

规则引导的知识图谱联合嵌入方法

姚思雨^{1,2} 赵天哲^{1,2} 王瑞杰^{1,2,3} 刘 均^{1,2}

¹(西安交通大学计算机科学与技术学院 西安 710049)
²(陕西省天地网技术重点实验室(西安交通大学) 西安 710049)
³(苏黎世大学计算机科学系 瑞士 8050)
(cheryl@stu.xjtu.edu.cn)

Rule-Guided Joint Embedding Learning of Knowledge Graphs

Yao Siyu^{1,2}, Zhao Tianzhe^{1,2}, Wang Ruijie^{1,2,3}, and Liu Jun^{1,2}

¹(College of Computer Science and Technology, Xi'an Jiaotong University, Xi'an 710049)
²(Shaanxi Key Laboratory of STN Technology R&D (Xi'an Jiaotong University), Xi'an 710049)
³(Department of Informatics, University of Zurich, Switzerland 8050)

Abstract In recent years, numerous research works have been devoted to knowledge graph embedding learning which aims to encode entities and relations of the knowledge graph in continuous low-dimensional vector spaces. And the learned embedding representations have been successfully utilized to alleviate the computational inefficiency problem of large-scale knowledge graphs. However, most existing embedding models only consider the structural information of the knowledge graph. The contextual information and literal information are also abundantly contained in knowledge graphs and could be exploited to learn better embedding representations. In this paper, we focus on this problem and propose a rule-guided joint embedding learning model which integrates the contextual information and literal information into the embedding representations of entities and relations based on graph convolutional networks. Especially for the convolutional encoding of the contextual information, we measure the importance of a piece of contextual information by computing its confidence and relatedness metrics. For the confidence metric, we define a simple and effective rule and propose a rule-guided computing method. For the relatedness metric, we propose a computing method based on the representations of the literal information. We conduct extensive experiments on two benchmark datasets, and the experimental results demonstrate the effectiveness of the proposed model.

Key words knowledge graph; representation learning; graph convolutional networks; contextual information; literal information

摘 要 近年来,大量研究工作致力于知识图谱的嵌入学习,旨在将知识图谱中的实体与关系映射到低维连续的向量空间中.且所学习到的嵌入表示已被成功用于缓解大规模知识图谱的计算效率低下问题.

收稿日期:2020-10-13;修回日期:2020-11-03
基金项目:国家重点研发计划项目(2018YFB1004500);国家自然科学基金项目(61672419,61672418,61877050,61937001);中国工程院咨询研究项目;国家自然科学基金创新研究群体(61721002);教育部创新团队(IRT_17R86);中国工程科技知识中心项目
This work was supported by the National Key Research and Development Program of China (2018YFB1004500), the National Natural Science Foundation of China (61672419, 61672418, 61877050, 61937001), the Consulting Research Project of Chinese Academy of Engineering, the Innovative Research Group Project of the National Natural Science Foundation of China (61721002), the Ministry of Education Innovation Research Team (IRT_17R86), and the Project of China Knowledge Centre for Engineering Science and Technology.

然而,大多数现有嵌入学习模型仅考虑知识图谱的结构信息.知识图谱中还包含有丰富的上下文信息和文本信息,它们也可被用于学习更准确的嵌入表示.针对这一问题,提出了一种规则引导的知识图谱联合嵌入学习模型,基于图卷积网络,将上下文信息与文本信息融合到实体与关系的嵌入表示中.特别是针对上下文信息的卷积编码,通过计算单条上下文信息的置信度与关联度来度量其重要程度.对于置信度,定义了一个简单有效的规则并依据该规则进行计算.对于关联度,提出了一种基于文本表示的计算方法.最后,在 2 个基准数据集上进行的实验结果证明了模型的有效性.

关键词 知识图谱;表示学习;图卷积网络;上下文信息;文本信息

中图法分类号 TP391

近年来,由于具有表达能力强、歧义性低、模式统一、且支持推理等优点,知识图谱已被广泛用于组织和发布各领域的结构化数据.通常,知识图谱由实体、实体所具有的属性以及实体间的关系所组成.例如,其中可能包含有实体中国、关系首都以及实体属性“China”.如图 1 所示,知识图谱的基础构成则是描述 2 个实体之间的关系或实体及其属性之间关系的三元组,如(中国,首都,北京)、(中国,英语标签,“China”).

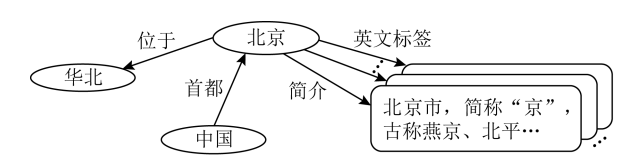


Fig. 1 Several triples which contain the entity Beijing and the related literals
图 1 包含实体“北京”的若干三元组及文本信息

目前,知识图谱已被广泛应用在智能问答^[1]、推荐系统^[2]和信息检索^[3]等任务中,其突出表现在学术与工业界均获得了广泛关注^[4].但是,受益于知识图谱所包含丰富信息的同时,其庞大的规模与数据稀疏性问题也给知识图谱的应用带来了挑战.例如,Freebase^[5], Yago^[6]和 Dbpedia^[7]等开放领域知识图谱中通常包含有数百万个实体,以及上亿条描述实体关系的三元组.将子图匹配等传统图算法应用在这些大规模知识图谱上往往存在计算低效性问题.为此,研究人员提出了知识图谱嵌入学习模型(knowledge graph embedding learning model),将知识图谱映射到低维、连续的向量空间中,学习实体与关系的嵌入表示^[8].

通过设计特定的表示学习机制,知识图谱的结构和语义等信息可被编码在所学习到的嵌入表示中.一方面,原本需要对大规模知识图谱进行频繁访问的操作,例如结构化查询构建(structured query

construction)^[9]、逻辑查询执行(logical query processing)^[10]和查询放缩(query relaxation)^[11],均可在所学习到的嵌入表示空间中通过数值计算完成,极大地提高了效率.另一方面,知识图谱的嵌入学习提供了一种抽取并高效表示知识图谱特征信息的方法,类似于自然语言处理领域中被广泛应用的词嵌入(word embedding),知识图谱的嵌入表示也为基于知识图谱的深度学习工作提供了极大的便利.

现有知识图谱嵌入学习模型大多仅关注知识图谱中以三元组表示的结构信息.例如, Bordes 等人提出了基于翻译机制(translation mechanism)的 TransE 模型^[12],其目标任务为链接预测(link prediction)与三元组分类(triple classification),概括而言就是判断知识图谱中给定的 2 个实体之间是否存在某个关系.因此 TransE 模型仅关注所学习到的嵌入表示对单条三元组结构信息的编码,其在嵌入学习过程中将知识图谱简化为互不相关的三元组的有限集合.因此, TransE 及其后续改进模型^[13-16]对知识图谱中上下文信息的编码能力非常弱,很难应用于语义相关的任务.针对这一问题,相继有一些基于上下文信息的嵌入表示模型被提出,如 GAKE^[17], RDF2Vec^[18].但是它们仍然仅关注知识图谱中由子图、路径等结构所表示的上下文信息.例如,在学习图 1 中实体北京的嵌入表示时,上述方法仅关注(中国,首都,北京)与(北京,位于,华北)等描述实体间关系的三元组,而忽略了北京的简介、英文标签等文本信息.显然,文本信息的缺失限制了所学到嵌入表示对语义信息的表达.

为解决这一问题,本文提出了一种规则引导的知识图谱联合嵌入学习模型.受 Vashishth 等人^[19]所提出的图卷积网络启发,模型首先通过多关系型图卷积将实体在知识图谱中的上下文信息编码到实体的嵌入表示中.与 Vashishth 等人的工作所不同的是,本文认为实体的多条上下文信息应该具有不同

的重要程度,并且某条上下文信息的重要程度取决于2个因素:该条上下文信息的置信度,以及其相对于实体的关联度.为此,本文提出了一条简单有效的规则引导上下文信息置信度的计算,并基于知识图谱中的文本信息表示提出了实体与其上下文信息之间关联度的计算方法.最后,模型将图卷积网络所编码的嵌入表示与文本信息的向量表示整合,以链接预测任务的结果作为训练目标,学习知识图谱中实体与关系的嵌入表示.

本文贡献主要体现在3个方面:

- 1) 基于图卷积网络,创新地提出了一种联合考虑知识图谱中上下文信息与文本信息,由规则引导的嵌入表示学习模型.
- 2) 针对上下文信息在图卷积中的重要程度,提出了应用规则以及知识图谱中文本信息来计算单条上下文信息置信度与关联度的新方法.
- 3) 在基准数据集上进行了充分的实验,并与相关的知识图谱嵌入学习方法进行了对比,实验结果验证了本文模型的有效性.

1 相关工作

本节对与本文工作较相关的知识图谱嵌入学习模型进行介绍,由于本文所提出的模型是基于图神经网络的,因此分别介绍基于图神经网络的知识图谱嵌入学习模型和其他非图神经网络的嵌入学习模型.

1.1 基于图神经网络的模型

基于图神经网络的模型主要包括 R-GCN^[20], W-GCN^[21], CompGCN^[19]等.该类模型通常将图卷积网络作为编码器,对图结构数据进行编码,并结合对应的解码器进行知识图谱上的链接预测、节点分类等任务.在 R-GCN 中,每层网络中节点与关系的特征利用权重矩阵进行计算,并通过领域聚合的方式传递至后续网络层.具体而言,R-GCN 利用基分解和块对角分解构造特定关系的权重矩阵,以处理不同类型的邻居关系,将其与邻居节点信息进行融合,并传递到目标实体上进行更新.W-GCN 在图卷积网络聚合过程中为每个权重矩阵分配可学习的权重参数,使模型获得更优的实体嵌入表示.CompGCN 则提出了针对中心节点的领域信息聚合方法,在理论上使用多种“实体-关系”组合算法对当前主流的基于多关系的图卷积网络模型进行了概括.

1.2 非图神经网络的模型

非图神经网络的嵌入学习模型类别较多,主要包括基于翻译机制的模型,如 TransE^[12]及其后续改进模型,包括 TransH^[13],TransR^[14],TransD^[15],TransAH^[16],基于上下文信息的模型,如 GAKE^[17],RDF2Vec^[18],基于张量分解的模型,如 ComplEx^[22], RESCAL^[23].

其中,基于翻译机制的模型应用较为广泛.该类模型通常仅关注知识图谱的结构信息,将实体之间的关系表示为嵌入向量空间中的某种翻译操作(translation operation).以 TransE 为例,其将知识图谱中的实体与关系都表示在同一个低维欧几里得空间中,以向量表示一个实体或关系.具体而言,对于知识图谱中的一条三元组 (h, r, t) ,TransE 将其中的关系 r 看作在欧几里得空间中从头实体 h 到尾实体 t 的平移操作,即其期望头实体所对应的向量 h 经过关系所对应的向量 r 的平移操作后可以非常逼近尾实体所对应的向量 t ,即 $h + r \approx t$.

TransE 的翻译机制较为简单,因此可以高效地应用于大规模知识图谱,但同时又限制了其模型的表达能力,使其难以处理一对多、多对一以及多对多类型的复杂关系^[14].为解决这一问题,TransE 之后相继有一些翻译机制更加复杂的模型被提出.例如,TransH^[15]相对于所给定三元组中关系的超平面空间设计翻译机制,TransR^[16]则针对知识图谱中的每一个关系额外学习一个矩阵,借助该矩阵将头、尾实体通过线性变换映射到相应的关系向量空间中,然后再计算其翻译机制的损失值.

2 联合嵌入表示学习

本节首先对知识图谱嵌入学习问题进行形式化定义,介绍相关概念的符号表示,然后详细介绍所提出的规则引导的联合嵌入学习模型.

2.1 问题定义

本文将知识图谱表示为 $\mathcal{G}=(\mathcal{E},\mathcal{R})$,其中 \mathcal{E},\mathcal{R} 分别代表知识图谱中的实体与关系集合.对于某个三元组 $(e_h,r,e_t)\in\mathcal{G}$,其中头尾实体均属于实体集合,即 $e_h,e_t\in\mathcal{E}$,其中关系属于关系集合,即 $r\in\mathcal{R}$.知识图谱的嵌入学习问题在于学习给定知识图谱 \mathcal{G} 中任意实体 $e\in\mathcal{E}$ 与任意关系 $r\in\mathcal{R}$ 的向量表示 $e,r\in\mathbb{R}^d$,其中 d 为嵌入表示的维度.本文通过链接预测任务评价所学习到的嵌入表示,该任务可能包括

2种情形:给定实体 $e \in \mathcal{E}$ 与关系 $r \in \mathcal{R}$, 基于它们的嵌入表示 $e, r \in \mathbb{R}^d$, 预测另外一个实体 $e' \in \mathcal{E}$, 使得存在三元组 $(e, r, e') \in \mathcal{G}$ 或 $(e', r, e) \in \mathcal{G}$; 或者给定2个实体 $e, e' \in \mathcal{E}$, 基于它们的嵌入表示 $e, e' \in \mathbb{R}^d$, 预测一个关系 $r \in \mathcal{R}$, 使得存在三元组 $(e, r, e') \in \mathcal{G}$ 或 $(e', r, e) \in \mathcal{G}$.

对于任意实体 $e \in \mathcal{E}$ 与关系 $r \in \mathcal{R}$, 本文将它们所对应的文本信息表示为 l_e 与 l_r . 对于实体 $e \in \mathcal{E}$, 本文将将其所有邻居三元组的集合 $\mathcal{N}(e)$ 视为 e 的上下文, 具体而言 $\mathcal{N}(e)$ 为集合 $\{(e, r, e') | (e, r, e') \in \mathcal{G}, e' \in \mathcal{E}\}$ 与集合 $\{(e', r, e) | (e', r, e) \in \mathcal{G}, e' \in \mathcal{E}\}$ 的并集, 且对于 $\mathcal{N}(e)$ 中任意邻居三元组, 本文认为其表达了节点 e 的一条上下文信息.

与 Vashishth 等人^[19]的做法类似, 本文也对知识图谱的关系集合进行扩充: $\mathcal{R} \leftarrow \mathcal{R} \cup \mathcal{R}_{\text{inverse}} \cup \text{SL}$, 其中 $\mathcal{R}_{\text{inverse}} = \{r^{-1} | r \in \mathcal{R}\}$ 为逆关系集合. 具体而言, 对于任意三元组 $(e_h, r, e_t) \in \mathcal{G}$, 本文在关系集合中增加一条逆关系 r^{-1} , 并相应地将三元组 (e_t, r^{-1}, e_h) 添加到知识图谱 \mathcal{G} 中, 即 $\mathcal{G} \leftarrow \mathcal{G} \cup \{(e_t, r^{-1}, e_h)\}$. SL 代表自环关系集合, 即对于任意实体 $e \in \mathcal{E}$, 在知识图谱 \mathcal{G} 中添加自环三元组, 即 $\mathcal{G} \leftarrow \mathcal{G} \cup \{(e, r_s, e)\}$.

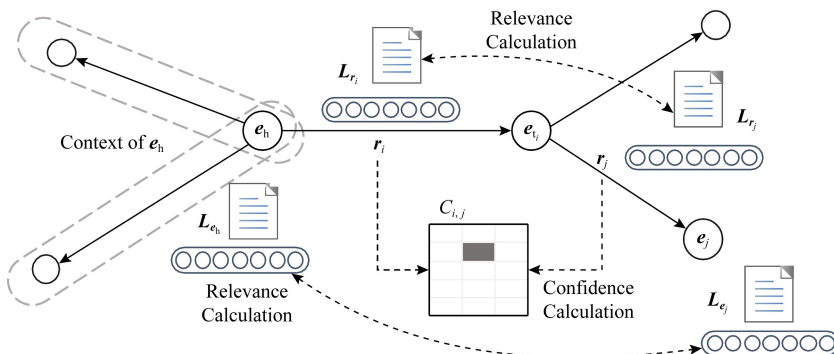


Fig. 2 An overview of the core part of the model

图2 模型核心部分框架图

对于关联度计算, 本文首先利用预训练语言模型对知识图谱中实体与关系的文本信息进行编码. 如图2所示, 对于实体 e_h 与关系 r_i 的文本 l_{e_h} 与 l_{r_i} , 它们的文本向量分别记为 L_{e_h} 与 L_{r_i} . 本文基于实体与关系的文本向量表示计算单条上下文信息与其对应实体之间的关联度, 如图2中标有关联度计算的虚线所示.

值得一提的是, 本文所提出的模型采用“编码器-解码器”框架 (encoder-decoder). 上述基于图卷积网络的上下文信息编码即为编码器的主要内容. 除此之外, 编码器还将上述过程学习到的实体与关系

$e)\}$, $r_s \in \text{SL}$. 此外, 本文使用 $\mathcal{N}_e(e)$ 代表实体 e 周围邻居实体的集合, $\mathcal{N}_r(e)$ 代表实体 e 周围邻居关系的集合, 例如对于图1中的实体北京, 其邻居实体集合为 {华北, 中国, ...}, 邻居关系集合为 {位于, 首都, 简介, 英文标签, ...}.

2.2 模型整体架构

本文基于知识图谱的上下文信息与文本信息联合学习实体与关系的嵌入表示. 图2展示了模型的整体架构, 针对节点 e_h , 其上下文信息 $\mathcal{N}(e_h)$ 由包括三元组 (e_h, r_i, e_t) 在内的所有邻居三元组表达. 本文模型利用图卷积网络, 基于 $\mathcal{N}(e_h)$, 将 e_h 的上下文信息编码到其嵌入表示中. 并且, 本文认为 e_h 的不同邻居三元组所各自表达的上下文信息具有不同的重要程度, 通过计算单条上下文信息的置信度与关联度对其重要程度进行度量.

对于置信度计算, 本文针对上下文信息中所包含的关系提出一条简单有效的规则, 并基于该规则在嵌入学习之前预先计算特定于一对关系的置信度矩阵 C , 并在图卷积过程中利用该矩阵计算某条上下文信息的置信度, 如图2中标有置信度计算的虚线所示.

的嵌入表示与它们的文本表示相结合. 本文模型的解码器则主要基于 ConvE 模型^[24]实现. 下面对模型的细节进行详细的介绍.

2.3 编码器

规则引导的置信度计算. 知识图谱中的关系并非相互独立. 对于一个实体 $e \in \mathcal{E}$, 当 e 拥有一条邻居关系 $r_1 \in \mathcal{N}_r(e)$ 时, 这可能暗示其还同时拥有另一条邻居关系 $r_2 \in \mathcal{N}_r(e)$. 例如, 当某个实体的一条邻居关系为首都时, 其很可能拥有另外一个邻居关系市长. 因为显然只有城市才可能是“首都”, 并且其往往拥有“市长”这一上下文信息. 本文认为这种

关系间的联系可用于对上下文信息的置信度进行估算.因此,本文提出以下规则:对于实体 $e \in \mathcal{E}$,当其某条上下文信息中包含有关系 $r_1 \in \mathcal{N}_r(e)$,并且 $\mathcal{N}_r(e)$ 中同时存在关系 $r_2, r_1 \neq r_2$.此时,对于知识图谱中任意实体 $e' \in \mathcal{E}, e \neq e'$,出现 $r_1, r_2 \in \mathcal{N}_r(e')$ 的概率越高,相对于 e 而言包含 r_1 的上下文信息置信度越高.例如,当某个实体拥有一条包含有关系市长的上下文信息时,如果该实体同时拥有关系首都,那么包含有市长的上下文信息置信度较高,因为首都与市长通常同时出现在实体的上下文中.

基于上述规则,本文在进行图卷积网络的训练前首先计算置信度矩阵 $\mathbf{C} \in \mathbb{R}^{|\mathcal{R}| \times |\mathcal{R}|}$, $|\cdot|$ 表示集合的大小.对于矩阵中的任意参数 $C_{i,j}, 0 \leq i, j \leq |\mathcal{R}| - 1$,其计算如下:

$$C_{i,j} = \frac{|\{e | e \in \mathcal{E}, r_i \in \mathcal{N}_r(e) \wedge r_j \in \mathcal{N}_r(e)\}|}{|\{e | e \in \mathcal{E}, r_i \in \mathcal{N}_r(e)\}|}, \quad (1)$$

其中,分母表示知识图谱中拥有邻居关系 r_i 的实体的个数,分子表示同时拥有邻居关系 r_i 与 r_j 的实体的个数, i 与 j 在此表示关系 r_i 与 r_j 在关系集合 \mathcal{R} 中的索引.

对于图 2 中实体 e_h ,当利用其邻居三元组 $\mathcal{N}(e_h)$,通过图卷积编码其向量表示 e_h 时,模型会首先基于置信度矩阵 \mathbf{C} 评价 e_h 的各个邻居三元组.例如,图 2 中邻居三元组 (e_h, r_i, e_{t_i}) 的置信度可以通过式(2)来计算:

$$\text{conf}_i = \frac{1}{|\mathcal{N}_r(e_{t_i})|} \sum_{r_j \in \mathcal{N}_r(e_{t_i})} C_{id(r_i), id(r_j)}, \quad (2)$$

其中, $\mathcal{N}_r(e_{t_i})$ 代表实体 e_{t_i} 邻居关系集合, $id(\cdot)$ 表示关系在 \mathcal{R} 中的索引.值得一提的是,在本文模型的实现中,置信度计算被整合到了关联度计算中对其进行详细介绍.

基于文本信息的关联度计算.考虑到知识图谱中实体与其不同上下文信息之间关联度的差异,在进行实体嵌入表示的卷积编码时,关联度高的上下文信息应该获得更多的关注.为此,本文借助实体和关系的文本描述作为辅助信息计算实体与其单条上下文信息之间的关联度,用于后续图卷积网络中实体嵌入表示的更新.本文首先将知识图谱中实体和关系的文本描述输入到预训练的 BERT^[25] 语言模型中,得到它们所对应的初始文本表示,然后再分别利用实体文本转换矩阵与关系文本转换矩阵计算它们的最终文本表示.具体而言,对于任意实体 $e \in \mathcal{E}$ 与关系 $r \in \mathcal{R}$,其对应文本分别为 l_e 与 l_r ,它们通过 BERT 得到的初始文本表示为 \mathbf{L}_e^0 与 \mathbf{L}_r^0 ,转换后的

最终文本表示为 \mathbf{L}_e 与 \mathbf{L}_r ,其中 $\mathbf{L}_e = \mathbf{L}_e^0 \cdot \mathbf{W}_{l_e}, \mathbf{L}_r = \mathbf{L}_r^0 \cdot \mathbf{W}_{l_r}$.

本文利用实体与关系的文本表示计算对于某一实体而言,其单条上下文信息的关联度.如图 2 所示,实体 e_h 的一条邻居三元组为 (e_h, r_i, e_{t_i}) ,本文计算参数 β_i 与 γ_i 来度量该条邻居三元组所表示的上下文信息与 e_h 之间的关联度,具体公式为:

$$\beta_i = \frac{1}{|\mathcal{N}_r(e_{t_i})|} \sum_{e_j \in \mathcal{N}_r(e_{t_i})} \mathbf{L}_{e_j}^\top \cdot \mathbf{L}_{e_h}, \quad (3)$$

其中, \mathbf{L}_{e_j} 与 \mathbf{L}_{e_h} 分别为实体 e_j 与 e_h 的文本表示, $\mathcal{N}_r(e_{t_i})$ 为实体 e_{t_i} 的邻居实体集合.

$$\gamma_i = \frac{1}{|\mathcal{N}_r(e_{t_i})|} \sum_{r_j \in \mathcal{N}_r(e_{t_i})} \mathbf{L}_{r_j}^\top \cdot \mathbf{L}_{r_i} \cdot C_{id(r_i), id(r_j)}, \quad (4)$$

其中, \mathbf{L}_{r_j} 与 \mathbf{L}_{r_i} 分别为关系 r_j 与 r_i 的文本表示, $\mathcal{N}_r(e_{t_i})$ 为实体 e_{t_i} 的邻居关系集合, $C_{id(r_i), id(r_j)}$ 如式(2)所定义.

基于图卷积网络的嵌入更新.本文采用 Vashishth 等人所提出的 CompGCN^[19] 模型作为图卷积网络的架构,对知识图谱上下文信息进行编码.

初始状态下,对于任意实体 $e \in \mathcal{E}$ 与关系 $r \in \mathcal{R}$,其嵌入表示为随机向量 \mathbf{e}^0 与 \mathbf{r}^0 ,且 $\mathbf{e}^0, \mathbf{r}^0 \in \mathbb{R}^d$.以图 2 中的实体 e_h 为例,其嵌入表示通过以下图卷积过程进行更新:

$$\mathbf{e}_h = \tanh\left(\sum_{(e_h, r_i, e_{t_i}) \in \mathcal{N}(e_h)} \alpha_i \cdot \mathbf{W}_{t(r_i)} \cdot f(\mathbf{r}_i^0, \mathbf{e}_{t_i}^0)\right), \quad (5)$$

其中, $\mathcal{N}(e_h)$ 表示实体 e_h 的邻居三元组集合, $(e_h, r_i, e_{t_i}) \in \mathcal{N}(e_h)$ 或 $(e_{t_i}, r_i, e_h) \in \mathcal{N}(e_h)$;针对包含有邻居关系 r_i 与邻居节点 e_{t_i} 的一条上下文信息, α_i 基于置信度计算与关联度计算度量该条上下文信息的重要程度,具体如下:

$$\alpha_i = \lambda_1 \beta_i + \lambda_2 \gamma_i, \quad (6)$$

其中, β_i 由式(3)计算得到, γ_i 则由式(4)计算得到; $\mathbf{W}_{t(r_i)}$ 为 CompGCN^[19] 中定义的关系类别矩阵,由于知识图谱中被加入了逆关系与自环关系, $\mathbf{W}_{t(r_i)} \in \mathbb{R}^{d' \times d}$ 可能为 3 种不同的表示,具体为:

$$\mathbf{W}_{t(r_i)} = \begin{cases} \mathbf{W}_0, & r_i \in \mathcal{R}, \\ \mathbf{W}_1, & r_i \in \mathcal{R}_{\text{inverse}}, \\ \mathbf{W}_s, & r_i \in \mathcal{SL}, \end{cases} \quad (7)$$

其中函数 $f(\cdot)$ 表示循环关联操作(circular correlation)^[26],可以将 2 个向量 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ 进行融合,得到 $\mathbf{x} \circ \mathbf{y} \in \mathbb{R}^d$,每个维度的数值计算为

$$[\mathbf{x} \circ \mathbf{y}]_j = \sum_{l=0}^{d-1} [\mathbf{x}]_l [\mathbf{y}]_{(l+1) \bmod d}. \quad (8)$$

遵循 CompGCN 框架,在对实体进行图卷积编码的同时,本文通过转换矩阵 $\mathbf{W}_r \in \mathbb{R}^{d' \times d}$ 更新关系的嵌入表示:

$$\mathbf{r}_i = \mathbf{W}_r \mathbf{r}_i^0. \quad (9)$$

最后,本文将任意实体 $e \in \mathcal{G}$ 与关系 $r \in \mathcal{R}$ 的文本表示加入到其嵌入表示中,具体为:

$$\mathbf{e} = \mathbf{e} + \mathbf{L}_e, \quad (10)$$

$$\mathbf{r} = \mathbf{r} + \mathbf{L}_r. \quad (11)$$

2.4 解码器

本文采用 ConvE^[24] 模型作为解码器,基于所学习到的嵌入表示进行链接预测,通过提高链接预测的表现更新模型参数.当知识图谱经过编码器编码后,对于某个任意构成的三元组 (e_h, r, e_t) ,可知其头尾实体 e_h 与 e_t 的嵌入表示为 \mathbf{e}_h 与 \mathbf{e}_t ,关系 r 的嵌入表示为 \mathbf{r} .ConvE 模型首先将 \mathbf{e}_h 与 \mathbf{r} 转换成二维形式,即 \mathbf{e}_h 与 \mathbf{r} ,然后计算该三元组的分数值:

$$sc(\mathbf{e}_h, \mathbf{r}, \mathbf{e}_t) = f'(\text{vec}(f'([\mathbf{e}_h; \mathbf{r}] * \omega))\mathbf{W}_{\text{cov}})\mathbf{e}_t, \quad (12)$$

其中, $[\cdot]$ 表示相连接, ω 表示卷积过滤器, $\text{vec}(\cdot)$ 为 ConvE 所定义的维度变换, \mathbf{W}_{cov} 为参数矩阵, $f'(\cdot)$ 为非线性函数.当式(12)计算得到的分数值越高, (e_h, r, e_t) 越有可能是正确的三元组.

3 实验

本节首先对实验所使用的数据集、对比模型和评价指标等进行说明,然后介绍本文所提模型的实验结果,并与其他基准模型进行比较与分析.

3.1 数据集及对比模型介绍

本文在 2 个广泛使用的数据集上进行试验,分别是 FB15K-237^[27] 和 WN18^[12],其统计数据如表 1 所示:

Table 1 Summary Statistics of Knowledge Graphs
表 1 数据集的统计信息

Dataset	FB15K-237	WN18
# Relation	237	18
# Entity	14 541	40 943
# Train	271 115	141 442
# Valid	17 535	2 500
# Test	20 466	2 500

为验证所提模型的有效性,本文广泛选取了当前被应用较多的知识图谱嵌入学习模型作为对比方法,具体包括 TransE^[11], DistMult^[28], ComplEx^[22],

R-GCN^[20], KBGAN^[29], ConvE^[24], ConvKB^[30], SACN^[21], Hyper^[31], RotatE^[32], ConvR^[33], VR-GCN^[34], CompGCN^[19].其中, TransE^[11] 为基于翻译机制的嵌入学习模型,上文已对其进行了详细介绍. DistMult^[28] 将实体表示为通过神经网络学习到的低维向量,将关系表示为双线性或线性映射函数. ComplEx^[22] 与 RESCAL^[23] 模型类似,属于基于矩阵/张量分解进行链接预测的模型. R-GCN^[20], VR-GCN^[34] 与 CompGCN^[19] 属于基于图卷积网络的嵌入表示模型,以 R-GCN^[20] 为例,其将知识图谱中的关系编码为矩阵,通过关系矩阵传递相邻实体的嵌入信息,并采用了多层图卷积网络. KBGAN 则应用了对抗生成网络 (generative adversarial network, GAN),在训练过程中生成更具迷惑性的负例来提高嵌入表示的训练效果.本文应用了 ConvE^[24] 模型作为解码器,在第 2 节中对其进行了详细介绍. ConvKB^[30], ConvR^[33], SACN^[21] 与 Hyper^[31] 均是基于卷积神经网络的方法.以 Hyper^[31] 为例,其可以生成简化的与关系相关的卷积过滤器,且可被构造为张量分解. RotatE^[32] 与 TransE^[11] 等基于翻译机制的模型类似,其将实体之间的关系表示为向量空间中从头实体到尾实体的旋转.

3.2 评价方法说明

本文通过链接预测任务来对模型的有效性进行评价.在实验中,针对被事先去掉头实体或尾实体的测试三元组,本文基于学习到的嵌入表示推测其被去掉的头实体或尾实体.具体对于每个测试三元组,本文选取知识图谱中的任意实体作为可能的预测结果,并计算利用该实体补全测试三元组后的分数值,如式(12)所示,最后对分数值进行排序.在此以缺失头实体的预测为例,对于测试集中每个三元组 (e_h, r, e_t) ,事先删除其头实体 e_h ,然后试图使用 \mathcal{G} 中的任意实体 $e_{hc} \in \mathcal{E}$ 补齐该测试三元组,从而产生候选三元组集合 $\{(e_{hc}, r, e_t) | e_{hc} \in \mathcal{E}\}$.最后,基于所学习到的嵌入表示计算候选三元组的分数值并进行排序,分数值越高表明学习到的模型,即嵌入表示,认为该结果更可靠,通过与真实结果进行比较从而判断所学习嵌入表示的优劣.

最后采用 MR (mean rank), MRR (mean reciprocal rank) 和 Hit@ k 作为评价指标^[12].其中, MR 与 MRR 均为预测结果平均排名的指标, Hit@ k 则指预测结果排在前 k 名中的比例,本文具体采用 Hit@10, Hit@3 和 Hit@1.总之,越好的预测结果,其 MR 值越低, MRR 值越高, Hit@ k 也越高.

3.3 实验设置

本文实验代码使用 Python 实现,在配置 Ubuntu 16.04.6 LTS 操作系统的服务器上完成,其 CPU 配置为 16 核 Intel Core i7-6900K 3.20 GHz, 内存 128 GB, GPU 配置为 4 张 GeForce GTX 1080 GPU 卡.

对于实体和关系文本表示向量的编码,本文借助 pretrained-bert-base-uncased 预训练模型^①,文本向量初始维度为 768,转换后的维度为 200.在图卷积网络中,实体和关系的初始化向量维度为 100,即 $d=100$,GCN 的维度为 200,即 $d'=200$.解码器中维度转换的高度和宽度分别为 10 和 20,卷积过滤器的大小为 7×7 ,数量为 200.利用 Adam 优化器对整体模型进行训练,批大小(batch size)为 256,学习率(learning rate)为 0.001.

本文对 TransE 模型进行了复现,其余模型则引用对比模型论文中所报告的结果.

3.4 实验结果分析

表 2 报告了本文模型与对比模型在链接预测任务中的实验结果.

通过表 2 可观察到如下结果:

1) 本文模型在各个评价指标上显著优于 TransE, DistMult 和 ComplEx 等基准模型,与 SACN, HyperER 和 CompGCN 等最新提出的模型十分接近,由此可证明本文模型的有效性.对于 FB15K-237 数据集,本文在 Hit@10 指标上排名第一.

2) 在 Hit@1 和 Hit@3 指标上也与 CompGCN, ConvR, SACN 相差极小.具体在 Hit@1 指标上仅比最高的 CompGCN 低 1.51%,在 MRR 指标上与 CompGCN 相比仅低 0.8%.而对于 WN18 数据集,本文模型在 MR 指标上排名第一,在 Hit@10 和 Hit@3 指标上也与第一名差距微小.具体在 Hit@10 指标上比 RotatE 低 0.2%,在 Hit@3 指标上比 ConvR 和 HyperER 仅低 0.9%.

3) 基于图神经网络的嵌入学习方法的表现普遍优于 TransE 等仅关注结构化信息的模型.就本文模型而言,由于其基于图卷积网络对知识图谱的上下文信息与文本信息进行了联合嵌入表示,显著提高了在链接预测任务中的表现.

Table 2 Link Prediction Results on FB15K-237 and WN18
表 2 链接预测在 FB15K-237 和 WN18 上的结果

Model	FB15K-237					WN18				
	MRR	MR	Hit@10	Hit@3	Hit@1	MRR	MR	Hit@10	Hit@3	Hit@1
TransE ^[12]	0.294	357	0.330	0.222	0.146	0.454	251	0.891	0.803	0.064
DistMul ^[28]	0.241	254	0.419	0.263	0.155	0.829		0.94	0.923	0.726
ComplEx ^[22]	0.247	339	0.428	0.275	0.158		844	0.94		
R-GCN ^[20]	0.248		0.417		0.151	0.773		0.944	0.889	0.65
KBGAN ^[29]	0.278		0.458			0.779		0.949		
ConvE ^[24]	0.325	244	0.501	0.356	0.237	0.943	374	0.956	0.946	0.935
ConvKB ^[30]	0.243	311	0.421	0.371	0.155					
SACN ^[21]	0.350		0.540	0.390	0.260					
HyperER ^[31]	0.341	250	0.520	0.376	0.252	0.951	431	0.958	0.955	0.947
RotatE ^[32]	0.338	177	0.533	0.375	0.241		309	0.959		
ConvR ^[33]	0.350		0.528	0.385	0.261	0.951		0.958	0.955	0.947
VR-GCN ^[34]	0.248		0.432	0.272	0.159	0.847		0.946	0.929	0.764
CompGCN ^[19]	0.355	197	0.535	0.390	0.264					
Our Methods	0.352	185.72	0.536	0.385	0.260	0.928	240	0.957	0.946	0.909

Note: The best performance is in bold.

4 总 结

现有多数知识图谱嵌入学习方法仅考虑由三元

组表示的知识图谱结构信息,而忽视了知识图谱中丰富的上下文信息与文本信息,限制了嵌入表示在链接预测等任务中的表现.针对现有方法的这一局限性,本文提出一种利用图卷积神经网络,结合知识

① <https://github.com/google-research/bert>

图谱的上下文信息与文本信息学习嵌入表示的方法.为了对上下文信息的重要程度进行细粒度分析,本文提出一条简单有效的规则来计算上下文信息的置信度,并基于文本信息的向量表示提出计算上下文信息关联度的方法,加强了对上下文信息的约束和引导.最后,通过在2个广泛使用的基准数据集上进行对比实验,验证了本文模型的有效性.

参 考 文 献

- [1] Hu Sen, Zou Lei, Yu J X, et al. Answering natural language questions by subgraph matching over knowledge graphs [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 30(5): 824-837
- [2] Palumbo E, Rizzo G, Troncy R. Entity2rec: Learning user-item relatedness from knowledge graphs for top- n item recommendation [C] //Proc of the 7th ACM Conf on Recommender Systems. New York: ACM, 2017: 32-36
- [3] Corcoglioniti F, Dragoni M, Rospocher M, et al. Knowledge extraction for information retrieval [C] //Proc of the European Semantic Web Conf. Berlin: Springer, 2016: 317-333
- [4] Zou Xiaohan. A survey on application of knowledge graph [J]. *JPhCS*, 2020, 1487(1): No.012016
- [5] Bollacker K, Evans C, Paritosh P, et al. Freebase: A collaboratively created graph database for structuring human knowledge [C] //Proc of the 2008 ACM SIGMOD Int Conf on Management of Data. New York: ACM, 2008: 1247-1250
- [6] Suchanek F M, Kasneci G, Weikum G. Yago: A core of semantic knowledge [C] //Proc of the 16th Int Conf on World Wide Web. New York: ACM, 2007: 697-706
- [7] Auer S, Bizer C, Kobilarov G, et al. Dbpedia: A nucleus for a Web of open data [M] //Proc of the Semantic Web. Berlin: Springer, 2007
- [8] Liu Zhiyuan, Sun Maosong, Lin Yankai, et al. Knowledge representation learning: A review [J]. *Journal of Computer Research and Development*, 2016, 53(2): 247-261 (in Chinese)
(刘知远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展[J]. *计算机研究与发展*, 2016, 53(2): 247-261)
- [9] Wang Ruijie, Wang Meng, Liu Jun, et al. Leveraging knowledge graph embeddings for natural language question answering [C] //Proc of Int Conf on Database Systems for Advanced Applications. Berlin: Springer, 2019: 659-675
- [10] Hamilton W, Bajaj P, Zitnik M, et al. Embedding logical queries on knowledge graphs [C] //Proc of Advances in Neural Information Processing Systems (NeurIPS 2018). Cambridge, Massachusetts: MIT Press, 2018: 2026-2037
- [11] Wang Meng, Wang Ruijie, Liu Jun, et al. Towards empty answers in SPARQL: Approximating querying with RDF embedding [C] //Proc of Int Semantic Web Conf. Berlin: Springer, 2018: 513-529
- [12] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data [C] //Proc of Advances in Neural Information Processing Systems (NIPS 2013). Cambridge, Massachusetts: MIT Press, 2013: 2787-2795
- [13] Wang Zhen, Zhang Jianwen, Feng Jianlin, et al. Knowledge graph embedding by translating on hyperplanes [C] //Proc of the 28th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2014, 14(2014): 1112-1119
- [14] Lin Yankai, Liu, Zhiyuan, Sun Maosong, et al. Learning entity and relation embeddings for knowledge graph completion [C] //Proc of the 29th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2015: 2181-2187
- [15] Ji Guoliang, He Shizhu, Xu Liheng, et al. Knowledge graph embedding via dynamic mapping matrix [C] //Proc of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th Int Joint Conf on Natural Language Processing (volume 1: Long papers). Stroudsburg, PA: ACL, 2015: 687-696
- [16] 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)
- [17] Feng Jun, Huang Minlie, Yang Yang, et al. GAKE: Graph aware knowledge embedding [C] //Proc of the 26th Int Conf on Computational Linguistics (COLING 2016). Stroudsburg, PA: ACL, 2016: 641-651
- [18] Ristoski P, Paulheim H. RDF2Vec: RDF graph embeddings for data mining [C] //Proc of the Int Semantic Web Conf. Berlin: Springer, 2016: 498-514
- [19] Vashishth S, Sanyal S, Nitin V, et al. Composition-based multi-relational graph convolutional networks [C/OL] //Proc of the Int Conf on Learning Representations. 2020: 1-15 [2020-09-10]. https://www.researchgate.net/publication/337158026_Composition-based_Multi-Relational_Graph_Convolutional_Networks
- [20] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks [C] //Proc of the 15th Int Conf on Extended Semantic Web Conf (ESWC 2018). Berlin: Springer, 2018: 593-607
- [21] Shang Chao, Tang Yun, Huang Jing, et al. End-to-end structure-aware convolutional networks for knowledge base completion [C] //Proc of the 33rd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2019: 3060-3067

[22] Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction [C] //Proc of the 33rd Int Conf on Machine Learning (ICML). New York; ACM, 2016: 2071-2080

[23] Nickel M, Tresc V, Krieger H P. A three-way model for collective learning on multi-relational data [C] //Proc of the 28th Int Conf on Machine Learning. Madison WI, USA: Omni Press, 2011: 809-816

[24] Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2D knowledge graph embeddings [C] //Proc of the 32nd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI Press, 2018: 1811-1818

[25] Devlin J, Chang Mingwei, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [C] //Proc of 2019 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 4147-4186

[26] Plate T A. Holographic reduced representations [J]. IEEE Transactions on Neural Networks, 1995, 6(3): 623-641

[27] Toutanova K, Chen Danqi. Observed versus latent features for knowledge base and text inference [C] //Proc of the 3rd Workshop on Continuous Vector Space Models and Their Compositionality. Stroudsburg, PA: ACL, 2015: 57-66

[28] Yang Bishan, Yih W T, He Xiaodong, et al. Embedding entities and relations for learning and inference in knowledge bases [C/OL] //Proc of the 3rd Int Conf on Learning Representations (ICLR 2015). 2015: 69-74 [2020-09-10]. https://www.researchgate.net/publication/269935407_Embedding_Entities_and_Relations_for_Learning_and_Inference_in_Knowledge_Bases

[29] Cai Liwei, Wang W Y. KBGAN: Adversarial learning for knowledge graph embeddings [C] //Proc of the 2018 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2018: 1470-1480

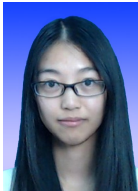
[30] Nguyen D Q, Nguyen T D, Nguyen D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network [C] //Proc of the 2018 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2018: 327-333

[31] Balažević I, Allen C, Hospedales T M. Hypernetwork knowledge graph embeddings [C] //Proc of the Int Conf on Artificial Neural Networks. Berlin: Springer, 2019: 553-565

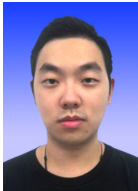
[32] Sun Zhiqing, Deng Zhihong, Nie Jianyun, et al. Rotate: Knowledge graph embedding by relational rotation in complex space [C/OL] //Proc of the 7th Int Conf on Learning Representations (ICLR 2019). 2019: 1-18 [2020-09-10]. https://www.researchgate.net/publication/331397037_RotatE_Knowledge_Graph_Embedding_by_Relational_Rotation_in_Complex_Space

[33] Jiang Xiaotian, Wang Quan, Wang Bin. Adaptive convolution for multi-relational learning [C] //Proc of the 2019 Conf of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA: ACL, 2019: 978-987

[34] Ye Rui, Li Xin, Fang Yujie, et al. A vectorized relational graph convolutional network for multi-relational network alignment [C] //Proc of the 28th Int Joint Conf on Artificial Intelligence. San Mateo, CA: Morgan Kaufmann, 2019: 4135-4141



Yao Siyu, born in 1992. PhD candidate. Student member of CCF. Her main research interests include knowledge graph embedding, and question answering.



Zhao Tianzhe, born in 1997. MSc. Student member of CCF. His main research interests include knowledge graph embedding, and pre-trained language models in course area.



Wang Ruijie, born in 1995. MSc. Student member of CCF. His main research interests include semantic Web, machine learning, and multimodal data.



Liu Jun, born in 1973. PhD, professor. IEEE senior member, senior member of CCF. His main research interests include text mining, data mining, and e-learning.