

基于知识图谱与图注意力网络的SG-CIM模型映射技术

李杏¹, 任小伟¹, 楼轶维^{2*}, 高士杰³, 葛鑫亮¹, 廖小琦¹

1. 国家电网有限公司大数据中心, 北京 100053

2. 北京大学计算机学院, 北京 100871

3. 北京中电普华信息技术有限公司, 北京 100085

摘要 为了更加智能化地实现国家电网公司设计的公共数据模型(SG-CIM)的业务需求和提高图谱数据本身的质量和共享程度, 提出一种基于图注意力网络的图结构映射模型。首先基于SG-CIM模型的数据构建SG-CIM模型知识图谱和数据库表知识图谱, 然后通过图注意力网络分别学习每个图中的实体嵌入, 将实体嵌入统一向量空间, 最后基于实体向量之间的距离计算相似度, 得到2个图的图结构映射结果。实验表明, 本文提出的模型在SG-CIM数据模型自动映射方面得到相当不错的结果。

关键词 SG-CIM模型; 知识图谱; 图注意力网络; 知识图谱映射

随着信息化平台建设需求不断增加, 电力行业不同省、市公司以及各专业各部门都针对自身的现实需求开发了一系列信息化应用, 例如电力营销系统、用电信息采集系统等。为方便各系统之间的数据交互, 统一数据模型(SG-CIM)是国家电网公司参考国际标准(IEC 61970/61968/62325)和行业最佳实践, 结合公司核心业务需求、在运系统数据字典等, 采用“业务需求驱动自顶向下”和“基于现状驱动自下向上”相结合的模式, 采用面向对象建模

技术而构建的企业数据模型^[1]。统一数据模型主要由逻辑模型和物理模型2部分构成。逻辑模型是从企业级视角统一定义公司核心业务对象及其属性字段、相互关联关系。物理模型是遵循逻辑模型, 依据业务应用的数据处理需要, 以选定数据库产品为基础而形成的数据库结构和表设计。本研究基于统一数据模型的逻辑模型和物理模型构建SG-CIM模型知识图谱和数据库表知识图谱。

知识图谱(knowledge graph, KG)以机器可读

收稿日期: 2022-10-18; 修回日期: 2022-11-24

作者简介: 李杏, 副高级工程师, 研究方向为大数据应用技术、数据模型、数据架构等, 电子信箱: 271434571@qq.com; 楼轶维(通信作者), 博士研究生, 研究方向为大数据应用技术, 电子信箱: cyfqlyw@gmail.com

引用格式: 李杏, 任小伟, 楼轶维, 等. 基于知识图谱与图注意力网络的SG-CIM模型映射技术[J]. 科技导报, 2023, 41(15): 124-132; doi: 10.3981/j.issn.1000-7857.2023.15.013

的格式表示人类知识,是结构化的语义知识库,描述了现实世界对象和事件或抽象概念的相互关联的集合,是人工智能和自然语言处理领域许多应用的重要基础。近年来,许多知识图谱被不同的组织构建,以为不同的领域提供结构化知识。其中一些知识图谱通常包含互补性的内容,吸引研究者将其整合成一个统一的知识图谱,以利于许多如信息提取和推荐知识驱动的任务^[2]。

近年来,多种基于嵌入的方法被提出用于图映射,包括基于翻译的模型(translation-based model for multilingual knowledge graph embeddings, MTransE)^[3]和联合属性保留嵌入的跨语言实体对齐模型(joint attribute-preserving embedding model, JAPE)^[4],它们的方法给定2个KG和1组预对齐实体,采用基于嵌入的方法将实体嵌入到低维向量空间中,通过计算它们的向量表示进行实体匹配。JE(joint embedding method)^[5]和ITransE(iterative alignment model)^[6]是基于嵌入的方法来匹配异构KG之间的实体,这些方法也可以用于基于嵌入的图映射的问题。Wang等^[7]提出的一种基于图卷积网络(graph convolutional networks, GNNs)的模型,也是一种基于嵌入的KG对齐方法,该方法使用图卷积网络直接建模实体之间的等价关系。MuGNN(multi-channel graph neural network model)^[2]是一种多通道图神经网络模型,它通过编码不同的KGs来学习面向对齐的嵌入,利用不同的通道对每个KG进行补全和剪枝,采用池化技术将不同的通道组合在一起,从不同的角度协调结构以增强实体嵌入。

为提高电力行业各系统的智能化交互,本研究主要研究SG-CIM模型知识图谱和数据库表知识图谱的映射技术方法,设计基于图注意力网络的统一数据模型——映射模型。图注意力网络(GAT)^[8],是一种新型神经网络体系结构,在图结构化数据上运行,采用注意力机制来对邻居节点进行聚合操作,得到不同的邻居权重,实现对其自适应分配,有效地提升了图神经网络模型的表达能力。

本研究提出一种基于知识图谱和图注意力网络的图结构映射模型,它使用图注意力网络编码不

同的知识图谱来学习实体和关系嵌入,生成实体表示向量。然后,通过相似度方法计算2个图谱的向量表示,得到图结构映射结果。

1 相关工作

1.1 知识图谱嵌入(Knowledge Graph Embedding)

知识图谱嵌入(knowledge graph embedding, KGE),也称基于知识图谱的知识表示学习(knowledge representation learning, KRL),通过将实体和关系映射到低维向量空间来间接捕获它们的语义,是对知识图谱中的实体和关系完成分布式表示的过程^[9]。嵌入通常是通过最小化知识图谱中所有实体和关系的全局损失函数来学习的,它可以进一步用于关系预测、问答系统、推荐系统、信息提取和其他一些任务。近年来,人们对知识图谱嵌入问题做了大量的研究,提出了多种知识图谱嵌入方法,如基于三元组(h, r, t)的模型MTransE、TransE,包括改进TransE的几种增强模型,这些模型通常只关注实体和实体之间的一跳关系,认定不同事实的三元组之间相互独立,在链路预测和三重分类等问题上取得了很好的结果。

基于图神经网络的模型,这些模型从邻域信息角度考虑,如GCN-Align^[7]是第一个用图神经网络进行实体对齐的模型,使用2个图卷积网络学习2个知识图谱的实体嵌入; Xu等^[10]提出图匹配神经网络(graph matching neural network, GMNN)模型,将实体对齐任务看作一个图匹配问题,同时引入主题实体图,采用GAT匹配主题实体图中的所有实体; Schlichtkrull等^[11]提出的R-GCN(relational graph convolutional network)模型将图卷积网络应用于关系数据的扩展方面。这些方法有些只关注实体特征,有些只考虑实体或关系的特征来学习KG嵌入。GCNs从实体的邻居收集信息,但不同邻居的权重是固定的,来自归一化的拉普拉斯矩阵。为了解决GCNs的缺点,引入了图注意力机制。图注意力网络(GATs)在图结构数据上运行利用隐藏的自注意力层来解决基于图卷积或其相近的现有方

法的缺点。通过堆叠层,让节点关注到其邻域的特征,然后指定不同的权值给邻域中的不同节点,这个过程不需要任何倒置等代价高昂的矩阵操作。本文采用的基于图注意网络的模型全面地捕捉了知识图谱中给定的实体节点的邻域中的多跳和语义相似关系。

1.2 基于嵌入的知识图谱对齐

近年来,有很多不同的KG对齐方法被提出,如JE、MTransE、JAPE、ITransE、基于GNN的KG对齐等^[12-17]。MTransE、JAPE和ITransE这些模型都依赖于TransE模型来学习实体嵌入,然后定义一些实体对齐的转换关系,都要求在KGs中对齐或共享。基于GCN的KG对齐方法将实体嵌入到一个统一的向量空间中,认为在这个空间中对齐的实体的距离应该尽可能接近,该方法只关注2个KGs中的实体匹配。

然而,本研究采用基于GAT的图结构映射方法,将实体节点间关系和相邻节点的特征结合到注意力机制中,不仅关注两个KGs中的实体匹配,也学习关系的嵌入。基于GAT的图结构映射方法可以总结为2个步骤:(1)基于图注意力机制学习KG嵌入,生成每个KGs中的实体和关系的低维嵌入;(2)将这些嵌入映射到一个统一的向量空间,通过对比相似度和预对齐实体对,建立图结构映射关系。

2 问题定义

知识图谱定义可用 $G=(E, R, T)$ 表示,其中 E, R, T 分别表示实体集、关系集和属性集。设 $G_1=(E_1, R_1, T_1)$ 是SG-CIM模型知识图谱, $G_2=(E_2, R_2, T_2)$ 是数据库表知识图谱, $S = \{(e_{1i}, e_{2i}) | e_{1i} \in E_1, e_{2i} \in E_2\}$ 是

G_1 和 G_2 之间的一组预对齐实体对,其中 e_{1i} 和 e_{2i} 分别表示 G_1 和 G_2 中的实体。本文的研究内容是在2个KGs(G_1, G_2)中寻找实体对齐,设计2个KGs的图结构映射方法。在SG-CIM模型知识图谱和数据库表知识图谱中,可以根据实体定义的相似度构建预先对齐的实体对集。在KG对齐的过程中,使用预先对齐的实体对集作为训练数据集。

3 SG-CIM模型知识图谱和数据库表知识图谱的构建

3.1 SG-CIM模型数据集

SG-CIM模型包括逻辑模型和物理模型,以树形结构存储在xml文件中。逻辑模型和物理模型存储形式相同,xml文件的根节点是XML,根节点的子节点包括XML.header和XML.content,子节点XML.content下有SG-CIM4.8的包,逻辑模型和物理模型的数据全部存储在这个包里,数据包具体结构如图1所示。

实体的英文名称存储在UML:Class标签中,即其属性name的值,实体的中文名、实体描述等信息存储在UML:ModelElement.taggedValue下,实体的所有属性信息存储在UML:Classifier.feature下;关系的信息,包括源实体的名称、目标实体的英文名称、关系类型等,存储在关系标签的UML:ModelElement.taggedValue下。

3.2 知识图谱构建技术

知识图谱(KG)以三元组 (h, r, t) 的形式存储真实世界的知识,其中 h, t 表示头尾实体, r 表示关系。本文基于SG-CIM模型的逻辑模型和物理模型采取自底向上法分别构建SG-CIM模型知识图谱和数据库表知识图谱,知识图谱构建流程如图2所示。

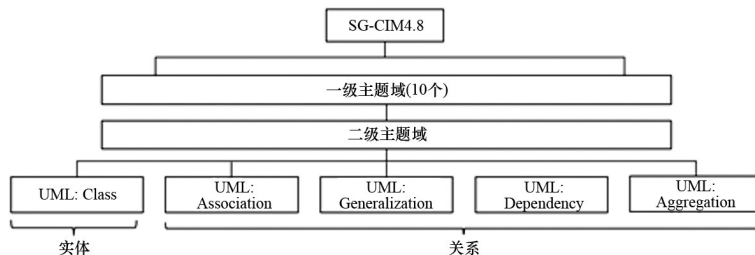


图1 SG-CIM数据包xml文件结构图

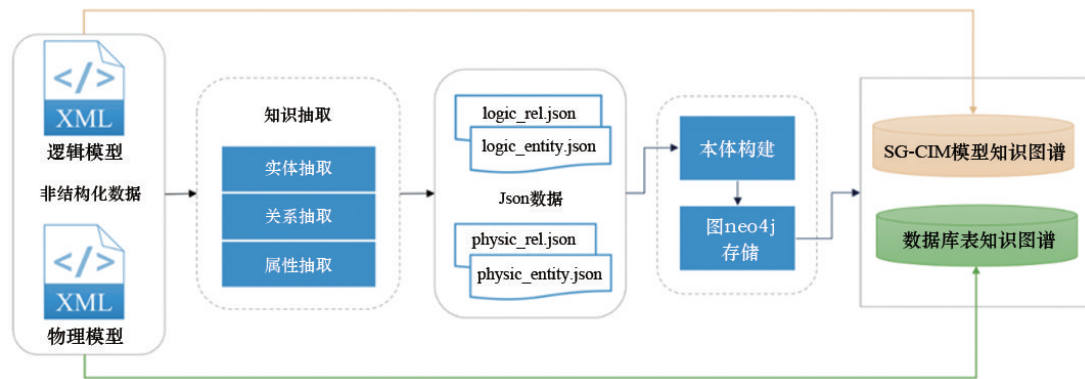


图2 知识图谱构建流程

1) 知识抽取。

知识抽取包含实体抽取、关系抽取和属性抽取。本文通过解析逻辑模型和物理模型的xml文件进行实体抽取和关系抽取,属性抽取包含在实体抽取中。本文通过遍历一级主题域以及其下的二级主题域,然后遍历所有实体和关系,得到实体数据集和关系数据集,导出到JSON存储文件。

2) 本体定义。

知识图谱由节点和边组成,每个节点表示现实世界中存在的实体,每条边表示实体与实体之间的关系。本文构建的知识图谱定义了两类节点:实体节点 entity_node、属性节点 attr_node;两类关系:实体节点与实体节点间的关系(4种:Association、Dependency、Generalization、Aggregation)、实体节点与属性节点间的关系 Entity_Attr。其中实体节点定义由唯一标识“id”、中文名称“ch”、英文名称“en”、字段描述“des”、一级主题域“area1”、二级主题域“area2”构成;属性节点定义由唯一标识“id”、中文名称“ch”、英文名称“en”、字段描述“des”构成;关系定义由源实体“source”、目标实体“target”、关系类型“relation type”构成。

3) 图谱构建。

根据抽取出来的实体和关系数据集,通过Python存储到Neo4j中得到数据图谱。图3为数据库表知识图谱部分示例,表1是SG-CIM模型知识图谱和数据库表知识图谱节点关系数量。

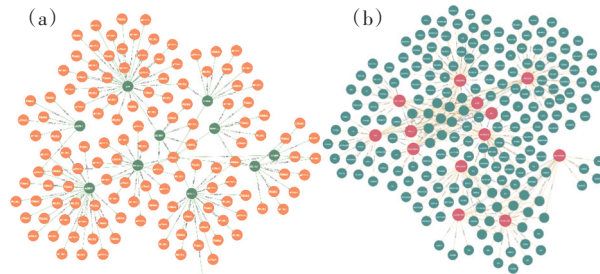


图3 SG-CIM模型知识图谱(a)和数据库表知识图谱(b)

表1 知识图谱节点关系数量

项目	SG-CIM模型知识图谱	数据库表知识图谱
实体节点	6759	4976
属性节点	87587	106776
实体节点间关系	3886	2348
Entity_Attr	87520	169722

4 基于图注意力网络的图结构映射模型

结合SG-CIM模型知识图谱和数据库表知识图谱的关联关系,通过图注意力网络(GAT)学习实体嵌入。然后,结合训练的实体嵌入列表,将实体嵌入到统一向量空间,采用余弦相似度方法计算两个图的实体表示向量之间的相似度,经过对比之后筛选出相似度最大的节点对,在其对应的两个图谱

实体之间建立映射关系,得出图结构映射模型。模型架构图具体如图4所示,模型核心算法如算法1所示。

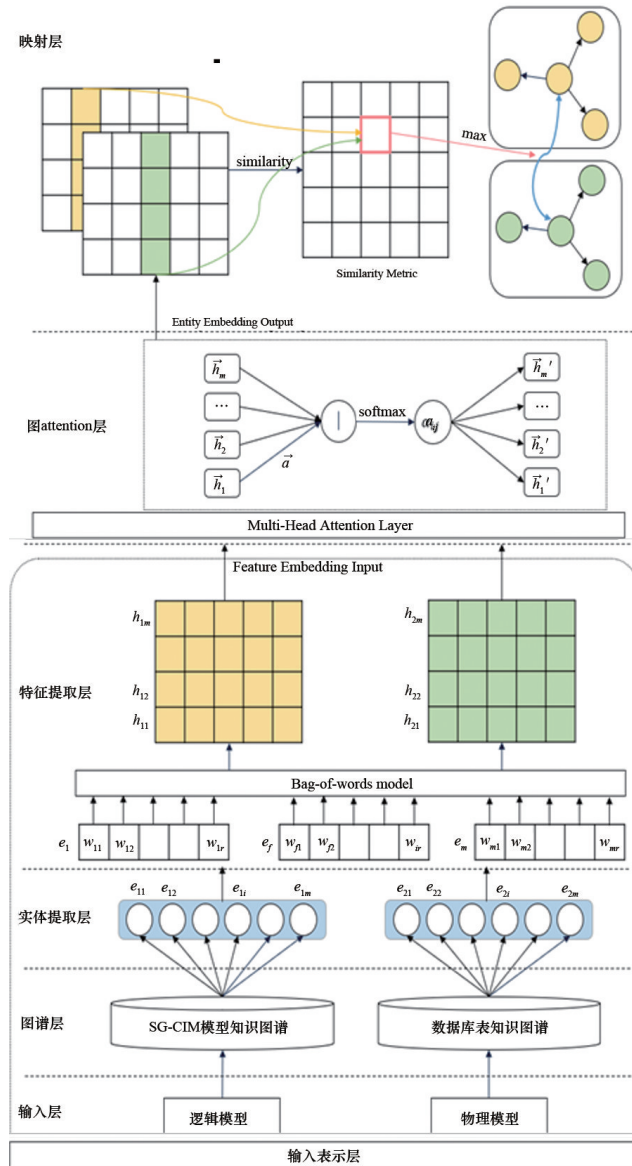


图4 基于图注意力网络的图结构映射模型架构

基于GAT的图结构映射模型包括以下3层。

1) 输入表示层。

这一层的目标是获得实体的初始特征向量嵌入,以使用GAT来学习这些嵌入,其种包括以下4层:(1)输入层:解析逻辑模型和物理模型的xml文件进行实体抽取和关系抽取;(2)图谱层:对已经抽取的实体和关系数据集进行知识融合、合并、加工,将更新后的数据集通过Python存储到Neo4j中

算法1 MappingGAT

Algorithm: MappingGAT

```

Input:  $F_{in}$  (in_feature), edges # 实体初始特征向量集合和图结构信息集合
 $F_{out}$  (out_feature) # 训练输出实体向量
T # the training epoch T
GraphAttentionLayer() # 图注意力训练算法
cosine_distance(x,y) # 两元素之间的距离
Output:  $S_k$ : a list of size k
Initialize:  $F_{out} = \emptyset$ 
for  $t=1, 2, \dots, T$  do
    do 输入 in_feature, edges 数据集到两层图注意力层训练
        GraphAttentionLayer(in_feature, edges)
        GraphAttentionLayer(in_feature, edges)
    until  $F_{in}$  全部训练, 输出  $F_{out}$ 
    for  $G_1$  中每一个实体向量 do
        cosine_distance( $G_1$  的实体向量,  $G_1$  的实体向量)  $\rightarrow$ 
        max_similarity,  $S_k$ 
    end for
    loss(pre_labels, source_labels) # 计算损失函数, 降低损失
end for

```

得到数据图谱,分别得到SG-CIM模型知识图谱(G_1)和数据库表知识图谱(G_2);(3)实体提取层: G_1 和 G_2 包含entity_node和attr_node,提取节点entity_node作为编码两个图的实体;(4)特征提取层:采用基于单词的方法将实体的定义信息转化为其初始特征向量,首先,提取 G_1 和 G_2 的所有实体的定义信息,构建词袋模型;然后分别对每个图的实体定义进行编码,得到2个图的实体初始特征向量表示列表。

2) 图attention层。

这一层是通过GAT^[8]对 G_1 和 G_2 分别来学习实体嵌入,本层以 G_1 为例来解释算法,主要包括以下步骤。

(1)输入:将输入表示层得到的实体初始特征向量表示列表作为图注意力层的输入,用 $h=\{h_1, h_2, \dots, h_n\}$, $h_i \in \mathbf{R}^F$ 来表示这一组节点初始特征,其中 N 为节点数, F 为每个节点中的特征数。

(2)自注意力机制(self-attention):计算任意2

个节点之间的注意力系数(attention coefficients), 使用一个共享的注意力计算函数 a , 公式为

$$e_{ij} = a(\mathbf{W}\mathbf{h}_i, \mathbf{W}\mathbf{h}_j) \quad (1)$$

式中, e_{ij} 表示节点 j 的特征对节点 i 的重要性; \mathbf{W} 为共享的参数矩阵。

在整个过程中, 需要计算出节点 i 的每一个邻居节点 j 对节点 i 的重要性。

(3) attention 归一化。用 softmax 函数将注意力系数归一化

$$\alpha_{ij} = \text{soft max}_j(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})} \quad (2)$$

式中, 注意力机制 α 是一个权重矩阵, 加入了 LeakyReLU 函数, 这里小于零, 斜率为 0.2。

综合式(1)、式(2), 得到完整的注意力机制:

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(a^T[\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j]\right)\right)}{\sum_{k \in N_i} \exp\left(\text{LeakyReLU}\left(a^T[\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_k]\right)\right)} \quad (3)$$

式中, \parallel 符号表示连接操作(concatenation operation)。

(4) 特征聚合: 得到归一化注意力系数, 计算与其对应的特征的线性组合, 得到每个节点的输出特征, 计算公式为

$$\mathbf{h}'_i = \sigma\left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W}\mathbf{h}_j\right) \quad (4)$$

式中, σ 为激活函数。

(5) 输出: 节点的输出特征, $\mathbf{h}' = \{\mathbf{h}'_1, \mathbf{h}'_2, \dots, \mathbf{h}'_n\}$, $\mathbf{h}'_i \in \mathbf{R}^F$, 即图 G_1 的实体嵌入表示, 为图注意力层的输出。

3) 映射层。

这一层是采用余弦相似度方法计算 G_1 和 G_2 中向量之间相似度, 构建图结构映射关系, 主要包括以下步骤。

(1) 输入: 通过图注意力层得到 G_1 实体嵌入列表 $\mathbf{h}'_1 = \{\mathbf{h}'_{11}, \mathbf{h}'_{12}, \dots, \mathbf{h}'_{1n}\}$, $\mathbf{h}'_{1i} \in \mathbf{R}^F$ 和 G_2 的实体嵌入列表 $\mathbf{h}'_2 = \{\mathbf{h}'_{21}, \mathbf{h}'_{22}, \dots, \mathbf{h}'_{2n}\}$, $\mathbf{h}'_{2i} \in \mathbf{R}^F$, 作为本层的输入。

(2) 相似度矩阵: 计算 G_1 、 G_2 的实体表示向量之间的相似度, 在这里对 G_1 中某一个实体和 G_2 中所有实体计算相似度, 采用余弦相似度的计算方

法, 得到 2 个图谱的实体表示向量之间的相似度矩阵

$$\cos(\mathbf{A}_1, \mathbf{A}_2) = \frac{\mathbf{A}_{1m \times p} \cdot \mathbf{A}_{2n \times p}^T}{|\mathbf{A}_1| \cdot |\mathbf{A}_2|} \quad (5)$$

式中, \mathbf{A}_1 、 \mathbf{A}_2 分别为 G_1 、 G_2 中所有实体嵌入的矩阵表示; m 、 n 分别为 G_1 和 G_2 中实体个数; p 为特征向量个数。

(3) 实体映射: 计算相似度矩阵中每一行相似度最大的位置, 映射到实体节点 id, 得到实体之间的映射关系。

(4) 输出: 三元组列表即 G_1 和 G_2 的图结构映射结果, 公式如下:

$$\left((e_{1i_id}, e_{2i_id}), \max_similarity\right), e_{1i} \in E_1, e_{2i} \in E_2 \quad (6)$$

式中, e_{1i_id} 为 G_1 中实体的 id; e_{2i_id} 为 G_2 中实体的 id; $\max_similarity$ 为 e_{1i_id} 、 e_{2i_id} 所代表实体的相似度。

5 实验

5.1 数据集

采用基于 GAT 的图结构映射技术, 首先通过实体的初始特征信息和图谱结构信息使用图注意力网络模型编码不同的知识图谱, 生成实体嵌入列表, 将 2 个 KGs 的实体嵌入统一向量空间; 然后, 结合预对齐实体对数据集, 比对得到基于相似度最大节点对的图结构映射模型。

5.1.1 GAT 数据集

为 SG-CIM 模型知识图谱 (G_1) 和数据库表知识图谱 (G_2) 依据其关联关系分别构建两个数据集。数据集将图谱中每个实体节点作为一个样本, 通过实体节点的属性“des”中实体的定义提取节点的初始特征, 每个实体节点都由一个 4579 维的词向量表示, 所以, 每个样本具有 4579 个特征。词向量中的每个元素都对应一个词, 而且该元素取值只有 0 或 1。若该元素对应的词不在实体定义中, 则取 0; 反之, 则取 1。所有的词来源于 G_1 、 G_2 中全部实体的定义。

图谱中的实体都存在所属一级主题域,数据集将10个一级主题域作为标签,即所有样本被分为10个类别。根据从SG-CIM模型的逻辑模型和物理模型中抽取的关系信息,图谱中实体节点存在边相连,即样本之间存在联系。数据集汇总于表2。

表2 实验中使用的数据集

项目	G_1 数据集	G_2 数据集
节点	4976	6759
边	2348	3802
节点特征	4579	4579
类	10	10
训练集	2976	4059
验证集	1000	1350
测试集	1000	1350

5.1.2 预对齐实体对集

本文根据SG-CIM模型知识图谱(G_1)和数据库表知识图谱(G_2)的实体节点属性信息的相似性,经过对比匹配构建1200个正样本和1200个负样本,作为预对齐实体对数据集。若实体节点属性信息相似性高,则表示为1;否则,表示为0。数

据集格式表示为(G_1 实体节点id, G_2 实体节点id,1或者0)。

5.2 结果与分析

首先,运用图注意力网络(GAT)来编码SG-CIM模型知识图谱(G_1)和数据库表知识图谱(G_2),得到每个图的实体嵌入列表。引入注意力机制,训练实体初始特征向量,将实体节点所属一级主题域作为类别标签, G_1 有4976个实体, G_2 有6759个实体,训练结果表示,SG-CIM模型知识图谱验证集和测试集准确率分别为86.74%、87.85%,数据库表知识图谱验证集和测试集准确率分别为88.10%、86.34%。然后,结合训练的实体嵌入列表,将实体嵌入到统一向量空间,采用余弦相似度方法计算2个图的实体表示向量之间的相似度,得到相似度矩阵,经过对比之后筛选出相似度最大的节点对,在其对应的2个图谱实体之间建立映射关系,得出图结构映射模型结果。图结构映射结果如表3所示,实体节点相似度最大的5个目标节点,其相似度对比情况见图5。

表3 图结构映射结果(部分示例)

源实体	目标实体1	目标实体2	目标实体3	目标实体4	目标实体5
10	4984/0.996961	4994/0.996901	10875/0.996883	5043/0.996838	4995/0.996719
13	4982/0.997252	5285/0.997193	4995/0.997189	5043/0.997134	5048/0.997094
713	5309/0.996667	7965/0.996396	10881/0.996350	7954/0.996143	5194/0.996128
1008	5634/0.997415	5807/0.997401	5793/0.997359	5676/0.997343	8536/0.997320
1017	5716/0.997266	6245/0.997207	5761/0.997199	5790/0.997159	8187/0.997158
2490	8004/0.997244	7954/0.997214	7992/0.997037	7991/0.996976	8010/0.996787
2743	7995/0.9970003	8056/0.996954	8003/0.996931	7993/0.996883	7954/0.996869
4739	11224/0.997183	8355/0.997115	11219/0.997087	8063/0.997027	8262/0.996985



图5 图结构映射id对相似度对比

表3中给出了逻辑模型和物理模型的部分映射结果,展示了源节点和与其相似度最大的5个目标节点及其相似度的值。图5展示了数据库表知识图谱中实体节点,与其相似度最大的5个目标节点的相似度对比情况。根据表中节点对(10, 4984),查询两个知识图谱得到,id“10”代表节点: (“d”: 10, “area1”: “人员”, “area2”: “人力资源公共包”, “ch”: “单位”, “en”: “DWD_HR_SgUnit”, “des”: “定义:主要用于记录组织机构规范中的单位信息”), id“4984”代表节点: (“id”: 4984, “en”: “Organization”, “ch”: “组织”, “area1”: “人员”, “area2”: “人力资源公共包”, “des”: “定义:主要用于记录组织机构规范中的单位信息”), 比对主题域和定义信息可知,2个节点映射关系是正确的。在这里,数据库表知识图谱的实体作为源节点,SG-CIM模型知识图谱的实体作为目标节点进行映射,其中相似度值最大的目标实体与源实体节点相似,则映射结果是准确的。

6 结论

提出一种基于知识图谱和图注意力网络的SG-CIM模型映射技术,首先构建基于逻辑模型和物理模型的知识图谱,然后基于图谱实体信息得到实体表示向量,通过图注意力网络训练得到实体嵌入列表,最后基于相似度建立两个图谱之间的实体映射关系,得到图结构映射模型。本文模型最大的优势在于结合图结构信息和实体属性信息特征的同时,学习了预对齐实体关系,进一步提高了逻辑模型和物理模型的映射效果,推进统一数据模型有效融合。

参考文献(References)

- [1] 万齐鸣, 王思宁, 何鑫. 数据中台SG-CIM模型应用方法[J]. 电信科学, 2020, 36(3): 136-143.
- [2] Cao Y X, Liu Z Y, Li C J, et al. Multi-channel graph neural network for entity alignment[DB/OL]. arXiv preprint: 1908.09898, 2019.
- [3] Chen M H, Tian Y T, Yang M H, et al. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. New York: ACM, 2017: 1511-1517.
- [4] Sun Z Q, Hu W, Li C K. Cross-lingual entity alignment via joint attribute-preserving embedding[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017: 628-644.
- [5] Hao Y C, Zhang Y Z, He S Z, et al. A joint embedding method for entity alignment of knowledge bases[C]//Chen H, Ji H, Sun L, et al. Knowledge Graph and Semantic Computing: Semantic, Knowledge, and Linked Big Data. CCKS 2016. Communications in Computer and Information Science. Singapore: Springer, 2016, 650: 3-14.
- [6] Zhu H, Xie R B, Liu Z Y, et al. Iterative entity alignment via joint knowledge embeddings[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. New York: ACM, 2017: 4258-4264.
- [7] Wang Z C, Lv Q S, Lan X H, et al. Cross-lingual knowledge graph alignment via graph convolutional networks [C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2018: 349-357.
- [8] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[DB/OL]. arXiv preprint: 1710.10903, 2017.
- [9] 徐有为, 张宏军, 程恺, 等. 知识图谱嵌入研究综述[J]. 计算机工程与应用, 2022, 58(9): 30-50.
- [10] Xu K, Wang L W, Yu M, et al. Cross-lingual knowledge graph alignment via graph matching neural network[DB/OL]. arXiv preprint: 1905.11605, 2019.
- [11] Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[DB/OL]. arXiv preprint: 1703.06103, 2017.
- [12] Mao X, Wang W T, Wu Y B, et al. From alignment to assignment: Frustratingly simple unsupervised entity alignment[C]//Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2021: 2843-2853.
- [13] 张富, 杨琳艳, 李健伟, 等. 实体对齐研究综述[J]. 计算机学报, 2022, 45(6): 1195-1225.
- [14] 车超, 刘迪. 基于双向对齐与属性信息的跨语言实体对齐[J]. 计算机工程, 2022, 48(3): 74-80.
- [15] 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. Stroudsburg: Association for Computational Linguistics, 2019: 4710–4723.
- [16] 赵丹, 张俊. 基于双重注意力和关系语义建模的实体对齐方法[J]. 计算机应用研究, 2022, 39(1): 64–69, 79.
- [17] Wu Y T, Liu X, Feng Y S, et al. Relation-aware entity alignment for heterogeneous knowledge graphs[C]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2019: 5278–5284.

An SG-CIM model mapping technology study via knowledge graph and graph attention network

LI Xing¹, REN Xiaowei¹, LOU Yiwei^{2*}, GAO Shijie¹, GE Xinliang¹, LIAO Xiaoqi¹

1. Big Data Center of State Grid Corporation of China, Beijing 100053, China
2. School of Computer Science, Peking University, Beijing 100871, China
3. Beijing Zhongdian Puhua Information Technology Co., Ltd., Beijing 100085, China

Abstract In order to intelligently meet the business requirements of the SG-CIM model designed by the State Grid Corporation of China and improve the quality and sharing of the graph data, this paper proposes a graph structure mapping model based on graph attention network(GAT). First of all, using the data of SG-CIM model, an SG-CIM knowledge graph and a database table knowledge graph are constructed. The entities embedding in each graph are learned separately through the GAN, which are then embedded into a unified vector space. Finally, the graph structure mapping results of the two graphs are obtained after calculating the similarities based on the distances between the entity vectors. Experiments show that the proposed model achieves good results in automatic mapping of SG-CIM model.

Keywords SG-CIM model; knowledge graph; graph attention network; knowledge graph mapping ●



(责任编辑 刘志远)