hierarchical attention,Knowledge graph completion

1 引言

随着信息技术和人工智能技术的快速发展与结合,知识图谱作为大数据时代的人工智能,已在自动问答系统^[1]、智能推荐系统^[2]以及自然语言处理^[3]等众多领域得到广泛关注。

知识图谱本质上是具有有向图结构的知识库,旨在建模、识别实体之间的复杂关系,通常以结构化三元组的形式表示。例如三元组(姚明,国籍,中国)描述了“姚明的国籍是中国”这个知识。然而,在知识图谱的构建过程中,大量知识信息来源于文档和网络,而文档信息量有限,网络数据来源分布广、格式

到稿日期:2022-10-09 返修日期:2023-05-12

基金项目:国家重点研发计划

This work was supported by the National Key Research and Development Program of China.

通信作者:陈廷伟(twchen@lnu.edu.cn)

多变、质量参差不齐等因素,导致知识图谱通常是不完整的,大量实体之间存在的隐含关系没有被充分挖掘,这将对知识分析与决策造成影响。因此,利用已获取的知识对缺失的实体关系进行预测,以达到实体类型或关系的补全,即知识图谱补全(Knowledge Graph Completion, KGC)^[4],成为了当前知识图谱领域的研究热点。

图1给出了关于章子怡的知识图谱,图谱中既包含了实例信息,又显示了实体类型信息(本体信息),如章子怡的类型集合为{演员,金马奖获得者,人},在不同关系中实体类型有所不同。通过关系预测可补全“章子怡是否出演过《卧虎藏龙》”的缺失信息。

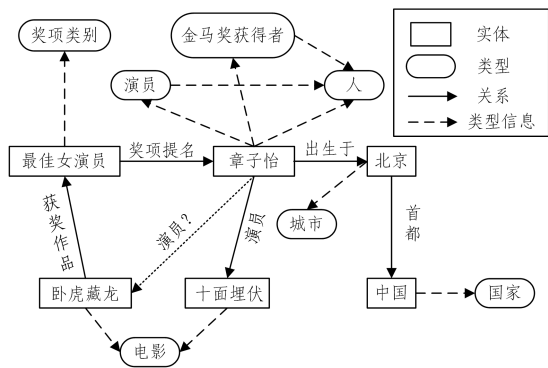


图1 知识图谱补全示例

Fig.1 Example of knowledge graph completion

现有的关系预测大多是基于知识图谱嵌入的方法, TransE^[5], DistMult^[6], ComplEx^[7], RotatE^[8], ConvE^[9]等常见的知识图谱嵌入模型通常先学习知识图谱中实体和关系的低维表示,再基于相应的线性表示或者神经网络模型进行链接预测。此类方法在聚合多跳邻居实体的消息传递过程中往往认为所有邻居实体具有相同的贡献,或者在所有任务中为邻域内的实体分配固定不变的权重,导致此类方法只能为实体和关系学习到单一的特征表示,而无法精准表述在当前上下文中的语义。实际上,现实知识图谱中很少存在孤立的实体和关系,其属性和特征表示取决于周围的相邻实体和对应关系的信息。将此类信息集成到实例级训练三元组中更详尽地了解实体的描述信息,可有效提高关系预测的准确性。

为此,本文将实体的类型信息和邻域信息视为层次结构,提出了一种基于贝叶斯规则的具有层次注意力的知识图谱补全方法(Bayesian Rule-based Knowledge Completion Method with Hierarchical Attention, BRHA)。本文的主要贡献如下:

(1)提出了一种多层注意力机制,按实体类型级、关系级以及实体级划分层次,并赋予它们不同的权重,有效利用实体类型及邻域信息增强了模型的预测效果。该机制为模型提供了一个细粒度的学习过程,增强了模型的可解释性。

(2)提出了基于贝叶斯规则的知识补全方法,在编码器-解码器框架中使用实体的类型信息和邻域信息来增强现有的嵌入模型。由于此类信息可独立于嵌入模型单独集成,因此无须额外训练,即可灵活应用于不同嵌入模型。

2 相关工作

近年来,基于知识图谱嵌入的方法被广泛应用于知识

图谱补全的任务中。嵌入模型将实体和关系映射为连续、低维向量空间中的向量进行表示,通过优化得分函数来学习实体和关系的嵌入,进而实现知识图谱补全。现有嵌入模型大致分为以下4类。

(1)基于翻译的模型。TransE模型^[5]将关系 r 视为在同一低维超平面上的从头实体 h 到尾实体 t 的平移,通过最小化损失函数来学习实体和关系的向量表示。TransH模型^[10]将三元组中的关系抽象成一个向量空间中的超平面,先将头实体或尾实体映射到超平面上,再通过超平面上的平移向量计算头尾实体的差值。TransR模型^[11]在实体空间和多个关系空间(关系特定的实体空间)中建模实体和关系,并在对应的关系空间中进行转换。

(2)基于张量分解的模型。RESCAL^[12]利用满秩矩阵表示关系,实体和关系之间通过矩阵运算实现深层次交互,具有较强的表现力。DisMult^[6]放松了对关系矩阵的约束,利用对角矩阵表示,并重新定义损失函数。ComplEx^[7]把DisMult扩展到复数空间,以解决对称及非对称关系的嵌入表示。

(3)基于神经网络的模型。ConvE^[9]将头实体和关系转换为二维向量,再利用卷积层和全连接层获取交互信息,通过与矩阵和尾实体的计算判断当前三元组的可信度。Wang等^[13]提出了一种自适应注意力网络用于知识图谱补全,引入自适应注意力建模每个特征维度对特定任务的贡献程度,为目标实体和关系生成动态可变的嵌入表示。

(4)基于旋转的模型。RotatE^[8]提出在复数空间中建模,把关系当作头尾实体之间的旋转,以建模和推断各种关系模式。QuatE^[14]扩展自ComplEx模型,其将基于复数的知识图谱嵌入拓展到超复数空间,采用四元数进行旋转,实现对称、反对称和反转关系的建模。

此外,现有方法试图通过将类型、语义等本体信息引入嵌入模型以及考虑类型之间的底层层结构等方式来提高知识补全效果。TKRL^[15]将实体的层次类型视为实体的投影矩阵,使用两种类型的编码器对层次结构进行建模。TransT^[16]通过实体类型构造关系类型,并利用相关实体、关系的类型语义相似性来捕获实体和关系的先验分布,然后使用先验分布生成实体的嵌入表示,并估计实体和关系预测的后验概率。JOIE^[17]是一种双视图知识图谱嵌入模型,该模型采用跨视图和视图内建模,学习跨视图关联模型以桥接本体概念及其相应实例实体的嵌入。HAKIE^[18]是一种层次感知知识图谱嵌入模型,可有效地对知识图谱中的语义层次结构进行建模,将实体映射到极坐标系中进行链接预测。TACT^[19]考虑了关系之间的语义相关性,以独立于实体的方式有效地利用关系之间的拓扑感知相关性进行链接预测。广义关系学习框架GRL^[20]通过关系之间的语义相关性来提高关系预测的效果。CFAG^[21]利用粗粒度聚合器和细粒度生成对抗网络中的两个粒度级别的关系语义来完成知识补全。GIE^[22]可以捕获更丰富的关系信息集,对关键推理模式进行建模,并实现跨实体的表达语义匹配。然而上述方法普遍存在集成操作高度依赖训练过程、忽略了实体固有的邻域信息在知识补全中的重要作用等问题。

实体邻域信息中隐含着大量实体间的继承、传递等关系,

将邻域信息引入嵌入模型,可进一步提升补全效果。R-GCN^[23]将图卷积网络扩展到关系图中,并利用本地邻域信息辅助完成链接预测。A2N^[24]自适应地结合实体的相关图邻域,以实现更准确的知识补全。Nathani 等^[25]提出了基于注意力的特征嵌入,其在模型中封装了关系集群和多跳关系,进一步区分了相邻节点的权重。RelGNN^[26]是一种基于消息传递的异构图注意力模型,其利用不同关系状态与节点状态对消息进行加权。然而目前鲜有方法将本体信息与实体信息联合考虑并独立集成地引入实体嵌入模型中。

3 基于贝叶斯规则的具有层次注意力的知识补全

本文提出的 BRHA 方法将实体类型信息和邻域信息视为层次结构,按关系进行分组,并独立计算组内类型信息以及实体和关系相应的注意力权重。BRHA 将类型信息和邻域信息的编码作为先验概率,将实例级信息编码作为似然概率,且按照贝叶斯规则将二者进行组合,这使得本文方法可更加灵活地使用任意嵌入模型。

3.1 BRHA 的整体框架

本文将知识图表示为 $\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{T}\}$, 其中 $\mathcal{E}$, $\mathcal{R}$ 和 $\mathcal{T}$ 分别是实体集、关系集和类型集。头尾实体 $e_h, e_t \in \mathcal{E}$, 关系 $r \in \mathcal{R}$, 类型 $t \in \mathcal{T}$ 。本文提出的模型遵循基于贝叶斯规则的编码器-

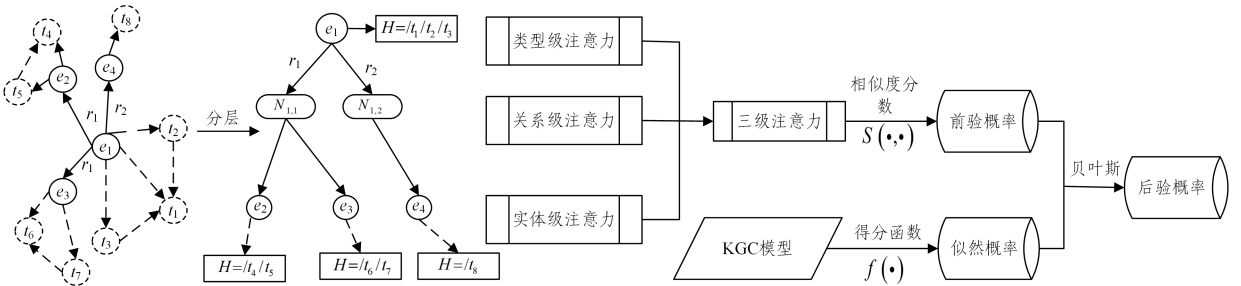


图2 BRHA 方法的整体框架

Fig. 2 Overall framework of BRHA approach

3.2 实体类型和邻域信息编码的先验模型

3.2.1 基于多层注意力机制的权重计算

本文使用分层注意力机制来为中心实体的类型信息和邻域信息赋权值,为模型提供一个细粒度的学习过程,进而增加模型的可解释性。类型信息是实体固有的信息,不受其他信息的影响,所以本文将实体的类型信息作为第一层注意力机制来赋权值;将邻域信息分为关系信息和相邻的实体信息,第二层注意力机制为关系信息赋权值;第三层则为相邻的实体信息赋权值。在这种设置下,具有相同关系的相邻实体的权重可以以组的方式进行训练,在集成各部分信息时既相互关联又相互独立,使得模型对中心实体的描述更准确,预测结果更稳定。

(1) 类型级注意力

研究发现,位于不同层次的类型权重在指示实体时的差异很大。例如,当我们想知道章子怡是否参演了《卧虎藏龙》时,章子怡的类型集合中演员类型比人的类型更具指示性。

对于实体 e_h , 类型级注意力表示每个类型在表示实体时的权重,定义为:

$$a_{e_h, r} = W_1 [e_h \parallel v_r] \quad (1)$$

解码器框架,如图2所示。

编码器首先将实体的类型信息和局部邻域信息视为层次结构。本文提出了一种多层注意力机制,包括:(1)类型级注意力,实体本身有许多固有的类型属性,这些属性按照特定的层次排列,不同层次的属性对指示实体具有不同的权重;(2)关系级注意力,实体之间具有多种不同关系,这些关系对于指示实体具有不同的权重;(3)实体级注意力,同一关系下不同实体对于指示实体具有不同的权重。从图2的左侧可以看出,给定一个实体 e , 它的类型集表示为 $T_e \subset \mathcal{T}$, 类型的子集之间的层次结构具有格式 $H = /t_1/t_2/\dots/t_k/t_K$ 。其中 $t_k \in T_e$, K 是层次结构级别的总数, t_K 是最具体的语义类型, t_1 是最抽象的语义类型, $N_{e_h, r}$ 表示实体 h 在关系 r 下的相邻实体集。其次,计算类型级注意力、关系级注意力和实体级注意力,并将3个注意力分数组合为一个三级注意力分数,该分数为中心实体的类型信息和邻域信息对其的权重。然后根据实体集与类型信息、邻域信息之间的相关性衡量实体和关系之间的语义相似性,通过语义相似性来计算模型的先验概率。

解码器通过嵌入模型的实例级信息计算似然概率。最后将先验概率和似然概率按贝叶斯规则进行组合以获取模型的后验概率。

$$a_{e_h, r} = \text{soft max}_t(a_{e_h, r}) = \frac{\exp(\sigma(p \cdot a_{e_h, r}))}{\sum_{t' \in T_{e_h}} \exp(\sigma(p \cdot a_{e_h, t'}))} \quad (2)$$

其中, $\parallel$ 表示连接操作, $e_h \in \mathbb{R}^d$ 是实体 h 的嵌入, d 是嵌入的维度; $W_1 \in \mathbb{R}^{d \times 2d}$, $v_r \in \mathbb{R}^d$ 和 $p \in \mathbb{R}^d$ 是训练参数, v_r 表示类型特征的类型特定参数; $\mathcal{T}_{e_h}$ 表示实体 e_h 的类型集, σ 表示负输入斜率为 0.2 的 LeakyReLU 函数。经过上面的计算,得到了类型级别的注意力分数 $a_{e_h, t}$, 它代表了类型 t 在表示实体 e_h 时的权重。

(2) 关系级注意力

关系级注意力旨在解决不同关系在指示实体时的权重不同的问题。例如,当表示《卧虎藏龙》这部电影时,导演关系比基于国家的关系更具指示性,因为在中国不止有这一部电影。

相邻实体和关系可以表示为多个 (e_h, r, e_t) 三元组。在本文中,对于实体 e_1 , 将三元组 (e_2, r, e_1) 转换为 (e_1, r^{-1}, e_2) 。这样,每个实体都可以充当邻域三元组的头实体,而本文只需将尾实体的信息聚合到头部实体即可。

对于实体 e_h , 关系级注意力表示每个关系在表示实体时的权重,定义为:

$$b_{e_h, t, r} = W_2 [a_{e_h, r} \parallel r] \quad (3)$$

$$\beta_{e_h, t, r} = \text{soft max}_r(b_{e_h, t, r}) = \frac{\exp(\sigma(q \cdot b_{e_h, t, r}))}{\sum_{r' \in N_{e_h}} \exp(\sigma(q \cdot b_{e_h, t, r'}))} \quad (4)$$

其中, $r \in \mathbb{R}^d$ 是关系 r 的嵌入; $\mathcal{N}_{e_h}$ 表示实体 e_h 的关系集; $W_2 \in \mathbb{R}^{d \times 2d}$ 和 $q \in \mathbb{R}^d$ 是训练参数; $b_{e_h, t, r}$ 可以看作是实体 e_h 在类型 t 和关系 r 下的表示; $\beta_{e_h, t, r}$ 是关系级别的注意力分数, 它代表了关系 r 在表示实体 e_h 时的权重。

(3) 实体级注意力

实体级注意力旨在解决同一关系下相邻实体的权重也可能不同的问题。例如,《卧虎藏龙》与不同演员存在联系,然而主演可能比其他演员更具指示性。

所提方法首先将同一关系下的相邻实体视为一个组,然后计算实体级注意力如下:

$$\gamma_{e_h, t, r, e_t} = W_3 [b_{e_h, t, r} \parallel e_t] \quad (5)$$

$$\lambda_{e_h, t, r, e_t} = \text{soft max}_{e_t}(\gamma_{e_h, t, r, e_t}) = \frac{\exp(\sigma(s \cdot \gamma_{e_h, t, r, e_t}))}{\sum_{e'_t \in N_{e_h, r}} \exp(\sigma(s \cdot \gamma_{e_h, t, r, e'_t}))} \quad (6)$$

其中, $e_t \in \mathbb{R}^d$ 是实体 e_t 的嵌入, 并且 e_t 是关系 r 下的尾部实体集; $\mathcal{N}_{e_h, r}$ 表示实体 e_h 在关系 r 下的尾部实体集; $W_3 \in \mathbb{R}^{d \times 2d}$ 和 $s \in \mathbb{R}^d$ 是训练参数; γ_{e_h, t, r, e_t} 可以看作是相邻三元组 (e_h, r, e_t) 的表示; λ_{e_h, t, r, e_t} 是实体级别的注意力分数, 表示实体 e_t 在表示 e_h 时在关系 r 下的所有尾部实体中的权重。

在得到类型级注意力、关系级注意力和实体级注意力之后, 将这 3 个分数进一步组合成一个三级注意力分数, 计算式如下:

$$\mu_{e_h, t, r, e_t} = \alpha_{e_h, r} \cdot \beta_{e_h, t, r} \cdot \lambda_{e_h, t, r, e_t} \quad (7)$$

其中, 三级注意力分数 μ_{e_h, t, r, e_t} 表示具有类型 t 的三元组 (e_h, r, e_t) 在表示 e_h 时的权重。分层注意力机制为注意力分数提供了细粒度的学习过程, 增加了模型的可解释性。此外, 值得注意的是, 关系 r 下的所有相邻三元组都显式共享了关系级注意力 $\beta_{e_h, t, r}$, 这有助于这些三元组之间的知识共享, 使得具有关系 r 的相邻三元组的权重能够以组的方式进行训练。

3.2.2 基于类型信息和邻域信息的先验概率

给定一个三元组 $(e_h, r, e_t) \in \mathcal{G}$, 本文根据其实体集与类型信息、邻域信息之间的相关性来衡量实体和关系之间的语义相似性^[16]。相似性函数 $s(\cdot, \cdot)$ 如下:

$$s(e_h, r) = \frac{\sum_{t \in \mathcal{T}_{r, \text{head}}} \sum_{r \in R \cap R_{e_h}} \sum_{e_t \in \mathcal{E} \cap \mathcal{E}_r} \mu_{e_h, t, r, e_t}}{\sum_{t \in \mathcal{T}_{r, \text{head}}} \sum_{r \in R} \sum_{e_t \in \mathcal{E}} \mu_{e_h, t, r, e_t}} \quad (8)$$

其中, $\mathcal{T}_{r, \text{head}} \cap \mathcal{T}_{e_h} = \{t | t \in \mathcal{T}_{r, \text{head}} \text{ and } t \in \mathcal{T}_{e_h}\}$, $0 \leq s(\cdot, \cdot) \leq 1$ 。然后根据相似度得分定义先验概率 $p(r | X(e_h, e_t, t, R))$ 为:

$$p(r | X(e_h, e_t, t, R)) = \frac{s(e_h, r)}{\sum_{r' \in R} s(e_h, r')} \quad (9)$$

其中, $X(e_h, e_t, t, R)$ 表示与实体对 (e_h, e_t) 和关系集 R 相关的类型信息和邻域信息。

给定来自第 i 个输入 (e_{h_i}, r_i, e_{t_i}) 的基本事实关系 r_i , 本文采用交叉熵来评估先验模型的损失, 计算式如下:

$$L_p = -\frac{1}{N} \sum_{i=1}^N \log p(r_i | X(e_{h_i}, e_{t_i}, t_i, R)) \quad (10)$$

其中, $p(r_i | X(e_{h_i}, e_{t_i}, t_i, R))$ 是关系 r_i 的先验概率。

3.3 基于实例级信息的似然模型

基于嵌入的模型表示连续嵌入空间中的关系和实体。将头尾实体的嵌入分别表示为 e_h 和 e_t , 关系 r 的嵌入是 r , 评分函数 $f_r(e_h, e_t)$ 通常定义为三元组 (e_h, r, e_t) 的度量。然后通过基于实例级三元组优化得分函数来学习嵌入。对于每个三元组 (e_h, r, e_t) , 根据它们对应的嵌入定义关系的似然概率为:

$$p(e_h, e_t | r) \triangleq \exp(f_r(e_h, e_t)) \quad (11)$$

本文尝试使用多种不同模型作为解码器, 发现使用 RotatE^[8] 可以获得最佳的性能, 因此本文使用 RotatE 作为似然模型, 其评分函数为:

$$f_r(e_h, e_t) = -\|e_h \circ r - e_t\| \quad (12)$$

其中, $e_h, e_t \in \mathbb{C}^d$, $r \in \mathbb{C}^d$ 。

3.4 后验模型

后验模型将类型信息、邻域信息与实例级信息进行集成。给定一个三元组 (e_h, r, e_t) , 根据类型信息和邻域信息获得关系 r 的先验概率, 即 $p(r | X(e_h, e_t, t, R))$, 基于嵌入模型获得关系的似然概率, 即 $p(e_h, e_t | r)$ 。然后将它们组合, 遵循贝叶斯规则获得关系 r 的后验概率, 如下所示:

$$p(r | e_h, e_t, X(e_h, e_t, t, R)) \propto p(e_h, e_t | r) p(r | X(e_h, e_t, r, R)) \quad (13)$$

后验概率 $p(r | e_h, e_t, X(e_h, e_t, t, R))$ 包含类型信息、邻域信息和实例级信息。

BRHA 模型的损失函数定义如下:

$$L = \sum_{(e_h, r, e_t)} -\frac{1}{N} \sum_{i=1}^N (y_{(e_h, r, e_t)} \cdot \log(g(f(e_h, r, e_t))) + (1 - y_{(e_h, r, e_t)}) \log(1 - g(f(e_h, r, e_t)))) \quad (14)$$

其中, $y_{(e_h, r, e_t)}$ 是三元组 (e_h, r, e_t) 的标签 (1 或 0); N 表示尾实体的候选数, g 是 sigmoid 函数。BRHA 的学习过程使用 Adam 优化器^[27] 进行。

本文提出的 BRHA 方法与现有方法将类型信息紧密编码到目标函数中有所不同, 其将实体类型信息和邻域信息的集成独立于基于嵌入的模型, 使得集成过程不依赖于训练过程, 提高了方法的可移植性。

4 实验验证

为评估 BRHA 方法的性能, 本文与其他先进的知识补全方法进行了性能对比, 并在 3 个不同基线数据集上进行了消融实验, 验证了基于多层注意力机制的权重、考虑类型信息和邻域信息以及使用贝叶斯框架对知识补全的重要性, 进一步验证了本文方法的有效性和准确性。

4.1 数据集

本文在 FB15k^[5], FB15k-237^[27] 和 YAGO26K-906^[17] 数据集上进行实验, 数据集统计信息如表 1 所列。

表 1 数据集的统计信息

Table 1 Statistical information of datasets

Dataset	$ \mathcal{E} $	$ \mathcal{R} $	Train	Valid	Test
FB15k	14 951	1 345	483 142	50 000	59 071
FB15k-237	14 541	237	272 115	17 535	20 466
YAGO26K-906	26 078	34	351 664	—	39 074

FB15k 是从 Freebase 中抽取的事实三元组, 其实体类型

信息的提取已在 TKRL^[15] 和 TransT^[16] 等中完成。FB15k-237 消除了数据集 FB15k 中的反向关系,使用的类型信息与 FB15k 一样。YAGO26K-906 是从 YAGO 中提取的三元组,它是一个包含显式本体信息的数据集。其中 $|\mathcal{E}|$ 表示实体数量, $|\mathcal{R}|$ 表示关系数量, $Train$ 表示训练集的三元组数量, $Valid$ 表示验证集的三元组数量, $Test$ 表示测试集的三元组数量。

在 FB15k 和 FB15k-237 数据集中,每个实体的类型都是非空的。在 YAGO26K-906 上,只有 8948 个实体具有非空的类型集,测试集中只有 4149 个测试三元组的实体具有类型信息;30839 个三元组只有头实体或尾实体的类型信息,4086 个三元组没有头尾实体的类型信息。对于实体 e 的类型集为空(e 可以是头实体 e_h 或尾实体 e_t)的情况,本文分配统一的相似度分数,即 $s(e,r')=1, \forall r' \in \mathcal{R}$ 。

4.2 实验设置

为了验证 BRHA 的有效性,本文将其与 TransE^[5], DistMult^[6], ComplEx^[7], RotatE^[8], ConvE^[9], TKRL^[15], TransT^[16], JOIE^[17], HAKE^[18], TACT^[19], GRL^[20], CFAG^[21], GIE^[22], R-GCN^[23], A2N^[24] 以及 Nathani^[25] 这 16 种基线进行比较。

在训练阶段,本文采用 BRHA 来训练实体和关系的嵌入。本文使用 Adam 作为优化器,采用多次实验来寻找最佳参数设置。对于编码器,实体和关系的嵌入维度在输入层和输出层都设为 100,将比率为 0.5 的 dropout 应用于编码器的输入层。对于解码器,将实体和关系的嵌入维度与编码器保持一致,将比率为 0.2 的 dropout 应用于特征图。在训练过程中,FB15k 和 FB15k-237 的学习率皆设置为 0.0005, YAGO26K-906 的学习率设置为 0.001。所有的训练参数应用 L2 正则化, $\lambda=0.0005$,并随机初始化所有的训练参数。

4.3 评估指标

在测试阶段,对于测试集中的每个三元组,依次用知识图谱中的所有实体来替换它们的头、尾实体,然后为所有损坏的三元组计算得分,并根据分数进行排名。本文使用以下指标将 BRHA 方法与基线进行比较:(1)MRR(平均倒数排名),MRR 值越大,表明排名越符合事实情况,嵌入的效果越好;(2)Hits@N(在链接预测中排名前 N 的三元组的平均占比),其中 $N=1,3,10$ 。MRR 和 Hits@N 值越大表示性能越好。

4.4 实验结果与分析

本文将所提方法与多种基线模型进行比较来验证其有效性,并通过消融实验验证基于多层注意力机制的权重、考虑类型信息和邻域信息以及使用贝叶斯框架对知识补全的影响。

4.4.1 知识补全性能对比

本文在不同数据集上与采用相应数据集的基线模型进行对比,表 2—表 4 列出了不同数据集上的知识补全实验结果。在 FB15K 数据集上,BRHA 获得了最好的 MRR, Hits@1, Hits@3 和 Hits@10。本文方法优于基线模型的原因是本文方法将实体的类型信息和邻域信息联合三元组进行建模,并基于多层注意力机制对各类信息赋予不同权重,有效提高了知识补全的效果。和 R-GCN 相比,BRHA 的 MRR 提高了 10.5%,Hits@10 提高了 5.5%,这说明 BRHA 比 R-GCN

使用的关系图卷积网络获得的信息更多。与融合了实体的类型信息的 TKRL 和 TransT 相比,BRHA 的实验结果更好,说明融合邻域信息能够更好地进行链接预测。

表 2 FB15k 数据集上的实验结果

Table 2 Experimental results on FB15k dataset				
Model	MRR	Hits@1	Hits@3	Hits@10
TransE	0.463	0.297	0.578	0.749
DistMult	0.798	—	—	0.893
ComplEx	0.692	0.599	0.759	0.840
RotatE	0.797	0.746	0.830	0.884
ConvE	0.657	0.558	0.723	0.831
TKRL	0.694	0.711	—	0.734
TransT	0.761	0.720	—	0.854
R-GCN	0.696	0.601	0.760	0.842
BRHA	0.801	0.775	0.841	0.897

表 3 FB15k-237 数据集上的实验结果

Table 3 Experimental resultson FB15k-237 dataset				
Model	MRR	Hits@1	Hits@3	Hits@10
TransE	0.294	—	—	0.465
DistMult	0.241	0.155	0.263	0.419
ComplEx	0.247	0.158	0.275	0.428
RotatE	0.338	0.241	0.375	0.533
ConvE	0.325	0.237	0.356	0.501
HAKE	0.346	0.250	0.381	0.542
TACT	0.830	0.741	—	—
GRL	0.326	0.248	—	0.502
CFAG	0.393	0.295	0.440	0.575
GIE	0.362	0.271	0.401	0.552
R-GCN	0.164	0.10	0.181	0.300
A2N	0.317	0.232	0.348	0.486
Nathani	0.518	0.460	0.540	0.626
BRHA	0.851	0.738	0.512	0.640

表 4 YAGO26K-906 数据集上的实验结果

Table 4 Experimental results on YAGO26K-906 dataset				
Model	MRR	Hits@1	Hits@3	Hits@10
TransE	0.491	0.116	0.874	0.915
RotatE	0.921	0.941	0.956	0.949
JOIE	0.888	0.708	0.938	0.893
BRHA	0.924	0.951	0.954	0.953

在 FB15k-237 数据集上,BRHA 获得了最好的 MRR, Hits@3 和 Hits@10。与 HAKE 相比,BRHA 的 MRR 提升了 50.5%,Hits@3 提升了 13.1%。与融合关系语义的 TACT,GRL,CFAG 和 GIE 相比,BRHA 方法融合了更丰富的信息,使得链接预测的结果更加准确。实验结果表明,BRHA 在大多数指标上都取得了较好的结果。

在 YAGO26K-906 数据集上,BRHA 获得了最好的 MRR,Hits@1 和 Hits@10。与 TransE 相比,BRHA 的 MRR 提升了 43.3%,Hits@10 提升了 3.8%。与双视图知识图谱嵌入模型 JOIE 相比,BRHA 方法聚合了实体的邻域信息,能够有效地增强知识补全的效果。

4.4.2 基于多层注意力机制的权重有效性

为了证明所提出的基于层次结构的类型权重和邻域权重(Hierarchy-based)的有效性,本文考虑与均匀权重(Uniform)进行比较,并根据具有均匀权重的类型信息和邻域信息计算先验概率,实验结果如表 5 所列。其中基于多层注意力机制权重的先验模型比具有均匀权重的先验模型实现了更好的

性能,特别是在 FB15K 数据集上。这是因为本文基于多层注意力的方法能够为链接预测中更重要的实体类型、关系和

实体赋予更高的权值,进而得到更准确的向量表示,而均匀权重无法区分优先级。

表 5 不同数据集上基于多层注意力机制的权重的有效性

Table 5 Validity of weights based on multi-layer attention mechanism on different datasets

Type and Neighborhood weights	FB15k				FB15k-237				YAGO26K-906			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
Uniform	0.415	0.476	0.574	0.769	0.524	0.547	0.625	0.784	0.378	0.496	0.450	0.756
Hierarchy-based	0.741	0.810	0.715	0.896	0.658	0.760	0.724	0.819	0.722	0.793	0.698	0.823

4.4.3 类型信息和邻域信息的有效性

为了研究类型信息和邻域信息的有效性,本文考虑仅用类型信息(T)来评估先验模型的链接预测性能、只用邻域信息(N)来评估先验模型的有效性,以及使用类型信息和邻域信息(T+N)来评估先验模型的有效性。YAGO26K-906 只

有一个三元组子集包含头部和尾部实体的类型信息(4149 个三元组),因此本文对这个子集进行评估。实验结果如表 6 所列,联合考虑类型信息和邻域信息时,先验模型取得了更优的性能,这是因为其能够对中心实体嵌入更丰富的描述信息,使得向量表示更加准确,有效提升知识补全的效果。

表 6 不同数据集上类型信息和邻域信息的有效性

Table 6 Validity of type information and neighborhood information on different datasets

Info.	FB15k				FB15k-237				YAGO26K-906			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
T	0.612	0.599	0.614	0.662	0.625	0.410	0.526	0.651	0.424	0.389	0.517	0.716
N	0.679	0.710	0.620	0.745	0.717	0.724	0.610	0.718	0.469	0.427	0.673	0.814
T+N	0.831	0.788	0.753	0.886	0.826	0.790	0.748	0.852	0.813	0.741	0.796	0.897

4.4.4 贝叶斯框架的有效性

为了研究 BRHA 使用的贝叶斯框架的有效性,本文考虑以下两种情况:(1)不使用贝叶斯框架(NO)整合先验模型和似然模型的信息验证链接预测的有效性;(2)使用贝叶斯规则(YES)验证链接预测的有效性。

实验结果如表 7 所列,可以看出,使用贝叶斯框架进行链接预测可以得到更好的实验结果。使用贝叶斯框架可以将先验模型和似然模型分开训练,而不使用贝叶斯的方法会导致实体类型和邻域信息被紧密地编码到目标函数,使得集成操作高度依赖于训练过程,导致知识补全的效果大大降低。

表 7 不同数据集上贝叶斯框架的有效性

Table 7 Validity of Bayesian framework on different datasets

Bayesian	FB15k				FB15k-237				YAGO26K-906			
	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10	MRR	Hits@1	Hits@3	Hits@10
NO	0.265	0.390	0.441	0.607	0.279	0.387	0.394	0.641	0.246	0.373	0.510	0.607
YES	0.481	0.492	0.562	0.696	0.393	0.510	0.598	0.701	0.399	0.512	0.622	0.714

结束语 本文提出了一种基于贝叶斯规则的具有层次注意力的知识补全方法(BRHA),BRHA 的多层注意力机制有效地将知识图谱中实体的类型信息和邻域信息进行整合,为模型提供了一个细粒度的学习过程,增强了模型的可解释性。BRHA 将实体类型信息和邻域信息编码为先验概率,将基于嵌入模型的实例级信息编码为似然概率,通过贝叶斯规则将其组合得到后验概率。本文框架可以独立于基于嵌入的模型,从而避免了训练过程中特征信息被密集编码的问题,增强了模型的灵活性。大量实验验证了 BRHA 方法的效率,其可以有效提高链接预测的准确性。

参 考 文 献

[1] HUANG X,ZHANG J,LI D,et al. Knowledge graph embedding based question answering [C] // Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining. 2019:105-113.

[2] WANG X,HE X,CAO Y,et al. Kgat: Knowledge graph attention network for recommendation[C]// Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019:950-958.

[3] KAPANIPATHI P,THOST V,PATEL S S,et al. Infusing knowledge into the textual entailment task using graph convolutional networks[C] // Proceedings of the AAAI Conference on Artificial Intelligence. 2020:8074-8081.

[4] WANG S,DU Z J,MENG X F. Research progress of large-scale knowledge graph completion technology[J]. Chinese Science: Information Science, 2020,50(4):551-575.

[5] BORDES A,USUNIER N,GARCIA-DURAN A,et al. Translating embeddings for modeling multi-relational data[C] // Proceedings of the Advances in Neural Information Processing Systems. 2013:2787-2795.

[6] YANG B,YIH W T,HE X,et al. Embedding entities and relations for learning and inference in knowledge bases[C] // Proceedings of the International Conference on Learning Representations. 2015.

[7] TROUILLON T,WELBL J,RIEDEL S,et al. Complex embeddings for simple link prediction[C]// Proceedings of the 33th International Conference on Machine Learning. 2016:2071-2080.

[8] SUN Z,DENG Z H,NIE J Y,et al. Rotate: Knowledge graph embedding by relational rotation in complex space[C] // Proceedings of the International Conference on Learning Representations.

tations, 2019.

[9] DETTMERS T, MINERVINI P, STENETROP P, et al. Convolutional 2d knowledge graph embeddings[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2018;1811-1818.

[10] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2014;1112-1119.

[11] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2015;2181-2187.

[12] NICKEL M, TRESP V, KRIEGEL H P. A three-way model for collective learning on multi-relational data[C]//Proceedings of the International Conference on Machine Learning. 2011;809-816.

[13] WANG J, LI X N, LI G Y. Knowledge Graph Completion Algorithm Based on Adaptive Attention Mechanism[J]. Computer Science, 2022, 49(7):204-211.

[14] ZHANG S, TAY Y, YAO L, et al. Quaternion knowledge graph embeddings[C]//Proceedings of the Advances in Neural Information Processing Systems. 2019;2731-2741.

[15] XIE R, LIU Z, SUN M. Representation Learning of Knowledge Graphs with Hierarchical Types[C]//Proceedings of the International Joint Conference on Artificial Intelligence. 2016;2965-2971.

[16] MA S, DING J, JIA W, et al. TransT: Type-based multiple embedding representations for knowledge graph completion[C]//Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. 2017;717-733.

[17] HAO J, CHEN M, YU W, et al. Universal representation learning of knowledge bases by jointly embedding instances and ontological concepts[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019;1709-1719.

[18] ZHANG Z, CAI J, ZHANG Y, et al. Learning hierarchy-aware knowledge graph embeddings for link prediction[C]//Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence. 2020;3065-3072.

[19] CHEN J, HE H, WU F, et al. Topology-Aware Correlations Between Relations for Inductive Link Prediction in Knowledge Graphs[C]//Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence. 2021;6271-6278.

[20] ZHANG Y, ZHANG X, WANG J, et al. Generalized Relation

Learning with Semantic Correlation Awareness for Link Prediction[C]//Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence. 2021;4679-4687.

[21] WANG C, ZHOU X, PAN S, et al. Exploring Relational Semantics for Inductive Knowledge Graph Completion[C]//Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence. 2022.

[22] CAO Z, XU Q, YANG Z, et al. Geometry Interaction Knowledge Graph Embeddings[C]//Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence. 2022.

[23] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]//Proceedings of the Extended Semantic Web Conference. 2018;593-607.

[24] BANSAL T, JUAN D C, RAVI S, et al. A2n: Attending to neighbors for knowledge graph inference[C]//Proceedings of the Association for Computational Linguistics. 2019;4387-4392.

[25] NATHANI D, CHAUHAN J, SHARMA C, et al. Learning attention-based embeddings for relation prediction in knowledge graphs[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019;4719-4723.

[26] QIN X, SHEIKH N, REINWALD B, et al. Relation-aware graph attention model with adaptive self-adversarial training[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021;5774-5782.

[27] KINGMA D P, BA J L. Adam: A method for stochastic optimization[C]//Proceedings of the International Conference on Learning Representations. 2014.



SHAN Xiaohuan, born in 1987, Ph.D candidate, is a student member of China Computer Federation. Her main research interests include graph data processing technology and knowledge graph data management, etc.



CHEN Tingwei, born in 1974, Ph.D, professor, master's supervisor, is a senior member of China Computer Federation. His main research interests include intelligent transportation and machine learning, etc.

(责任编辑:何杨)