

# 基于联合嵌入的工业设备故障知识图谱推理诊断方法

张永伟<sup>1\*</sup> 罗 川<sup>1</sup> 张迎春<sup>1</sup>

ZHANG Yongwei LUO Chuan ZHANG Yingchun

## 摘 要

工业设备故障诊断是保障生产连续性、降低运维成本的核心环节，在多个子系统构成的工业设备管理系统中，知识共享与协作对故障诊断至关重要。然而，传统故障诊断方法局限于封闭系统，忽略开放世界实体信息，存在依赖专家经验、数据利用率低、复杂场景适应性差等问题，且难以有效学习稀疏故障知识的实体嵌入表示。针对上述问题，文章提出一种基于联合嵌入的工业设备故障知识图谱推理诊断方法。该方法将大模型生成的逻辑规则注入实体与关系中，并以大模型获取的开放世界知识为导向进行联合嵌入，以提升故障诊断性能。实验结果表明，在构建的工业故障知识图谱推理数据集与通用知识图谱数据集上，该方法性能显著优于现有方法，验证了其在复杂工业系统故障诊断中的有效性。

## 关键词

大模型；知识图谱；故障诊断；联合嵌入

doi: 10.3969/j.issn.1672-9528.2026.06.003

## 0 引言

随着工业互联网、物联网等技术的广泛应用，工业设备运维系统的复杂度呈指数级攀升。设备运行过程中产生的海量多源异构数据，如传感器监测数据、日志记录、维修报告等蕴含着丰富的故障信息。这些数据具有非结构化、碎片化的特点，导致传统方法难以提取有价值的知识。这些数据中隐藏的故障关联模式、潜在故障原因等关键信息，有待通过更先进的技术方法实现设备故障的快速定位和精准诊断。因此，如何实现高效、精准的设备故障诊断，已成为保障设备稳定运行，推动工业智能化转型的核心问题。

传统的故障诊断方法，如基于规则的推理（RBR）<sup>[1]</sup>和基于案例的推理（CBR）<sup>[2]</sup>，在相对封闭的工业系统中曾发挥过一定作用。然而，在开放的场景中，面对设备故障知识的海量性、动态性与复杂性，上述方法的局限性显著。一是依赖预定义规则或历史案例，难以适配设备故障模式的持续演变与新型故障类型，也无法充分利用开放世界中丰富的实体信息及潜在知识关联；二是设备运维系统的故障数据普遍存在稀疏性问题，使得传统机器学习方法在学习稀疏实体的有效嵌入表示时面临巨大挑战，进而导致诊断模型的泛化能力与准确性降低<sup>[3]</sup>。

知识图谱技术的兴起，为设备故障诊断带来了新的发展方向。然而，在处理工业领域设备故障知识时，仍存在自动化程度、知识抽取准确率、推理效率等均较低的问题<sup>[4]</sup>。近

年来，以 GPT（generative pretrained transformer）系列为代表的大语言模型（LLM）凭借其在大规模无监督数据上预训练所获得的强大语义理解、知识生成与跨领域迁移能力，在自然语言处理、知识问答等领域取得了突破性进展，也为解决工业设备故障诊断中的知识获取与推理难题提供了新的思路。文献[5]探讨了大模型与知识图谱结合在工业智能中的应用潜力，为后续研究奠定了基础。文献[6]提出利用大模型生成设备故障描述文本的语义向量，辅助知识图谱的实体对齐与关系抽取，但该方法未充分考虑工业领域知识的专业性与复杂性，知识图谱的构建质量有待提高。Luo 等<sup>[7]</sup>运用大模型对故障案例文本进行解析，生成初步的故障知识图谱，但在推理过程中未能有效利用知识图谱的结构化信息，推理准确性与可解释性不足。此外，文献[8]尝试通过大模型优化故障知识图谱的查询效率，但在复杂故障场景下的泛化能力较弱；文献[9]基于大模型进行故障知识图谱的动态更新，却难以处理非结构化数据中的隐含知识；文献[10]探索利用大模型进行跨领域故障知识迁移，然而在知识图谱的领域适配方面存在显著误差。

当前在实现大模型与工业故障知识图谱的深度融合、构建高效可靠的故障推理机制方面仍存在较大提升空间，尤其在解决工业系统中知识稀疏与开放世界知识利用问题上，尚未形成成熟有效的解决方案。基于此，提出一种基于联合嵌入的工业设备故障知识图谱推理诊断方法 ChatFD（chat fault diagnosis）。将大模型生成的逻辑规则注入实体和关系中，同时利用大模型从开放世界获取丰富知识作为导向，进行联合嵌入学习，从而构建一个能够充分利用开放知识、有效应

1. 南京南瑞智慧交通科技有限公司 江苏南京 210000

[基金项目] 国家电网科技项目

对数据稀疏问题的故障推理模型。实验表明,在工业设备的故障诊断中,ChatFD 优于现有方法,能够实现更精准、高效的故障定位与诊断。

## 1 相关工作

### 1.1 知识图谱推理

知识图谱推理(knowledge graph reasoning, KG Reasoning)是知识图谱应用的核心问题,主要目标是基于已有的实体和关系推理出潜在的知识。早期的推理方法多采用逻辑规则和路径查询。例如,Horn 逻辑规则<sup>[11]</sup>在知识推理中通过识别明确的关系路径和实体联系,构建显式的逻辑推理框架。然而,这类方法要求专家手工定义推理规则,局限于小规模知识图谱,难以扩展至大规模知识图谱或处理知识不完备的场景。知识图谱嵌入(knowledge graph embedding, KGE)技术则利用嵌入空间的几何关系进行推理,将实体和关系映射到低维向量空间,被广泛应用于知识图谱推理。经典的嵌入模型包括 TransE<sup>[12]</sup>、TransH<sup>[13]</sup>等。这些方法在知识图谱推理任务中表现出色,在处理大规模图谱时显著提升了推理效率。

大语言模型(LLM)近年来在自然语言处理领域展现出强大的语义理解和生成能力,为知识图谱推理提供了新的研究方向。LLM 通过大规模预训练,能够在上下文中捕获丰富的常识和语义信息,可以直接应用于知识生成、常识推理和问答系统。基于 LLM 的知识图谱推理通常采用语言生成模型来扩展图谱的知识表示。COMET<sup>[14]</sup>是基于 GPT 的常识生成模型,通过知识生成弥补知识图谱中的常识性缺失,使得图谱推理更加全面。KG-LLM<sup>[15]</sup>进一步提出了一种结合 LLM 和 KG 的多跳推理框架,通过将 KG 中的关系信息作为 LLM 的条件输入,实现对复杂关系的隐式推理,展现了 LLM 在处理多跳推理中的潜力。此外,KG-BERT<sup>[16]</sup>将 LLM 与知识图谱嵌入结合,用于知识图谱问答和推理任务。KG-BERT 通过将实体对和关系嵌入自然语言上下文中,使得模型能够直接在自然语言空间进行推理,从而在知识图谱问答任务上取得了显著的性能提升。这些研究表明,LLM 在处理复杂语义和跨领域推理方面具有显著优势,但在开放世界和稀疏数据条件下仍有优化空间。

### 1.2 工业领域知识图谱推理

工业知识图谱(industrial knowledge graphs)主要应用于设备管理、故障诊断和生产监控等领域,为工业系统提供数据驱动的智能支持。工业知识图谱的推理通常面临数据噪声、数据不完整以及复杂时空依赖性等。为了增强工业领域的故障诊断能力,研究者们提出 GNNs 增强异构图嵌入的方法。例如,IAGNN<sup>[17]</sup>应用于工业异构图中的异常检测,通过 GNNs 的扩展建模多关系信息,用于检测设备故障或生产过

程中的异常模式。然而,传统 GNNs 模型在处理工业复杂数据时仍存在瓶颈,无法有效建模时序依赖性和多级关系。为了解决这一问题,Zhou 等<sup>[18]</sup>提出了一种基于时序工业知识图谱的推理方法,能够对工业场景中的故障和维保数据进行时序分析,为故障诊断提供时序上下文支持。

随着 LLM 的发展,基于 LLM 的工业知识图谱推理方法开始引起关注。刘沛丰等<sup>[19]</sup>提出了一种将 LLM 与工业故障知识图谱结合的推理系统,通过 LLM 的生成能力生成解释性分析,为工业设备故障的诊断提供自动化的知识生成和故障解释。Jose 等<sup>[20]</sup>则进一步将 LLM 用于工业场景的问答系统,通过大语言模型的常识推理能力生成设备故障的预测和维护建议,为工业知识图谱提供了新的智能推理方式。在工业领域,LLM 和工业故障知识图谱的结合为设备故障诊断和维护提供了新思路,具备良好的应用前景。

## 2 技术方法

### 2.1 问题定义

为了更好地表示工业领域的故障机制,本文构建了工业设备故障知识图谱,它是一种领域知识图谱<sup>[21]</sup>,其中实体和关系为故障实体和故障关系。

设定工业故障知识图谱  $G$  是一种典型的异构图,其中包含大量的故障实体关系三元组,定义为  $G = \{(s, r, o) | \{s, o\} \in E, r \in R\}$ 。其中  $s$ 、 $o$  分别表示故障头实体和尾实体, $r$  表示头实体和尾实体之间的故障关系, $E$  表示故障实体集合, $R$  表示故障关系集合。工业领域故障知识图谱中,不同类型的故障实体可以通过多条路径进行连接。例如  $s \rightarrow \text{path} \rightarrow o$ ,其中  $\text{path}$  为多条路径的组合,用  $\{r_1, r_2, \dots, r_n\}$  表示,  $r_i \in R$ 。

在领域知识图谱中,规则作为描述实体及其关系的核心要素,支持知识推理任务的开展。规则通过定义设备与故障之间的因果关系,提供了知识驱动的推理能力。规则不仅需要高效覆盖领域知识图谱中的事实,还须具备一定的泛化性和可靠性。定义工业领域规则  $p$  表示为一个逻辑关系,其形式为:

$$p: \text{body}(p) \rightarrow \text{head}(p) \quad (1)$$

式中:  $\rightarrow$  为推理符号,  $\text{body}(p)$  为规则的前提条件,由实体与关系的逻辑路径构成;  $\text{head}(p)$  为规则的推理结果,通常表示为一个故障实体关系三元组。例如,规则  $p$  的形式可以具体化为:

$$(e_1, r_1, e_2) \wedge (e_2, r_1, e_i) \wedge \dots \wedge (e_{n-1}, r_1, e_n) \rightarrow (e_1, r_1, e_n) \quad (2)$$

式中:  $e_i$  为故障实体,  $e_i \in E$ 。

工业领域的推理任务是基于知识图谱推理进行故障诊断,推理设备实体存在的故障问题。该任务可进一步分为三个子任务,分别是头实体预测( $?, r, o$ )、尾实体预测( $s, r, ?$ )

以及关系预测  $(s, ?, o)$ ，其中  $?$  表示需要推理的故障知识。知识图谱推理通过链接预测中的三元组打分函数进一步构成损失函数推理出故障知识，实现工业领域故障诊断。

## 2.2 推理总体框架

基于联合嵌入的工业设备故障图谱推理框架 ChatFD，其整体架构如图 1。在图 1 (1) 中，输入故障知识图谱，并利用大模型生成逻辑规则。在图 1 (2) 中，将实体经过开放世界知识导向得到实体描述，并和原始实体进行联合嵌入。在图 1 (3) 中，将图 1 (1) 处生成的规则迭代的注入图 1 (2) 中的联合嵌入中，进行实体和关系嵌入更新。子模块图 1 (1) (2) 的外部知识以及规则生成底层表示为 ChatFD 解码层。

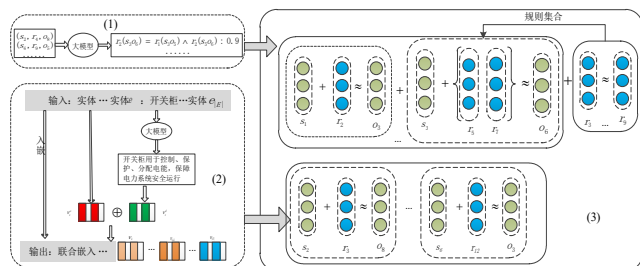


图 1 ChatFD 模型整体架构图

### (1) ChatFD 模型底层结构

ChatFD 模型的底层架构如图 2 所示，分为输入层、ChatFD 解码层以及输出层。首先是输入层，将实体集合以及规则提示模板作为输入，并经过嵌入层对输入文本进行初步嵌入，得到  $\text{Init\_proj}$  向量。

ChatFD 解码层，是底层架构的核心层。为解决初始嵌入向量尺度不均衡问题，需要使用 RMS Norm 函数对  $\text{Init\_embedding}$  进行向量归一化处理得到隐藏层向量。对于向量  $\mathbf{x} = [x_1, x_2, \dots, x_n]$ ，RMS Norm 函数的计算公式为：

$$\text{RMS}(\mathbf{x}) = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2} \quad (3)$$

式中： $n$  是嵌入的向量维度， $x_i$  是向量的分量。

归一化得到的隐藏层向量通过三个不同的线性映射得到  $\mathbf{Q}$ 、 $\mathbf{K}$ 、 $\mathbf{V}$  向量，将  $\mathbf{Q}$  和  $\mathbf{K}$  通过注意力机制计算得到注意力权重，并分配到  $\mathbf{V}$  向量得到  $\text{Attention\_proj}$ 。其次，将  $\text{Init\_proj}$  和  $\text{Attention\_proj}$  进行向量拼接后输入另一个 RMS Norm 函数中进行归一化后，继续输入 (swish gated linear unit, SwiGLU) 激活函数中得到隐藏层向量  $\text{hidden\_proj}$ ，SwiGLU 由 Swish 激活函数和 (gated linear unit, GLU) 结合而成。SwiGLU 的具体表达式为：

$$\text{SwiGLU}(\mathbf{x}) = (\text{Swish}(\mathbf{x}\mathbf{W}_1 + b_1)) \cdot (\mathbf{x}\mathbf{W}_2 + b_2) \quad (4)$$

式中： $\mathbf{W}_1$ 、 $\mathbf{W}_2$  是线性变换矩阵， $b_1$ 、 $b_2$  是偏置项。

输出层，由  $\text{hidden\_proj}$  和  $\text{Attention\_proj}$  进行拼接后的

向量输入 RMS Norm 后，通过线性层并经 Softmax 计算后即可得到最终的输出，分别为实体的外部世界实体描述以及工业领域故障知识图谱中的故障规则。

### (2) 基于大模型的规则生成器

传统的逻辑规则挖掘方法通常使用知识图谱的实体信息，而忽略了图谱中的关系信息。而规则生成器则利用大模型中大量训练的知识语义和结构信息来生成推理故障信息的规则。

为了使大模型能够充分理解 KG 的结构进行逻辑规则挖掘，采用深度优先搜索 (depth first search, DFS) 从 KG 中采样一些路径，这些路径是实例化的逻辑规则。给定一个三元组  $(e_1, e_m, e_n)$ ，封闭路径  $p$  定义为连接知识图谱中  $e_1$  和  $e_n$  的关系序列  $(r_1, r_2, \dots, r_n)$ 。

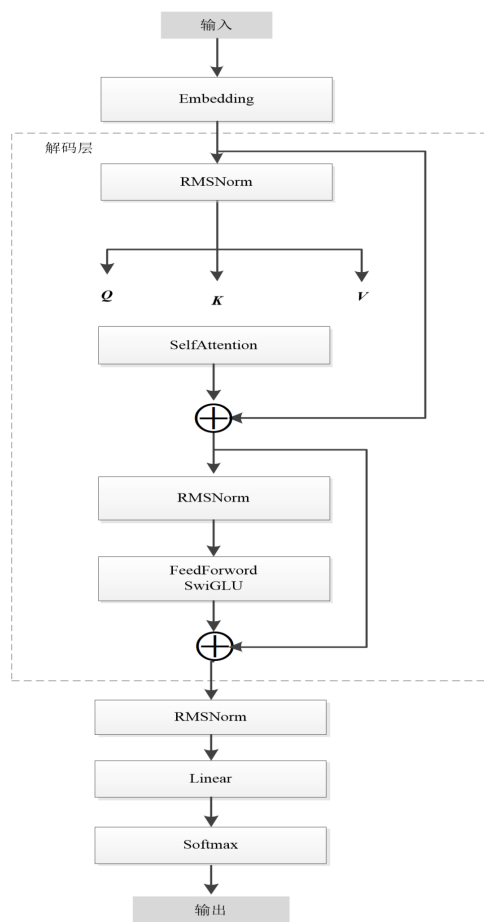


图 2 ChatFD 模型底层架构图

例如：给定一个工业设备中的故障三元组（接地电缆，故障等级，4 级），可以找到路径  $p$  为：

$$p = \text{接地电缆} \xrightarrow{\text{所属类型}} \text{电源设备} \xrightarrow{\text{故障等级}} 4 \text{级} \quad (5)$$

通过将三元组（接地电缆，故障等级，4 级）作为规则头，挖掘的路径  $p$  作为规则体，可以得到式 (6) 所示的规则。

$$\text{故障等级 (接地电缆, 4级)} \leftarrow \text{所属类型 (接地电缆, 电源设备)} \wedge \text{故障等级 (电源设备, 4级)} \quad (6)$$

式(6)表示接地电缆的所属类型是电源设备,电源设备的故障等级是4级,即接地电缆的故障等级同为4级。

给定关系 $r_m$ ,首先从KG中选择锚三元组集合 $(e, r_m, e')$ ,并使用DFS算法从这些锚三元组集合中采样一组长度小于 $m$ 的路径集合 $\{p\}$ ,并构建出规则集合 $\{\psi\}$ 。接下来使用变量代替规则 $p$ 中的实体和关系得到规则样本 $R_m$ 。当规则样本输入到LLM中,其中的结构信息可以促进新规则的生成。对于原始关系的逆,通过添加“inv”来表示。

为了生成更多关于目标关系的逻辑规则,将处理后的规则样本放入提示模板中,以此利用LLM生成丰富的潜在逻辑规则。然而,LLM在生成规则时不可避免地受到“幻觉”问题的影响,表现为生成的规则可能不符合实际逻辑或背景知识。这种问题在开放领域推理中尤为突出,因为LLM生成的规则可能会因语料中的偏差或上下文的模糊性而产生不正确的逻辑或不相关的知识。因此,仅依赖LLM生成的规则,可能导致推理的准确性和可靠性下降,进而影响工业故障诊断的效果。

### (3) 联合嵌入

为了扩展KG的语义表达能力,充分利用开放世界的外部知识提升工业故障诊断中的推理效果。首先提取KG中所有实体的文本嵌入,同时通过查询LLM获取开放世界中相关的知识补充。因此,提出一种基于层次聚类结构的联合嵌入方法,逐层融合图谱内部和开放世界中的知识,从而实现多源知识的语义统一表达。故障知识图谱中的实体集合为 $E = \{e_1, e_2, \dots, e_{|E|}\}$ ,对于每个实体 $e_i$ ,使用大模型提示模板“在工业设备故障领域,‘实体’是什么”的格式简要描述‘实体’,该提示词让大模型根据知识图谱内部的知识,生成开放世界对实体的自然语言描述 $d_i$ 。为了证实ChatFD模型的有效性,在欧氏空间 $\mathcal{O}$ 中得到KG内部的实体嵌入 $v_i^e \in R^{\dim(\mathcal{O})}$ 、关系嵌入 $r_j \in R^{\dim(\mathcal{O})}$ ,以及LLM对实体生成的开放世界自然描述嵌入 $v_i^d = \mathcal{O}(d_i) \in R^{\dim(\mathcal{O})}$ 。其中, $R^n$ 表示 $n$ 维空间。最后,将开放世界和封闭知识图谱的嵌入进行拼接得到联合实体嵌入 $v_i$ ,用公式表示为:

$$v_i = [v_i^d; v_i^e] \quad (7)$$

其次,为将联合嵌入 $v_i$ 整合到模型中,本节将实体嵌入 $e_i$ 初始化为欧氏空间中的随机嵌入 $e_i'$ 和切片嵌入 $v_i = [v_i^e; \frac{n}{2}; v_i^e; \frac{n}{2}]$ 的线性组合, $e_i$ 的具体表达式为:

$$e_i = \lambda e_i' + (1 - \lambda) v_i' \quad (8)$$

式中: $\lambda$ 为平衡实体随机嵌入和联合嵌入的超参数。

## 3 实验分析

### 3.1 数据集

目前工业领域尚无公开的知识图谱故障数据集,文章构建了一个工业设备故障诊断数据集RTKG。同时,在FB15k-237和YAGO3-10数据集上,进行了ChatFD模型评估。RtKG中的实体关系本体包括25种关系,20个实体类型。包括“故障原因”“故障类型”“设备名称”“设备类型”“位置”等实体类型,以及“上级位置”“上级专业”“故障现象”等实体之间的关系,表1列举了其中的部分实体关系本体结构。

表1 RTKG 实体关系表

| 主体   | 关系   | 客体   | 例子                          |
|------|------|------|-----------------------------|
| 位置   | 上级位置 | 位置   | (隔离开关室, 维护, 站台层)            |
| 专业系统 | 上级专业 | 专业系统 | (DCS 系统, 上级专业, ATP/ATO 子系统) |
| 组件   | 故障现象 | 故障现象 | (消防泵控制柜, 故障现象, 电源线)         |
| 组件   | 包含   | 组件   | (地面电子单元, 包含, 处理器板)          |
| 故障现象 | 故障原因 | 原因   | (蓄水池溢水, 故障原因, 梳齿板处有异物)      |
| 规程   | 使用工具 | 工具   | (车辆部, 使用工具, 螺丝刀)            |
| 故障现象 | 故障等级 | 故障   | (管道漏水, 故障等级, 4 级)           |
| 组件   | 所属专业 | 专业系统 | (开关柜, 所属专业, 750 V 系统)       |
| 设备类型 | 父级节点 | 设备类型 | (梯形轨枕, 父级节点, 轨枕)            |

为了与RTKG保持同步,本文从FB15k-237和YAGO3-10中随机抽取了100条数据来测试ChatFD的故障泛化推理能力。RTKG、FB15k-237以及YAGO3-10的相关数据统计信息如表2所示。

表2 RTKG、FB15k-237 以及 YAGO3-10 数据集相关统计信息

| 数据集       | 实体      | 关系  | 训练集       | 验证集    | 测试集    |
|-----------|---------|-----|-----------|--------|--------|
| RTKG      | 8 413   | 25  | 18 671    | 2 067  | 2 068  |
| FB15k-237 | 14 541  | 237 | 272 115   | 17 535 | 20 466 |
| YAGO3-10  | 123 182 | 37  | 1 079 040 | 5 000  | 5 000  |

### 3.2 评价指标

本文所涉及的工业知识图谱故障诊断任务通过实体排序来进行评估。具体来说,对于工业知识图谱中的三元组 $(h, r, t)$ ,在给定头实体 $h$ 和关系 $r$ 的条件下,在给定的实体集合中预测尾实体并进行排序。并使用两个评价指标来评估故障诊断的效果,分别是Hits@ $k$  ( $k \in \{1, 3, 10\}$ )和平均倒数排名(MRR),MRR和Hits@ $n$ 越大,表示预测越准确,具体的计算公式分别为:

$$MRR = \frac{1}{m} \sum_{i=1}^m \frac{1}{\text{rank}_i} \quad (9)$$

$$\text{Hits}@n = \frac{1}{m} \sum_{i=1}^m \mathbb{I}(\text{rank}_i \leq n) \quad (10)$$

式中： $m$  表示集合中有  $m$  条三元组， $\text{rank}_i$  表示正确的实体在实体集合中的排名， $\mathbb{I}$  表示指示函数，如条件为真，则函数值为 1，否则为 0。

### 3.3 实验设置

ChatFD 实验使用 RoBERTa-www-ext 预训练模型嵌入实体描述，参数中的 requires\_grad 设置为 True，使模型在训练学习时参数可微调。模型的学习率 lr 设置为 0.001，训练集以及测试集的训练批次 Batch\_Size 选取为 64，模型的最大迭代次数 epochs 为 100 次。使用通义千问大模型的 API 来实现 ChatFD。对比的模型包括三类方法：（1）基于传统深度学习的方法：包括 Complex<sup>[22]</sup>、RGCN<sup>[23]</sup>、ConvE<sup>[24]</sup> 以及 RotatE<sup>[25]</sup> 模型。（2）基于神经 - 符号推理的方法，将知识图谱中存在的规则和深度学习方法进行有效结合。包括 KALE<sup>[26]</sup>、IterE<sup>[27]</sup> 以及 RPJE<sup>[28]</sup> 模型。（3）基于其他大模型提示下的 ChatFD 模型，包括阿里通义千问（Qianwen）、LLaMA、星火（Xinghuo）大模型和清华 ChatGLM 大模型。

### 3.4 结果分析

基于知识图谱的故障诊断对比实验结果如表 3 所示。

首先，在与传统的深度学习方法（Complex-RGCN,ConvE,RotatE）进行对比时，ChatFD 取得了最优的结果。原因主要有两点：一是传统的深度学习方法只利用了封闭知识图谱内部的三元组知识，而 ChatFD 充分利用了外部世界实体的丰富语义，提高了故障诊断的效果；二是这些方法并未考虑到知识图谱内部的逻辑规则，而逻辑规则包含丰富的背景信息，有利于知识获取和推理。使得这些嵌入模型无法得到更为准确的实体、关系嵌入表示，效果上也明显差于本文提出的 ChatFD 模型。

其次，在两个数据集，ChatFD 也都优于基于神经 - 符号的推理方法（KALE,IterE,RPJE）。虽然这些方法利用了知识图谱内部逻辑规则中的背景信息，但是 ChatFD 利用大模型从外部世界获取的实体描述信息丰富了知识图谱中实体的嵌入。最终 ChatFD 优于这些神经 - 符号推理方法，同时也表明了知识图谱中引入外部知识可以提高故障诊断的效果。

### 3.5 案例分析

为了进一步验证 ChatFD 模型，文章对三个具体案例进行了案例分析。如表 4 所示，对于故障诊断任务，本文给出模型在测试集上的输出结果。具体来说，给定正确的故障设备和关系、LLM 提示模板、故障规则以及大模型根据模板给出的提示结果，输出正确实体在候选实体集合中的排名变化。

第一个案例通过故障诊断模型推理故障设备“消防稳压装置”的故障类型。首先，将已有的规则“故障类型 ( $e_1, e_2$ ) → 故障类型 ( $e_2, e_1$ )”和大模型规则提示模板“工业逻辑规则”，定义两个实体： $x$  和  $y$  之间的关系，如果右边（规则主体）的条件得到满足，即左边（规则头部）的语句成立。现有以下规则示例：

表 3 在 RTKG 和 FB15k-237 数据集上的实验对比

| 模型             | RtKG         |              |              |              | FB15k-237    |              |              |              |
|----------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                | Hit@1        | Hit@3        | Hit@10       | MRR          | Hit@1        | Hit@3        | Hit@10       | MRR          |
| Complex        | 0.053        | 0.088        | 0.121        | 0.075        | 0.194        | 0.297        | 0.450        | 0.278        |
| RGCN           | 0.146        | 0.195        | 0.252        | 0.184        | 0.100        | 0.181        | 0.300        | 0.164        |
| ConvE          | 0.088        | 0.102        | 0.182        | 0.091        | 0.225        | 0.341        | 0.497        | 0.312        |
| RotatE         | 0.121        | 0.274        | 0.293        | 0.199        | 0.241        | 0.375        | 0.533        | 0.338        |
| KALE           | 0.151        | 0.174        | 0.212        | 0.169        | 0.245        | 0.361        | 0.516        | 0.299        |
| IterE          | 0.204        | 0.296        | 0.318        | 0.261        | 0.179        | 0.262        | 0.392        | 0.247        |
| RPJE           | 0.251        | 0.337        | 0.381        | 0.309        | 0.417        | 0.531        | 0.625        | 0.470        |
| ChatGLM-ChatFD | 0.116        | 0.157        | 0.216        | 0.142        | 0.204        | 0.267        | 0.334        | 0.241        |
| LLaMA-ChatFD   | 0.135        | 0.161        | 0.223        | 0.159        | 0.231        | 0.293        | 0.367        | 0.273        |
| Xinghuo-ChatFD | 0.168        | 0.189        | 0.242        | 0.174        | 0.199        | 0.308        | 0.452        | 0.257        |
| Qianwen-ChatFD | <b>0.278</b> | <b>0.361</b> | <b>0.408</b> | <b>0.334</b> | <b>0.419</b> | <b>0.553</b> | <b>0.657</b> | <b>0.481</b> |

表 4 故障诊断具体案例分析

| 任务输入            | LLM 提示模板               | LLM 提示结果                                    | 故障规则                                                            | 排名变化                |
|-----------------|------------------------|---------------------------------------------|-----------------------------------------------------------------|---------------------|
| （消防稳压装置，故障类型，？） | 在工业设备故障领域，“消防稳压装置”是什么？ | 消防稳压装置用于保持消防系统管网内的水压稳定，确保在火灾发生时能提供足够的水量和压力。 | 故障类型 ( $e_1, e_2$ ) → 故障类型 ( $e_2, e_1$ )                       | 硬件故障<br>rank: 145→2 |
| （电源设备，故障等级，？）   | 在工业设备故障领域，“电源设备”是什么？   | 电源设备是用于为各种机械设备和系统提供稳定电力的装置，确保生产和运行的连续性和安全性。 | 所属类型 ( $e_1, e_2$ ) ∧ 故障等级 ( $e_2, e_3$ ) → 故障等级 ( $e_1, e_3$ ) | 4 级<br>rank: 75→5   |
| （照明末端设备，故障原因，？） | 在工业设备故障领域，“照明末端设备”是什么？ | 照明末端设备是指用于提供光源和照明控制的设备，确保生产环境的照明需求。         | 所属类型 ( $e_1, e_2$ ) ∧ 故障原因 ( $e_2, e_3$ ) → 故障原因 ( $e_1, e_3$ ) | 灯管老化<br>rank: 86→2  |

故障类型  $(e_1, e_2) \rightarrow$  故障类型  $(e_2, e_1)$ 、故障类型  $(e_1, e_2) \wedge$  故障类型  $(e_2, e_3) \rightarrow$  故障类型  $(e_1, e_3)$ 、所属类型  $(e_1, e_2) \wedge$  故障等级  $(e_2, e_3) \rightarrow$  故障等级  $(e_1, e_3)$ 。

基于上述规则，为规则头故障类型  $(e_1, e_2)$  生成尽可能多的最重要的规则。输入 ChatFD 解码层得到新生成的规则“故障等级  $(e_1, e_2) \wedge$  故障类型  $(e_2, e_1) \rightarrow$  故障类型  $(e_1, e_2)$ 、故障等级  $(e_1, e_3) \wedge$  故障等级  $(e_3, e_2) \rightarrow$  故障类型  $(e_1, e_2)$  等”。

其次，将大模型外部知识提示模板“在工业设备故障领域，消防稳压装置是什么？”输入到 ChatFD 解码层得到故障设备“消防稳压装置”的外部知识“消防稳压装置用于保持消防系统管网内的水压稳定，确保在火灾发生时能提供足够的水量和压力”。通过将外部知识的嵌入和知识图谱内部节点的嵌入进行拼接得到故障设备的联合嵌入。

最后，将已有的规则和 ChatFG 提示得到 RTKG 训练集中实体与实体之间的关系没有包含这些外部知识和规则，导致正确实体“硬件故障”的排名最初时较低。然而，通过外部知识“消防稳压装置用于保持消防系统管网内的水压稳定……”和提示得到的规则，使得 ChatFD 模型可以判断出该装置为硬件，从而进一步判断该故障为“硬件故障”。故障规则以及联合嵌入注入此前公式中，进行迭代训练。最终输出正确实体“硬件故障”在所有候选实体集合中的排名。由于在 RTKG 训练集中实体与实体之间的关系没有包含这些外部知识和规则，导致正确实体“硬件故障”的排名最初时较低。然而，通过外部知识“消防稳压装置用于保持消防系统管网内的水压稳定……”和提示得到的规则，使得 ChatFD 模型可以判断出该装置为硬件，从而进一步判断该故障为“硬件故障”。

第二个案例通过 ChatFD 模型推理故障设备“电源设备”的“故障等级”。同样地，在训练集中没有包含这样的信息，使得正确实体“4级”的排名较低。然而，通过故障规则“所属类型  $(e_1, e_2) \wedge$  故障等级  $(e_2, e_3) \rightarrow$  故障等级  $(e_1, e_3)$ ”和训练集中的已知三元组推理出“故障等级”为“4级”。

第三个案例在给出故障设备“照明末端设备”的前提下，需要推理出设备出现故障的“故障原因”。在该案例中，通过外部知识“照明末端设备是指用于提供光源和照明控制的设备……”和故障规则“所属类型  $(e_1, e_2) \wedge$  故障原因  $(e_2, e_3) \rightarrow$  故障原因  $(e_1, e_3)$ ”进行联合推理得出“照明末端设备”的故障原因是“灯管老化”，正确实体的排名也得到了巨大提升。

以上案例证明了通过大模型得到的外部世界知识和故障规则可以在工业领域有效地提升故障推理的性能。

#### 4 总结

本文提出了基于联合嵌入的工业设备故障知识图谱推理

方法，用于处理工业领域设备故障诊断中的多项关键问题。通过引入逻辑规则和外部知识，ChatFD 实现了知识图谱内部和开放世界知识的联合嵌入，提升了知识图谱推理的精度和鲁棒性。实验表明，ChatFD 方法显著优于传统方法和神经-符号结合方法，验证了其在复杂工业设备故障诊断中的实用性和优越性。

#### 参考文献：

- [1] Davis R, Buchanan B, Shortliffe E H. Production rules as a representation for a knowledge-based consultation program[J]. Artificial Intelligence, 1977, 8(1): 15-45.
- [2] KOLODNER J L. Case-based reasoning[M]. San Mateo: Morgan Kaufmann Publishers, 1993.
- [3] Zhao Zhibin, Zhang Qiyang, Yu Xiaolei, et al. Applications of unsupervised deep transfer learning to intelligent fault diagnosis: a survey and comparative study[J]. IEEE Transactions on Industrial Informatics, 2021, 70: 1-16.
- [4] Xue Bingcong, Zou Lei. Knowledge graph quality management: a comprehensive survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(5): 4969-4988.
- [5] Brown T B, Mann B, Ryder N, et al. Language models are few-shot learners[PP/OL]. (2020-05-28)[2026-05-29]. <https://arxiv.org/abs/2005.14165>.
- [6] Liao Xingming, Chen Chong, Wang Zhuowei, et al. Large language model assisted fine-grained knowledge graph construction for robotic fault diagnosis[J]. Advanced Engineering Informatics, 2025, 65: 103134.
- [7] Luo Linhao, Li Yuanfang, Haffari R, et al. Reasoning on graphs: faithful and interpretable large language model reasoning[PP/OL]. (2024-02-24)[2026-05-29]. <https://doi.org/10.48550/arXiv.2310.01061>.
- [8] Xie Xin, Wang Junbo, Han Yu, et al. Knowledge graph-based in-context learning for advanced fault diagnosis in sensor networks[J]. Sensors, 2024, 24(24): 8086.
- [9] Lin Junhong, Wang Song, Guo Xiaojie, et al. Temporal reasoning with large language models augmented by evolving knowledge graphs[PP/OL]. (2025-09-18)[2026-05-29]. <https://doi.org/10.48550/arXiv.2509.15464>.
- [10] Xiao Chengyong, Liu Xiaowei, Wulamu A, et al. KG-SR-LLM: knowledge-guided semantic representation and large language model framework for cross-domain bearing fault

- diagnosis[J]. *Sensors* 2025, 25(18): 5758.
- [11]Horn A. On Sentences Which are True of Direct Unions of Algebras[J]. *Journal of Symbolic Logic*,1951,16(1):14-21.
- [12]Bordes A, Usunier N, Garcia - Duran A, et al. Translating embeddings for modeling multi - relational data[C]//NIPS'13: Proceedings of the 27<sup>th</sup> International Conference on Neural Information Processing Systems -Volume 2. NewYork: ACM, 2013: 2787-2795.
- [13]Wang Zhen, Zhang Jianwen, Feng Jianlin, et al. Knowledge graph embedding by translating on hyperplanes[C]// Proceedings of the AAAI conference on artificial intelligence. Palo Alto: AAAI Press, 2014:1112-1119.
- [14]Bosselut A, Rashkin H, Sap M, et al. COMET: commonsense transformers for automatic knowledge graph construction[C]// Proceedings of the 57<sup>th</sup> Annual Meeting of the Association for Computational Linguistics. Chicago: Association for Computational Linguistics, 2019: 4762-4779.
- [15]Shu Dong, Chen Tianle, Jin Mingyu, et al. Knowledge graph large language model (KG-LLM) for link prediction[EB/OL].(2025-06-16)[2026-05-29].<https://doi.org/10.48550/arXiv.2403.07311>.
- [16]Yao Liang, Mao Chengsheng, Luo Yuan. KG-BERT: BERT for knowledge graph completion[PP/OL].(2019-09-11)[2026-05-29].<https://doi.org/10.48550/arXiv.1909.03193>.
- [17] Li Tianfu, Zhou Zheng , LiSinan, et al. The emerging graph neural networks for intelligent fault diagnostics and prognostics: a guideline and a benchmark study[J].*Mechanical Systems and Signal Processing*, 2022, 168: 108653.
- [18]Zhou Bin, Shen Xingwang, Lyu Yuqian, et al. Semantic-aware event link reasoning over industrial knowledge graph embedding time series data[J].*International Journal of Production Research*, 2022, 60(22): 6687-6704.
- [19] 刘沛丰. 基于知识图谱的航空装配知识推理研究 [D]. 武汉: 华中科技大学, 2024.
- [20]Jose S, Nguyen K T P, Medjaher K,et al. Advancing multimodal diagnostics: integrating industrial textual data and domain knowledge with large language models[J]. *Expert Systems With Applications*, 2024, 255(PD): 124603.
- [21] 刘烨宸, 李华昱. 领域知识图谱研究综述 [J]. *计算机系统应用*, 2020, 29(6):1-12.
- [22]Trouillon T, Welbl J, Riedel S, et al. Complex embeddings for simple link prediction[C]//Proceedings of the 33<sup>rd</sup> International Conference on Machine Learning. New York: PMLR, 2016: 2071-2080.
- [23]Schlichtkrull M, Kipf T N, Bloem P, et al. Modeling relational data with graph convolutional networks[C]//Proceedings of the 15<sup>th</sup> International Conference on The Semantic Web (ESWC 2018). Berlin: Springer, 2018: 593-607.
- [24]Dettmers T, Minervini P, Stenetorp P, et al. Convolutional 2D knowledge graph embeddings[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18). Palo Alto:AAAI,2018: 1811-1818.
- [25]Sun Zhiqing, Deng Zhihong, Nie Jianyun, et al. RotatE: knowledge graph embedding by relational rotation in complex space[PP/OL].(2019-02-26)[2026-05-29].<https://doi.org/10.48550/arXiv.1902.10197>.
- [26]Jiang Yanbei, Ehinger K A, Lau J H. KALE: an artwork image captioning system augmented with heterogeneous graph[PP/OL].(2024-09-17)[2026-05-29].<https://doi.org/10.48550/arXiv.2409.10921>.
- [27]Zhang Wen, Paudel B, Wang Liang,et al. Iteratively learning embeddings and rules for knowledge graph reasoning[C]// WWW'19: The World Wide Web Conference.NewYork: ACM, 2019: 2366-2377.
- [28]Niu Guanglin, Zhang Yongfei, Li Bo, et al. Rule-guided compositional representation learning on knowledge graphs[C]// Proceedings of the 34<sup>th</sup> AAAI Conference on Artificial Intelligence (AAAI-20). Palo Alto: AAAI Press, 2020: 2950-2958.

#### 【作者简介】

张永伟（1988—），通信作者（email: zhangyongwei0611@163.com），男，河南商丘人，硕士，高级工程师，研究方向：人工智能、知识图谱。

罗川（1986—），男，四川成都人，硕士，高级工程师，研究方向：人工智能、大数据、大模型。

张迎春（1978—），男，江苏徐州人，硕士，高级工程师，研究方向：信息技术。

（收稿日期：2025-11-11 修回日期：2026-06-08）