计算机科学
COMPUTER SCIENCE

基于知识增强的企业实体关系预测模型

王家祺, 李文根, 关佶红, 邢婷, 魏小敏, 邵冰清, 付宠洁

引用本文

王家祺, 李文根, 关佶红, 邢婷, 魏小敏, 邵冰清, 付宠洁. 基于知识增强的企业实体关系预测模型[J]. 计算机科学, 2023, 50(10): 146-155.

WANG Jiaqi, LI Wengen, GUAN Jihong, XING Ting, WEI Xiaomin, SHAO Bingqing, FU Chongjie. Knowledge Enhanced Relationship Prediction Model for Enterprise Entities[J]. Computer Science, 2023, 50(10): 146-155.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221000063

基于知识增强的企业实体关系预测模型

王家祺[1] 李文根[1] 关佶红[1] 邢 婷[2] 魏小敏[2] 邵冰清[2] 付宠洁[2]
1 同济大学电子与信息工程学院 上海 201804
2 北京上奇数字科技有限公司 北京 100084
(wangjq@tongji.edu.cn)

摘 要 随着知识图谱的不断发展,大量应用于工业界的产业知识图谱应运而生。然而,这些产业知识图谱经常缺乏充足的企业关联关系,如上下游关系、供应关系、合作关系、竞争关系等,导致其应用范围受到极大限制。现有企业关系预测研究大多仅关注知识图谱中三元组本身的结构信息,未能充分利用企业文本描述和企业关联实体的描述等多视角信息。为解决该问题,提出了一种基于知识增强的企业实体关系预测模型KERP。模型首先通过多视角实体特征三元组学习,完善企业实体特征表示;其次,利用图注意力网络获取实体的高阶语义表示,并与TransR模型学习的实体关系低阶语义表示进行融合,进一步增强企业实体及其关系的特征表示;最后,通过二维卷积解码器ConvE实现对企业实体关系的预测。在新能源汽车产业知识图谱数据上的实验分析表明,与现有主流实体关系预测模型相比,KERP在预测企业关系上具有更好的效果,在F1值上有6.7%的提升。此外,在多个公开实体关系预测数据集上的实验结果表明,KERP模型在一般化的实体关系预测任务上也具有较好的通用性。

关键词:产业知识图谱;企业实体关系;知识补全;链路预测;知识增强
中图法分类号 TP391

Knowledge Enhanced Relationship Prediction Model for Enterprise Entities

WANG Jiaqi[1],LI Wengen[1],GUAN Jihong[1],XING Ting[2],WEI Xiaomin[2],SHAO Bingqing[2] and FU Chongjie[2]
1 College of Electronic and Information Engineering,Tongji University,Shanghai 201804,China
2 Beijing Shangqi Digital Technology Co.,Ltd.,Beijing 100084,China

Abstract With the development of knowledge graphs,a variety of industrial knowledge graphs have come into being. However,these industrial knowledge graphs lack sufficient relationships among enterprises,such as up-down stream relationship,supply relationship,cooperation and competition relationship,which greatly affects their applications. Most existing methods for predicting the enterprise entity relationships focus on the fact triples and cannot fully utilize multiple perspectives such as enterprise descriptions and associated entity descriptions. To solve this problem,KERP,a knowledge enhanced relationship prediction model for enterprise entities is proposed. The model first improves enterprise features representations using a multi-view entity feature learning module,then uses graph attention network to obtain higher-order semantic representations of entities and fuses lower-order semantic representations learned by TransR for knowledge enhancement,and finally predicts enterprise entity relationships by a convolutional decoder ConvE. Experimental results on the new energy automobile industrial knowledge graph show that KERP has better results in predicting the relationships between enterprises with a improvement of 6.7% in terms of F1 value compared with the existing models. Generalization is also evaluated on multiple datasets,and the experimental results demonstrate that KERP has good generality for generalized entity relationship prediction tasks.

Keywords Industrial knowledge graph,Enterprise entity relationship,Knowledge completion,Link prediction,Knowledge enhancement

1 引言

近年来,知识图谱已经被广泛应用于各个行业领域[1],诸如农业知识图谱、中医药知识图谱、阿里金融知识图谱[2]等,能够为用户提供智能化的搜索和决策支持服务。然而,在实际应用中,由于缺少数据支撑或知识提取不充分,大量

到稿日期:2022-10-10 返修日期:2023-02-20
基金项目:国家自然科学基金(U1936205,62202336);上海"科技创新行动计划"软科学研究项目(22692194100)
This work was supported by the National Natural Science Foundation of China(U1936205,62202336) and Shanghai Soft Science Research Program(22692194100).
通信作者:关佶红(jhguan@tongji.edu.cn)

实体关系缺失，导致知识图谱的完整性不足[1]。因此，亟需提出高效的实体关系预测方法，通过发掘实体之间的未知关系来完善知识图谱，提高图谱的完整性[3]。特别地，面向产业知识图谱的企业实体关系预测可以完善产业供应链合作体系，有利于企业之间建立潜在的合作关系，并能及时应对供应链突发的断链等情况，提高整个产业供应链体系的稳健性。

现有企业实体关系预测方法主要有基于复杂网络的链路预测方法和基于知识补全的链路预测方法。基于复杂网络的链路预测方法通常根据供应链网络拓扑结构进行实体未知关系的预测。例如，Wang 等[4]利用供应链网络中节点间现有的合作伙伴关系预测“未知合作”关系和“未来合作”关系；Lu 等[5]提出基于投影和时间事件的链路预测模型，并将其用于预测动态供应链网络中的企业合作关系。基于知识补全的链路预测方法通常对实体关系进行向量嵌入，定义实体关系三元组评分函数，得到实体关系三元组的置信度，进而判定实体之间是否存在某种关系。例如，翻译模型 TransR[6]将实体空间映射到关系向量空间，定义评分函数为实体和关系向量的距离函数，距离越近表示实体之间具有某种关系的可能性越大。二维卷积模型 ConvE[7]将实体和关系向量进行二维混合后，使用卷积运算评分函数预测实体关系。

基于复杂网络的链路预测方法能够对企业关系网络结构进行刻画，但涉及的实体类型较为单一，大多是企业供应商、经销商等，并且缺乏对实体文本描述语义信息的嵌入[8]。产业知识图谱中不仅包含许多头实体-关系-尾实体(head，relation，tail)形式的事实三元组，还有大量实体文本描述信息。图 1 展示了产业知识图谱中实体关系三元组以及实体的文本描述信息，其中椭圆代表实体节点，方框代表实体文本描述信息，实线代表已知关系信息，虚线代表需要预测的关系。显然，实体文本描述包含了对实体的大量语义信息，而且能够在实体关联关系较少的情况下提供更多的实体特征信息。基于知识图谱的链路预测方法虽然能够融入实体文本描述信息，但是只能使用实体自身的文本描述信息，无法利用关联实体的文本描述信息来增强自身特征表示。如图 1 所示，现有方法无法有效利用与企业关联的专利实体的文本描述信息来增强企业实体表示。

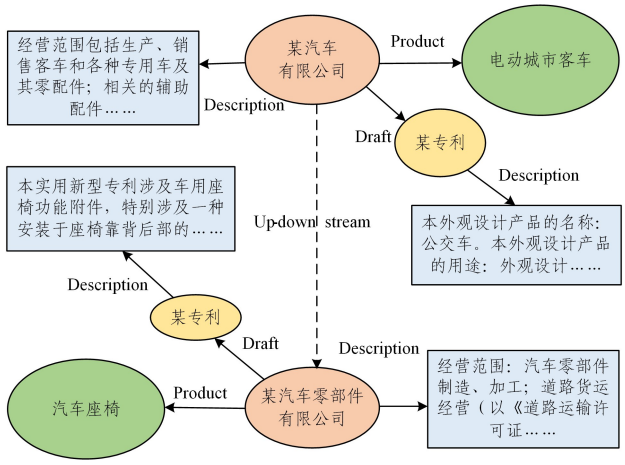

图 1 企业上下游关系的预测

Fig. 1 Prediction of enterprise up-down stream relationship

针对上述问题，提出一种基于知识增强的企业实体关系预测模型 KERP(Knowledge Enhanced Relationship Prediction Model)。首先，KERP 通过多视角实体特征三元组学习，从企业实体描述信息和它关联的专利实体描述信息中得到企业的环节类别特征、技术特征等三元组特征表示；其次，通过图注意力网络学习知识图谱中实体的高阶语义信息，并与 TransR 模型学习的实体关系低阶语义信息进行融合，增强企业实体及其关系的特征表示；最后，将融合后的实体和关系嵌入表示输入解码器 ConvE 计算事实三元组的置信度，并通过该置信度判断企业间是否存在特定关系。

本文主要贡献如下：

1)提出基于知识增强的企业实体关系预测模型 KERP，该模型能够准确预测产业知识图谱中企业之间的各类关系。

2)提出多视角实体三元组特征学习方法，其不仅可以对实体本身的文本描述信息进行特征学习，而且能够利用关联实体的文本描述信息完善自身特征表示，并将学习的特征信息转化为实体的特征三元组，提高实体文本描述信息的利用率。

3)提出融合低阶语义信息的图注意力网络知识补全方法，通过图注意力网络有效捕捉实体和多跳邻居节点之间的拓扑结构，获取实体的高阶语义表示，并通过融入 TransR 模型学习的实体关系低阶语义表示增强模型的可靠性和稳定性。

4)在真实产业知识图谱上对 KERP 模型进行了充分的实验验证。实验结果表明，与现有实体关系预测模型相比，KERP 在预测企业实体关系的准确性上有显著提升。同时，通过在多个公开数据集进行实验，证明模型具有很好的通用性。

2 相关工作

2.1 基于复杂网络的实体关系预测

此类方法主要通过供应商、销售商等不同企业角色预测企业之间的关系。Wang 等[4]利用供应链网络中节点间现有的合作伙伴关系预测“未知合作”关系和“未来合作”关系。Lu 等[5]提出一种基于投影和时间事件的链路预测模型，用于预测动态供应链网络中的企业合作关系。Lu 等[9]根据供应链网络结构发现供应链网络中的潜在合作伙伴。Xie 等[10]将

供应链网络结构表示为拉普拉斯矩阵，通过张量分解的方法为汽车制造商推荐优质零部件供应商。

基于复杂网络的链路预测方法根据供应链的网络结构进行实体关系预测，涉及的实体类型较为单一，无法有效利用实体的文本描述信息，预测准确性通常较差。

2.2 基于知识补全的实体关系预测

基于知识补全的实体关系预测方法通过计算实体关系三元组的置信度来判断头尾实体是否具有特定关系。这类方法可以进一步分为传统知识补全方法和基于知识表示学习的补全方法。

传统知识补全方法针对不同网络结构的知识图谱设计不同的逻辑规则和概率似然函数来推断实体之间的关系，具有较强的可解释性，在规模较小的数据集上取得了较好的效果。典型的传统知识补全方法有概率图模型和图计算模型。概率图模型方法常用贝叶斯网。Kersting 等[11]提出的贝叶斯逻辑程序模型将贝叶斯网和正定子句逻辑结合，实现了对实体之间关系的预测。在图计算模型方面，Lao 等[12]提出了路径排序算法(Path Ranking Algorithm，PRA)，其使用随机游走获取实体之间的路径，并使用获得的路径特征训练实体关系预测模型。

知识表示学习的本质是对知识图谱中的实体和关系进行嵌入编码[13]，得到实体和关系的低维向量表示。知识表示学习保留了知识图谱的结构，在大规模知识图谱补全任务上表现优异，并且具有良好的扩展性，提高了知识推理、知识补全、知识问答等应用的准确性。知识表示学习方法又可分为翻译模型、语义匹配模型、神经网络模型和多源信息融合模型。根据词向量平移不变性，Bordes 等提出了知识补全中的经典翻译模型 TransE[14]，其使用实体之间的距离评分函数计算三元组置信度。之后，基于 TransE 的改进模型不断出现，如 TransR[6]和 TransD[15]。语义匹配模型主要基于实体的潜在语义和关系向量空间表示对三元组的可信度进行打分，代表方法有 RESCAL 模型[16]和 ComplEx 模型[17]。早期，基于神经网络模型的知识表示学习方法主要采用卷积神经网络(Convolutional Neural Network，CNN)[18]。例如，ConvE 和 ConvKB[19]模型均采用卷积计算实体和关系向量之间的相互作用，从而得到三元组的置信度。目前，大多主流知识表示学习方法采用图神经网络(Graph Neural Networks)进行知识补全。Schlichtkrull 等[20]提出能够处理多关系特征的图卷积网络(Relational Graph Convolutional Networks，R-GCN)。Shang 等[21]结合图卷积网络和 ConvE 模型，提出端到端的结构感知卷积网络(Structure-Aware Convolutional Networks，SACN)。KB-GAT[22]模型将注意力集中在识别实体邻域中的重要信息上，在知识补全任务中实现了 SOTA 性能。多源信息融合模型旨在使用实体的文本描述语义进行嵌入，从而增强实体的语义特征，如 SSP[23]，DKRL[24]和 AKRL[25]。Lin 等[26]提出一种使用卷积融合内部结构特征和外部文本特征的知识图谱关系预测新方法 ConvF。Zhai 等[27]基于 LSTM 文本编码器提出融入实体描述的知识表示模型。但是，这类方法缺乏多视角实体描述信息的融入以及图拓扑结构的表达，用于特定产业知识图谱实体关系预测时效果不佳。

综上所述，基于知识补全的链路预测方法通常基于网络结构和实体自身文本描述信息学习实体特征，视角单一，无法充分利用关联实体的文本描述信息来增强自身特征表示。

3 问题描述与数据集

3.1 问题描述

给定产业知识图谱 $G=(E,R)$，其中 E 和 R 分别表示实体和关系集合。三元组 (e_i,r_k,e_j) 表示 G 中实体 e_i 和 e_j 之间存在关系 r_k。对于企业实体关系预测问题，给定三元组 $t=(e_i,r_k,e_j)$，旨在判断实体 e_i 和 e_j 之间是否存在关系 r_k。

3.2 数据集

本文以新能源汽车企业上下游关系预测为例来研究产业知识图谱中的企业关系预测问题。如图 1 所示，虚线代表需要判断两个企业之间是否存在上下游关系。新能源汽车产业大致可以分为 5 个环节：原材料加工制造、汽车零部件、汽车整车制造、汽车相关服务和汽车相关设施制造。相邻两个环节存在上下游关系。例如，处于原材料加工制造环节的企业 A 生产的产品提供给处于新能源汽车零部件环节的企业 B 使用，那么企业 A 和 B 即存在上下游关系。

本文使用的新能源汽车产业知识图谱(New Energy Automobile Industry Knowledge Graph，NEAI-KG)中存在的实体关系三元组如表 1 所列，包含企业、城市、产品、专利等不同类型实体 27346 个，其中有 3599 个企业实体。关系类型有企业之间的投融资关系、上下游关系，以及企业与专利的起草关系等 8 种关系，共 260376 个实体关系三元组，其中包括 96372 个已知的企业上下游关系三元组。

表 1 新能源汽车产业知识图谱中的实体关系信息

Table 1 Entity relationships in NEAI-KG dataset

Head	Relation	Tail
Enterprise	Up-down stream	Enterprise
Enterprise	Invest	Enterprise
Enterprise	Locate	City
Enterprise	Product	Production
City	Locate	Province
Enterprise	Link	Industry
Enterprise	Draft	Patent

4 KERP 模型

图 2 给出了基于知识增强的企业实体关系预测模型 KERP 的整体架构。KERP 模型由多视角实体特征三元组学习模块和融合低阶语义信息的图注意力网络知识补全模块组成。首先，KERP 模型通过多视角实体特征三元组学习，从企业实体描述信息及其关联的专利实体描述信息中学习企业环节类别特征、技术特征等三元组特征表示；其次，通过图注意力聚合实体多跳邻居节点的信息，学习知识图谱中实体的高阶语义信息，并与 TransR 模型学习的实体关系低阶语义信息进行融合；最后，将融合特征表示输入 ConvE 解码器中，实现企业实体关系预测。

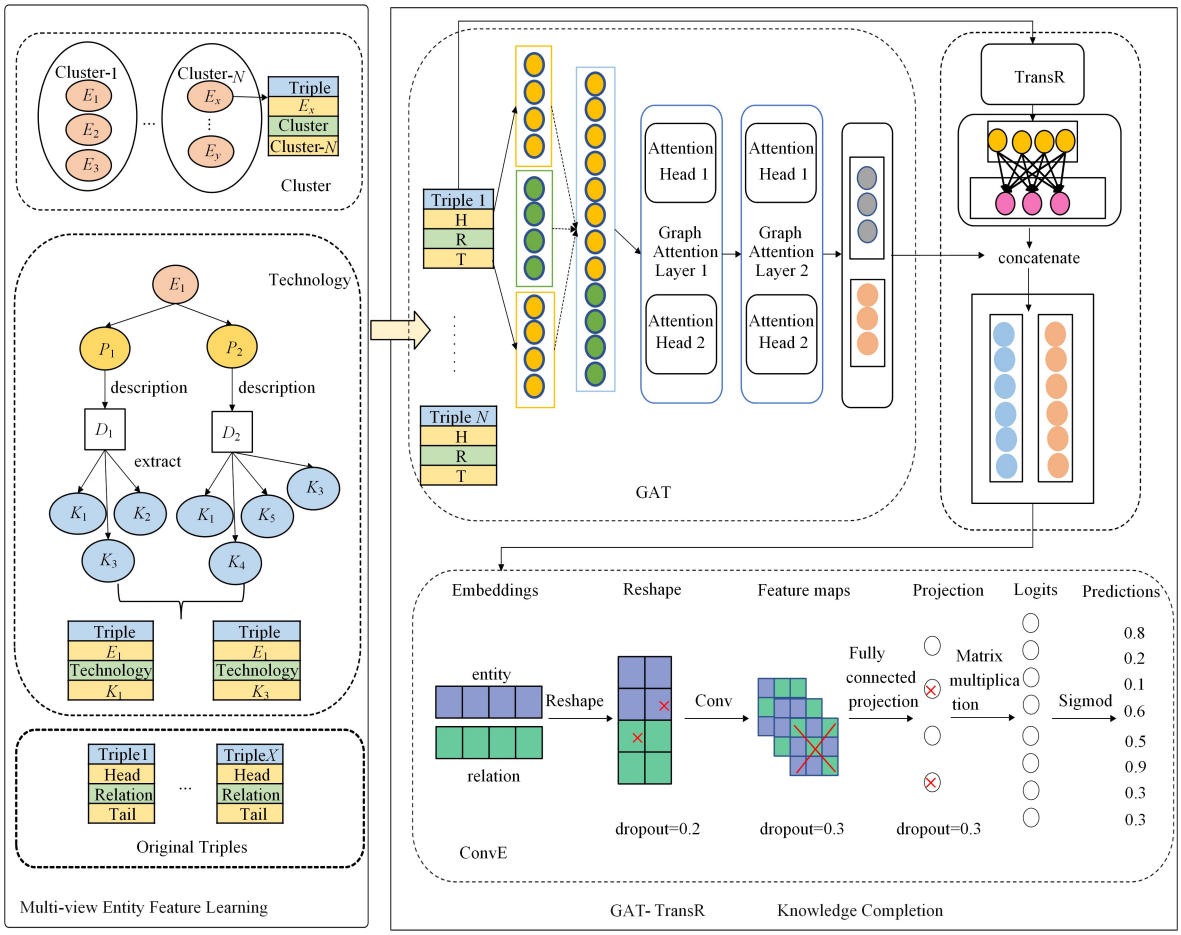

图 2 KERP 模型架构

Fig. 2 Structure of KERP model

4.1 多视角实体特征三元组学习

在预测企业实体关系时,企业及其关联专利的文本描述信息对企业关系表示具有重要作用。现有方法对实体特征表示仅能利用实体本身属性的文本信息,无法利用关联实体的文本描述信息增强自身实体表示。为解决该问题,提出多视角实体特征三元组学习。具体而言,为融合企业实体本身的文本描述信息,提出基于主题聚类的企业环节类别特征学习方法;为融合企业关联专利实体的文本描述信息,提出基于词嵌入的企业技术特征学习方法。最终,得到企业实体的多视角特征三元组表示。

4.1.1 基于主题聚类的企业环节类别特征学习

处于相同生产环节的企业的文本描述通常包含相似的企业经营范围、企业生产产品等信息描述。因此,本文基于主题聚类对企业文本描述信息进行聚类,挖掘企业实体非结构化文本信息中的产业环节类别特征,得到企业环节类别特征三元组。采用的主题聚类方法为 LDA[28] 模型,将每个企业的文本描述信息看作一篇文章,对每个企业按照文本描述进行主题聚类。

算法 1 展示了产业知识图谱中每个企业的文本描述信息 $\boldsymbol{W}$ 的生成过程,可以根据单词分布的最大似然估计,通过变分贝叶斯估计方法估计模型参数。

算法 1 企业环节类别特征学习

输入:参数 α,参数 β,单词个数 N

输出:企业文本描述信息 **W**

1. 企业主题分布 θ←从 Dirichlet 分布 α 取样生成
2. for k=1 to N do
3. 主题 z_k←从多项式分布 θ 中生成
4. 词语分布 Ø←从 Dirichlet 分布 β 取样生成
5. 词语 w_k←从多项式分布 Ø 中生成
6. $\boldsymbol{W}=(w_1, w_2, \cdots, w_N)$

具体而言,首先将所有企业文本描述信息转换为大小为 $N\times D$ 的词频矩阵 $\boldsymbol{W}$,其中 N 为企业数量,D 为所有词语的数量。设置主题个数后,使用词频矩阵 $\boldsymbol{W}$ 训练 LDA 模型。由于新能源汽车产业有 5 个主要环节,因此选择聚类主题数目为 5,其中各主题的关键词如表 2 所列。

表 2 聚类主题关键词

Table 2 Clustered topic keywords

主题	关键词
0	汽车前后组合灯 室内加热器 汽车前后雾灯
1	机加工 热处理 冷锻 组装 齿轮 热锻 传动轴
2	纯电动厢式运输车 纯电动城市客车 客车
3	驱动电机 锂离子电池 组装 整车控制器
4	胎压监测系统 驾驶辅助系统组件 门锁控制器

最终的聚类效果如图 3 所示,其中横轴代表 5 个主题数目,纵轴代表每个主题类在企业文本描述中的权重系数。对比预测结果和企业真实生产环节类别标签可以发现,算法能够对相同环节的企业实现有效聚类。例如,处于整车制造

环节的天津广通汽车有限公司和四川江淮汽车有限公司被聚为了一类。完成聚类后，在知识图谱中增添(Enterprise，Cluster，Cluster-X)企业环节类别特征三元组，其中三元组头实体是某企业，关系是 Cluster，尾实体是某聚类标签。

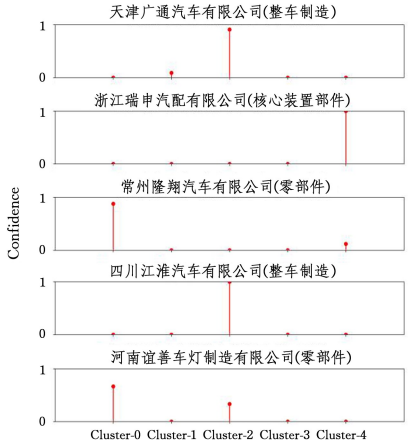

图 3 企业文本描述信息的聚类结果

Fig. 3 Clustering results of enterprise text descriptions

4.1.2 基于词嵌入的企业技术特征学习

企业申请的专利数据对描述企业技术特征和识别产业环节具有重要作用[29]。如图 1 所示，在知识图谱中，企业以实体关联的方式连接其所起草的专利。本文提出基于词嵌入的企业技术特征学习方法，从专利文本描述信息中提取技术词汇，并将关键技术词添加到关联专利的企业技术特征三元组中，增强企业实体特征表示。

首先，对专利文本数据进行中文分词和停用词处理。其次，使用 word2vec 算法[30]的 CBOW 模型学习词的向量表示。模型包含输入层、隐藏层和输出层，其中隐藏层向量 $\boldsymbol{h}$ 是对输入词向量 $\boldsymbol{x}_{ik}$ 进行权重矩阵为 $\boldsymbol{W}$ 的转换，并取平均，具体计算方式如下：

$$\begin{aligned}\boldsymbol{h}&=\frac{1}{C}\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{x}_{1k}+\boldsymbol{x}_{2k}+\cdots+\boldsymbol{x}_{Ck})\\&=\frac{1}{C}(\boldsymbol{v}_{w_1}+\boldsymbol{v}_{w_2}+\cdots+\boldsymbol{v}_{w_C})\end{aligned}\tag{1}$$

其中，C 是上下文单词数量，$\boldsymbol{v}_{w_i}$ 代表词 w_i 的向量。

给定中心词 w_o 的上下文词 $w_1\cdots w_C$ 的情况下，最大化观察到词 w_o 的条件概率损失函数 E：

$$E=-\boldsymbol{v}'^{\mathrm{T}}_{w_O}\boldsymbol{h}+\log\sum_{j'=1}^{V}\exp(\boldsymbol{v}'^{\mathrm{T}}_{w_j}\cdot\boldsymbol{h})\tag{2}$$

训练 CBOW 模型时，由于专利文本描述中词语数量较多，因此设置模型词向量维度为 256。在大量专利文本数据中，标注了一批与企业生产相关的技术词汇，从中可以发现技术词汇对近距离上下文词语依赖较强，对远距离词语不会产生过多依赖，因此将模型的最大滑动窗口(w)设置为 3。滑动窗口大小代表了中心词上下文的距离，即训练模型时输入数据包含中心词的前 w 个词和后 w 个词。负采样个数设置为 5，即在模型训练的反向更新时，仅更新每个训练样本的一小部分权重，包含正样本和 5 个负样本的权重系数矩阵。

模型训练结束后得到所有词的向量表示。图 4 给出了 PCA 降维处理后的部分词向量分布。从图中可以看出，模型能较好地学习词语之间的相似关系。

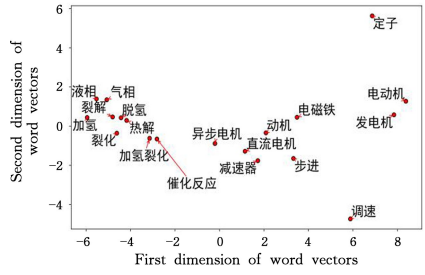

图 4 专利文本词向量分布

Fig. 4 Distribution of patent text word vectors

基于标注的技术词汇，根据余弦相似度计算每个专利文本中与标注样本向量最相近的关键词，用以加强专利文本的实体表示。最后，从企业起草的所有专利的技术词中提取出现频率较高的关键词汇作为企业的技术特征，生成技术特征三元组(Enterprise，Technology，Keyword)。

4.2 融合低阶语义信息的图注意力网络知识补全

4.2.1 编码器

知识图谱可以表示为有向图，节点之间的关联信息通过图编码表示。使用图注意力网络作为编码器，通过邻居节点的信息传播，使用注意力机制赋予邻居节点不同的权重系数，更新中心实体的编码表示，能够捕捉知识图谱实体的高阶语义信息。

原始图注意力网络(Graph Attention Network，GAT)[31]不适合知识图谱，因为它忽略了关系特征，而关系是知识图谱中不可分割的一部分。为解决该问题，本文将关系和相邻节点特征结合到注意机制中[22]。

编码器定义了一个注意力层作为基础模块。每个注意力层的输入是两个嵌入矩阵，即实体和关系嵌入矩阵。实体嵌入由矩阵 $\boldsymbol{H}\in R^{N_e\times D}$ 表示，其中第 i 行是实体 e_i 的嵌入向量，N_e 代表实体的数量，D 是每个实体嵌入向量的维度；类似地，关系嵌入用矩阵 $\boldsymbol{G}\in R^{N_r\times P}$ 表示，N_r 代表关系类型数量，P 是关系嵌入向量的维度。该层输出相应的嵌入矩阵：$\boldsymbol{H}'\in R^{N_3\times D'}$ 和 $\boldsymbol{G}'\in R^{N_r\times P'}$。为了获得实体 e_i 的新表示嵌入，学习与 e_i 关联的所有关系和实体的三元组表示。对于与实体 e_i 相关联的特定三元组 $t_{kij}=(e_i,r_k,e_j)$，采取串联实体和关系向量的方式获取三元组的新表示：

$$\boldsymbol{c}_{ijk}=\boldsymbol{W}_1[\boldsymbol{h}_i\parallel\boldsymbol{h}_j\parallel\boldsymbol{g}_k]\tag{3}$$

其中，$\boldsymbol{c}_{ijk}$ 是三元组 t_{kij} 的向量表示，向量 $\boldsymbol{h}_i$，$\boldsymbol{h}_j$ 和 $\boldsymbol{g}_k$ 分别表示实体 e_i，e_j 和关系 r_k 的嵌入。此外，$\boldsymbol{W}_1$ 表示线性变换矩阵。类似于 GAT，模型学习每个三元组 t_{kij} 的重要性 b_{ijk}。首先将 $\boldsymbol{c}_{ijk}$ 与权重矩阵 $\boldsymbol{W}_2$ 相乘，然后使用 LeakyRelu 函数获得三元组的绝对注意力值：

$$b_{ijk}=\mathrm{LeakyReLU}(\boldsymbol{W}_2\boldsymbol{c}_{ijk})\tag{4}$$

如图 5 所示，为获得相对注意力值，将 softmax 应用于 b_{ijk}：

$$\alpha_{ijk}=\mathrm{softmax}(b_{ijk})\tag{5}$$

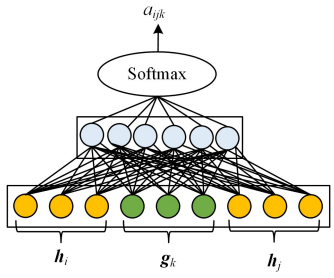

图 5 三元组相对注意力值计算

Fig. 5 Triple relative attention calculation

KERP 模型使用多头注意力机制来稳定训练过程,允许模型在不同的表示空间学到节点邻居的信息,并且使用 M 个独立的注意力头串联的方式融合计算实体的嵌入 $\boldsymbol{h}_i'$:

$$\boldsymbol{h}_i' = \Big\|_{m=1}^{M} \sigma\Big(\sum_{j\in\mathcal{N}_i} \alpha_{ijk}^m c_{ijk}^m\Big) \tag{6}$$

对知识图谱关系矩阵 $\boldsymbol{G}$ 执行线性变换得到 $\boldsymbol{G}'$,由权重矩阵 $\boldsymbol{W}^R$ 对其参数化:

$$\boldsymbol{G}' = \boldsymbol{G}\boldsymbol{W}^R \tag{7}$$

模型使用了两个图注意力层,第一个注意力层捕获关于中心节点相邻一跳邻居的信息,第二个注意力层能够捕获中心节点相邻两跳邻居的信息。在第二个注意力层,多个注意力头以平均的方式输出最终的嵌入向量。

4.2.2 融合实体关系低阶信息

为融合实体关系的低阶语义信息,KERP 使用知识补全中的翻译模型 TransR 给解码器提供实体在关系向量空间内的平移语义表示。图 6 为 TransR 模型的示意图。TransR 在低维空间内通过单个三元组学习实体的向量表示,能够有效弥补多个图注意力层过于关注实体的高阶语义信息而忽略了实体关系的低阶语义信息的缺陷。

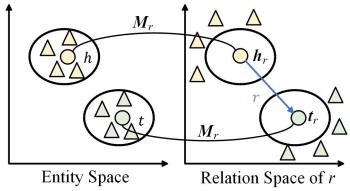

图 6 TransR 模型

Fig. 6 TransR model

在 TransR 中,对于每个三元组 (e_i, r_k, e_j),实体嵌入的空间是 $\boldsymbol{h}, \boldsymbol{t} \in R^k$,并且关系嵌入的空间为 $\boldsymbol{r} \in R^d$。对于每个关系 r,建立一个投影矩阵 $\boldsymbol{M}_r \in R^{k\times d}$,将实体 h 和 t 从实体空间投影到关系空间:

$$\boldsymbol{h}_r = \boldsymbol{h}\boldsymbol{M}_r, \boldsymbol{t}_r = \boldsymbol{t}\boldsymbol{M}_r \tag{8}$$

TransR 的距离评分函数定义为:

$$f_r(\boldsymbol{h}, \boldsymbol{t}) = \| \boldsymbol{h}_r + \boldsymbol{r} - \boldsymbol{t}_r \|_2^2 \tag{9}$$

TransR 模型将实体和关系投影到不同空间,并在关系空间中进行实体的平移翻译操作,能够更好地表示具有多种关系或者属性的实体。将训练集中的三元组数据送入 TransR 模型,训练得到实体和关系的向量表示。记输出实体的向量矩阵为 $\boldsymbol{H}^i$,关系向量矩阵为 $\boldsymbol{R}^i$。最后,使用系数矩阵 $\boldsymbol{W}^E$ 对实体向量进行线性变换,和编码器的实体输出 $\boldsymbol{H}^e$ 进行拼接,得到融合实体向量 $\boldsymbol{H}''$:

$$\boldsymbol{H}'' = [\boldsymbol{W}^E \boldsymbol{H}^i \,\|\, \boldsymbol{H}^e] \tag{10}$$

同理,对关系向量进行类似操作:

$$\boldsymbol{R}'' = [\boldsymbol{W}^R \boldsymbol{R}^i \,\|\, \boldsymbol{R}^e] \tag{11}$$

$\boldsymbol{H}''$ 和 $\boldsymbol{R}''$ 是融合高低阶信息的实体和关系向量嵌入表示,本质依旧是实体和关系的表示学习。然而,单纯的表示学习无法直接对实体关系进行预测,需要采用解码器将上述实体关系表示解码,进而完成实体关系预测任务。

4.2.3 解码器

KERP 选择表达能力强、正负三元组区分度高[32]的 ConvE 模型作为解码器。ConvE 使用二维卷积嵌入实体和关系向量来预测知识图谱中的缺失链接。ConvE 模型由一个卷积层、一个对嵌入维度的投影层和一个内积层组成。

给定输入三元组 (e_i, r_k, e_j),ConvE 将实体 e_i, e_j 映射到它们的嵌入表示 $\boldsymbol{h}_i, \boldsymbol{h}_j \in R^k$,将关系映射为 $\boldsymbol{g}_k \in R^k$,其中 $\boldsymbol{h}_i, \boldsymbol{h}_j$ 来自式(10)的实体表示向量矩阵 $\boldsymbol{H}''$,$\boldsymbol{g}_k$ 来自式(11)的 $\boldsymbol{R}''$。在关系 r_k 下的两个实体嵌入 $\boldsymbol{h}_i, \boldsymbol{h}_j$ 的评分函数 ψ_k 定义如下:

$$\psi_k(\boldsymbol{h}_i, \boldsymbol{h}_j) = f(vec(f([\overline{\boldsymbol{h}_i}; \overline{\boldsymbol{g}_k}] * \omega))\boldsymbol{W})\boldsymbol{h}_j \tag{12}$$

其中,$\overline{h_i}$ 和 $\overline{g_k}$ 分别表示 $\boldsymbol{h}_i$ 和 $\boldsymbol{g}_k$ 的二维混合。

ConvE 对实体和关系的嵌入矩阵 $\boldsymbol{H}''$ 和 $\boldsymbol{R}''$ 进行向量查找,得到三元组对应的实体、关系嵌入向量 $\boldsymbol{h}_i$ 和 $\boldsymbol{g}_k$,将两者重新连接,和卷积核 ω 进行二维卷积计算。卷积计算后,返回特征图张量 $\boldsymbol{T} \in R^{c\times m\times n}$,其中 c 是尺寸为 $m \times n$ 的二维特征图的数量。张量被重塑为矢量 $vec(\boldsymbol{T}) \in R^{cmn}$,使用由矩阵 $\boldsymbol{W} \in R^{cmn\times k}$ 参数化的线性变换将其投影到 k 维空间,并通过内积与尾实体嵌入 $\boldsymbol{h}_j$ 匹配。损失函数采用交叉熵函数:

$$\mathcal{L}(p, t) = -\frac{1}{N}\sum_i (t_i \cdot \log(p_i) + (1-t_i) \cdot \log(1-p_i)) \tag{13}$$

其中,$p = \sigma(\psi_k(\boldsymbol{h}_i, \boldsymbol{h}_j))$ 表示头尾实体 e_i, e_j 存在关系 r_k 的概率;N 是实体个数;$\boldsymbol{t}$ 是标签向量,取值为 1 表示存在关系,0 表示不存在关系。

KERP 模型设定阈值 θ,进而判断企业之间是否存在特定关系。当两家企业存在某种关系 r 的概率大于 θ 时,可以判定两者具有该关系,否则判定没有该关系。根据 Hit@50 指标计算 θ 的初始值,不断增大 θ,进而选择使 F1 值最大的 θ 作为三元组可信度阈值。

5 实验评估

首先,将 KERP 模型与翻译模型 TransR、卷积网络模型 ConvE 和目前 SOTA 效果的图注意力网络模型 KB-GAT 进行对比分析,结果表明 KERP 模型显著优于现有模型;其次,通过消融实验证明 KERP 模型各模块的有效性;最后,在多个公开知识补全数据集上进行实验,验证了 KERP 模型的通用性。

5.1 实验设置

5.1.1 数据集

利用新能源汽车产业知识图谱数据集 NEAI-KG 来分析模型在企业上下游关系预测问题上的性能。为证明模型的

通用性，选用知识补全任务中 WN18RR[7] 和中文百科知识图谱 CN-Dbpedia[33] 等多个主流据集进行实验验证，其中 WN18RR 数据集来自于 WN18[14]。各数据集的实体数目，关系种类，训练集、验证集、测试集三元组数目如表 3 所列。

表 3 数据集统计信息

Table 3 Datasets statistics

Dataset	# Entities	# Relation types	# Triples		
			Train	Valid	Test
NEAI-KG	27346	8	260376	500	500
CN-Dbpedia	111377	6667	199000	500	500
WN18RR	40943	11	86835	3034	3134

5.1.2 评价指标

在实体关系预测任务中，对每个测试三元组 (e_i, r_k, e_j)，去除尾实体 e_j，再由任意 e_j' 替换该实体生成三元组 (e_i, r_k, e_j')，通过评分函数计算新三元组得分，并根据得分高低进行排序，其中原三元组的评分也会计入排名。同样，对头实体 e_i 也进行类似操作。最终计算测试集中所有三元组的平均排名，记为 Mean Rank(MR)。根据三元组得分排名计算排名前 N 中实际存在于测试集的三元组个数的比例，记为 Hit@N。N 的常见取值通常有 3,5,10。

大部分现有实体关系预测模型最终输出三元组评分 $\psi(e_i, r_k, e_j)$，评分越高表明三元组可信度越大。为了便于模型对比，将该评分转化为头尾实体存在关系 r 的概率 $p=\sigma(\psi(e_i, r_k, e_j))$，其中 σ 为 sigmoid 函数。此外，Trans 系列模型(如 TransR 和 TransE)输出的是头尾实体在某种关系下的距离度量值 $D(e_i, r_k, e_j)$，D 越小表明三元组可信度越大。针对这类模型，定义评分函数 $\psi(e_i, r_k, e_j)=-D(e_i, r_k, e_j)$，计算三元组可信度概率。最终通过可信度阈值 θ 判断两家企业是否存在特定关系，使用 F1 值作为模型的评估指标。

5.1.3 实验参数

在训练阶段，KERP 模型的编码器采用两层 GAT 训练实体和关系的嵌入，两层多头注意力机制的头数均设置为 2，并将 $dropout=0.3$ 应用于每层 GAT 的输入。在解码器中，输入层的 $dropout=0.2$，特征图层的 $dropout=0.3$，隐藏层的 $dropout=0.3$，卷积核大小为 3×3。训练采用的优化器是 Adam[34]，损失函数采用交叉熵函数。此外，如表 4 所列，根据数据集的不同，也对 KERP 模型进行了针对性的参数设置。

表 4 不同数据集实验的参数设置

Table 4 Parameter setting of different datasets

Data	Input-dim	Decoder-input	Kernels	Learn rate
NEAI-KG	[100,100]	[150,150]	30	0.0005
CN-Dbpedia	[200,200]	[250,250]	50	0.0010
WN18RR	[200,200]	[250,250]	50	0.0010

5.2 企业上下游关系预测结果和分析

图 7 展示了 KERP 模型在预测企业上下游关系上 *Precision*，*Recall* 和 *F1* 值随着三元组置信度阈值 θ 变化的情况，最终选取使得 *F1* 达到最高值的 $\theta=0.82$。

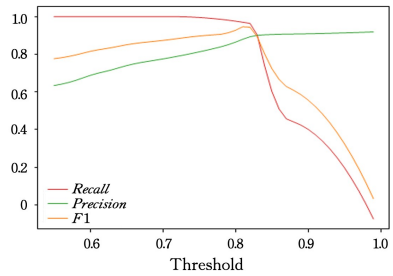

图 7 置信度阈值对精度、召回率和 F1 的影响

Fig. 7 Influence of threshold on Recall, Precision and F1

表 5 展示了 KERP, ConvE, TransR 和 KB-GAT 模型在新能源汽车产业知识图谱数据集上的实验结果。在 F1 指标上，KERP 比经典翻译模型 TransE 高出 25%，比单个解码器 ConvE 高 14%，表明了模型的 GAT 编码器模块的有效性，其能够从中心实体节点的局部邻域学习到实体高阶语义信息。KERP 比同样基于图注意力网络的 KB-GAT 模型高 7%，其原因在于 KB-GAT 使用 ConvKB 作为解码器，其中大量三元组最终得分相似甚至一致[32]，导致正确三元组和错误三元组的区分度小；而 KERP 模型融入 TransR 学习的实体关系低阶语义信息并采取 ConvE 作为解码器，能够更好地区分正例和负例，从而取得更优的预测效果。

表 5 企业上下游关系预测结果

Table 5 Results of enterpris eup-down stream relationship prediction

Model	*F1*	*Hit@5*
TransR	0.621	0.782
ConvE	0.737	0.859
KB-GAT	0.803	0.950
KERP	**0.870**	**0.961**

通过分析预测企业关系的正确例子，KERP 能够有效捕获以下几方面的信息：同一聚类标签的企业容易拥有相似的上下游企业，起草相同专利的企业容易拥有相似的上下游企业，被同一家投资机构投资的两家企业容易拥有相似的上下游企业。例如，如图 8 所示，广西双英集团股份有限公司和中颖电子股份有限公司都被九鼎投资所投资，中颖电子股份有限公司和广西汽车集团有限公司容易被预测为上下游关系。如图 9 所示，四川江淮汽车有限公司和天津广通汽车有限公司聚类标签都为 2，更容易预测天津广通汽车有限公司和浙江瑞申汽配有限公司有上下游关系。

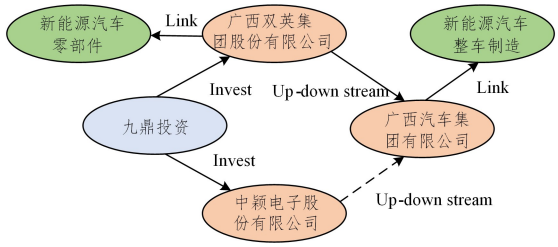

图 8 同一机构投资的两家企业的上下游预测结果分析

Fig. 8 Prediction results for two companies invested by the same institution

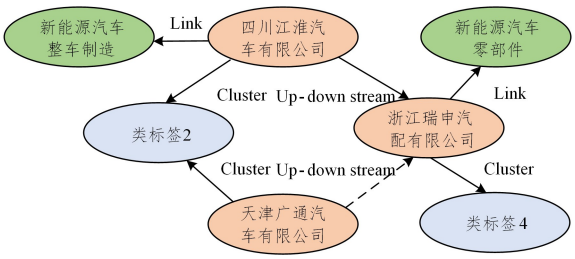

图 9 同一聚类标签的企业上下游预测结果

Fig. 9 Prediction results for two companies with the same cluster label

5.3 消融实验

为证明 KERP 模型各模块的有效性，在新能源汽车产业知识图谱上进行消融实验。从表 6 中的消融实验结果可以看出，加入 GAT 编码器后，$F1$ 有近 14%的提升，证明 GAT 能够有效捕捉实体的高阶语义表示。加入 TransR 后，KERP 模型的 $F1$ 有近 8%的提升，证明了融合 TransR 学习的实体关系低阶信息的必要性。加入企业生产环节类别特征信息(Cluster)和企业技术特征(Technology)后，$F1$ 值分别提升 34%和 16%，说明加入实体多视角特征三元组学习对企业实体关系预测十分有效。

表 6 新能源汽车产业知识图谱上的消融实验结果

Table 6 Ablation experiments on NEAI-KG

Model	$F1$
KERP	0.870
KERP w/o GAT	0.737
KERP w/o TransR	0.862
KERP w/o Cluster	0.533
KERP w/o Technology	0.710

5.4 通用性评估

为验证模型的通用性，在多个主流实体关系预测数据集上对 KERP 进行了实验验证，并将其与 TransE，Complex，ConvE，R-GCN 和 KB-GAT 等模型方法进行横向对比，结果如表 7 所列。除 NEAI-KG 以外的 2 个数据集仅使用事实三元组。因此，验证 KERP 通用性能实际上是验证 KERP 的融合低阶语义信息图注意力网络知识补全模块的通用性。首先，与除 KB-GAT 外的 4 个基线模型在多个数据集上进行对比，KERP 模型在 Mean Rank(MR)，Hit@3 和 Hit@10 方面均取得最好效果。这表明，KERP 模型采取融合低阶语义信息的图注意力网络知识补全技术，能更好地捕捉实体之间的关系。

表 7 WN18RR 和 CN-Dbpedia 数据集上的对比实验

Table 7 Comparative experiments on WN18RR and CN-Dbpedia datasets

Model	WN18RR			CN-Dbpedia		
	MR	*Hit@3*	*Hit@10*	*MR*	*Hit@3*	*Hit@10*
TransE	3384	—	0.501	6.59	0.312	0.892
ComplEx	5261	0.460	0.510	—	—	—
ConvE	4187	0.440	0.520	—	—	—
R-GCN	6700	0.137	0.207	—	—	—
KB-GAT	**1940**	**0.483**	**0.581**	5.63	0.577	0.921
KERP	2362	0.470	0.553	**5.51**	**0.612**	**0.933**

同时，与在知识补全任务中具有 SOTA 性能的 KB-GAT 模型相比，KERP 模型在 CN-Dbpedia 数据集上的 3 个指标(MR，Hit@3，Hit@10)均优于 KB-GAT 模型；此外，在 WN18RR 数据集上 KERP 略逊于 KB-GAT，这主要是因为 WN18RR 的训练集数目较少，无法提供较多的实体之间的关联关系等先验知识，导致 KERP 在训练数据较少的情况下难以取得最优效果。由于 2 个通用数据集仅使用事实三元组，因此 KERP 的通用性能没有远超 SOTA 模型 KB-GAT。

5.5 讨论与分析

5.5.1 模型性能讨论

KERP 首次提出多视角实体三元组特征学习方法，不仅可以对实体本身的文本描述信息进行特征学习，而且能够充分利用关联实体的文本描述信息完善自身特征表示，并将学习的特征信息转化为实体的特征三元组，提高了实体文本描述信息的利用率；其他模型不具备这种能力。

KERP 首次提出融合低阶语义信息的图注意力网络知识补全方法，使用图注意力网络获取实体的高阶语义表示，并且融合实体关系的低阶语义表示，大大增强了模型的可靠性和稳定性。

在 ENAI-KG 新能源汽车产业数据集上，将 KERP 与翻译模型 TransR、卷积网络模型 ConvE 和在知识补全链路预测任务上 SOTA 效果的图注意力网络模型 KB-GAT 进行企业实体上下游关系预测的对比分析，结果表明 KERP 模型显著优于现有模型，在 F1 值上相比 KB-GAT 提升了 6.7%；然后，通过消融实验证明 KERP 模型各模块的有效性；最后，在多个公开知识补全数据集上进行实验。KERP 在 CN-Dbpedia 数据集上的 3 个指标(MR，Hit@3，Hit@10)均优于 KB-GAT 模型，验证了 KERP 模型的通用性。

5.5.2 模型复杂性分析

假定数据集共有 N_e 个实体、N_r 种关系、N_t 个训练事实三元组、N_c 个企业实体，每个实体文本描述平均词语量为 L_c，不重复词语总量为 L_w。KERP 中主题聚类个数为 N_C，CBOW 词输入维度为 D_w，隐藏层输入维度为 D_h，输出层输出维度为 D_w。KERP 中，GAT 编码实体嵌入维度为 D_e，关系嵌入维度为 D_r，TransR 编码实体嵌入维度为 $D_{e'}$，关系嵌入维度为 $D_{t'}$，ConvE 实体关系重塑矩阵维度为 $H\times W$，卷积窗口为 $h\times w$，卷积核数目为 N_k。

1)KERP 的多视角实体三元组特征学习模块参数计算

主题聚类模型参数量：文档词频矩阵为 $N_c\times L_c$，文档主题分布参数矩阵为 $N_c\times N_C$，词语主题分布参数矩阵为 $L_w\times N_C$；词嵌入模型的参数量：词向量矩阵为 $L_w\times D_w$，权重参数矩阵为 $D_w\times D_h+D_h\times D_w$。KERP 的多视角实体三元组特征学习模块参数量为：

$$N_c\times(L_c+N_C)+L_w\times(N_C+D_w)+2\times D_w\times D_h \tag{14}$$

2)KERP 的融合低阶语义信息的图注意力网络知识补全模块参数计算

首先对实体和关系进行参数随机初始化，实体和关系嵌入矩阵为：$N_e\times D_e+N_r\times D_r$。式(3)共享权重参数 $\boldsymbol{W}_1$ 的维度大小为 $D_e\times(2\times D_e+D_r)$。每层 GAT 使用两个图注意力头，故式(4)中注意力权重矩阵 $\boldsymbol{W}_2$ 的大小为 $1\times D_e\times 2$。式(7)中的权重矩阵 $\boldsymbol{W}^R$ 采用单位矩阵，不计入计算参数量；TransR

需要单独的实体和关系嵌入矩阵参数，实体和关系嵌入矩阵为：$N_e \times D_{e'} + N_r \times D_{r'}$ 和，投影矩阵 $\boldsymbol{M}_r$ 为：$N_r \times D_{r'} \times D_{e'}$。实际应用中，式(10)中的矩阵 $\boldsymbol{W}^E$ 和式(11)中的 $\boldsymbol{W}^R$ 采用单位矩阵，不计入参数量；解码器 ConvE 参数量包括卷积计算和全连接投影层，卷积核所需参数为：$N_k \times h \times w + N_k$，最后全连接投影层所需参数为：$N_k \times (H-h) \times (W-w) \times D_e$。记：

$$P_1 = N_e \times (D_e + D_{e'}) + N_r \times (D_r + D_{r'}) \tag{15}$$

$$P_2 = D_e \times (2 \times D_e + D_r + 2) + N_r \times D_{r'} \times D_{e'} \tag{16}$$

$$P_3 = N_k \times (h \times w + 1 + (H-h) \times (W-w) \times D_e) \tag{17}$$

KERP 融合低阶语义信息的图注意力网络知识补全模块参数所需参数量为：$P_1 + P_2 + P_3$

在 NEAI-KG 数据集上，计算得到 KERP 参数量约为 6.3×10^6，是 ConvE 参数量(4.7×10^6)的约 1.3 倍，是 TransR 参数量(2.8×10^6)的 2.25 倍。KERP 所需计算的参数量略大。

结束语 针对产业知识图谱中企业实体关系缺失问题，提出一种新的基于知识增强的企业实体关系预测模型 KERP。该模型由多视角实体特征三元组学习模块和融合低阶语义信息的图注意力网络知识补全模块组成，能有效利用相邻实体的描述信息完善企业实体特征，并融合实体的高阶语义信息和实体关系的低阶语义信息增强企业实体及其关系的特征表示。实验分析表明，和现有实体关系预测模型相比，KERP 模型具有更好的预测性能。

未来工作主要有两个方面：1)融合可用的外部知识库来增强知识图谱中三元组实体的语义信息；2)KERP 计算参数量较大，后续将在保证模型准确度的前提下降低计算复杂度，提高模型的可用性。

参 考 文 献

[1] CHEN Z, WANG Y, ZHAO B, et al. Knowledge Graph Completion: A Review[J]. IEEE Access, 2020, 8: 192435-192456.

[2] LI F L, CHEN H, XU G, et al. AliMeKG: Domain knowledge graph construction and application in e-commerce[C]// Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020: 2581-2588.

[3] MOON C, JONES P, SAMATOVA N F. Learning Entity Type Embeddings for Knowledge Graph Completion[C]// Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. 2017: 2215-2218.

[4] WANG J J, LIU J G, LI Z K. Research on Partnership of Supply Chain Based on Complex Network[J]. Chinese Journal of Systems Science, 2021, 29(3): 110-115130.

[5] LU Z Z, CHEN Q. Link Prediction of Enterprise Cooperation Relationship in Dynamic Supply Chain Network[J]. Computer Engineering and Applications, 2022, 58(2): 265-273.

[6] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion[C]// Twenty-ninth AAAI Conference on Artificial Intelligence. 2015.

[7] DETTMERS T, MINERVINI P, STENETORP P, et al. Convolutional 2d knowledge graph embeddings[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2018.

[8] ZHANG J Z, HU Y M. Uncovering the Mechanism of Co-Authorship Network Evolution by Link Prediction[J]. Information Science, 2017, 35(7): 75-81.

[9] LU Z G, CHEN Q. Discovering Potential Partners via Projection-Based Link Prediction in the Supply Chain Network[J]. International Journal of Computational Intelligence Systems, 2020, 13(1): 1253-1264.

[10] XIE M, WANG T, JIANG Q, et al. Higher-Order Network Structure Embedding in Supply Chain Partner Link Prediction [C]// CCF Conference on Computer Supported Cooperative Work and Social Computing. Singapore: Springer, 2019: 3-17.

[11] KERSTING K, RAEDT L D. Adaptive Bayesian logic programs [C]// International Conference on Inductive Logic Programming. Berlin: Springer, 2001: 104-117.

[12] LAO N, COHEN W W. Relational retrieval using a combination of path-constrained random walks[J]. Machine Learning, 2010, 81(1): 53-67.

[13] WANG Q, MAO Z, WANG B, et al. Knowledge graph embedding: A survey of approaches and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2017, 29(12): 2724-2743.

[14] BORDES A, USUNIER N, GARCIA-DURAN A, et al. Translating embeddings for modeling multi-relational data[C]// Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 2. 2013: 2787-2795.

[15] 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). 2015: 687-696.

[16] NICKEL M, TRESP V, KRIEGEL H P. A three-way model for collective learning on multi-relational data[C]// ICML. 2011.

[17] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction[C]// International Conference on Machine Learning. PMLR, 2016: 2071-2080.

[18] GU J, WANG Z, KUEN J, et al. Recent advances in convolutional neural networks[J]. Pattern Recognition, 2018, 77: 354-377.

[19] NGUYEN D Q, NGUYEN T D, NGUYEN D Q, et al. A novel embedding model for knowledge base completion based on convolutional neural network[J]. arXiv: 1712.02121, 2017.

[20] SCHLICHTKRULL M, KIPF T N, BLOEM P, et al. Modeling relational data with graph convolutional networks[C]// European Semantic Web Conference. Cham: Springer, 2018: 593-607.

[21] SHANG C, TANG Y, HUANG J, et al. End-to-end structure-aware convolutional networks for knowledge base completion [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2019: 3060-3067.

[22] NATHANI D, CHAUHAN J, SHARMA C, et al. Learning attention-based embeddings for relation prediction in knowledge graphs[J]. arXiv: 1906.01195, 2019.

[23] XIAO H, HUANG M, MENG L, et al. SSP: semantic space projection for knowledge graph embedding with text descriptions

[C]// Thirty-First AAAI Conference on Artificial Intelligence. 2017.

[24] XIE R,LIU Z,JIA J,et al. Representation learning of knowledge graphs with entity descriptions[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2016.

[25] ZHANG Z,CAO L,CHEN X,et al. Representation learning of knowledge graphs with entity attributes[J]. IEEE Access,2020,8:7435-7441.

[26] LIN Z F,OU S Y. Research on Relation Prediction in Knowledge Graphs by Fusing Structure and Text Features[J]. Library and Information Service,2020,64(21):99-110.

[27] ZHAI S P,WANG S H,SHANG D R,etal. An adaptive model for knowledge representation with entity description[J]. Journal of Chinese Information Processing,2021,35(01):43-53.

[28] BLEI D M,NG A Y,JORDAN M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research,2003,3(Jan):993-1022.

[29] SUN X L,ZHUANG W H,LI B,et al. Research on Characteristics and Cooperation Prediction of Industry-University-Research Institute Collaboration Based on Patents inRegional Equipment Manufacturing Industry[J]. Science and Management,2020,40(1):31-40.

[30] MIKOLOV T,CHEN K,CORRADO G,et al. Efficient estimation of word representations in vector space[J]. arXiv:1301.3781,2013.

[31] VELICKOVIC P,CUCURULL G,CASANOVA A,et al. Graph Attention Networks[J]. arXiv:1710.10903,2017.

[32] SUN Z,VASHISHTH S,SANYAL S,et al. A Re-evaluation of Knowledge Graph Completion Methods[C]// Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020:5516-5522.

[33] XU B,XU Y,LIANG J,et al. CN-DBpedia:A never-ending Chinese knowledge extraction system[C]// International Conference on Industrial,Engineering and Other Applications of Applied Intelligent Systems. Springer,2017:428-438.

[34] KINGMA D P,BA J. Adam:A method for stochastic optimization[J]. arXiv:1412.6980,2014.

WANG Jiaqi, born in 2000, doctoral student, is a member of China Computer Federation. His main research interests include knowledge graph and data mining.

GUAN Jihong, born in 1969, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include machine learning and bioinformatics.

(责任编辑:柯颖)