



# 基于异构图注意力的工业设备故障诊断知识图谱补全方法

马亚杰<sup>1\*</sup>, 刘洋<sup>1</sup>, 姜斌<sup>1</sup>, 冒泽慧<sup>1</sup>, 刘剑慰<sup>1</sup>, 刘文静<sup>2,3</sup>

1. 南京航空航天大学自动化学院, 南京 211106

2. 北京控制工程研究所, 北京 100190

3. 空间智能控制技术重点实验室, 北京 100190

\* 通信作者. E-mail: yajiema@nuaa.edu.cn

收稿日期: 2022-11-25; 修回日期: 2023-01-31; 接受日期: 2023-03-11; 网络出版日期: 2024-02-05

科技创新 2023 — “新一代人工智能” 重大项目 (批准号: 2020AAA0109305) 资助

**摘要** 针对工业设备故障诊断知识图谱故障实体属性残缺、故障关系链接缺失的问题, 本文提出了一种基于知识图谱异构图注意力网络 (knowledge graph heterogeneous graph attention network, KGHAN) 模型的工业设备故障诊断知识图谱补全方法, 通过对故障实体概念补全和故障关系链接补全完善了工业设备故障诊断知识图谱. 所提 KGHAN 模型在异构图注意力网络模型的基础上, 融合了故障知识结构信息和故障图结构信息, 有效地表征了故障实体和故障关系的嵌入表示, 提高了故障实体概念补全任务的准确率和故障关系链接补全任务的命中率. 将所提工业设备故障诊断知识图谱补全方法应用在国内某企业的工业设备故障运维数据上, 结果表明, 故障实体概念补全任务的准确率提高了约 10%, 故障关系链接补全任务的命中率提高了约 37%, 验证了方法的有效性.

**关键词** 知识图谱补全, 知识图谱, 图神经网络, 故障诊断, 工业设备

## 1 引言

随着现代自动化技术水平的不断提高, 工业设备系统的复杂性急剧增加, 在恶劣的工作环境和长时间的运行压力下, 工业设备很容易发生故障, 造成难以预料的经济损失<sup>[1,2]</sup>. 因此, 基于知识图谱的工业设备智能故障诊断技术受到了高度重视. 知识图谱 (knowledge graph, KG) 以强大的语义处理能力和开放互联能力, 将互联网表达成更接近人类认知的形式, 解决了数据孤岛、烟囱式开发等问题<sup>[3]</sup>, 使得在“Web 3.0”时代提出的“知识之网”成为了可能. 知识图谱的出现改变了传统的知识获取模式, 将知识工程“自上而下”方式转变为挖掘数据、抽取知识的“自下而上”方式<sup>[4~7]</sup>. 因此, 工

**引用格式:** 马亚杰, 刘洋, 姜斌, 等. 基于异构图注意力的工业设备故障诊断知识图谱补全方法. 中国科学: 信息科学, 2024, 54: 354-372, doi: 10.1360/SSI-2022-0440  
Ma Y J, Liu Y, Jiang B, et al. Knowledge graph completion method for industrial equipment fault diagnosis based on heterogeneous graph attention (in Chinese). Sci Sin Inform, 2024, 54: 354-372, doi: 10.1360/SSI-2022-0440

业界与学术界将知识图谱引入工业设备系统故障诊断领域,构建了各自领域的故障诊断知识图谱,例如,文献[8]构建了流程工业智能制造故障诊断知识图谱,文献[9]构建了电信领域智能故障诊断知识图谱,文献[10]构建了变电站设备故障诊断知识图谱,文献[11]构建了工业过程监视系统的故障诊断知识图谱,文献[12]构建了工业设备故障处置知识图谱,文献[13]构建了电信信息收集系统的故障诊断知识图谱.上述工业设备故障诊断知识图谱充分利用了故障知识本身丰富的语义以及故障知识之间巨大的潜在联系,提高了系统的运行安全性和可靠性,为设备智能维护提供有力的解决方案.但是,对于现阶段由人工或者半自动得到的不完整工业设备故障诊断知识图谱,其故障实体属性、故障关系链接缺失严重,需要有效的知识图谱补全方法,从给定的工业故障诊断知识图谱中推导出故障实体属性和故障关系链接,完成对故障知识的深度分析和挖掘.

知识图谱补全的核心是知识嵌入表示,包含两个方面的补全:概念补全、链接补全<sup>[14~16]</sup>.概念补全通过实体分类实现,为了识别出知识图谱中指定实体的类型;链接补全通过链接预测实现,为了预测出知识三元组(头实体,关系,尾实体)中缺失的部分<sup>[17]</sup>.学者们为了预测单一的关系链接模式,最早提出了翻译模型,通过关系嵌入表示将头实体嵌入表示转换为尾实体嵌入表示<sup>[18~22]</sup>,但是翻译模型只能建模单一的关系类型.为了预测出对称与反对称、反演和合成的关系链接模式,进一步在翻译模型的基础上提出了乘法模型,通过张量积形成实体和关系的嵌入表示<sup>[23~27]</sup>,但是乘法模型无法对高阶关系进行建模.虽然知识三元组的知识嵌入表示形式受到了广泛的认可,但存在计算效率低下、数据稀疏等问题,且未能实现对实体类型的补全,无法得到实际应用<sup>[28]</sup>.

近年来,以图神经网络<sup>[29]</sup>(graph neural network, GNN)为代表的深度学习技术将实体和关系的语义信息表示为稠密低维的实值向量,在低维空间中高效计算实体之间复杂的语义关联,解决了翻译模型计算效率低下、数据稀疏的问题,同时实现了对知识图谱实体类型的补全,是现阶段实现知识嵌入表示的最好方法之一.图神经网络借鉴了卷积神经网络、循环神经网络和深度自编码器的思想,定义和设计了用于处理图的神经网络结构,有着较好的性能和可解释性.文献[30]提出了图卷积神经网络(graph convolutional network, GCN)模型,通过知识图谱的线性缩放来学习局部图结构,以此编码节点的特征表示,实现邻域的信息传递;文献[31]提出了 GraphSAGE 模型,对图中每个节点的邻居节点进行随机采样,聚合邻居节点的信息,得到图中各节点的嵌入表示;文献[32]提出了图注意力网络(graph attention network, GAT)模型,采用了自注意力机制,为不同的节点分配不同的权重;文献[33]提出了异构图注意力网络(heterogeneous graph attention network, HAN)模型,包含节点级注意力和语义级注意力,同时考虑节点和元路径的重要性,该模型相对于一条元路径的节点数具有线性复杂度,可以应用于大规模异构图;文献[34]提出了 R-GCN 模型,引入了权值共享和系数约束,使得 GCN 模型可以应用于链接众多的关系网络;文献[35]提出了 KGAT 模型,将知识图谱的关系信息及用户与商品的交互信息融合到一个图空间中,实现了对高阶关系的建模.由于图神经网络可以捕捉到邻居节点的信息,目前有很多研究将其运用于知识图谱补全.但是,大部分研究重点利用了实体和关系的名称、描述、类别等信息,忽略了知识图谱中仍存在的大量知识结构信息和图结构信息,这些被忽略的信息可以为知识图谱补全提供重要依据.

本文针对工业设备故障诊断知识图谱故障实体属性残缺、故障关系链接缺失的问题,提出了一种基于知识图谱异构图注意力网络(knowledge graph heterogeneous graph attention network, KGHAN)模型的工业设备故障诊断知识图谱补全方法,提高了故障实体概念补全任务的准确率和故障关系链接补全任务的命中率,主要贡献如下:

(1) 构建了一种新的图神经网络 KGHAN 模型,充分融合了故障知识结构信息和故障图结构信息,利用异构图注意力网络模块学习不同路径下邻居故障实体的知识表示,得到了故障实体和故障关

系在复杂多元关系下的特征表达, 实现了工业设备故障诊断知识图谱的知识嵌入表示。

(2) 提出了一种新的基于 KGHAN 模型的工业设备故障诊断知识图谱补全方法, 有效地处理了工业设备故障诊断知识图谱的复杂多元关系, 在已建立的工业设备故障诊断知识图谱上进行了实验验证, 相比于基于 HAN 模型的补全方法, 使得故障实体概念补全的准确率提高了约 10%, 故障关系链接补全的命中率提高了约 37%。

本文其余章节结构如下: 第 2 节阐述了工业设备故障诊断知识图谱及其相关概念的定义, 工业设备故障诊断知识图谱补全任务的定义, 以及本文的问题与目标; 第 3 节给出了知识图谱异构图注意力网络模型的具体构建过程, 并设计了工业设备故障诊断知识图谱补全任务; 第 4 节给出了实验条件及实验结果, 验证了本文所提的工业设备故障诊断知识图谱补全方法的有效性; 第 5 节总结全文。

## 2 问题描述

本节将详细介绍工业设备故障诊断知识图谱及其相关概念的定义, 工业设备故障诊断知识图谱补全任务的定义, 并对本文的研究问题与目标进行阐述。

### 2.1 工业设备故障诊断知识图谱及其相关概念定义

为了更加合理地表征工业设备故障诊断机制, 本文构建了工业设备故障诊断知识图谱。工业设备故障诊断知识图谱属于领域知识图谱, 与通用知识图谱架构类似, 本质上是一种语义网络, 是对故障事实的结构化表示, 由故障实体、故障关系和故障描述构成。然而不同的是, 通用知识图谱面向大众领域, 包含了大量的常识性知识, 强调的是知识的广度; 而领域知识图谱则面向专业领域, 包含了某个领域的特定知识, 强调的是知识的深度。

工业设备故障诊断知识图谱是一种典型的异构有向属性图, 被描述为  $G = \{E, R, F\}$ , 由故障实体集  $E$ 、故障关系集  $R$  和故障知识三元组集  $F$  组成。故障实体类型由映射函数  $\phi: E \rightarrow \mathcal{E}$  定义, 故障关系类型由映射函数  $\varphi: R \rightarrow \mathcal{R}$  定义, 其中,  $\mathcal{E}$  和  $\mathcal{R}$  表示预先定义的故障实体类型集合和故障关系类型集合, 并且满足不等式  $|\mathcal{E}| + |\mathcal{R}| \geq 2$ 。故障知识三元组由  $\langle h, r, t \rangle \in F$  定义, 其中,  $h, r$  和  $t$  表示一个故障知识三元组的头实体、故障关系和尾实体<sup>[36]</sup>。在工业设备故障诊断知识图谱中, 两个不同类型的故障实体可以通过不同的语义路径进行关联, 这条路径由  $\Phi: e_0 \xrightarrow{r_1} e_1 \xrightarrow{r_2} \dots \xrightarrow{r_l} e_l$  定义, 其中,  $\{e_0, e_1, \dots, e_l\} \in E$  是工业设备故障诊断知识图谱中一条路径上的故障实体,  $\{r_1, r_2, \dots, r_l\} \in R$  是工业设备故障诊断知识图谱中一条路径上的故障关系。

### 2.2 工业设备故障诊断知识图谱补全任务定义

工业设备故障诊断知识图谱补全任务是为了解决故障事实不完备、专家知识稀疏的问题, 使工业设备故障诊断知识图谱变得更加完整, 其包含两个方面的补全任务: 故障实体概念补全任务、故障关系链接补全任务。

故障实体概念补全通过实体分类实现, 是为了识别出工业设备故障诊断知识图谱中指定故障实体的类型, 并将识别出的故障实体划分给预先给定的不同类别。针对工业设备故障诊断知识图谱中类型缺失的故障实体  $e \in E$ , 通过学习一个映射函数  $\phi: E \rightarrow \mathcal{E}$  预测出其原本的类型  $\mathcal{E}_e$ , 实现故障实体概念补全。故障关系链接补全通过链接预测实现, 是为了预测出故障知识三元组中缺失的部分。故障关系链接补全任务可分成 3 个子任务: 头实体预测  $\langle ?, r, t \rangle$ 、关系预测  $\langle h, ?, t \rangle$  和尾实体预测  $\langle h, r, ? \rangle$ , 其中, ? 代表着故障知识三元组缺失的部分。针对工业设备故障诊断知识图谱中故障知识三元组  $\langle h, r, t \rangle \in F$ ,

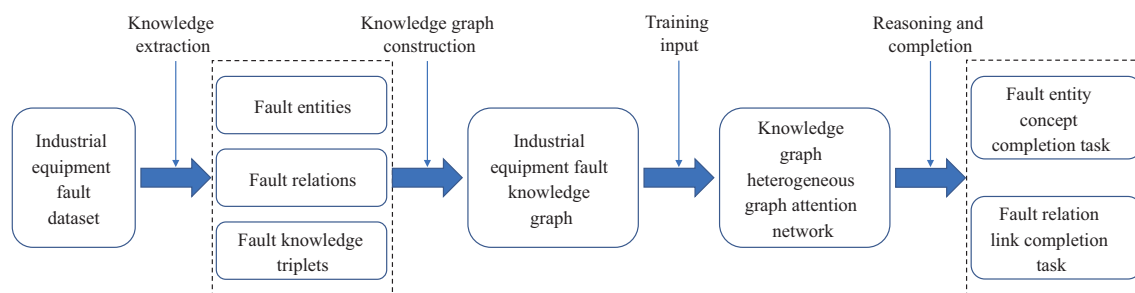


图 1 (网络版彩图) 基于 KGHAN 模型的工业设备故障诊断知识图谱补全框架图

Figure 1 (Color online) Framework of industrial equipment fault diagnosis knowledge graph completion based on KGHAN model

通过学习一个预测函数  $\pi$  预测出原本缺失的部分, 实现故障关系链接补全.

### 2.3 问题与目标

本文的目标是设计一种工业设备故障诊断知识图谱补全方法框架, 解决目前工业设备故障诊断知识图谱故障实体属性残缺、故障关系链接缺失的问题, 其架构如图 1 所示. 首先, 对故障实体、故障关系和故障知识三元组进行 one-hot 词向量编码, 实现工业设备故障诊断知识图谱的数据预处理; 其次, 基于 HAN 模型构建 KGHAN 模型, 在 HAN 模型的基础上增加故障知识结构信息和故障图结构信息, 充分学习到故障实体和故障关系的深层表达; 最后, 通过故障实体概念补全和故障关系链接补全实现对工业设备故障诊断知识图谱补全任务.

## 3 基于知识图谱异构图注意力网络模型的工业设备故障诊断知识图谱补全设计

本节将构建知识图谱异构图注意力网络 KGHAN 模型, 获得故障实体和故障关系的嵌入表示. 然后, 基于知识图谱异构图注意力网络 KGHAN 模型设计工业设备故障诊断知识图谱补全方法框架, 以故障实体概念补全和故障关系链接补全两个任务实现工业设备故障诊断知识图谱补全.

### 3.1 知识图谱异构图注意力网络模型构建

本小节基于异构图注意力网络 HAN 模型构建知识图谱异构图注意力网络 KGHAN 模型, 主要包含 3 个部分: 故障知识结构信息融合模块、故障图结构信息融合模块以及异构图注意力网络模块, 其结构如图 2 所示, 输入是故障实体的词向量、故障关系的词向量以及工业设备故障诊断知识图谱, 输出是故障实体的嵌入向量和故障关系的嵌入向量, 该嵌入向量是每个故障实体和故障关系的表征.

#### 3.1.1 故障知识结构信息融合模块设计

为了便于表示文本型故障知识数据, 在工业设备故障诊断知识图谱的数据预处理中, 以 one-hot 词向量表示故障实体和故障关系的嵌入. 然而, 当使用 one-hot 词向量表示一个故障知识三元组的头实体、尾实体和故障关系时, 会导致故障知识结构信息被破坏. 因此, 为了保留工业设备故障诊断知识图谱的故障知识结构信息, 在头实体、尾实体和故障关系的词向量中融入经过知识图谱翻译模型 TransD 转换而成的故障知识三元组的嵌入表示.

模块的结构如图 2 的故障知识结构信息融合模块所示, 使用两个实值向量表示故障知识三元组的头实体、尾实体和故障关系, 第一个向量表示自身的实际语义, 另一个向量用于构造映射矩阵. 头实体

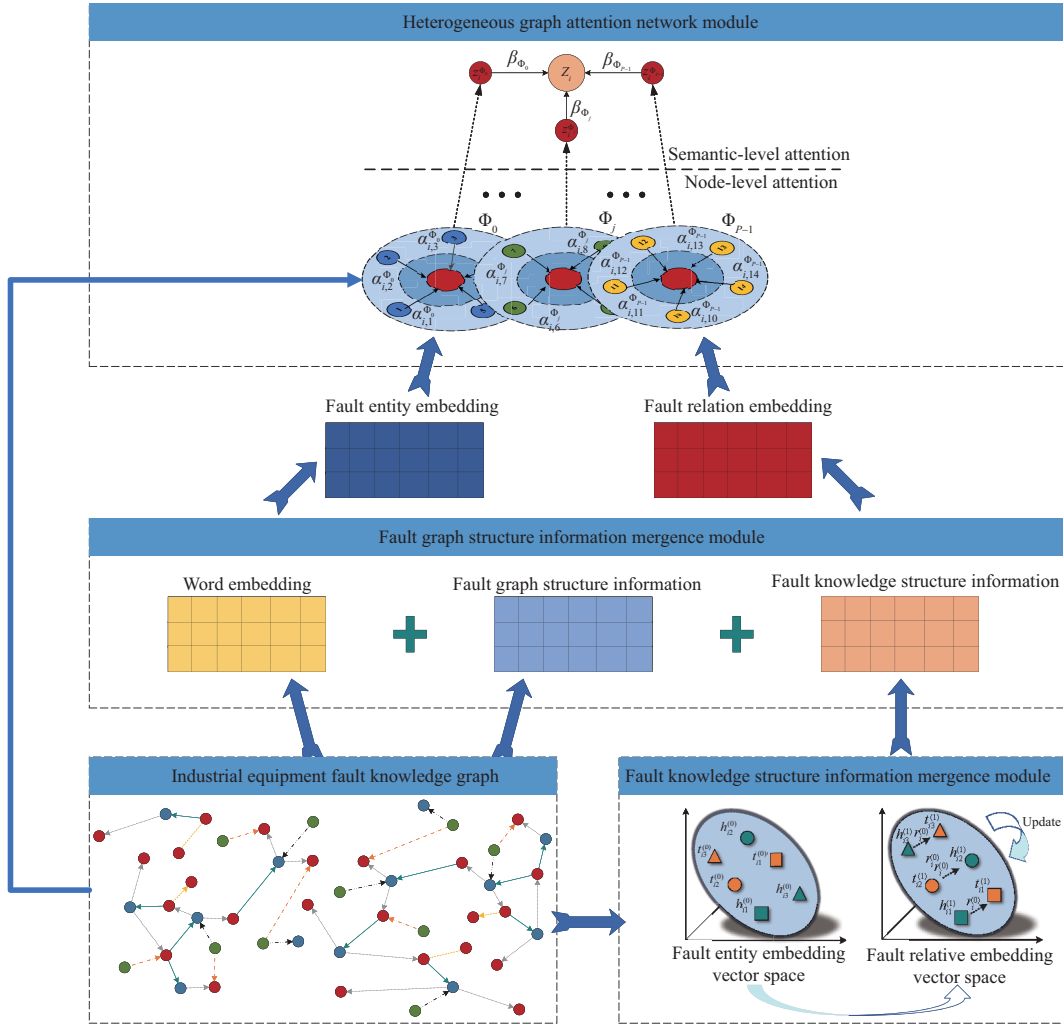


图 2 (网络版彩图) KGHAN 模型的结构示意图

Figure 2 (Color online) Structure of KGHAN model

和尾实体的映射矩阵  $M_{rh_i}$ ,  $M_{rt_i}$  定义为

$$M_{rh_i} = r_i^{(0)} \left( h_i^{(0)} \right)^T + I^{m \times n}, \quad (1)$$

$$M_{rt_i} = r_i^{(0)} \left( t_i^{(0)} \right)^T + I^{m \times n}, \quad (2)$$

其中,  $r_i^{(0)}$  是故障关系词向量,  $h_i^{(0)}$ ,  $t_i^{(0)}$  是头实体和尾实体词向量,  $m$  是故障关系词向量的维度,  $n$  是头实体和尾实体词向量的维度,  $I^{m \times n} \subseteq \mathbb{R}^{m \times n}$  是单位矩阵.

通过映射矩阵将头实体和尾实体的词向量映射到对应的故障关系词向量空间中, 以此可以表征故障知识结构信息:

$$h_i^{(1)} = M_{rh_i} h_i^{(0)}, \quad (3)$$

$$t_i^{(1)} = M_{rt_i} t_i^{(0)}, \quad (4)$$

其中,  $h_i^{(1)}, t_i^{(1)}$  是更新后头实体和尾实体的嵌入向量.

为了方便表示, 将头实体和尾实体的词向量统一成故障实体的词向量  $x_i^{(0)}$ , 将头实体和尾实体的嵌入向量统一成故障实体的嵌入向量  $x_i^{(1)}$ . 为了融合故障知识结构信息, 将故障实体和故障关系的嵌入向量拼接到其各自的词向量中, 得到

$$x_i^{(2)} = x_i^{(0)} \parallel x_i^{(1)}, \quad (5)$$

$$r_i^{(2)} = r_i^{(0)} \parallel r_i^{(1)}, \quad (6)$$

其中,  $x_i^{(2)}, r_i^{(2)}$  是融合故障知识结构信息后的第  $i$  个故障实体和故障关系的嵌入向量,  $\parallel$  是向量拼接运算.

### 3.1.2 故障图结构信息融合模块设计

在工业设备故障诊断知识图谱中, 仅依据故障知识结构信息表示故障实体的嵌入是不够的, 因此需要引入故障图结构信息进一步表达故障实体的嵌入. 而节点中心性<sup>[37]</sup> 是理解故障图结构信息的重要信号, 能很好地衡量故障实体在工业设备故障诊断知识图谱中的重要程度. 因此, 为了保留工业设备故障诊断知识图谱的故障图结构信息, 在故障实体的词向量中融入故障实体的度, 作为其自身嵌入向量的附加信号, 其结构如图 2 的故障图结构信息融合模块所示, 对于每个故障实体, 在其词向量中加入了该故障实体的中心性编码, 根据故障实体入度和出度为其分配两个实值嵌入向量. 融合故障图结构信息后实体的嵌入向量  $x_i$  定义为

$$x_i = x_i^{(2)} \parallel (z_i^{\text{deg}^+} \parallel z_i^{\text{deg}^-}), \quad (7)$$

其中,  $z_i^{\text{deg}^+}, z_i^{\text{deg}^-}$  是第  $i$  个故障实体的入度向量和出度向量.

### 3.1.3 异构图注意力网络模块设计

工业设备故障诊断知识图谱属于典型的异构有向属性图, 图中包含不同类型的节点或边. 在前期研究中, 为了简化工业设备故障诊断知识图谱中复杂的关系, 将异构有向属性图转化为同构有向属性图, 采用 GAT 模型完成工业设备故障诊断知识图谱补全任务. 虽然 GAT 模型可用于有向图且时间复杂度为  $O(|V|FF' + |E|F')$ , 但是对于异构图的嵌入表示却无法适用. 因此, 为了实现了工业设备故障诊断知识图谱的嵌入表示, 设计了异构图注意力网络模块, 结构如图 2 的异构图注意力网络模块所示, 主要包含节点级注意力和语义级注意力.

(1) 节点级注意力. 节点级注意力主要学习实体及其邻近实体间的权重, 通过聚合工业设备故障诊断知识图谱中不同路径下邻居故障实体的语义信息来学习目标故障实体的嵌入表示, 以此学习不同路径下邻居故障实体对于目标故障实体的重要性.

由于工业设备故障诊断知识图谱的异构特性, 不同类型的故障实体具有不同的嵌入向量空间. 因此, 需要设计不同类型的投影变换矩阵将不同类型的故障实体的嵌入向量投影到相同的嵌入向量空间中. 投影变换定义为

$$x'_i = M_{\Phi} x_i, \quad (8)$$

其中,  $x'_i$  是投影后第  $i$  个故障实体的嵌入向量,  $M_{\Phi}$  是第  $i$  个故障实体所属类型的投影变换矩阵.

为了学习不同类型的故障实体之间的权重, 引入自注意力机制. 对于任一通过路径  $\Phi$  连接的故障实体对  $\langle i, j \rangle$ , 通过权重系数  $e_{i,j}^{\Phi}$  表示故障实体  $j$  对故障实体  $i$  的重要性. 权重系数  $e_{i,j}^{\Phi}$  定义为

$$e_{i,j}^{\Phi} = \text{attn}_{\text{node}}(x'_i, x'_j; \Phi), \quad (9)$$

其中,  $\text{attn}_{\text{node}}$  是计算节点级注意力的深层神经网络,  $\Phi$  是故障实体  $j$  到故障实体  $i$  的路径.

因此, 对于任一通过路径  $\Phi$  连接的故障实体对  $\langle i, j \rangle$ , 故障实体  $j$  对故障实体  $i$  的注意力系数  $\alpha_{i,j}^\Phi$  定义为

$$\alpha_{i,j}^\Phi = \text{softmax}_j (\sigma (e_{i,j}^\Phi)) = \frac{\exp(\sigma(a_\Phi^\top (M_\Phi x_i \parallel M_\Phi x_j)))}{\sum_{k \in N_i^\Phi} \exp(\sigma(a_\Phi^\top (M_\Phi x_i \parallel M_\Phi x_k)))}, \quad (10)$$

其中,  $\sigma(\cdot)$  是激活函数, 采用 LeakyReLU 非线性函数,  $N_i^\Phi$  为第  $i$  个故障实体的邻居故障实体集合,  $\exp(x)$  表示为数学运算  $e^x$ ,  $\parallel$  表示量拼接运算,  $a_\Phi^\top$  是映射向量的转置, 为模型的可训练参数.

对于任一路径  $\Phi$  下的故障实体  $i$ , 其嵌入向量  $z_i^\Phi$  由该路径下邻居故障实体的嵌入向量  $x'_j$  和相应的注意力系数  $\alpha_{i,j}^\Phi$  聚合而成. 由于异构图具有无标度的特性, 得到的故障实体嵌入向量的方差很大, 因此, 需要采用多头注意力机制使得训练过程更加稳定. 多头注意力机制分别对每个故障实体调用  $K$  次相互独立的节点级注意力运算, 得到故障实体  $i$  的嵌入向量  $z_i^\Phi$  为

$$z_i^\Phi = \parallel_{k=1}^K \sigma \left( \frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i^\Phi} \alpha_{i,j}^\Phi x'_j \right), \quad (11)$$

其中,  $\sigma(\cdot)$  是激活函数, 采用 ELU 非线性函数, 其目的是增强模型的表达能力.

综上, 对于工业设备故障诊断知识图谱的路径集合  $\{\Phi_1, \Phi_2, \dots, \Phi_P\}$ , 通过节点级注意力运算可以得到  $P$  组特定路径下故障实体的嵌入表示, 定义为  $\{Z_{\Phi_1}, Z_{\Phi_2}, \dots, Z_{\Phi_P}\}$ , 其中,  $P$  为路径个数.

(2) 语义级注意力. 工业设备故障诊断知识图谱中包含着不同的、复杂的、有意义的故障语义信息, 这些信息通常通过路径来反映, 不同的路径可以提取出不同的故障语义信息. 为了学习到更全面的故障实体嵌入表示, 引入了语义级注意力, 融合各条路径上蕴含的特定故障语义信息, 学习不同路径的重要性. 路径  $\Phi_i$  的权重  $\beta_{\Phi_i}$  可以反映路径  $\Phi_i$  的重要性, 定义为

$$\beta_{\Phi_i} = \text{attn}_{\text{sem}}(Z_{\Phi_i}), \quad \Phi_i \in \Lambda, \quad (12)$$

其中,  $\text{attn}_{\text{sem}}$  是计算语义级注意力的深层神经网络,  $\Lambda = \{\Phi_1, \Phi_2, \dots, \Phi_P\}$  是所有路径的集合,  $Z_{\Phi_i}$  是路径  $\Phi_i$  上故障实体的嵌入向量.

为了衡量每条路径的重要性, 首先通过变换矩阵  $W$  将每条路径上故障实体的嵌入表示变换到特定故障语义的嵌入空间中, 然后通过语义级注意力向量  $q$  综合考虑特定路径的故障语义嵌入表示. 对于任一路径  $\Phi_i$ , 其重要性指标  $w_{\Phi_i}$  定义为

$$w_{\Phi_i} = \frac{1}{|\Lambda|} \sum_{\Phi_i \in \Lambda} q^\top \tanh(W z_i^{\Phi_i} + b), \quad (13)$$

其中,  $|\Lambda|$  代表着工业设备故障诊断知识图谱中路径的数量,  $b$  是偏置.

为了定义更合理的梯度, 引入 softmax 函数对重要性指标进行归一化得到重要性权重  $\beta_{\Phi_i}$ , 以此获得更合理的概率分布. 对于任一路径  $\Phi_i$ , 其重要性权重  $\beta_{\Phi_i}$  定义为

$$\beta_{\Phi_i} = \frac{\exp(w_{\Phi_i})}{\sum_{i=1}^P \exp(w_{\Phi_i})}. \quad (14)$$

最后, 将学习到的路径重要性权重  $\beta_{\Phi_i}$  作为系数, 与路径  $\Phi_i$  上故障实体的嵌入表示  $Z_{\Phi_i}$  进行融合, 得到最终的故障实体嵌入表示  $Z$ :

$$Z = \sum_{i=1}^P \beta_{\Phi_i} Z_{\Phi_i}. \quad (15)$$

**Algorithm 1** The algorithm strategy for KGHAN

---

**Input:** Industrial equipment fault diagnosis knowledge graph  $G$ , fault entity set  $E$ , fault relation set  $R$ , fault knowledge triplet set  $F$ , path set  $\Lambda$ , and number of attention heads  $K$ ;

**Output:** Fault entity embedding set  $Z$ ;

- 1: Initialize all parameters;
- 2: **for**  $\langle h, r, t \rangle \in F$  **do**
- 3:   Encode the word vectors for the fault entities  $h$ ,  $t$ , and fault relation  $r$  using the one-hot function;
- 4: **end for**
- 5: **for**  $\langle h, r, t \rangle \in F$  **do**
- 6:   Sample a minibatch of true and false fault knowledge triplet from  $G$ ;
- 7:   Merge fault knowledge structure information;
- 8:   Update fault entity embeddings  $h$ ,  $t$ , and fault relation embedding  $r$ ;
- 9: **end for**
- 10: **for**  $\langle h, r, t \rangle \in F$  **do**
- 11:   Merge fault graph structure information;
- 12:   Update fault entity embeddings  $h$ ,  $t$ , and fault relation embedding  $r$ ;
- 13:   Integrate fault entity embedding  $x = \{h, t\}$ ;
- 14: **end for**
- 15: **for**  $\Phi \in \Lambda$  **do**
- 16:   **for**  $e \in E$  **do**
- 17:     Calculate the weight coefficient of fault entity  $e$  under path  $\Phi$  using  $e_{i,j}^\Phi = \text{attn}_{\text{node}}(x'_i, x'_j; \Phi)$ ;
- 18:     Calculate the attention coefficient of fault entity  $e$  under path  $\Phi$  using  $\alpha_{i,j}^\Phi = \text{softmax}_j(\text{LeakyReLU}(e_{i,j}^\Phi))$ ;
- 19:     Update the embedding  $z_i^\Phi$  of the fault entity  $e$  under path  $\Phi$  using the multi-head attention mechanism;
- 20:   **end for**
- 21:   Calculate the importance weight of the path  $\Phi$  using  $\beta_{\Phi_i} = \text{attn}_{\text{sem}}(Z_{\Phi_i})$ ;
- 22:   Get and update fault entity embedding set  $Z$ ;
- 23: **end for**

---

### 3.1.4 算法描述

本文提出了知识图谱异构图注意力网络 KGHAN 模型, 该模型在保留工业设备故障诊断知识图谱的故障知识结构信息和故障图结构信息的基础上, 通过异构图注意力网络模块同时考虑邻居故障实体和路径的重要性, 实现了工业设备故障诊断知识图谱中故障实体和故障关系的嵌入表示. 具体的算法描述如算法 1 所示.

## 3.2 工业设备故障诊断知识图谱补全任务设计

本小节在训练知识图谱异构图注意力网络 KGHAN 模型的基础上, 设计故障实体概念补全和故障关系链接补全两个任务, 解决故障实体属性残缺、故障关系链接缺失的问题, 实现工业设备故障诊断知识图谱补全.

### 3.2.1 故障实体概念补全任务设计

为了完成故障实体概念补全任务, 在获取工业设备故障诊断知识图谱中故障实体嵌入表示  $Z^l$  的基础上, 使用多层感知机 MLP 获得故障实体的预测类型标签的概率  $Y^l$ :

$$Y^l = \text{MLP}(\text{MLP}(\cdots \text{MLP}(Z^l))) = \text{MLP}^L(Z^l), \quad (16)$$

其中,  $\text{MLP}(Z^l) = \sigma(W_{\text{MLP}}Z^l + b_{\text{MLP}})$  是单层感知机,  $W_{\text{MLP}}$  是单层感知机的权重参数,  $b_{\text{MLP}}$  是偏置,  $\sigma(\cdot)$  是非线性激活函数.



将故障实体的预测类型标签概率  $Y^l$  与真实类型标签的交叉熵函数作为损失函数:

$$L_{EC} = - \sum_{l \in y_L} Y^l \ln(C \cdot Z^l), \quad (17)$$

其中,  $C$  是分类器的参数,  $y_L$  是训练集故障实体的索引集. 通过自适应矩估计优化器 Adam 最小化损失函数  $L_{EC}$  训练知识图谱异构图注意力网络 KGHAN 模型, 实现故障实体类型的预测, 完成工业设备故障诊断知识图谱的故障实体概念补全任务.

### 3.2.2 故障关系链接补全任务设计

故障关系链接补全任务的核心是预测出故障知识三元组中缺失的部分. 通过计算工业设备故障诊断知识图谱中故障实体之间的相似度选择特定的故障关系, 实现故障关系链接补全. 为了评价故障知识三元组中头实体和尾实体嵌入向量之间的相关性, 引入向量距离误差  $s_i$ , 定义故障知识三元组  $\langle h_i, r_i, t_i \rangle$  的信誉评分  $\text{score}_{ri}$  为

$$\text{score}_{ri}(s_i) = \|h_i + r_i - t_i\|_2^2, \quad (18)$$

其中,  $\|\cdot\|_2$  是向量的 2 范数, 用来表示两个故障实体向量之间的欧氏距离.

为了优化故障关系链接补全任务的训练过程, 在工业设备故障诊断知识图谱中对正确的故障知识三元组  $\langle h_i, r_i, t_i \rangle \in F$  进行随机头、尾实体替换以实现负样本采样. 错误的故障知识三元组定义为  $\langle h'_i, r_i, t'_i \rangle \in F'$ , 其向量距离误差为  $s'_i$ . 故障关系链接补全任务的损失函数采用 Margin 损失函数, 定义为

$$L_{LP} = \sum_{s \in F} \sum_{s' \in F'} \max(0, (\gamma + \text{score}_{ri}(s'_i) - \text{score}_{ri}(s_i))), \quad (19)$$

其中,  $\gamma$  是正确的故障知识三元组得分和错误的故障知识三元组得分的裕度. 通过自适应矩估计优化器 Adam 最小化损失函数  $L_{LP}$  训练知识图谱异构图注意力网络 KGHAN 模型, 实现故障知识三元组中缺失部分的预测, 完成工业设备故障诊断知识图谱的故障关系链接补全任务.

## 4 实验结果与分析

为了验证本文提出的基于知识图谱异构图注意力网络 KGHAN 模型的工业设备故障诊断知识图谱补全方法的有效性, 使用了 4 个工业设备故障维修数据集构建了 4 个不同的工业设备故障诊断知识图谱, 并与多种方法进行了对比.

### 4.1 工业设备故障诊断知识图谱数据集

本文处理了国内某企业数控机床设备故障维修记录、上汽工业产线故障维修记录、国产某型号挖掘机故障工单和实用数控机床故障诊断及维修技术手册<sup>1)</sup>, 保留了与故障实体以及故障关系相关的语义信息, 形成了 4 个工业设备故障诊断知识图谱数据集, 分别命名为 FDI-CNCM, FDI-CAR, FDI-EXC, FDI-CNC, 数据集的精确信息统计如表 1 所示. 4 个工业设备故障诊断知识图谱数据集的详细介绍如下所示:

(1) 数据集 FDI-CNCM 主要描述了国内某企业数控机床设备在工业生产活动中机床设备的故障情况、目前状态等故障信息;

1) 此公开数据集来自第八届中国软件杯大赛“基于移动端在线设备故障诊断平台”的参赛作品, 公开数据集的地址链接详见 Github: <https://github.com/wangrenyisme/Shukongdashi>.

表 1 工业设备故障诊断知识图谱数据集精确信息统计

Table 1 The accurate information statistics of industrial equipment fault diagnosis knowledge graph datasets

Dataset	Entities	Relations	Entity types	Relation types	Train	Validation	Test
FDI-CNCM	9238	9943	3	2	5541	1848	1849
FDI-CAR	9146	12376	5	4	5490	1830	1826
FDI-EXC	3824	7281	7	16	2298	765	761
FDI-CNC	1971	5346	5	7	1185	394	392

(2) 数据集 FDI-CAR 主要描述了上汽工业产线在工业生产活动中机械臂的故障现象、3 级故障原因、维修方法等故障信息;

(3) 数据集 FDI-EXC 主要描述了国产某型号挖掘机在工业生产活动中零部件的故障现象、故障原因、故障程度、故障位置、维修方法、辅助工具等故障信息;

(4) 数据集 FDI-CNC 主要描述了实用数控机床在工业生产活动中机床设备的故障位置、故障现象、故障操作、故障原因等故障信息.

#### 4.2 故障实体概念补全实验过程

(1) 实验评价指标. 为了量化模型性能, 本文采用准确率 Accuracy 作为故障实体概念补全实验的性能评价指标. 准确率 Accuracy 用于描述故障实体概念预测正确的比例, 定义为

$$\text{Accuracy} = \frac{N_{\text{True}}}{N_{\text{Total}}}, \quad (20)$$

其中,  $N_{\text{True}}$  表示正确预测概念类型的故障实体个数,  $N_{\text{Total}}$  表示所有故障实体的个数.

(2) 参数设置. 为了保证实验的公平性, 限定工业设备故障诊断知识图谱中故障实体的最终嵌入表示为 64 维. 为了避免在训练过程中出现过拟合现象, 利用 L2 正则化对故障实体的嵌入表示进行约束, 使故障实体特征的先验分布满足高斯分布, 保证模型的稳定性和参数的平衡性. 同时使用 dropout 策略, 忽略一部分特征检测器, 可以减少特征检测器间的相互作用, 进一步减少过拟合现象. 设置注意力头数  $K$  为 8, 隐藏层单元数为 8, 特征层的 dropout 比率为 0.1, 注意力层的 dropout 比率为 0.1, 学习速率为 0.001, 权重衰减率为 0.001, 最大训练次数为 200.

(3) 实验结果. 实验基于 4 个工业设备故障维修数据集构建了 4 个不同的工业设备故障诊断知识图谱, 分别使用准确率指标 Accuracy 对各种模型的性能进行评价, 准确率指标 Accuracy 如表 2 所示, 故障实体概念补全任务训练过程展示如图 3 所示. 故障实体概念补全结果如图 4~7 所示. 为了评价本文设计的故障知识结构信息融合模块和故障图结构信息融合模块是否有效, 实验中分别对仅使用故障知识结构信息融合模块的 KGHAN-cat 模型、仅使用故障图结构信息融合模块的 KGHAN-deg 模型、同时使用故障知识结构信息融合模块和故障图结构信息融合模块的 KGHAN-catdeg 模型进行了验证.

(4) 实验分析. 结合表 2 的数据可知, 相较于其他基准模型, 本文所提的 KGHAN 模型使故障实体概念补全任务的准确率提高了约 10%; 对比图 4~7 的 (f)~(h) 可知, 相比于 HAN 模型, 设计的故障知识结构信息融合模块和故障图结构信息融合模块使 KGHAN 模型在故障实体概念补全任务上的准确率提高了约 7%, 表明本文设计的故障知识结构信息融合模块和故障图结构信息融合模块强化了 HAN 模型的表达能力. 同时, 对比 4 个工业设备故障维修数据集发现, 工业设备故障诊断知识图谱的故障实体类型和故障关系类型越多, 本文所提方法对故障实体概念补全任务准确率的提升就越大. 除此之外, 分析图 3 可知, 本文所提的 KGHAN 模型只需要 50 次训练就能达到平稳状态, 其训练速度明显优

表 2 故障实体概念补全任务准确率指标

Table 2 The accuracy for the fault entity concept completion task

Dataset	FDI-CNCM	FDI-CAR	FDI-EXC	FDI-CNC
GCN	0.9822	0.6851	0.6058	0.6046
GAT	0.8740	0.7207	0.7070	0.7372
GATv2	0.9114	0.7251	0.7346	0.7577
GraphSAGE	0.9998	0.7306	0.7569	0.7730
HAN	0.9989	0.7913	0.7898	0.8929
KGHAN-cat	1.0000	0.8959	0.8870	0.9847
KGHAN-deg	0.9996	0.9819	0.9185	0.9592
KGHAN-catdeg	1.0000	0.9847	0.9317	0.9923

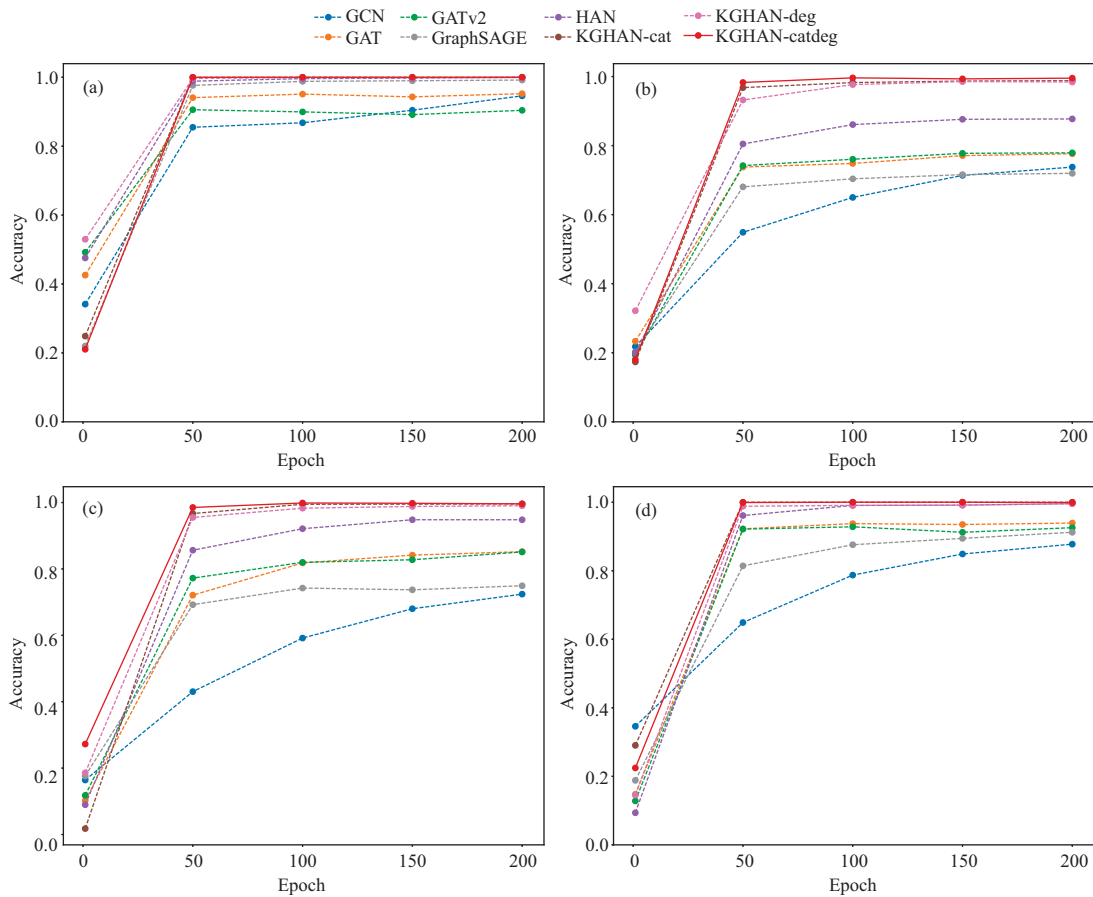


图 3 (网络版彩图) 故障实体概念补全任务训练过程

Figure 3 (Color online) Training process for the fault entity concept completion task. (a) FDI-CNCM dataset; (b) FDI-CAR dataset; (c) FDI-EXC dataset; (d) FDI-CNC dataset

于其他模型, 表明本文所提方法的快速性. 据分析, 本文所提的 KGHAN 模型在故障实体概念补全任务的性能上优于其他基准模型的原因在于:

- (1) 设计的故障知识结构信息融合模块能够保留工业设备故障诊断知识图谱的故障知识结构信息,

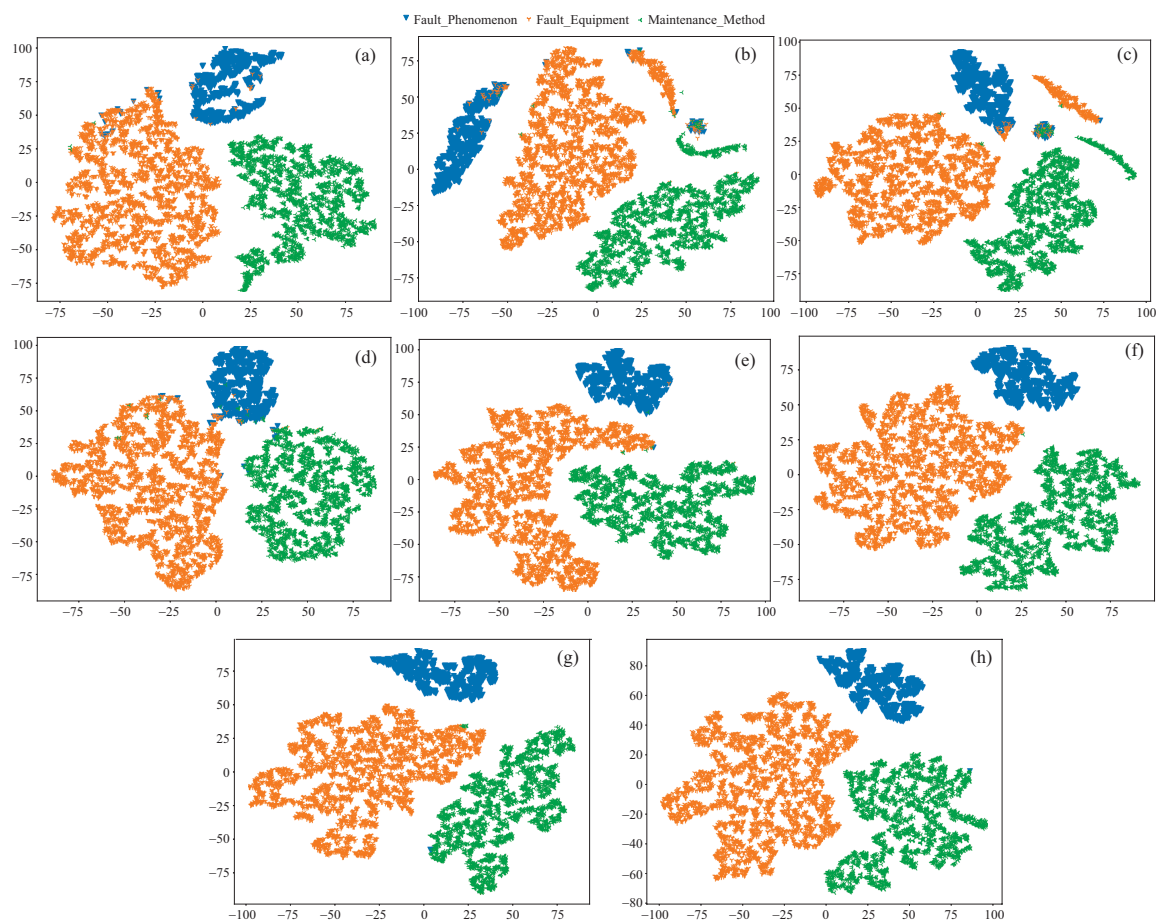


图 4 (网络版彩图) 数据集 FDI-CNCM 故障实体概念补全任务结果

**Figure 4** (Color online) Results for the fault entity concept completion task on the FDI-CNCM dataset. (a) GCN; (b) GAT; (c) GATv2; (d) GraphSAGE; (e) HAN; (f) KGHAN-cat; (g) KGHAN-deg; (h) KGHAN-catdeg

不仅考虑到了故障关系的多样性,也考虑到了故障实体的多样性;

(2) 设计的故障图结构信息融合模块能够保留工业设备故障诊断知识图谱的故障图结构信息,可以充分获得异构有向属性拓扑图的固有特征;

(3) 设计的异构图注意力网络模块综合考虑了邻居故障实体和元路径的重要性,充分学习到了故障实体和故障关系的深层表达。

综上,本文所提方法能够较好地推导出故障实体概念,更适合完成实体概念补全任务。

### 4.3 故障关系链接补全实验过程

(1) 实验评价指标. 为了量化模型性能,本文采用平均倒数排名 MRR 和命中率 Hits@n 评分作为故障关系链接补全实验的性能评价指标。

平均倒数排名指标 MRR 关心的是预测的故障关系链接是否摆在所有故障关系最前面的位置,强调故障关系链接的顺序性,定义为

$$\text{MRR} = \frac{1}{|F|} \sum_{i=1}^{|F|} \frac{1}{\text{rank}_i}, \quad (21)$$

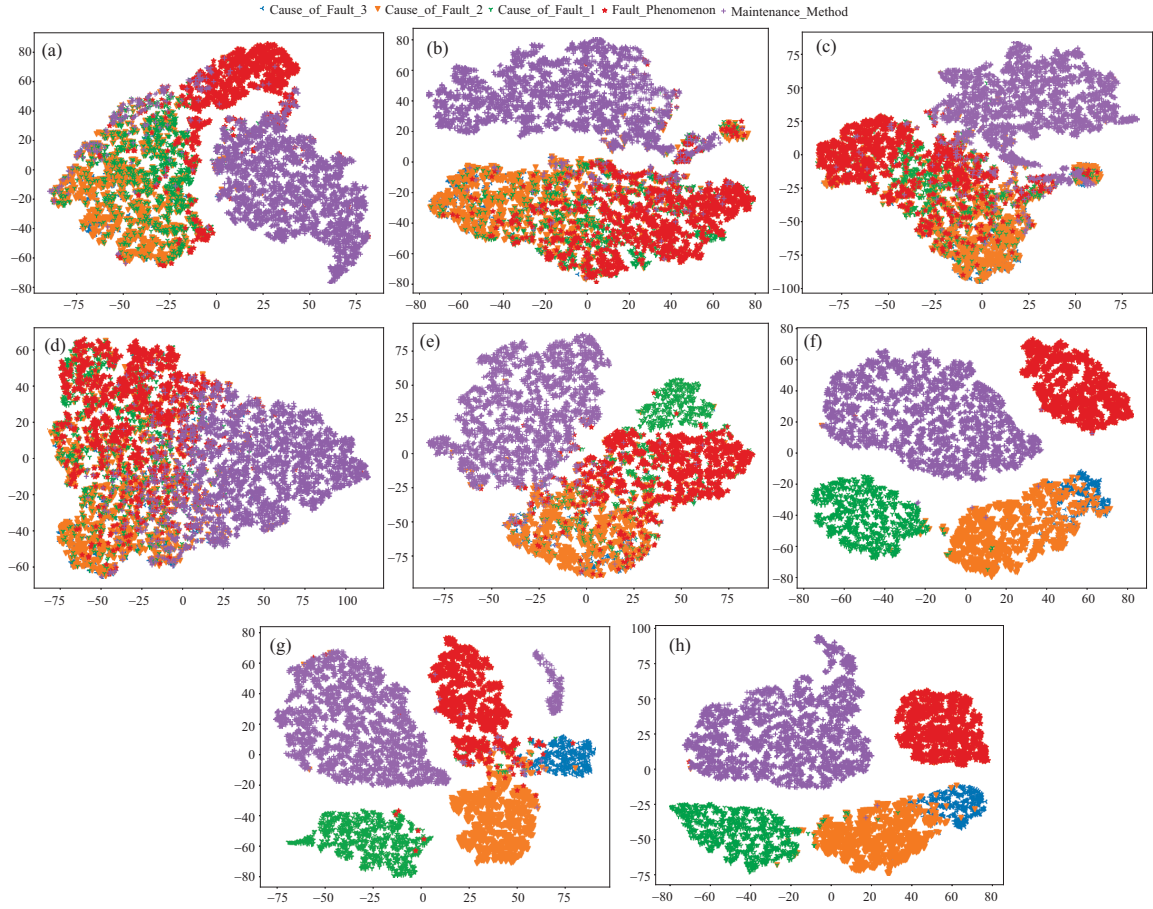


图 5 (网络版彩图) 数据集 FDI-CAR 故障实体概念补全任务结果

Figure 5 (Color online) Results for the fault entity concept completion task on the FDI-CAR dataset. (a) GCN; (b) GAT; (c) GATv2; (d) GraphSAGE; (e) HAN; (f) KGHAN-cat; (g) KGHAN-deg; (h) KGHAN-catdeg

其中,  $\text{rank}_i$  是第  $i$  个故障关系链接预测排名,  $|F|$  表示故障关系总个数.

命中率指标  $\text{Hits}@n$  是指在故障关系链接预测中排名小于等于  $n$  的故障关系链接的平均占比, 定义为

$$\text{Hits}@n = \frac{1}{|F|} \sum_{i=1}^{|F|} \text{Index}(\text{rank}_i \leq n), \quad (22)$$

其中, 当故障关系链接预测排名小于等于  $n$  时,  $\text{Index}(\text{rank}_i \leq n)$  为 1, 否则为 0.

(2) 参数设置. 为了保证实验的公平性, 在故障关系链接补全任务中, 限定工业设备故障诊断知识图谱中故障实体和故障关系的嵌入表示为 128 维, 设置注意力头数  $K$  为 8, 隐藏层单元数为 8, 特征层 dropout 比率为 0.1, 注意力层 dropout 比率为 0.1, 学习速率为 0.001, 权重衰减率为 0.001, Margin 损失函数的裕度  $\gamma$  为 1.0, 最大训练次数为 600.

(3) 实验结果. 分别对数据集 FDI-CNCM, FDI-CAR, FDI-EXC 和 FDI-CNC 进行故障关系链接预测实验, 完成工业设备故障诊断知识图谱的故障关系链接补全任务. 故障关系链接预测实验的评价指标如表 3 所示.

(4) 实验分析. 由表 3 可知, 相比于其他基准模型, 本文所提的 KGHAN 模型使 4 个工业设备故

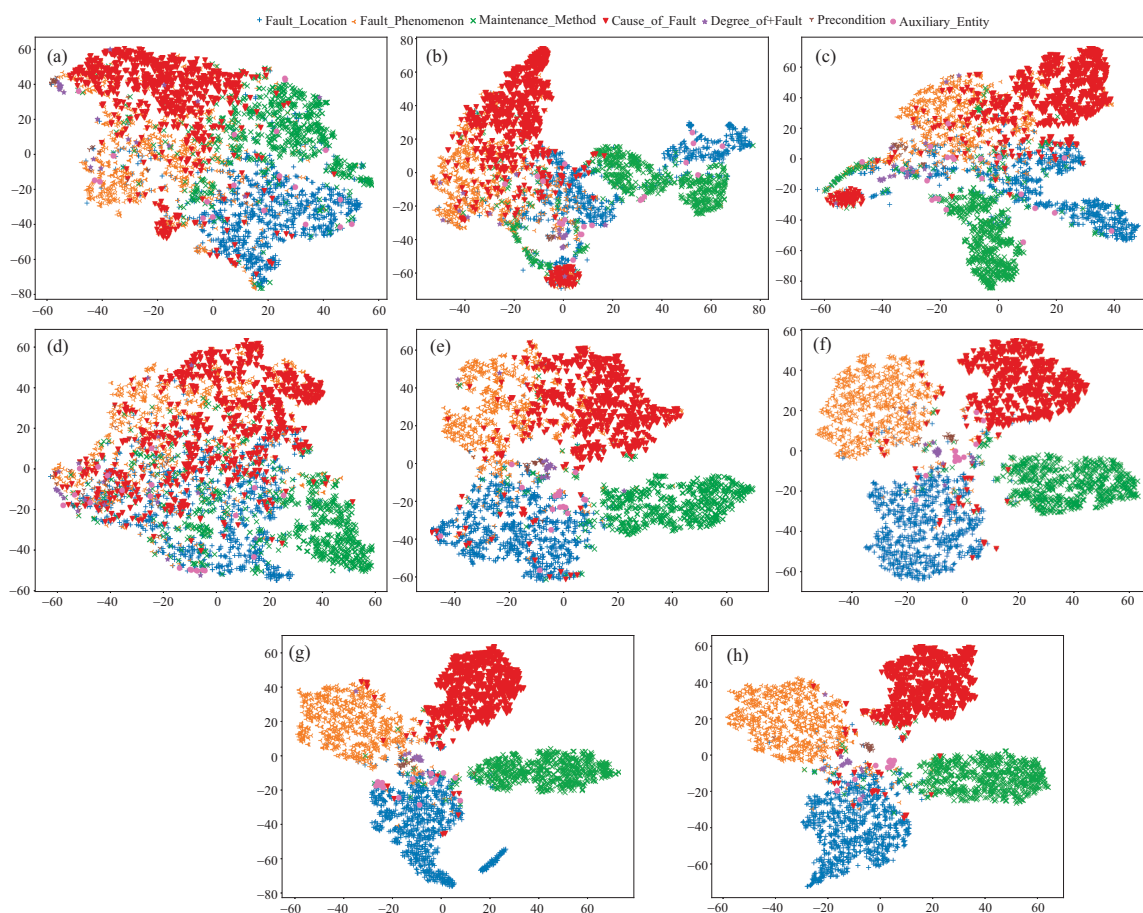


图 6 (网络版彩图) 数据集 FDI-EXC 故障实体概念补全任务结果

**Figure 6** (Color online) Results for the fault entity concept completion task on the FDI-EXC dataset. (a) GCN; (b) GAT; (c) GATv2; (d) GraphSAGE; (e) HAN; (f) KGHAN-cat; (g) KGHAN-deg; (h) KGHAN-catdeg

障诊断知识图谱数据集的故障关系链接补全任务的命中率  $\text{Hits}@n$  指标提高了约 37%。明确的说,使得故障关系链接预测任务的命中率  $\text{Hits}@3$  指标提高了约 37%, 命中率  $\text{Hits}@5$  指标提高了约 39%, 命中率  $\text{Hits}@10$  指标提高了约 34%。同时,平均倒数排名 MRR 指标相较于基准模型也有所提升,约为 15%,说明已有部分故障关系链接预测的排名有所提升。对比 4 个工业设备故障维修数据集发现,工业设备故障诊断知识图谱的故障实体类型和故障关系类型越多,本文所提方法对故障关系链接补全任务的平均倒数排名 MRR 和命中率  $\text{Hits}@n$  的提升就越大。

因此,本文所提的 KGHAN 模型在故障关系链接补全任务的性能上优于其他基准模型,能够较好地推导出故障关系链接,更适合完成对工业设备故障诊断知识图谱的故障关系链接补全任务。但是,数据集 FDI-EXC 在故障关系链接补全任务中 MRR 指标和  $\text{Hits}@n$  指标较低,原因在于数据集 FDI-EXC 中存在大量的故障实体与故障关系拥有多种不同的类型标签,给故障关系链接预测带来较大的误差。

#### 4.4 对比与分析

基于知识图谱异构图注意力网络模型的工业设备故障诊断知识图谱补全方法采用了模块并行串接框架,考虑了工业设备故障诊断知识图谱复杂多元关系。本文所提模型不仅考虑到了故障实体和故

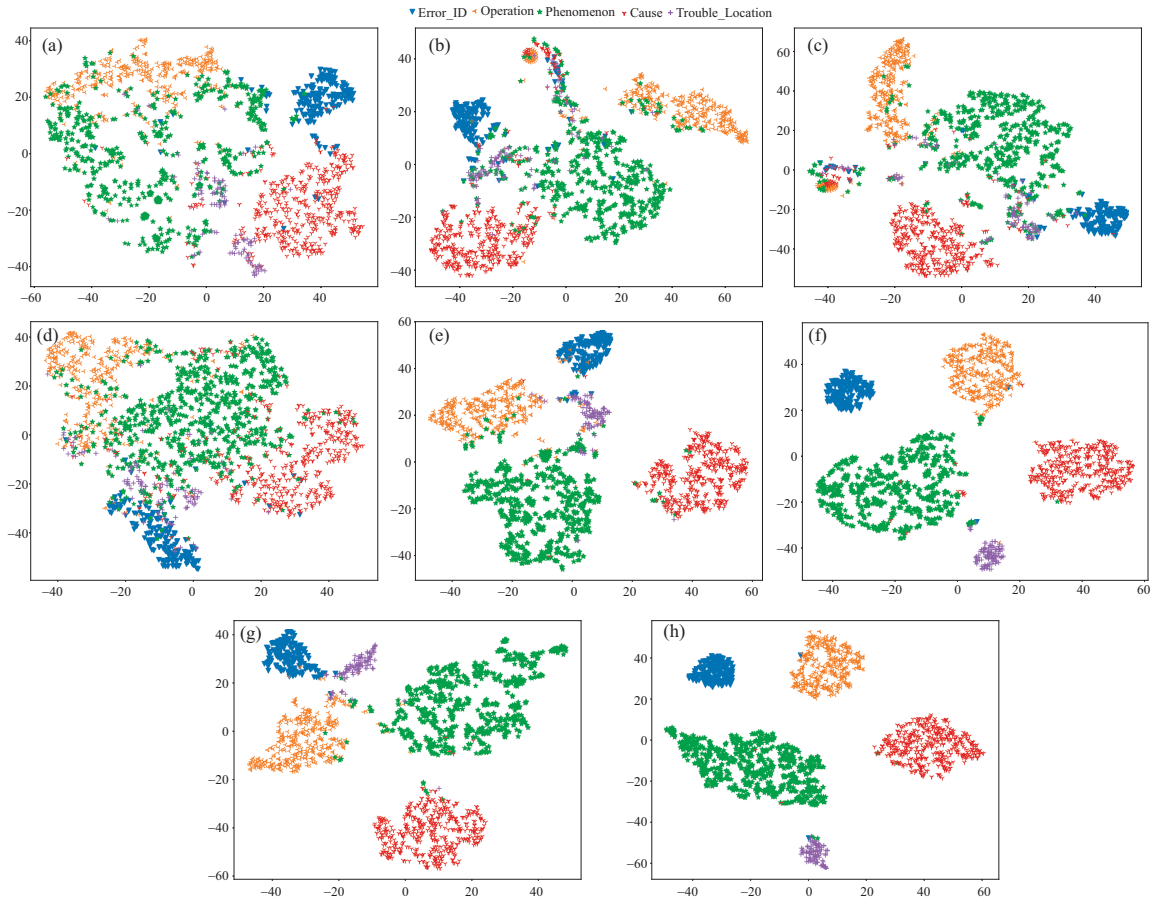


图 7 (网络版彩图) 数据集 FDI-CNC 故障实体概念补全任务结果

Figure 7 (Color online) Results for the fault entity concept completion task on the FDI-CNC dataset. (a) GCN; (b) GAT; (c) GATv2; (d) GraphSAGE; (e) HAN; (f) KGHAN-cat; (g) KGHAN-deg; (h) KGHAN-catdeg

障关系的多样性, 而且充分利用了邻居故障实体和元路径的重要性, 除此之外, 还充分获取了知识图谱的固有特征, 以此来学习故障实体和故障关系的深层表达.

该设计方式使得故障实体概念补全任务的准确率指标达到了约 95%, 同时只需要 50 次训练就能达到平稳状态; 使得故障关系链接补全任务的命中率 Hits@3 指标达到了约 64%, 命中率 Hits@5 指标达到了约 77%, 命中率 Hits@10 指标达到了约 86%. 同时, 平均倒数排名 MRR 指标相较于基准模型也有所提升, 约为 0.34, 说明已有部分故障关系链接预测的排名有所提升.

然而, 知识图谱异构图注意力网络 KGHAN 模型在设计上只关注工业设备故障诊断知识图谱中故障实体的局部特征, 获取其全局特征的效果不佳, 无法进一步提升故障实体概念补全任务的准确率以及故障关系链接补全任务的命中率和平均倒排名指标.

## 5 结论

本文面向工业设备故障诊断领域提出了一种改进的知识图谱补全方法, 有效地解决了故障实体属性残缺、故障关系链接缺失的问题, 实现了工业设备故障诊断知识图谱的故障实体概念补全和故障关

表3 故障关系链接补全任务评价指标

Table 3 The evaluations for the fault relation link completion task

Dataset	Model	MRR	Hits@3	Hits@5	Hits@10
FDI-CNCM	TransE	0.2647	0.2907	0.3520	0.4198
	TransH	0.2629	0.2830	0.3330	0.3860
	TransD	0.2148	0.2280	0.2630	0.3160
	HAN	0.1552	0.2150	0.2700	0.3840
	KGHAN-catdeg	0.3817	0.7760	0.8550	0.9120
FDI-CAR	TransE	0.0851	0.0850	0.0950	0.1060
	TransH	0.0751	0.0700	0.0910	0.1200
	TransD	0.0523	0.0520	0.0700	0.0910
	HAN	0.1251	0.1600	0.2440	0.3580
	KGHAN-catdeg	0.3477	0.6660	0.7970	0.8640
FDI-EXC	TransE	0.2299	0.2470	0.3200	0.3990
	TransH	0.2273	0.2480	0.3060	0.4140
	TransD	0.2251	0.2380	0.3020	0.4290
	HAN	0.1426	0.1730	0.2880	0.4530
	KGHAN-catdeg	0.2482	0.4130	0.5460	0.7090
FDI-CNC	TransE	0.2269	0.2685	0.3700	0.4493
	TransH	0.3628	0.4170	0.4920	0.5900
	TransD	0.4122	0.4490	0.5000	0.5990
	HAN	0.3190	0.5280	0.7300	0.9050
	KGHAN-catdeg	0.3728	0.7060	0.8940	0.9680

系链接补全. 所提知识图谱异构图注意力网络 KGHAN 模型保留了工业设备故障诊断知识图谱拥有的故障知识结构信息和故障图结构信息, 同时, 采用了 HAN 模型作为异构图注意力网络模块, 共同考虑了邻居故障实体和路径的重要性. 所提工业设备故障诊断知识图谱补全方法在深度挖掘故障实体属性与故障实体间潜在关系的基础上, 对工业设备故障诊断知识图谱进行推理与动态推断, 有效地提升了故障实体概念补全任务的准确率和故障关系链接补全任务的命中率.

本文采用图注意力机制构建知识图谱异构图注意力网络 KGHAN 模型, 然而, 图注意力机制是注意力机制的简化变形版本, 在设计上只关注故障实体的局部特征, 获取其全局特征的效果不佳, 无法进一步提升故障实体概念补全任务的准确率以及故障关系链接补全任务的命中率和平均倒排名指标. 因此, 本文后续将基于完全注意力机制开展工业设备故障诊断知识图谱补全方法的研究.

## 参考文献

- 1 Huang D R, Hua X X, Mi B, et al. Incipient fault diagnosis on active disturbance rejection control. *Sci China Inf Sci*, 2022, 65: 199202
- 2 Huang D R, Li S Q, Zhang Z Y, et al. Design and analysis of longitudinal controller for the platoon with time-varying delay. *IEEE Trans Intell Transp Syst*, 2022, 23: 23628–23639
- 3 Zhou Y C, Wang W J, Qiao Z Y, et al. A survey on the construction methods and applications of sci-tech big data knowledge graph. *Sci Sin Inform*, 2020, 50: 957–987 [周园春, 王卫军, 乔子越, 等. 科技大数据知识图谱构建方法及应用研究综述. *中国科学: 信息科学*, 2020, 50: 957–987]



- 4 Tian L, Zhang J C, Zhang J H, et al. Knowledge graph survey: representation, construction, reasoning and knowledge hypergraph theory. *J Comput Appl*, 2021, 41: 2161–2186 [田玲, 张谨川, 张晋豪, 等. 知识图谱综述 – 表示、构建、推理与知识超图理论. *计算机应用*, 2021, 41: 2161–2186]
- 5 Zhang B, Zhu J, Su H. Toward the third generation of artificial intelligence. *Sci Sin Inform*, 2020, 50: 1281–1302 [张钹, 朱军, 苏航. 迈向第三代人工智能. *中国科学: 信息科学*, 2020, 50: 1281–1302]
- 6 Qin C, Zhu H S, Zhuang F Z, et al. A survey on knowledge graph-based recommender systems. *Sci Sin Inform*, 2020, 50: 937–956 [秦川, 祝恒书, 庄福振, 等. 基于知识图谱的推荐系统研究综述. *中国科学: 信息科学*, 2020, 50: 937–956]
- 7 Wang J P, Zhang W S, Wang Y F, et al. Constructing and inferring event logic cognitive graph in the field of big data. *Sci Sin Inform*, 2020, 50: 988–1002 [王军平, 张文生, 王勇飞, 等. 面向大数据领域的事理认知图谱构建与推断分析. *中国科学: 信息科学*, 2020, 50: 988–1002]
- 8 Gui W H, Zeng Z H, Chen X F, et al. Knowledge-driven process industry smart manufacturing. *Sci Sin Inform*, 2020, 50: 1345–1360 [桂卫华, 曾朝晖, 陈晓方, 等. 知识驱动的流程工业智能制造. *中国科学: 信息科学*, 2020, 50: 1345–1360]
- 9 Liu R H, Xie G Q, Yuan Z G, et al. Research on intelligent fault diagnosis based on knowledge graph. *Desig Techn Post Telecommun*, 2020, 10: 30–35 [刘瑞宏, 谢国强, 苑宗港, 等. 基于知识图谱的智能故障诊断研究. *邮电设计技术*, 2020, 10: 30–35]
- 10 Xiao F L, Wu Y Z, Shen X H, et al. Intelligent fault diagnosis of substation equipment on the basis of deep learning and knowledge graph. *Electric Power Construction*, 2022, 43: 66–74 [肖发龙, 吴岳忠, 沈雪豪, 等. 基于深度学习和知识图谱的变电站设备故障智能诊断. *电力建设*, 2022, 43: 66–74]
- 11 Liu Q, Chai T Y, Qin S Z, et al. Progress of data-driven and knowledge-driven process monitoring and fault diagnosis for industry process. *Control Decision*, 2010, 25: 801–807, 813 [刘强, 柴天佑, 秦泗钊, 等. 基于数据和知识的工业过程监视及故障诊断综述. *控制与决策*, 2010, 25: 801–807, 813]
- 12 Qu Z H, Hu J P, Huang Z Q, et al. Research on the construction and application of knowledge graph for industrial equipment fault disposal. *Comput Eng Appl*, 2022. doi: 10.3778/j.issn.1002-8331.2208-0186 [瞿智豪, 胡建鹏, 黄子麒, 等. 工业设备故障处置知识图谱构建与应用研究. *计算机工程与应用*, 2022. doi: 10.3778/j.issn.1002-8331.2208-0186]
- 13 Feng Y, Zhai F, Li B F, et al. Research on intelligent fault diagnosis of power acquisition based on knowledge graph. In: *Proceedings of the 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)*, Xiamen, 2020. 1737–1740
- 14 Wang S, Du Z J, Meng X F. Research progress of large-scale knowledge graph completion technology. *Sci Sin Inform*, 2020, 50: 551–575 [王硕, 杜志娟, 孟小峰. 大规模知识图谱补全技术的研究进展. *中国科学: 信息科学*, 2020, 50: 551–575]
- 15 Luo M Q, Zhang C X, Peng C, et al. Knowledge graph completion based on parsing graph embedding and a weighted graph convolutional network. *Sci Sin Inform*, 2022, 52: 2037–2057 [罗妹秋, 张春霞, 彭成, 等. 基于解析图嵌入和加权图卷积网络的知识图谱补全. *中国科学: 信息科学*, 2022, 52: 2037–2057]
- 16 Zhang J, Chen B, Zhang L, et al. Neural, symbolic and neural-symbolic reasoning on knowledge graphs. *AI Open*, 2021, 2: 14–35
- 17 Wang H, Lin H Z, Lu L Y. Knowledge graph inference algorithm based on Att\_GCN model. *Comput Eng Appl*, 2020, 56: 183–189 [王红, 林海舟, 卢林燕. 基于 Att\_GCN 模型的知识图谱推理算法. *计算机工程与应用*, 2020, 56: 183–189]
- 18 Bordes A, Usunier N, Garcia-Durán A, et al. Translating embeddings for modeling multi-relational data. In: *Proceedings of the 26th International Conference on Neural Information Processing Systems*. Lake Tahoe: Curran Associates Inc., 2013. 2787–2795
- 19 Wang Z, Zhang J W, Feng J L, et al. Knowledge graph embedding by translating on hyperplanes. In: *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, 2014. 1112–1119
- 20 Lin Y K, Liu Z Y, Sun M S, et al. Learning entity and relation embeddings for knowledge graph completion. In: *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, 2015. 2181–2187
- 21 Ji G L, He S Z, Xu L H, et al. Knowledge graph embedding via dynamic mapping matrix. In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, 2015. 687–696
- 22 Ji G L, Liu K, He S Z, et al. Knowledge graph completion with adaptive sparse transfer matrix. In: *Proceedings of*

- the 30th AAAI Conference on Artificial Intelligence, 2016. 985–991
- 23 Sun Z Q, Deng Z H, Nie J Y, et al. RotatE: knowledge graph embedding by relational rotation in complex space. 2019. ArXiv:1902.10197
- 24 Yang B S, Yih W T, He X D, et al. Embedding entities and relations for learning and inference in knowledge bases. 2014. ArXiv:1412.6575
- 25 Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning, 2016. 2071–2080
- 26 Nickel M, Rosasco L, Poggio T. Holographic embeddings of knowledge graphs. In: Proceedings of the 30th AAAI Conference on Artificial Intelligence, 2016. 1955–1961
- 27 Kazemi S M, Poole D. Simple embedding for link prediction in knowledge graphs. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, 2018. 4289–4300
- 28 Rao Z Y, Zhang Y, Liu J T, et al. Recommendation methods and systems using knowledge graph. *Acta Automa Sin*, 2021, 47: 2061–2077 [饶子昀, 张毅, 刘俊涛, 等. 应用知识图谱的推荐方法与系统. *自动化学报*, 2021, 47: 2061–2077]
- 29 Xu K, Hu W H, Leskovec J, et al. How powerful are graph neural networks? 2018. ArXiv:1810.00826
- 30 Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. 2016. ArXiv:1609.02907
- 31 Xu D, Ruan C W, Korpeoglu E, et al. Inductive representation learning on temporal graphs. 2020. ArXiv:2002.07962
- 32 Veličković P, Cucurull G, Casanova A, et al. Graph attention networks. 2017. ArXiv:1710.10903
- 33 Wang X, Ji H Y, Shi C, et al. Heterogeneous graph attention network. In: Proceedings of the World Wide Web Conference, 2019. 2022–2032
- 34 Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks. In: Proceedings of European Semantic Web Conference, 2018. 593–607
- 35 Wang X, He X N, Cao Y X, et al. KGAT: knowledge graph attention network for recommendation. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019. 950–958
- 36 Ji S X, Pan S R, Cambria E, et al. A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Trans Neural Netw Learn Syst*, 2022, 33: 494–514
- 37 Ying C X, Cai T L, Luo S J, et al. Do Transformers really perform bad for graph representation? 2021. ArXiv:2106.05234

## Knowledge graph completion method for industrial equipment fault diagnosis based on heterogeneous graph attention

Yajie MA<sup>1\*</sup>, Yang LIU<sup>1</sup>, Bin JIANG<sup>1</sup>, Zehui MAO<sup>1</sup>, Jianwei LIU<sup>1</sup> & Wenjing LIU<sup>2,3</sup>

1. *College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;*

2. *Beijing Institute of Control Engineering, Beijing 100190, China;*

3. *Science and Technology on Space Intelligent Control Laboratory, Beijing 100190, China*

\* Corresponding author. E-mail: yajiema@nuaa.edu.cn

**Abstract** To address the issue of a serious lack of fault entity attributes and fault relation links in an industrial equipment fault diagnosis knowledge graph, this paper develops an industrial equipment fault diagnosis knowledge graph completion scheme based on a knowledge graph heterogeneous graph attention network (KGHAN) model. By combining fault knowledge structure information and fault graph structure information in a heterogeneous graph attention network (HAN) model, the developed KGHAN model effectively represents the embedding representations of fault entities and fault relations, which enhances the accuracy of the fault entity concept completion task and the hit rate of the fault relation link completion task. We apply our developed KGHAN model-based industrial equipment fault diagnosis knowledge graph completion scheme to the industrial equipment fault operation and maintenance data of a local enterprise. The results show that the accuracy of the fault entity concept completion task and the hit rate of the fault relation link completion task increased by about 10% and 37%, respectively, which confirms the effectiveness of our method.

**Keywords** knowledge graph completion, knowledge graph, graph neural network, fault diagnosis, industrial equipment