

DOI: 10.19650/j.cnki.cjsi.J2514095

图谱 RAG 赋能的航天器机电设备故障诊断 *

赵春州¹, 于劲松², 周金涵², 高占宝², 唐荻音²

(1. 北京航空航天大学仪器科学与光电工程学院 北京 100191; 2. 北京航空航天大学自动化科学与电气工程学院 北京 100191)

摘要:航天器机电设备因其部件高耦合性及故障级联隐蔽性,对故障诊断的推理效率与可解释性提出了严苛要求。在文本知识驱动的智能故障诊断中,针对传统知识图谱(KG)构建成本高、通用大语言模型(LLM)对特定领域诊断知识专业性不足、检索增强生成(RAG)技术关联推理能力有限的问题,故提出一种本体约束驱动的图谱RAG故障诊断方法。一方面,构建了4层故障诊断本体框架,通过本体注入的提示学习实现LLM对多源诊断知识的规范化抽取,并基于字符比较与嵌入模型的双层相似度校准实现知识图谱的动态集成更新,自主构建一体化的诊断知识图谱基座。另一方面,基于LLM与词向量联合的实体模糊检索,结合幂编码的图谱即时蒸馏方法,在结合图谱节点可视化故障传播路径的同时,针对性地融合故障子图结构特征与上下文知识,提升通用LLM故障溯源与维修策略生成的逻辑完备性。以太阳翼驱动机构(SADA)诊断文本、FMEA表格为验证对象,通过本体注入的提示词,借助通用LLM抽取诊断知识图谱,进一步结合可视化图谱进行诊断问答,结果表明,相比于传统RAG方法,结合故障子图的图谱RAG方法在智能故障诊断问答中的关键词F1分数提高了70.88%,语义相似度提高了11.60%,其回答的准确性、可解释性均优于仅LLM方法与RAG方法,为航天器机电设备的智能化故障诊断提供了理论支撑与技术路径。

关键词:航天器机电设备;故障诊断;知识图谱;检索增强生成;大语言模型

中图分类号: TH17 文献标识码: A 国家标准学科分类代码: 460.40

Fault diagnosis for spacecraft electromechanical devices enhanced by knowledge graph-RAG

Zhao Chunzhou¹, Yu Jinsong², Zhou Jinhan², Gao Zhanbao², Tang Diyin²

(1. School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China;
2. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China)

Abstract:The high coupling of components and the concealed nature of cascading faults in spacecraft electromechanical devices impose stringent demands on the reasoning efficiency and interpretability of fault diagnosis systems. To address the challenges of high construction costs associated with traditional knowledge graphs (KG), the lack of domain-specific expertise in general-purpose large language models (LLM), and the insufficient associative reasoning capability of retrieval-augmented generation (RAG) in textual knowledge-driven intelligent fault diagnosis, this study proposes an ontology-constrained knowledge graph-RAG fault diagnosis method. Firstly, a four-layer fault diagnosis ontology framework is constructed. Utilizing ontology-injected prompt learning, the LLM achieves standardized extraction of multi-source diagnostic knowledge. A dynamic integration and updating mechanism for the knowledge graph, based on dual-layer similarity calibration involving character comparison and an embedding model, is implemented to autonomously build an integrated diagnostic knowledge graph base. Secondly, leveraging entity fuzzy retrieval that combines LLM and word embeddings, along with a power-encoding-based instant knowledge graph distillation method, the approach incorporates fault subgraph structural features and contextual knowledge while visualizing fault propagation paths via graph nodes. This significantly enhances the logical completeness of fault root cause analysis and maintenance strategy generation by the general-purpose LLM. Validation using diagnostic texts and FMEA tables of the solar array drive assembly (SADA) shows that, compared with traditional RAG methods, the proposed KG-

RAG method combined with fault subgraphs improves the keyword *F1-score* by 70.88% and semantic similarity by 11.60% in intelligent diagnostic Q&A. The results show superior accuracy and interpretability over using LLM or RAG alone, providing substantial theoretical support and a technical pathway for intelligent fault diagnosis of spacecraft electromechanical equipment.

Keywords: spacecraft electromechanical devices; fault diagnosis; knowledge graph; retrieval-augmented generation (RAG); large language models (LLM)

0 引言

航天器机电设备(如姿态控制力矩陀螺、太阳翼驱动机构(solar array drive assembly, SADA)等)作为航天器在轨能量转换与运动控制的核心执行单元,其可靠性直接决定航天器全生命周期任务成功率。太空中严苛的力、热空间环境导致这类设备的故障诊断面临特殊挑战:其内部能量转换与机械运动的强耦合特性,使得故障传播具有高度隐蔽性和级联性——如润滑失效可能引发轴承卡滞,进而导致机构锁死甚至姿态失控。然而,当前航天器机电设备在设计与维护阶段产出的可靠性分析文本资料繁多、文档风格多样,传统依赖人工翻阅资料的分析策略难以快速满足航天器复杂的故障诊断任务,亟需构建一种知识管理与故障诊断推理方法以实现航天器的智能化管理^[1]。

当前文本知识驱动的智能故障诊断领域主要存在2种技术路线:基于知识图谱(knowledge graph, KG)的方法与基于检索增强生成^[2](retrieval-augmented generation, RAG)的方法。

面向特定知识领域,基于知识图谱的故障诊断通过本体框架结合特定模型完成实体与关系的抽取^[3-4],构建起诊断知识图谱以满足可溯源、解释的智能故障查询与问答需求。张鑫鑫等^[5]通过关联固化卫星系统多维专家知识,并综合运用本体建模与多种知识抽取模型构建了卫星故障领域知识图谱,辅助卫星故障诊断。许驹雄等^[6]在柴油发动机故障领域,通过本体建模,使用BERT和BiLSTM-CRF模型抽取实体及关系并构建了图谱诊断原型系统。杨同智等^[7]将航天器供配电图、故障预案、故障时序动作等多维度专家知识关联,构建航天器多学科故障诊断图谱,并依靠图谱的因果推理辅助故障诊断,提升航天器故障诊断效率、可解释性与处置能力。上述方法主要关注多源知识的结构化表达,同时依靠图谱的故障传播可视化图结构保证了故障诊断结论的可解释性。然而,传统的知识图谱构建方法需要基于标注语料学习或人工提取,构建成本较大;此外,面对航天器机电设备故障机理强耦合性特点,基于静态查询的知识图谱推理缺乏对级联故障传播路径的动态建模能力,难以满足复杂故障诊断的深度推理需求。

基于 RAG 的故障诊断方法首先将故障诊断文本知

识切分后向量化存储构建诊断知识库,而后基于现场排故人员的问询,通过余弦相似度匹配检索相关诊断文本,最后连同用户的问询作为上下文一并给大语言模型(large language model, LLM),LLM 针对问询与上下文语段生成诊断结果^[8],降低了排故人员对多源海量健康知识的认知诊断难度。洪亮等^[9]以瓦斯超限煤矿安全隐患知识为数据源,并将其切分与向量化后构建知识库,使用词嵌入模型完成向量检索,建立了煤矿安全辅助决策智能问答模型。肖雪迪等^[10]以历史故障案例与故障知识构建知识库,利用 LLM 的检索、理解与归纳能力建立了故障定位问答交互系统。江颉等^[11]引入 RAG 技术,利用外部知识库和内部经验库提升了智能体渗透测试结果的准确性和可靠性。刘文婷等^[12]面向宽带网络设备广泛分布且高度互联,故障根源难融合等问题,引入领域知识库,利用 RAG 技术实现智能问答系统,辅助网络故障的定位。上述 RAG 技术通过利用外部知识,可在一定程度减少 LLM 幻觉,增强知识推理决策过程的逻辑性与自主性^[13]。然而,面向来源广泛且故障级联传播明显的可靠性文本资料,RAG 技术较难感知多源知识间关联关系,在文本切分时易截断部分关联语义^[14],导致检索匹配到的文本内容与用户问询的相关性较低,严重限制了诊断推理回答的深度与准确度。因此,现有方法难以有效应对航天器机电设备故障所具有的强隐蔽性和级联效应,无法满足可解释性故障推理分析的高要求。

因此,面向航天器运维保障中高可靠性故障诊断业务需求,本研究提出一种图谱 RAG 赋能的故障诊断方法。首先,利用本体注入的 LLM 提示学习,赋能知识图谱技术,进而有效实现知识的挖掘与关联分析,并基于图谱动态更新机制快速实现多源文档的高性能知识整合与应用,构建规范化、关联可推理的故障诊断图谱知识库;其次,通过知识图谱即时蒸馏方法与故障子图嵌入的 RAG 技术进一步联合,补充诊断回答的知识来源与全局故障传播细节,增强故障推理结果的知识完整性、可解释性与可溯源性。

1 本体约束的图谱 RAG 故障诊断架构

本体是共享概念模型的、明确的、形式化规范说明,它包含一组概念、关系以及它们之间约束的定义^[15]。在航天器机电设备的故障诊断中,本体建模能够形式化地

定义知识空间,规范航天器机电设备知识之间的概念及关系,有效实现多源健康知识的规范化整合,并支撑可解释性诊断推理。

根据知识本体六元素^[16]表示法,同时考虑到航天器机电设备的故障诊断业务场景主要集中在设备部件、部件故障传播机理等方面,对知识中的实体类型及实体关系的定义也应集中于上述方面。首先针对实体类型构建 4 层故障诊断本体实体框架。

1) 物理实体层:按设备分解结构如电机、轴系等核心部件及其层级关系,对应“设备结构类”实体;

2) 监测感知层:角度传感器、功率输出监视等与物理实体的监测关系,对应“测点信号类”实体;

3) 故障演化层:表征“故障原因”、“故障现象”之间的传播演化关系,如建立:机构不能转动→太阳翼驱动机构输出轴不转动…→影响太阳翼驱动机构能源获取等故障演化关系,对应“故障模式类”实体;

4) 维修策略层:映射故障模式与工艺过程控制,如电阻导通测试,控制多余物等应对措施,对应“应对措施类”实体。

进一步,对航天器机电设备健康知识本体中实体类型的说明如表 1 所示。

表 1 面向航天器机电设备故障诊断的本体实体类型定义

Table 1 Definition of ontology entity types for spacecraft electromechanical device fault diagnosis

实体类型	主要属性	说明	示例
设备结构类	系统名称、部件名称等	系统组成、功能单元/部件等	功率导电环、旋转变压器…
测点信号类	信号名称、信号状态等	与产品检测或排故相关的传感信号等	旋变角度数据、电池阵功率输出…
故障模式类	故障原因、故障现象	所有与故障相关的表征	触点开路、电噪声增加…
应对措施类	工具清单、维修操作等	针对故障现象或故障原因作出的应对措施	切换备份、严格控制多余物…

在上述实体类型定义的基础上,构建故障诊断知识本体(fault diagnosis ontology,FDO)为:

$$FDO = \langle C, A^C, R, A^R, I \rangle \quad (1)$$

式中: C 为不同概念的集合; A^C 为概念属性的集合; R 为不同关系的集合; A^R 为关系属性的集合; I 为概念实例集合。其中不同概念集合 C 为:

$$C = \{C_1, C_2, C_3, C_4\} \quad (2)$$

式中: C_1 为设备结构类本体; C_2 为测点信号类本体; C_3 为

故障模式类本体; C_4 为应对措施类本体。FDO 中不同本体概念间的关系概念集合 R 为:

$$R = \{R_1, R_2, R_3, R_4, R_5\} \quad (3)$$

式中: R_1 为故障模式位于某个设备结构处, $R_1 = R_{\text{位于}}(C_3, C_1)$; R_2 为测点监测某处设备结构, $R_2 = R_{\text{监测}}(C_2, C_1)$; R_3 为故障模式导致某种故障模式出现, $R_3 = R_{\text{导致}}(C_3, C_3)$; R_4 为针对故障模式采取应对措施建议, $R_4 = R_{\text{建议}}(C_4, C_3)$; R_5 为设备部件之间的归属关系, $R_5 = R_{\text{属于}}(C_1, C_1)$ 。

各实体类型及其之间的相互关系如图 1 所示。

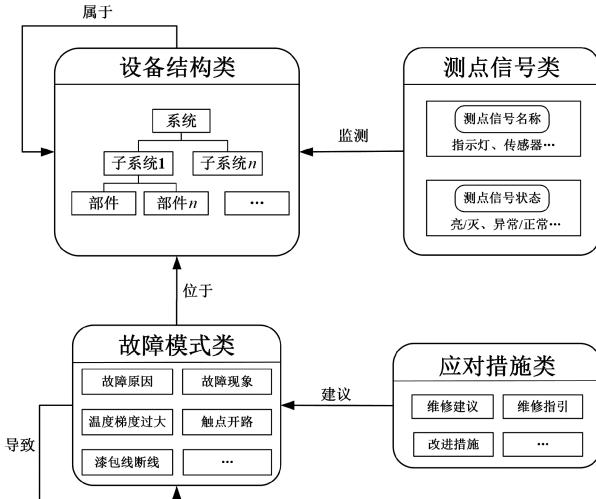


图 1 故障诊断知识本体的构建

Fig. 1 Construction of fault diagnosis knowledge ontology

基于上述本体约束,构建如图 2 所示的图谱 RAG 架构。

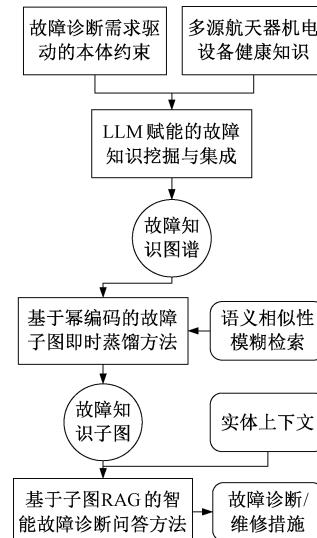


图 2 本体约束的图谱 RAG 架构

Fig. 2 Ontology-constrained knowledge graph-RAG architecture

首先,面向航天器机电设备的可解释性故障诊断业务构建本体,约束多源健康知识,通过领域本体注入的提示学习实现 LLM 的语义校准,解决知识抽取的可靠性问题,并基于图谱的动态更新机制形成本体约束下的全局知识图谱,实现健康知识的挖掘与集成;其次,通过 LLM 与词向量联合模糊检索定位故障实体,并以故障实体为中心,利用幂编码的子图即时蒸馏方法,结合本体关系传播路径,实现故障子图的快速精准定位与可视化验证;最后,为进一步降低维修人员的故障推理负担,开发基于子图 RAG 推理的智能故障问答方法,融合多跳关联与多维度上下文知识,形成人机协同闭环,提供更为全面细致的故障诊断策略,提升复杂故障的可溯源与可解释能力,降低知识检索的认知负荷。

2 LLM 赋能的故障知识挖掘与集成方法

知识挖掘主要指从自然语段到实体知识的挖掘过程。传统基于正则匹配的实体挖掘只能作用于特定的文本结构,而无监督学习等途径通过识别各类较短语义的名词挖掘实体,易导致知识文本被碎片化拆解,且精确度有限,降低知识图谱的可用性^[17]。

因此,针对航天器机电设备多源知识语义表述的多样性,设计基于 LLM 的知识挖掘方法。

如图 3 所示,该方法包含知识抽取与图谱集成两大核心模块:1)将规范多源航天器机电设备诊断知识的本体约束注入到 LLM 的提示词中,规范化抽取多源知识中的实体及关系;2)将抽取出的碎片分散化实体及其关系通过多维度语义相似性比较实现知识的结构化整合集成。

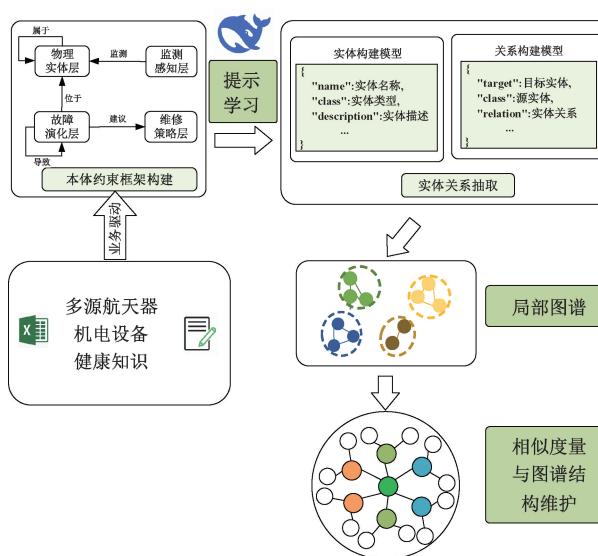


图 3 LLM 赋能的故障知识挖掘与集成方法

Fig. 3 LLM-enabled fault knowledge mining and integration method

2.1 本体注入的提示学习故障知识挖掘方法

通过创建与精炼提示,将本体设计注入到 LLM 中,并构造实体与关系挖掘模版,实现 LLM 在航天器机电设备故障诊断领域的情境学习与知识的系统化抽取。

所说提示,是指添加到 LLM 输入中的信息,通过提示,可以激活模型部分特定参数,提高其面向特定任务的泛化能力。针对离散提示^[18]而言,可通过人工设计离散提示激发大语言模型的能力,包括上下文学习^[19]和思维链^[20]。通过少量上下文学习,LLM 可以直接完成自然语言处理任务而无需更新任何模型参数;而思维链则是通过零样本或者小样本上下文提示,细化指令或中间推理步骤以提升 LLM 的复杂推理能力。

充分利用上述“明确指令”、“思维链”、“格式化输入输出”等原则,在本体约束的框架下,设计的提示词模板为:

1) 指令

给定一个航天器机电设备故障诊断相关的文本文档,首先确定文本中本体框架下的所有实体。接下来,根据本体框架中实体之间的关系,约束并配对实体之间的所有关系。

2) 本体框架

本体框架的实体类型包括设备结构、故障模式(故障原因、故障现象)、测点信号、应对措施,实体间关系包括位于、属于、监测、导致、建议。

3) 思维链步骤 1

进行实体构建,识别所有实体(每个实体可以是一个完整事件),对于每个已识别的实体,提取实体名称、类别、上下文、描述等信息,并将每个实体格式化为特定的字典形式。

4) 思维链步骤 2

进行关系构建,对于 3) 中识别的实体,根据实体间关系列表判定实体间的关系,提取源实体、目标实体、关系类别、关系描述、关系强度等信息,并将每个关系格式化为特定的字典形式。最后,以中文、列表套字典的形式,返回思维链步骤 1 和 2 中确定的所有实体和实体关系。

5) 输出样例

```
{
  "nodes": [ { "name": <entity_name>, "class": <entity_class>, "context": <entity_context>, "description": <entity_description> }, ... ],
  "edges": [ { "source": <source_entity>, "target": <target_entity>, "class": <relation_class>, "context": <relation_context>, "description": <relationship_description>, "strength": <relationship_strength> }, ... ]
}
```

上述基于提示学习的实体挖掘方法专注于高裕度知

识的实体挖掘,可抽取出本体框架下实体类型对应的完备语义表达,每个实体均是一个完整“事件”,可有效避免“事件”内部“实体”再次被切分抽取带来的图谱复杂性大幅度提升的问题。同时在知识挖掘中保留实体的上下文信息便于知识溯源。

2.2 双层相似度检测的图谱动态集成方法

为实现不同知识来源的整合、外推与迭代,为航天器机电设备提供规范精简的知识基础,针对不同知识源提取的局部图谱之间存在相似表达实体的问题,提出双层相似度检测的图谱动态集成方法向全局图谱增量融合。

一方面,在图谱动态更新的过程中,首先利用基于字符串相似性比较的方法快速实现对简单相似实体的初次融合,而后对剩余实体基于语义指纹向量的余弦相似度比较实现二次融合,兼顾实体融合的效率与准确性。另一方面,在局部图谱向全局图谱合并的过程中,融合相似实体的同时需保留被融合实体的边关系以保持图谱结构的一致性。

具体而言,若设局部子图为 $G_l = (V_l, E_l)$, 全局主图为 $G_g = (V_g, E_g)$, 其中 V 为实体集合, E 为关系集合。则定义知识图谱的动态更新需解决实体同义消解与图谱结构一致性维护两个核心问题。

1) 实体同义消解

对 $\forall v \in V_l, \forall u \in V_g$, 若实体 v 与实体 u 的相似度大于一定阈值, 则将两者合并。因而定义 Jaccard 相似度与基于语义指纹向量的余弦相似度实现相似实体的去冗合并。

(1) Jaccard 相似度

$$J(u, v) = \frac{|\mathcal{C}(v) \cap \mathcal{C}(u)|}{|\mathcal{C}(v) \cup \mathcal{C}(u)|} \quad (4)$$

其中, \mathcal{C} 表示将字符串转换为字符集合,通过对两字符串实体 v, u 集合的并集与交集计算可快速简单实现相似实体的相似度度量。

(2) 语义指纹向量余弦相似度

基于字符串比较的方法简单但缺乏语义信息,而在多来源知识文档中,文字书写风格往往不统一,如“电机疲劳”与“电机老化”表述不同但表意接近,此时则需借助词嵌入模型进行知识融合,词嵌入模型通过大规模语料库来学习词的向量表示,该向量表示捕捉了词汇的语义信息和语境信息,语义上相似的词在向量空间中的距离也相近。

但仅靠实体,表意有限,因而需充分借助各实体的本体描述、上下文片段及实体的邻接关系与邻接实体获取更为全面的综合性表达。对图谱内某实体 v ,生成语义指纹向量 e_v ,即:

$$e_v = \text{Embedding}(\phi_{\text{ont}}(v) \parallel \phi_{\text{ctx}}(v) \parallel \phi_{\text{rel}}(v)) \quad (5)$$

式中: Embedding 表示嵌入模型; \parallel 表示向量拼接; ϕ_{ont} 为本体描述; ϕ_{ctx} 为上下文片段; ϕ_{rel} 为实体邻接的三元组。则对另一实体 u 其语义指纹向量为 e_u , 则实体 v 与实体 u 的余弦相似度度量为:

$$C_{\cos}(e_v, e_u) = \frac{e_v^T e_u}{\|e_v\| \cdot \|e_u\|} \quad (6)$$

通过上述基于语义指纹向量的余弦相似度度量可实现对复杂实体间的相似度度量。

2) 图谱结构一致性维护

在局部子图向全局图谱合并的过程中,需保证图谱结构中实体及关系的一致性,因而构建映射函数 ϕ 使得其满足:

(1) 实体映射

$$\phi: V_l \rightarrow V_g \cup V_{\text{new}} \quad (7)$$

式中: V_{new} 为新增节点集合,也即局部图谱中无法与全局图谱匹配的节点集合,生成规则为:

$$V_{\text{new}} = \{v \in V_l \mid \nexists u \in V_g : \text{sim}(v, u) > \theta\} \quad (8)$$

式中: sim 为式(4)与(6)代表的相似度度量。对每个局部节点 $v \in V_l, \phi(v)$ 的取值只能是以下两种之一,即:

a. 全局图谱中的某个节点(即 $\phi(v) \in V_g$),表示 v 与全局图谱中的节点匹配。

b. 新增节点集合中的节点(即 $\phi(v) \in V_{\text{new}}$),表示 v 是局部图谱独有的新实体。

(2) 关系映射约束

$$\phi(E_l) = \{(\phi(s), \phi(t)) \mid (s, t) \in E_l\} \quad (9)$$

局部图谱中的每条边 (s, t) 在映射后变为 $(\phi(s), \phi(t))$,所有映射后的边构成 $\phi(E_l)$,这些边必须存在于合并后的全局图谱中。

约束条件:边映射后的两端节点必须有效,即 $\phi(s)$ 和 $\phi(t)$ 必须属于 $V_g \cup V_{\text{new}}$ 。

通过上述方法可使得局部图谱的原始关系在全局图谱中完整保留,具体算法流程如算法 1 所示。

算法 1: 图谱动态集成更新机制

输入: 局部图谱 G_l , 全局图谱 G_g , Jaccard 阈值 θ_j , 余弦阈值 θ_c

输出: 更新后的全局图谱 G'_g

1: 初始化映射

$$\phi \leftarrow \emptyset, V_{\text{new}} \leftarrow \emptyset$$

2: 实体对齐阶段

for each $v \in V_l$ **do**

 matched \leftarrow False

for each $u \in V_g$ **do**

$$J(u, v)$$

if $J > \theta_j$ **then**

$$\phi[v_{\text{id}}] \leftarrow u_{\text{id}}$$

 matched \leftarrow True

```

break
end if
end for
if  $\neg$  matched then
     $V_{\text{new}} \leftarrow V_{\text{new}} \cup \{v\}$ 
end if
end for
3: 语义增强对齐
if  $V_{\text{new}} \neq \emptyset$  then
     $E_l \leftarrow \text{EMBED}(\cdot)$ ,  $E_g \leftarrow \text{EMBED}(\cdot)$ 
    for each  $v \in V_{\text{new}}$  do
        for each  $u \in V_g$  do
             $C_{\cos}(e_v, e_u)$  ( $e_v = E_l[v]$ ,  $e_u = E_g[u]$ )
            if  $C_{\cos} > \theta_c$  then
                 $\phi[v_{\text{id}}] \leftarrow u_{\text{id}}$ 
                 $V_{\text{new}} \leftarrow V_{\text{new}} \setminus \{v\}$ 
            break
        end if
    end for
end for
end if
4: 实体合并
 $V'_g \leftarrow V_g \cup (V_l \setminus \text{dom}(\phi))$ 
(其中,  $\text{dom}(\phi) = \{v \in V_l \mid \phi(v) \in V_g\}$ )
5: 关系重构
 $E'_g \leftarrow E_g \cup \{(\phi(s), \phi(t)) \mid (s, t) \in E_l \wedge \{\phi(s), \phi(t)\} \subseteq V'_g\}$ 
return  $G'_g(V'_g, E'_g)$ 
end procedure

```

通过上述双层相似度检测的实体融合方法及图谱结构一致性维护机制,可实现多源文档抽取出的局部图谱间的实体消融与集成,实现同名故障实体、相似故障模式的整合,构建精简完备的知识图谱,服务后续全局性故障诊断与推理。

3 基于幂编码的故障子图即时蒸馏方法

海量多源知识经过规范化知识抽取构建起大规模诊断图谱后,为降低维修排故人员对全局图谱的认知负荷,提高诊断排故效率,设计如图 4 所示的基于幂编码的故障子图即时蒸馏方法。

具体而言,根据维修排故人员针对某一故障原因,故障现象等模糊认知从全局图谱中定位相关实体,并进一步以该相关实体为中心,从全局图谱中蒸馏以该故障现象/维修操作为中心的故障子图,为维修排故人员快速提供可解释的可视化验证。

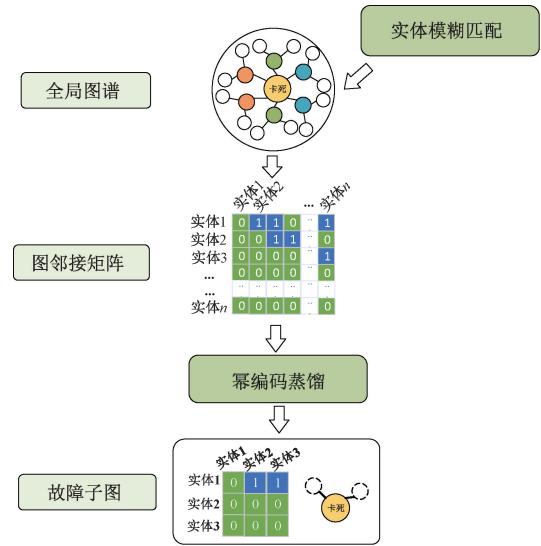


图 4 基于幂编码的故障子图即时蒸馏方法

Fig. 4 On-demand fault subgraph distillation via power encoding

3.1 LLM 与词向量联合的实体模糊检索

完备知识图谱中包含本体约束下的所有实体与相互关系,正则匹配等本质为基于字符串对比的语义检索方法,虽然可以快速筛选出备选实体,但不具备语义信息,面对节点繁多的知识图谱,其筛选的范围也会随之增大,增大维修测试人员定位意向实体的筛选难度。

此外,传统 RAG 技术中仅通过将用户的问询向量化后与文段向量进行相似度匹配定位相关文档,用户询问的关键语义较难与文段精准匹配,造成回答效果下降。因而,为提升维修测试人员与知识图谱的交互与排故效率,设计 LLM 与词向量联合的模糊检索方法快速定位需查询实体。

具体而言,如图 5 所示,为在故障诊断图谱中更为细粒度精准定位用户关心的实体,首先利用 LLM 从用户“输入描述”抽取本体约束下的实体,而后利用嵌入模型将“抽取实体”向量化为 m 维,通过语义向量相似度对比图谱中 n 个实体向量实现模糊检索。

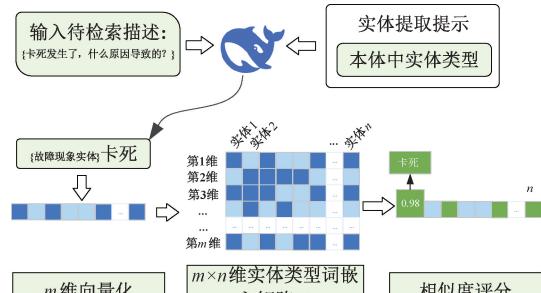


图 5 LLM 与词向量联合的实体模糊检索

Fig. 5 Fuzzy entity retrieval via joint LLM and word embeddings

具体计算过程可表述为:

$$\mathbf{s} = \left[\frac{\mathbf{q}^T \cdot \mathbf{w}_1}{\|\mathbf{q}\| \cdot \|\mathbf{w}_1\|}, \frac{\mathbf{q}^T \cdot \mathbf{w}_2}{\|\mathbf{q}\| \cdot \|\mathbf{w}_2\|}, \dots, \frac{\mathbf{q}^T \cdot \mathbf{w}_n}{\|\mathbf{q}\| \cdot \|\mathbf{w}_n\|} \right] \quad (10)$$

式中: $\mathbf{q} \in \mathbb{R}^{m \times 1}$ 为“抽取实体”的向量化表示; $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n] \in \mathbb{R}^{m \times n}$ 为知识图谱实体嵌入矩阵(包含 n 个实体向量); 通过余弦相似度计算, 生成相似度评分向量 $\mathbf{s} \in \mathbb{R}^{1 \times n}$, 通过评分排序生成候选实体集合 $C = \{e_i | s_i > \tau\}_{i=1}^k$; 其中阈值 τ 可用于动态调整检索范围, 帮助维修测试人员选定关心实体, 从而有效弥补维修人员对航天器机电设备知识体系中潜在实体关联的认知盲区。

3.2 基于幂编码蒸馏的节点多跳传播建模

在面向知识图谱的航天器机电设备故障诊断中, 在选定关心实体后, 仍需以该实体为中心, 通过邻接矩阵对多跳关系建模, 构建故障子图以实现故障的定位与原因、维修操作等快速查询。

具体而言, 邻接矩阵以矩阵形式精确刻画节点间的连接关系, 对于图谱 $G = (V, E)$, 其中, V 为实体集合, E 为关系集合, 其邻接矩阵 \mathbf{A} 是 $|V| \times |V|$ 的方阵, \mathbf{A} 中的元素 A_{ij} 满足:

$$A_{ij} = \begin{cases} 1, & \text{存在边 } (v_i, v_j) \in E \\ 0, & \text{其他} \end{cases} \quad (11)$$

在知识图谱的图结构中, 定义布尔半环运算:

$$\begin{cases} a \oplus b = a \vee b, & \text{逻辑或} \\ a \otimes b = a \wedge b, & \text{逻辑与} \end{cases} \quad (12)$$

则可进一步定义, 用于描述图中所有顶点对之间的 n 跳可达性矩阵 $\mathbf{A}^{(n)}$ 为:

$$\mathbf{A}^{(n)} = (\mathbf{I} \oplus \mathbf{A}) \otimes (\mathbf{I} \oplus \mathbf{A})^{n-1}, \text{ 逻辑与布尔矩阵运算} \quad (13)$$

式中: \mathbf{I} 为单位矩阵; $\mathbf{A}^{(n)}$ 元素 $A_{i,j}^{(n)} = 1 \Leftrightarrow \exists$ 路径 $i \rightarrow j$ 的传播关系且长度 $\leq n$ 。

为利用 $\mathbf{A}^{(n)}$ 实现对图谱中意向实体周边指定 n 跳数的邻接子图计算, 通过快速幂分解定理实现大规模图谱的指定跳数的即时蒸馏。

快速幂分解定理: 对任意正整数 n , 存在唯一二进制分解, 即:

$$n = \sum_{k=0}^m 2^{d_k}, d_k \in N \quad (14)$$

令分解后的幂编码集 $D_n = \{d_1, d_2, \dots, d_m\}$, 则矩阵幂 $\mathbf{A}^{(n)}$, 可分解为:

$$(\mathbf{I} \oplus \mathbf{A})^n = \prod_{d \in D_n} (\mathbf{I} \oplus \mathbf{A})^{2^d} \quad (15)$$

采用快速幂算法计算 $\mathbf{A}^{(n)}$ 时, 需要 $\lfloor \log_2 n \rfloor$ 次矩阵平方操作, k 次矩阵乘法操作。其中, $\lfloor \log_2 n \rfloor = \max\{j \in \mathbb{Z} | 2^j \leq n\}$, $k \leq \lfloor \log_2 n \rfloor + 1$ 。因而其时间复杂度 T_{fast} 有:

$$T_{fast}(n) = O(\lfloor \log_2 n \rfloor + k) \cdot N^3 = O(\log n \cdot N^3) \quad (16)$$

因而, 通过快速幂计算, 利用二进制分解和平方迭代策略, 可充分利用图谱邻接矩阵的稀疏性特征, 极大程度降低 n 跳可达性矩阵 $\mathbf{A}^{(n)}$ 的计算时间复杂度, 从而实现对大规模图谱的即时蒸馏, 获取故障子图以满足故障定位、邻接故障传播及相应维修操作的快速查询与可解释的可视化验证。

4 基于子图 RAG 的智能故障诊断问答方法

知识图谱对于故障模式、设备结构与测点等故障实体具有全局认知能力, 面向航天器机电设备的故障诊断、故障排故归零的实际需求, 提出基于 LLM 的子图 RAG 推理的智能故障问答方法, 通过问答实现故障子图及上下文知识向 LLM 的动态嵌入, 实现可解释图结构与问答描述的混合问答。从而解决传统 RAG 技术在复杂装备故障诊断中存在的知识割裂性与推理不可追溯问题。

如图 6 所示, 首先通过第 3 章的语义模糊检索定位及幂编码蒸馏获取节点的多跳故障子图实现知识检索, 基于本体约束, 一方面获取故障子图的结构化图传播特征, 另一方面获取各实体的原文描述, 为用户提供可解释、可溯源的知识分析。

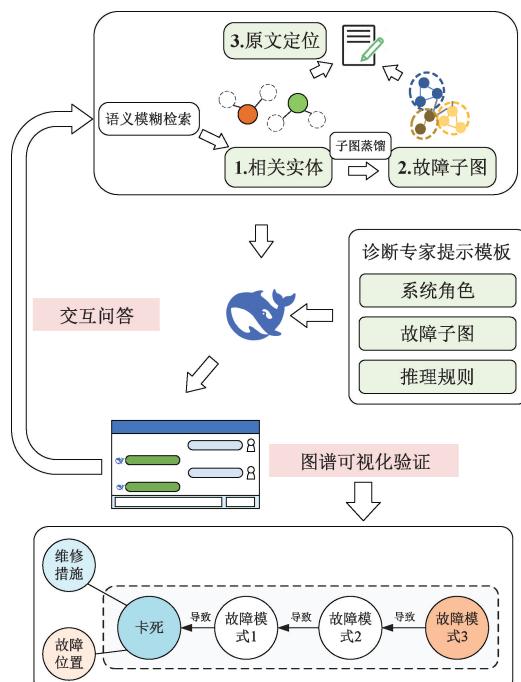


图 6 基于子图 RAG 的智能故障问答方法

Fig. 6 Intelligent fault diagnosis Q&A method via subgraph-RAG

其次, 为最大化 LLM 对图谱的认知关系, 设计诊断专家提示词并将故障子图及其上下文作为 LLM 输入前缀, 引导 LLM 进行知识推理分析, 通过 LLM 的语义理解、

分析能力给出全面综合可解释溯源的回答,进一步面向诊断结果,与维修人员实时交互,通过“交互式问答”的方式明晰故障传播逻辑及其应对措施等。诊断专家提示词模板设计为:

1) 系统角色

定义该角色为航天器机电设备故障诊断领域的专家,专长包括分析复杂系统组件关系、追踪故障传播路径、推荐维护策略。

2) 故障子图及上下文

故障子图及上下文包含实体列表(每个实体包含名称、类型、描述等信息)、关系列表(每个关系包含源实体、目标实体、类型等信息)以及实体和关系的上下文信息。

3) 推理规则

如果问题涉及故障传播,输出加粗的关键实体及其路径;如果故障子图无法完成故障推理服务,根据故障子图,向用户提出支持推理的所需内容或者疑问,如若追加故障子图中不存在的知识,则必须标明。

该模板通过结构约束与语义增强的混合提示策略,引导大语言模型实现故障路径的推理、确定情境下的参数追问以及应对措施的推荐。

5 实验结果分析

为验证本研究方法在航天器机电设备故障诊断中的有效性选取太阳翼驱动机构作为验证对象。作为航天器中用于控制太阳翼姿态和角度的核心部件,太阳翼驱动机构承担着太阳翼姿态调控与能量传输的双重功能,其能量转换与机械运动的强耦合使得故障传播具有高度级联性,作为航天器中不能够采取整机冗余备份的连续运行装置,它从根本上影响着航天器的可靠性和寿命。

面向其可靠性知识文本与故障模式、影响和危害性分析表格^[21-23],在上述故障诊断本体约束下,利用 LLM 挖掘与图谱动态集成方法,自动构建如图 7 所示的太阳翼驱动机构故障诊断图谱(部分)。

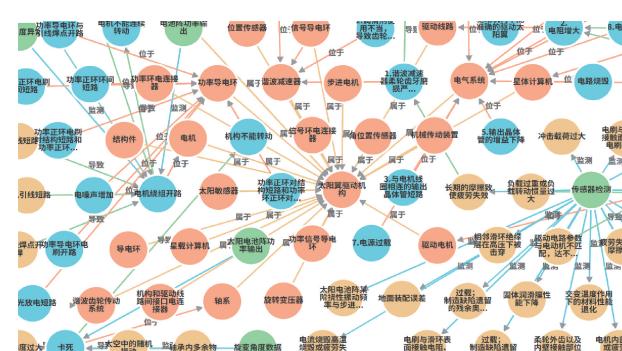


图 7 某太阳翼驱动机构故障诊断知识图谱(部分)

Fig. 7 Partial SADA fault diagnosis knowledge graph

基于浏览器/服务器(browser/server, BS)架构,利用 Flask 后端框架, Vue2.0 前端框架, Axios 前后端交互方式,结合 D3.js 构建图数据库,开发的智能故障诊断辅助问答原型系统如图 8 所示。诊断排故人员通过询问发生的故障现象,利用语义检索的模糊匹配定位故障实体,并以该实体为中心利用幂编码的矩阵运算快速实现子图蒸馏,展示子图中实体之间的可解释传播路径关系。

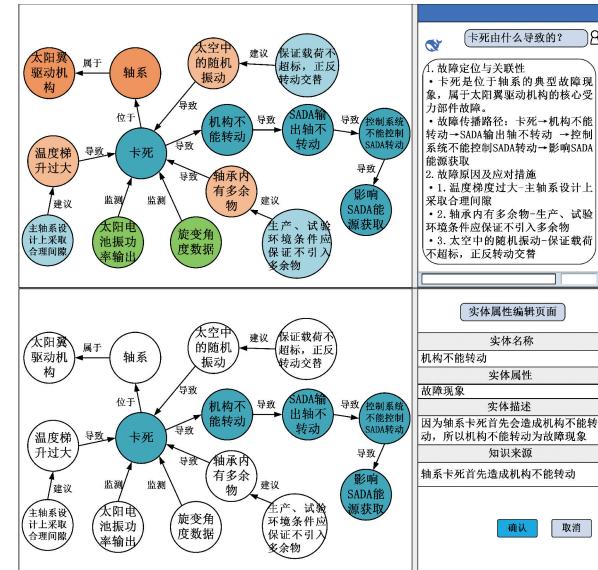


图 8 智能故障诊断问答界面

Fig. 8 Interactive interface for intelligent fault diagnosis Q&A

如以“卡死”这一故障现象为例,定位该实体的同时获取包含其邻接关系及节点的故障子图,在该故障子图中,LLM 根据故障子图中“设备结构”、“故障原因”、“故障现象”节点之间的传播关系定位“卡死”这一故障位于“轴系”中,故障原因为“温度梯度过大/支撑轴承内有多余物/太空中的随机振动”等,并可针对不同的故障原因提出相应的维修建议。故障子图展现出了完整的故障传播链路:卡死→机构不能转动→SADA 输出轴不能转动→控制系统不能控制 SADA 转动→影响 SADA 能源获取,进一步点击图谱中各个实体,可查看该实体的知识来源。通过图 8 结构对故障传播路径的直观展示与原文定位,可满足太阳翼驱动机构故障诊断中的可解释性,可溯源性需求,验证了方法的有效性。

5.1 图谱挖掘完备性实验

对基于 LLM 规范挖掘与图谱动态集成机制得到的太阳翼驱动机构故障诊断图谱中各本体约束下的节点及关系统计如表 2 与 3 所示,共计 414 个实体与 931 条关系。

为综合验证图谱抽取的完备性,通过人工对太阳翼驱动机构诊断图谱内容作验证与查漏补缺,得到下述完备知识图谱的实体及关系统计如表 4 与 5 所示,共计 437 个实体和 1 018 条关系。

表 2 LLM 抽取的故障图谱中的实体

Table 2 Entities in LLM-extracted fault knowledge graphs

实体类型	数量/个	
设备结构	31	
测点信号	9	
故障	故障原因	109
模式	故障现象	183
应对措施	82	

表 3 LLM 抽取的故障图谱中的关系

Table 3 Relationships in LLM-extracted fault knowledge graphs

关系类型	数量/个
属于	62
位于	208
监测	84
导致	406
建议	171

表 4 人工校核后故障图谱中的实体

Table 4 Entities in human-verified fault knowledge graphs

实体类型	数量/个	
设备结构	31	
测点信号	9	
故障	故障原因	116
模式	故障现象	197
应对措施	84	

表 5 人工校核后故障图谱中的关系

Table 5 Relationships in human-verified fault knowledge graphs

关系类型	数量/个
属于	62
位于	213
监测	87
导致	479
建议	177

因而图谱的综合实体抽取完备率为 94.74%, 关系抽取完备率为 91.45%, 验证了 LLM 赋能的图谱挖掘与图谱动态更新方法的有效性。

5.2 图谱蒸馏效率实验

为满足实际应用中快速可视化的需求, 进一步衡量

故障诊断图谱蒸馏的效率, 定义单跳蒸馏平均耗时 $T_{Distill}$ 评价图谱蒸馏效果。

$$T_{Distill} = \frac{\sum_{i \in V} \frac{T_i}{\max_{p_j \in paths}(length(P_j))}}{|V|} \quad (17)$$

式中: V 为诊断图谱中的实体集合; T_i 是从故障诊断全图中以第 i 个实体为中心蒸馏案例子图所用时间; $paths$ 为第 i 个案例子图中所有实体间的传播路径集合; $length(P_j)$ 是路径 P_j 的长度, 即路径上节点的数量, 反映蒸馏深度。

通过对故障全图中 437 个实体逐一进行子图蒸馏, 记录各子图的蒸馏时间与最大传播路径长度, 并计算单跳蒸馏平均耗时 $T_{Distill}$, 结果如表 6 所示。

表 6 图谱蒸馏实验结果

Table 6 Experimental results of knowledge graph distillation

实体类型	$T_{Distill}/ms$
设备结构	13.33
测点信号	9.69
故障模式	14.40
应对措施	11.56

从表 6 可以看出, 毫秒级的子图蒸馏单跳传播时间可为现场诊断排故人员快速提供图谱节点多跳传播建模的可视化验证。

5.3 图谱 RAG 与 RAG 对比实验

为进一步评估子图定位嵌入的图谱 RAG 问答与 RAG 问答效果, 面向故障诊断图谱中故障现象实体构造问题模板:“{故障现象}发生的原因是什么? 位于哪个设备结构中? 针对此问题又该如何采取针对性措施?”, 该问题模板面向故障诊断业务场景, 关切故障现象的故障原因, 故障位置, 故障应对方法, 调用同一 LLM, 分别使用 RAG 方法, 图谱 RAG 方法与仅 LLM 回答方法对构造的问题进行回答。

例如, 针对问题:“电机定子或转子或定子绕组失效发生的原因是什么? 位于那个设备结构中? 针对此问题又该如何采取针对性措施?”可给出 4 种回答结果, 即:

1) 标准答案: 电机定子或转子或定子绕组失效发生的原因是疲劳失效或高温烧毁, 位于步进电机中。针对此问题, 可以采取定子绕组备份; 选用热阻较小的定子绕组材料; 散热设计等针对性措施。

2) 图谱 RAG 回答结果: 电机定子或转子或定子绕组失效的原因是疲劳失效或高温烧毁。这个问题位于步进电机中。针对此问题, 可以采取的针对性措施为:(1) 定

子绕组备件;选用热阻较小的定子绕组材料;散热设计。(2)信号检测,监测疲劳失效或高温烧毁。

3) RAG 回答结果:电机定子或转子或定子绕组失效的原因有:(1)漆包线断线;(2)漆包线焊点开焊;(3)漆包线短路;(4)电机引线短路。位于电机内部结构中。针对此问题,可采取的措施为:(1)设计上采用高可靠性材料和工艺;(2)生产、装配阶段严格质量控制,确保焊接牢固;(3)加强绝缘保护,防止短路;(4)定期检测和维护,及时发现隐患。在轨状态下,可使用切换备份的方法解决。

4) 仅 LLM 回答结果:电机定子、转子或定子绕组失效的原因主要有:绝缘老化、过载运行、机械振动或冲击、制造缺陷等。这些部件位于 SADA 的驱动电机结构中。针对此问题可采取以下措施:选用高可靠性航天级电机材料、加强绝缘防护设计、优化电机负载匹配、实施振动环境试验验证、进行冗余备份设计等。

同时,就回答结果设计关键词准确率 P 、关键词召回率 R 、关键词 $F1$ 分数、标准化交叉熵分数(normalized cross-entropy score, NCES)和语义相似度 Sim_{sem} 这 5 个指标以综合评估问答效果。

$$P = \frac{|M \cap S|}{|M|} \quad (18)$$

$$R = \frac{|M \cap S|}{|S|} \quad (19)$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \quad (20)$$

式中: M 与 S 分别为系统生成答案 SA 及标准答案 GA 的关键词集合; P 用于衡量 SA 中与 GA 相关的关键词比例,反映回答的精准性; R 用于评估 SA 对 GA 关键词的覆盖程度,体现答案的全面性; $F1$ 为综合精确率与召回率的调和平均数,用于平衡系统回答的准确性与完整性。

为避免仅依靠关键词带来的评估不完备性,从词频概率分布角度进一步量化生成答案与标准答案之间的相似度,设计基于标准化交叉熵的评估指标 NCES。对两个答案 SA 和 GA 进行分词处理,得到词集合 $W = \{w_1, w_2, \dots, w_n\}$ 。对于每个答案,计算其中各个词语的概率分布:

$$P_g(w) = \frac{\text{count}_{GA}(w)}{\sum_{w' \in W} \text{count}_{GA}(w')} \quad (21)$$

$$P_s(w) = \frac{\text{count}_{SA}(w)}{\sum_{w' \in W} \text{count}_{SA}(w')} \quad (22)$$

式中: $P_g(w)$ 表示词语 w 在标准答案中的概率; $P_s(w)$ 表示词语 w 在系统答案中的概率; $\text{count}_{GA}(w)$ 和 $\text{count}_{SA}(w)$ 分别表示词语 w 在标准答案和系统答案中的出现次数;实际交叉熵 $H(P_g, P_s)$ 的计算如式(23)所示。

$$H(P_g, P_s) = - \sum_{w \in W} P_g(w) \log P_s(w) \quad (23)$$

式中: $P_s(w) = \max(P_s(w), \epsilon)$ 是经过平滑处理的概率分布; $\epsilon = 1 \times 10^{-10}$ 是平滑因子。则标准化交叉熵分数(NCES)为:

$$\text{NCES} = 1 - \frac{H(P_g, P_s) - H(P_g, P_g)}{H_{\max} - H(P_g, P_g)} \quad (24)$$

式中: $H(P_g, P_g) = - \sum_{w \in W} P_g(w) \log P_g(w)$ 为自信息熵,表示理想情况下(完全匹配)的交叉熵; $H_{\max} = -\log \epsilon$ 为最大交叉熵,表示最差情况下(完全不匹配)的交叉熵。则当 $\text{NCES} = 1$ 表示生成答案与标准答案的词频分布完全一致, $\text{NCES} = 0$ 表示两个分布完全不同。

进一步,从问答的语义相似度考虑,引入语义相似度 Sim_{sem} ,通过向量空间中的语义距离评估生成答案与标准答案的语义一致性,调用语义向量模型将各问答结果表示为 1 024 维向量并按式(25)比较系统生成答案与标准答案之间的语义相似性。

$$\text{Sim}_{\text{sem}} = \cos(\mathbf{v}_a, \mathbf{v}_g) = \frac{\mathbf{v}_a \cdot \mathbf{v}_g}{\|\mathbf{v}_a\| \cdot \|\mathbf{v}_g\|} \quad (25)$$

式中: \mathbf{v}_a 为系统生成答案的语义向量表示; \mathbf{v}_g 为标准答案的语义向量表示。

从表 7 的实验结果可以看出,基于故障子图嵌入的图谱 RAG 方法在回答关键词的准确率,召回率, $F1$ 分数及回答的语义相似性等方面均优于 RAG 与仅 LLM 回答方法,体现了方法的优越性。

表 7 智能问答实验结果

Table 7 Experimental results of intelligent Q&A

诊断方法	P	R	$F1$	Sim_{sem}	NCES
图谱 RAG	0.847	0.770	0.798	0.943	0.834
RAG	0.580	0.439	0.467	0.845	0.584
仅 LLM	0.504	0.196	0.263	0.809	0.500

6 结 论

针对航天器机电设备故障诊断中存在的知识来源广泛、RAG 方法知识推理深度不足等问题,以构建可解释、高效率、知识关联性强的故障诊断框架为目标,开展了本体约束下知识图谱与 RAG 的协同诊断问答研究。

1) 本体驱动的知识挖掘与集成架构设计:通过 4 层本体框架(FDO),结合 LLM 的提示学习、嵌入模型的语义校准能力与图谱动态更新机制,实现多源知识的规范化抽取与融合,解决知识挖掘可靠性低、知识关联性弱的难题。

2) LLM 与词向量联合的幂编码子图蒸馏方法:通过 LLM 提取用户问询中的关键实体与模糊检索定位图谱实

体，并以该实体为中心通过幂编码的图谱高效蒸馏，实现多跳故障子图的精准定位与可视化验证，提升大规模知识图谱的子图定位效率。

3) 可解释性问答机制：开发多维 RAG 诊断代理，通过故障子图结构特征与上下文知识的动态嵌入，提供可解释的故障传播路径，支持故障的可溯源分析。

通过本体约束下的知识图谱与 RAG 深度协同，以太阳翼驱动机构为验证对象，整合其故障诊断知识，实现了知识推理深度与可解释性的提升，开发了智能故障诊断辅助问答原型系统，可为航天器机电设备的智能化运维提供理论支撑与技术路径。

后续研究一方面引入更多航天器机电设备对象，聚焦于航天领域典型设备故障诊断图谱的大规模构建，一方面结合图论与逻辑推理开发故障诊断图谱领域的推理框架，而非仅依赖 LLM 推理，从而进一步提高故障诊断推理的可解释性。

参考文献

- [1] 慕晓冬, 魏轩, 曾昭菊. 基于注意力残差网络的航天器测控系统故障诊断[J]. 仪器仪表学报, 2022, 43(9): 81-87.
- [2] MU X D, WEI X, ZENG ZH J. Fault diagnosis method of spacecraft tracking telemetry and control system based on the attention residual network[J]. Chinese Journal of Scientific Instrument, 2022, 43(9): 81-87.
- [3] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[J]. Advances in Neural Information Processing Systems, 2020, 33: 9459-9474.
- [4] 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[J]. 电力建设, 2022, 43(3): 66-74.
- [5] XIAO F L, WU Y ZH, SHEN X H, et al. Intelligent fault diagnosis of substation equipment on the basis of deep learning and knowledge graph[J]. Electric Power Construction, 2022, 43(3): 66-74.
- [6] XIAO X D, HE Y, ZHANG W, et al. Improved fault localization method for spacecraft using large language model[J]. Spacecraft Engineering, 2024, 33(5): 15-21.
- [7] JIANG J, CAI CH X, LI M D, et al. Research on building an automated penetration testing framework based on LLM and RAG [J]. Telecommunications Science, 2025, 41(9): 119-132.
- [8] 刘文婷, 徐昶, 朱茂中, 等. 基于 AI Agent 技术实现宽带故障的智能诊断应用研究[J]. 电信工程技术与标准化, 2025, 38(5): 30-36.
- [9] LIU W T, XU CH, ZHU M ZH, et al. Intelligent fault diagnosis is implemented based on AI large model agent technology [J]. Telecom Engineering Technics and Standardization, 2025, 38(5): 30-36.
- [10] PROCKO T T, OCHOA O. Graph retrieval-augmented on domain knowledge graph [J]. Aerospace Shanghai (Chinese & English), 2023, 40(3): 151-158.
- [11] 许驹雄, 李敏波, 刘孟珂, 等. 发动机故障领域知识图谱构建与应用[J]. 计算机系统应用, 2022, 31(7): 66-76.
- [12] XU J X, LI M B, LIU M K, et al. Construction and application of knowledge graph in diesel engine fault field[J]. Computer Systems & Applications, 2022, 31(7): 66-76.
- [13] 杨同智, 张鑫鑫, 董房, 等. 基于知识图谱的航天器健康管理技术研究[J]. 航天器工程, 2023, 32(2): 61-69.
- [14] YANG T ZH, ZHANG X X, DONG F, et al. Research on spacecraft health management technology based on knowledge graph [J]. Spacecraft Engineering, 2023, 32(2): 61-69.
- [15] YU H, GAN AO R, ZHANG K, et al. Evaluation of retrieval-augmented generation: A survey [J]. ArXiv preprint arXiv:2405.07437, 2024.
- [16] 洪亮, 郭瑶, 刘兴丽, 等. 基于 RAG 的煤矿安全智能问答模型[J]. 黑龙江科技大学学报, 2024, 34(3): 487-492.
- [17] HONG L, GUO Y, LIU X L, et al. Intelligent Q&A model of coal mine safety based on RAG[J]. Journal of Heilongjiang University of Science and Technology, 2024, 34(3): 487-492.
- [18] 肖雪迪, 何宇, 张伟, 等. 应用大语言模型的航天器故障定位改进方法[J]. 航天器工程, 2024, 33(5): 15-21.
- [19] XIAO X D, HE Y, ZHANG W, et al. Improved fault localization method for spacecraft using large language model[J]. Spacecraft Engineering, 2024, 33(5): 15-21.
- [20] 江颉, 蔡辰旭, 李明达, 等. 基于大语言模型和 RAG 构建的自动化渗透测试框架研究[J]. 电信科学, 2025, 41(9): 119-132.
- [21] JIANG J, CAI CH X, LI M D, et al. Research on building an automated penetration testing framework based on LLM and RAG [J]. Telecommunications Science, 2025, 41(9): 119-132.
- [22] 刘文婷, 徐昶, 朱茂中, 等. 基于 AI Agent 技术实现宽带故障的智能诊断应用研究[J]. 电信工程技术与标准化, 2025, 38(5): 30-36.
- [23] LIU W T, XU CH, ZHU M ZH, et al. Intelligent fault diagnosis is implemented based on AI large model agent technology [J]. Telecom Engineering Technics and Standardization, 2025, 38(5): 30-36.
- [24] PROCKO T T, OCHOA O. Graph retrieval-augmented

- generation for large language models: A survey [C]. 2024 Conference on AI, Science, Engineering, and Technology, 2024; 166-169.
- [14] ARSLAN M, GHANEM H, MUNAWAR S, et al. A survey on RAG with LLMs [J]. Procedia Computer Science, 2024, 246: 3781-3790.
- [15] GRUBER T R. A translation approach to portable ontology specifications [J]. Knowledge Acquisition, 1993, 5(2): 199-220.
- [16] NAING M M, LIM E P, HOE-LIAN D G. Ontology-based web annotation framework for hyperlink structures[J]. Proceedings of the Third International Conference on Web Information Systems Engineering, 2002, 2002; 184-193.
- [17] 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 [J]. 计算机应用, 2021, 41(8): 2161-2186.
- TIAN L, ZHANG J CH, ZHANG J H, et al. Knowledge graph survey: Representation, construction, reasoning and knowledge hypergraph theory [J]. Journal of Computer Applications, 2021, 41(8): 2161-2186.
- [18] TAY Y, DEHGHANI M, TRAN V Q, et al. UL2: Unifying language learning paradigms[J]. ArXiv preprint arXiv:2205.05131, 2022.
- [19] LUO L H, LI Y F, HAFFARI G, et al. Normalizing flow-based neural process for few-shot knowledge graph completion [C]. 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2023: 900-910.
- [20] ZOPH B, BELLO I, KUMAR S, et al. ST-MoE: Designing stable and transferable sparse expert models[J]. ArXiv preprint arXiv:2202.08906, 2022.
- [21] 龚渐安. 卫星太阳翼驱动机构可靠性建模与分析 [D]. 成都: 电子科技大学, 2013.
- GONG ZH AN. Reliability modeling and analysis of solar array drive assembly in satellite[D]. Chengdu: University of Electronic Science and Technology of China, 2013.
- [22] 石磊. 太阳翼驱动机构的可靠性分析 [D]. 成都: 电子科技大学, 2011.
- SHI L. Reliability analysis of solar array drive assembly[D]. Chengdu: University of Electronic Science and Technology of China, 2011.
- [23] 黄意贤. 航天器太阳翼驱动机构的时变可靠性研究 [D]. 成都: 电子科技大学, 2018.
- HUANG Y X. The time varying reliability of solar array drive assembly [D]. Chengdu: University of Electronic Science and Technology of China, 2018.

作者简介



赵春州, 2023 年于武汉理工大学获得学士学位, 现为北京航空航天大学博士研究生, 主要研究方向为故障预测与健康管理技术。

E-mail: ChunZhou_Zhao@buaa.edu.cn

Zhao Chunzhou received his B. Sc. degree from Wuhan University of Technology in 2023. He is currently a Ph. D. candidate at Beihang University. His main research interest is prognostics and health management (PHM) technologies.



于劲松, 1989 年于南京航空航天大学获得学士学位, 1996 年与 2004 年于北京航空航天大学分别获得硕士和博士学位。现为北京航空航天大学自动化科学与电气工程学院教授, 主要研究方向为自动化测试和故障预测与健康管理技术。

E-mail: yujs@buaa.edu.cn

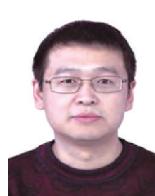
Yu Jinsong received his B. Sc. degree from Nanjing University of Aeronautics and Astronautics in 1989, and his M. Sc. and Ph. D. degrees both from Beihang University in 1996 and 2004, respectively. He is currently a professor in the School of Automation Science and Electrical Engineering at Beihang University. His main research interests include automated testing and prognostics and health management (PHM) technologies.



周金洽, 分别在 2020 与 2023 年于北京航空航天大学获得学士学位和硕士学位, 现为北京航空航天大学博士研究生, 主要研究方向为故障预测与健康管理技术。

E-mail: jinhan_zhou@buaa.edu.cn

Zhou Jinhan received his B. Sc. and M. Sc. degrees both from Beihang University in 2020 and 2023, respectively. He is currently a Ph. D. candidate at Beihang University. His main research interest includes prognostics and health management (PHM) technologies.



高占宝(通信作者), 分别在 1996 年、2002 年和 2006 年于北京航空航天大学获得学士学位、硕士学位和博士学位, 现为北京航空航天大学讲师, 主要研究方向为故障诊断与健康管理, 数据挖掘。

E-mail: gaozhanbao@buaa.edu.cn

Gao Zhanbao (Corresponding author) received his B. Sc. , M. Sc. , and Ph. D. degrees all from Beihang University in 1996, 2002, and 2006, respectively. He is currently a lecturer at Beihang University. His main research interests include fault diagnosis and health management, datamining.