冯雅凡, 曹从咏, 陈萌,等. 基于 LLMs 的高速铁路信号设备故障领域知识图谱构建方法研究[J]. 智能计算机与应用,2026,16(6):103-111. DOI:10.20169/j.issn.2095-2163.24093001

基于 LLMs 的高速铁路信号设备故障领域知识图谱构建方法研究

冯雅凡, 曹从咏, 陈 萌, 申缘琪
(南京理工大学 自动化学院, 南京 210094)

摘 要: 为了解决高速铁路设备故障处理中遇到的知识存储和应用难题,以及文本数据利用与检索的挑战,本文提出一种基于大语言模型(Large Language Models,LLMs)和指令微调技术的高速铁路设备故障领域知识图谱构建方法。首先,自顶向下构建知识图谱本体模型,明确本文涉及到的实体与关系的类型及含义;其次,以高铁信号设备故障案例文本数据为原始语料数据,基于提示学习进行语料标注;接着,通过大语言模型的低阶适应技术(Low-Rank Adaptation of Large Language Models, LoRA)对 LLMs 进行基于故障文本信息抽取的指令微调与性能测试,通过验证集验证,模型抽取了 310 个实体,构建关系 548 条,信息抽取的精确度为 66.06%,召回率为 85.99%,调和平均值 *F*1 为 74.64%;最后,构建领域知识图谱,包含 6 903 个实体,19 215 条关系,以进行可视化展示与检索。实验结果表明,本文的知识图谱能够实现设备关联查询、故障溯因等功能,其构建方法有效提升了文本数据的处理效率和质量,为铁路安全运营与保障提供了技术支撑。

关键词: 高铁设备; 大语言模型(LLMs); 提示学习; LoRA; 信息抽取; 领域知识图谱
中图分类号: TP391.1 **文献标志码**: A **文章编号**: 2095-2163(2026)06-0103-09

Research on the construction method of knowledge graph in the field of high-speed railway signaling equipment fault based on LLMs

FENG Yafan, CAO Congyong, CHEN Meng, SHEN Yuanqi
(School of Automation, Nanjing University of Science and Technology, Nanjing 210094, China)

Abstract: To solve the problems of knowledge storage and application encountered in the fault handling of high-speed railway equipment, as well as the challenges of text data utilization and retrieval, this paper proposes a method for constructing a knowledge graph of the field of high-speed railway equipment faults based on Large Language Models (LLMs) and instruction fine-tuning technology. Firstly, a knowledge graph ontology model is top-down constructed to clarify the types and meanings of entities and relationships involved in the paper. Secondly, based on the original text data of high-speed railway signal equipment fault case studies, text annotation is carried out using prompt learning. Then, the LLMs are fine-tuned based on the Low-Rank Adaptation of Large Language Models (LoRA) for fault text information extraction using instruction fine-tuning and performance testing. Through the validation set verification, the model extracted 310 entities, built 548 relationships, and the precision of information extraction was 66.06%, the recall was 85.99%, and the *F*1 harmonic average was 74.64%. Finally, a domain knowledge graph is constructed, containing 6 903 entities and 19 215 relationships, for visualization and retrieval. The experimental results show that the knowledge graph proposed in this paper can realize equipment-related query and fault causation tracing functions, and the construction method effectively improves the efficiency and quality of text data processing, providing technical support for railway safe operation and protection.

Key words: high-speed rail equipment; Large Language Models (LLMs); prompt learning; LoRA; information extraction; domain knowledge graph

0 引 言

随着社会发展,中国高铁建设取得了显著成就,已基本建成规模庞大、技术先进、信息化程度高的运营网络[1]。然而,随着设备老化和设备服役时间的增加,高铁的智能运营维护和安全应急防控压力日

作者简介: 冯雅凡(1999—),女,硕士研究生。主要研究方向:交通信息工程及控制; 陈 萌(1999—),女,硕士研究生。主要研究方向:交通信息工程及控制; 申缘琪(2000—),男,硕士研究生。主要研究方向:交通信息工程及控制。
通信作者: 曹从咏(1964—),男,博士,教授。主要研究方向:交通信息工程及控制。Email:jy108@njust.edu.cn。

收稿日期: 2024-09-30

益加剧[2],知识智慧应用的需求愈发突出。

目前,铁路系统积累了大量面向路网运营与安全的故障数据,特别是非结构化的故障案例文本数据[3]。这些数据承载着路网安全的关键信息,如故障表现、处置方法、故障原因、解决方案等。然而,以自然语言形式存在的数据,数据量巨大,人工处理成本高、效率低,计算机识别处理也有难题亟待解决[4]。

在这一背景下,知识图谱为复杂知识转化为人类认知注入了新的活力,LLMs[5-6]与指令微调技术为处理和分析非结构化文本数据提供了新的可能。知识图谱具有强大的语义解析能力,被广泛应用于知识检索、问答、对话及推荐等多个方面[7-12]。在铁路领域,杨连报[13]基于协同过滤算法,以铁路故障数据为样本构建故障知识图谱,实现故障事件关联分析。林海香等学者[14]针对高速铁路道岔设备故障处置问题,构建知识图谱以展示故障关系,实现了故障数据关联查询及设备预防性维护。关于信息抽取的最新研究大多都基于 LLMs 与微调技术[15-16],通过海量数据进行预训练以掌握语言规律和相关知识,高效提取并分析文本关键信息。孙熠衡等学者[17]利用提示学习对 LLMs 注入领域知识,基于 LoRA 微调,设计提示训练旁路与标书提示模板,实现了标书文本信息的有效抽取。杨燕等学者[18]通过 LangChain 部署本地 LLMs,集成数字孪生水利领域知识,提出一种基于提示学习微调 LLMs 的数字孪生水利领域知识抽取方法,优化了实体抽取结果。

因此,本文以高铁信号设备故障数据为样本,基于 LLMs 进行指令微调,结合知识图谱进行挖掘分析,提出基于 LLMs 的知识图谱构建方法,为铁路安全运营与保障提供技术支撑。

1 高速铁路信号设备故障领域知识图谱

知识图谱,被称为语义网络的知识库[19],通过“实体-关系-实体”的三元组结构[20-21]描述知识及其关系,可分为通用知识图谱和领域知识图谱。高速铁路信号设备故障知识图谱属于领域知识图谱,使用自顶向下[22]的方法进行构建。知识图谱构建如图 1 所示。

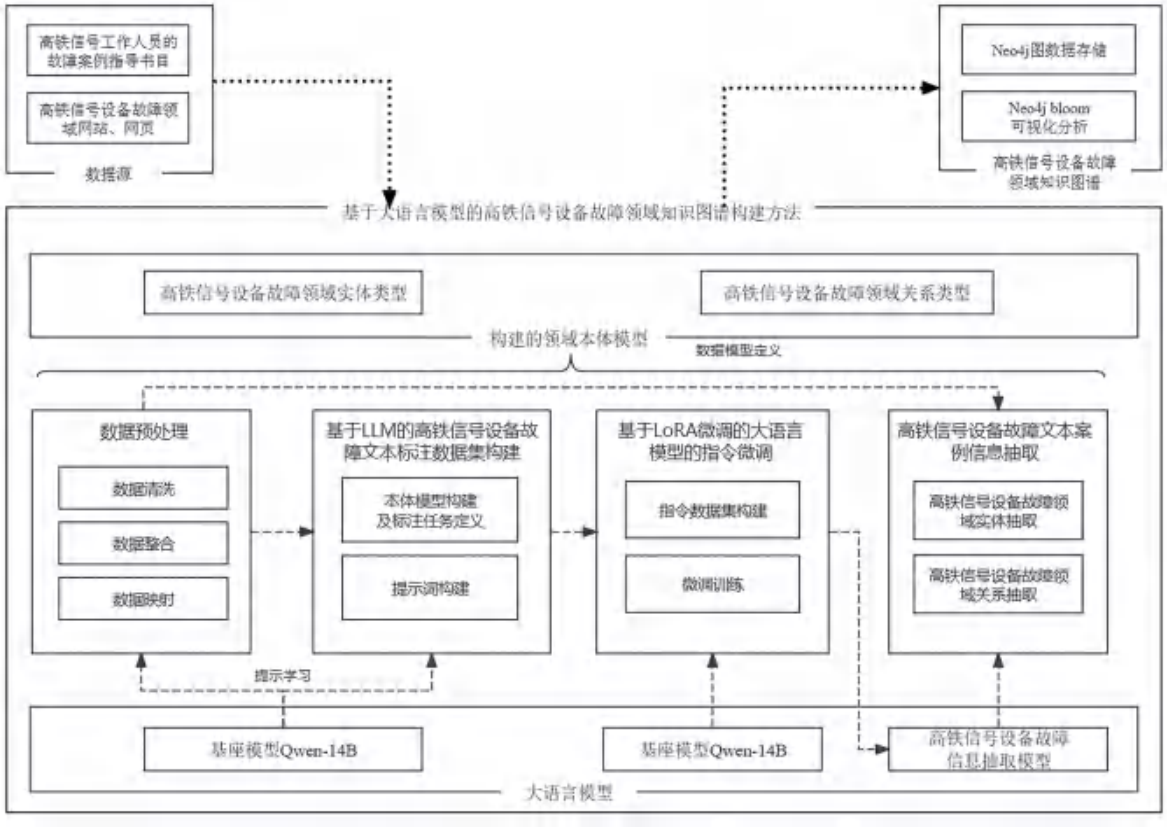


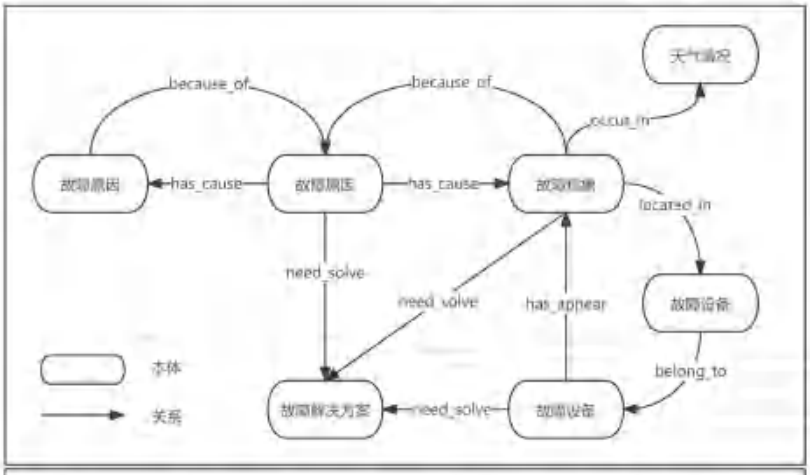
图 1 知识图谱构建

Fig. 1 Knowledge graph construction

1.1 本体模型

本体是对客观情况进行抽象化,对抽象概念及关系进行清晰定义。本文结合专家经验构建顶层概念,逐步细化为信号设备故障领域的实体和关系,确

定实体类型及关系类型，为后续数据组织和知识提取提供标准语义框架。高速铁路信号设备故障领域知识图谱本体模型如图 2 所示。



故障设备：故障发生的设备名称　　故障原因：故障发生的原因　　天气情况：故障发生时的天气情况
故障现象：故障发生时可以直接观察到的现象　　故障解决方案：针对故障原因、现象的解决方法

关系名称	<头节点,尾节点>	含义
belong_to	<故障设备,故障设备>	某个故障设备属于某个（些）故障设备
need_solve	<故障设备,故障解决方案>	某种设备发生故障需要相应的某种故障解决方案
need_solve	<故障现象,故障解决方案>	某种故障现象需要某种（些）故障解决方案
need_solve	<故障原因,故障解决方案>	某种故障原因需要某种（些）故障解决方案
has_cause	<故障原因,故障现象>	某种故障原因导致某种(些)故障现象
has_cause	<故障原因,故障原因>	某个故障原因会导致其他的故障原因
because_of	<故障现象,故障原因>	发生某种故障现象由于某种(些)故障原因
because_of	<故障原因,故障原因>	某个故障原因发生是由于某种(些)故障原因
has_appear	<故障设备,故障现象>	某种故障设备出现某种（些）故障表现
located_in	<故障现象,故障设备>	某种故障现象发生在某个(些)故障位置
occur_in	<故障现象,天气情况>	发生某种故障现象是在某种天气情况

图 2　高速铁路信号设备故障领域知识图谱本体模型

Fig. 2　Ontology model of fault domain knowledge map of high-speed railway signal equipment

1.2　领域数据集构建

本文以《高速铁路信号设备典型故障案例》等铁路工作人员故障案例教材为原始数据。对故障文本进行去除无效数据、删除重复数据、去除异常数据等操作，得到 1 821 个信号设备故障案例，以 txt 形式存储。设备故障案例如图 3 所示。

本文基于 LLMs，设计提示词（Prompt）引导 LLMs 标注标签，结合人工校验，实现领域数据自动标注。自动化数据标注策略如图 4 所示。

1.3　领域信息抽取及知识可视化

信息抽取是文本关键信息的获取手段，主要分为实体抽取和关系抽取。本文针对高铁信号设备故障领域，指导和微调 LLMs，使其专注于信息抽取，包括 3 个关键步骤：指令微调数据集构建；基于 LoRA 指令微调 LLMs；模型部署与应用。基于 LoRA 指令微调大模型的信息抽取方法如图 5 所示。

知识图谱以图结构对知识进行建模和表示。Neo4j 有操作界面友好、代码开源、查询方便、兼容广泛等优点。本文采用 Neo4j 对构建的知识图谱进行可视化展示。



图 3　设备故障案例

Fig. 3　Equipment failure cases

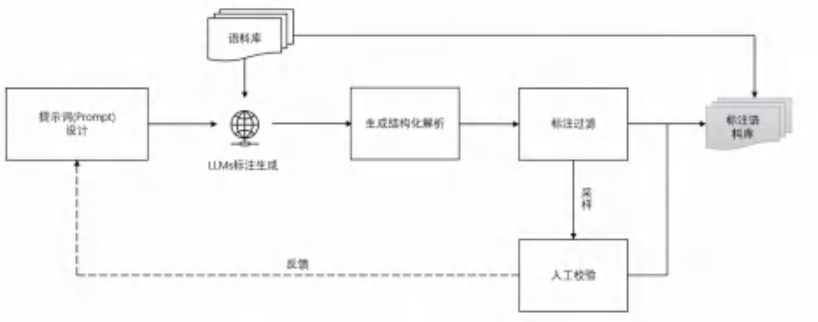


图 4 自动化数据标注策略

Fig. 4 Automated data annotation strategy

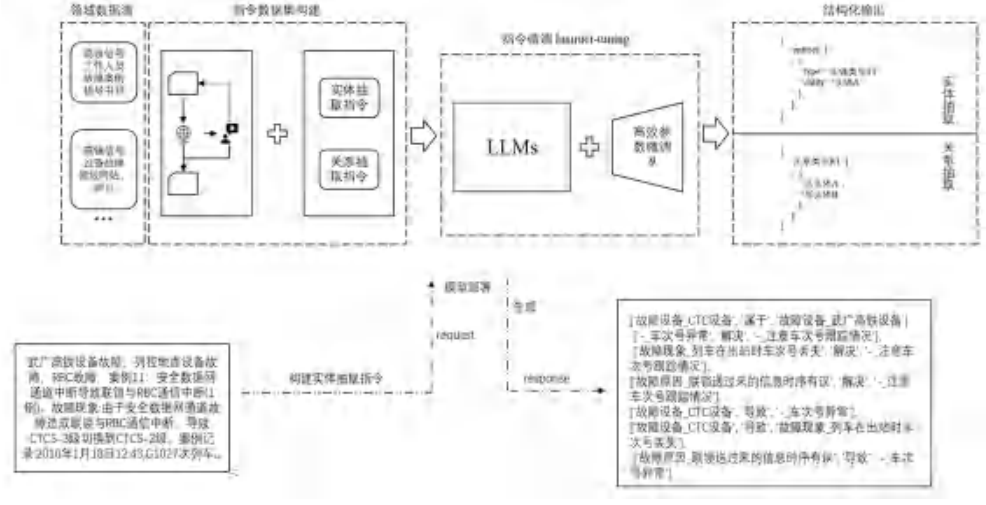


图 5 基于 LoRA 指令微调大模型的信息抽取方法

Fig. 5 Fault domain information extraction method based on LoRA command fine-tuning large model

2 基于 LLMs 的高铁信号设备故障文本数据标注方法

2.1 任务定义

本文针对故障文本数据标注任务设计特定提示模板 t，令 $p_t(x,d)$ 表示填充了 x 和 d 到提示模板 t 后的结果，$f(p_t(x,d)) \in F$ 表示将 LLMs 应用于 $p_t(x,d)$ 后的输出字符串；设计结构化解析器 R，R 定义为一个函数 $R(x,d,f(p_t(x,d)))$，将输入 x、标注规范 d 和 LLMs 的输出 $f(p_t(x,d))$ 映射到特定的输出空间 ϑ，利用 LLMs 的提示学习完成标注任务。基于 LLMs 的结构化数据自动标注方法如图 6 所示。

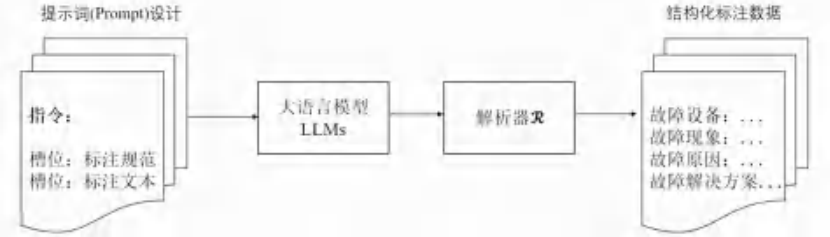


图 6 基于 LLMs 的结构化数据自动标注方法

Fig. 6 Automatic annotation method of structured data based on LLMs

2.2 基于 LLMs 提示学习的文本标注实验

为进一步提升 LLMs 的性能，研究人员提出了面向 LLMs 的提示(prompt)学习方法[23]，将任务统一为生成式自然语言理解框架，构造训练语料，定义提示函数，进行少样本或零样本学习与微调以提升 LLMs 的性能[24]。具体步骤[25-26]为：构造提示，将标注规范和待标注的故障文本填充至提示模板中，形成一个完整的提示；将提示输入到 LLMs 进行概率性填充，实现信息补全，得到含标注标签的输出结果；利用解析器将 LLMs 的输出结果转换为结构化

故障文本标注数据。高铁信号设备故障领域自动标注提示模板具体如下。

prompt=你是一名高铁信号设备故障诊断领域的专家，请根据以下文本提取文本中的实体，参考“样例”JSON 文件，其中实体的类型为：故障设备、故障现象、故障原因、故障解决方案、天气情况。最后以 JSON 的格式输出。

实体的定义如下：<槽位 1>

实体识别说明：　<槽位 2>

给定故障文本：　<槽位 3>

请识别：

<槽位 1>示例：故障设备是指故障发生的设备名称

<槽位 2>示例：不要改动原字段...

<槽位 3>示例：武广高铁设备故障，列控地面...

通过结构化输出提示，引导 LLMs 生成符合本文要求的输出文本；采用约束性提示限制 LLMs 的输出，使其满足固定的语法结构和逻辑关系。

本文选用 Qwen-14B 为基座模型，进行提示学习。Qwen-14B 是阿里云基于 Transformer 架构训练的大型语言模型，具备强大的并行计算和文本处理能力。实验成功生成构建了高铁信号设备故障文本数据实体标注数据集，数据集包含 5 种高铁信号设备故障实体类别，总计 1 821 个标注文档和 6 903 个标注实体。

3　基于 LoRA 微调的高铁信号设备故障信息抽取模型

3.1　任务定义

研究中通过给定一个包含 n 个词元的输入文本 $X=[x_1,\cdots,x_n]$、任务指令 J 以及目标提取序列 $y=[y_1,\cdots,y_n]$，在一个自回归公式得到最大化条件概率，计算公式如下：

$$p_\theta(y \mid X,P)=\prod_{i=1}^{m} p_\theta(y_i \mid X,J,y_{<i}) \quad (1)$$

其中，θ 表示 LLMs 的参数；任务指令 $J=J(T,p,s)$，这里 T 表示信息抽取任务类型，p 表示候选信息类型；s 表示任务约束。

3.2　指令数据集的构建

构建指令微调数据集涉及 3 个核心步骤：数据标注、设计指令模板与数据格式转换。

(1)数据标注：通过构建一种基于 LLMs 的高铁信号设备故障文本数据标注方法，将自动标注与人工校验相结合，确保标注数据的完整和准确，并根据人工反馈优化提示设计。

(2)设计指令模板：指令模板由任务描述、标签清单、输出规范组成。模板明确了 LLMs 需要完成的任务、必须识别的标签类别和输出数据的格式。信息抽取任务指令模板分述如下。

① 实体抽取：prompt=你是一名高铁信号设备故障诊断领域的专家，请根据以下文本提取文本中的实体，实体的类型为：故障设备、故障现象、故障原因、故障解决方案、天气情况。最后以 JSON 的格式输出。

② 信息抽取：prompt=你是一名高铁信号设备故障诊断领域的专家，请根据以下文本提取文本中的关系，其中关系的类型为：belong_to、need_solve、has_cause、because_of、has_appear、located_in、occur_in。最后以 JSON 的格式输出。

(3)数据格式转换：将标注数据的格式按照要求转换。

3.3　LoRA 微调方法

本文选择 Qwen-14B 作为基座模型，针对信息抽取任务，采用了低秩自适应(LoRA)微调技术，实现对高铁信号设备故障信息抽取任务的适应。LoRA 微调技术结构如图 7 所示。

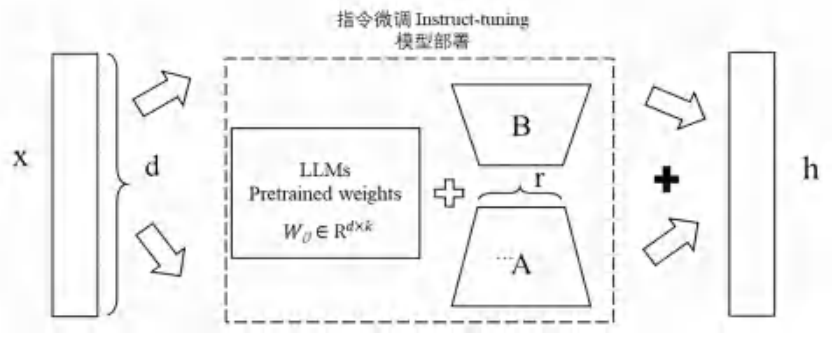


图 7　LoRA 微调技术结构

Fig. 7　LoRA fine-tuning technical structure

LoRA 微调技术允许冻结预训练模型的权重，保持预训练的权重不变，通过引入 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 矩阵替代在微调过程中需要改变的参数，来训练神经网络中的密集层，降低模型训练的参数与资源。由此得到公式如下：

$$W_0 + \Delta W = W_0 + \boldsymbol{BA},\ \boldsymbol{B} \in R^{d\times r},\ \boldsymbol{A} \in R^{r\times k} \quad (2)$$

在此基础上，推得的数学公式为：

$$h = W_0 x$$

其中，x 表示上下文特征，在对下游任务进行微调时，仅需更新 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 矩阵，以极小的参数规模实现对大模型的间接训练，前向传播更新参数的过程可表示为：

$$h = W_0 x + \Delta W x = W_0 x + \boldsymbol{BA} x \quad (3)$$

在前向过程中，W_0 和 ΔW 都乘以相同的 x 后再相加，在初始化的时候对 $\boldsymbol{A}$ 使用随机高斯初始化，对 $\boldsymbol{B}$ 使用零初始化，训练过程可如下公式来表示：

$$\Delta W = \boldsymbol{BA} \quad (4)$$

3.4 评估指标

本文将数据划分为训练集（1 276 篇信号设备故障案例）和验证集（45 篇信号设备故障案例），以供模型参数学习、微调以及性能评估。

在信息抽取任务中，常以精确度（Precision，P）、召回率（Recall，R）以及调和平均值 $F1$ 分数（$F1$ - Score）作为评估标准。其中，精确率表示正确识别词数占总词数比例，反映实体识别效果。召回率表示正确识别数占原始文本总词数比例。两者分别反映模型识别的精确性和全面性。调和平均值是对两者的综合考虑，避免差异过大，客观反映提取效果。研究推得的数学公式具体如下：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

其中，TP 表示预测结果与实际答案均为正；FP 表示预测结果为正而实际答案为负；FN 表示预测结果为负而实际答案为正。

3.5 案例分析

基于 LoRa 微调的 LLMs 被用于从包含 45 篇信号设备故障案例的验证集中进行信息抽取。实验环境与模型训练参数设置见表 1、表 2。

经上述操作，LLMs 能够实现对故障案例文本的信息抽取，LoRA 微调技术对 LLMs 理解高铁故障文本和提高故障知识抽取的准确性产生了积极的作用。

表 1 实验环境配置

Table 1 Experimental environment configuration

实验配置	版本
CPU	NVIDIA 4090
Memory	24 GB
OS	Ubuntu 22.04.3 LTS
Pytorch	2.2.1
CUDA	12.2
transformers	4.40.0

表 2 模型训练的参数设置

Table 2 Parameter setting of model training

参数	值
batch_size	1
gradient_accumulation_steps	4
epochs	10
learning_rate	le-4
weight_decay	0
adam_beta2	0.9
optimizer	AdamW
max_source_length	100
max_target_length	256
lora_r	8
ora_alpha	32
lora dropout	0.1

信息抽取实验结果见表 3。由表 3 可知，在信息抽取的任务中，基于 LoRA 微调的高铁信号设备故障信息抽取模型的实体识别精确度为 60.65%，召回率为 74.90%，调和平均值 $F1$ 分数为 67.03%；还可得到，关系抽取精确度为 72.99%，召回率为 95.01%，调和平均值 $F1$ 分数为 82.56%。

实体关系联合抽取的多 LoRA 微调的 LLMs 模型部署使得信息抽取结果更为精准。联合抽取的三元组中，本文成功预测 TP 值为 362，说明模型能够较为准确地识别高铁信号设备中的关键实体及其关系；FN 值为 59，表明模型在处理某些复杂或隐晦关系时仍有待提高空间。同时，模型识别 FP 值为 186，表明模型在区分真实三元组与非三元组时存在一定的误差。模型的关系抽取精确度为 66.06%，表明模型在识别复杂三元组时保持了一定的准确性；召回率为 85.99%，表明模型能够识别 85.99% 的真实三元组，在发现三元组方面表现良好，但与此同时也导致了较高的 FP 比例；调和平均值 $F1$ 分数

为 74.64%，说明模型综合表现良好。

表 3　信息抽取实验结果展示（验证集）

Table 3　Experimental results of information extraction (verification set)

内容	TP	FP	FN	P/%	R/%	$F1$ - Score/%
三元组	362	186	59	66.06	85.99	74.64
实体	188	122	63	60.65	74.90	67.03
关系	400	148	21	72.99	95.01	82.56

经过多 LoRA 微调的 LLMs 在高铁信号设备故障领域信息抽取任务中展现了显著的有效性，也在复杂文本分析和知识发现中显示出应用价值。

4　知识图谱可视化展示与应用

将上述实体间的关系转化为“实体-关系-实体”三元组后，导入 Neo4j 图数据库中，包含 6 903 个实体，19 215 条关系，形成了高速铁路信号设备故障领域故障图谱。高速铁路信号设备故障领域故障图谱如图 8 所示。

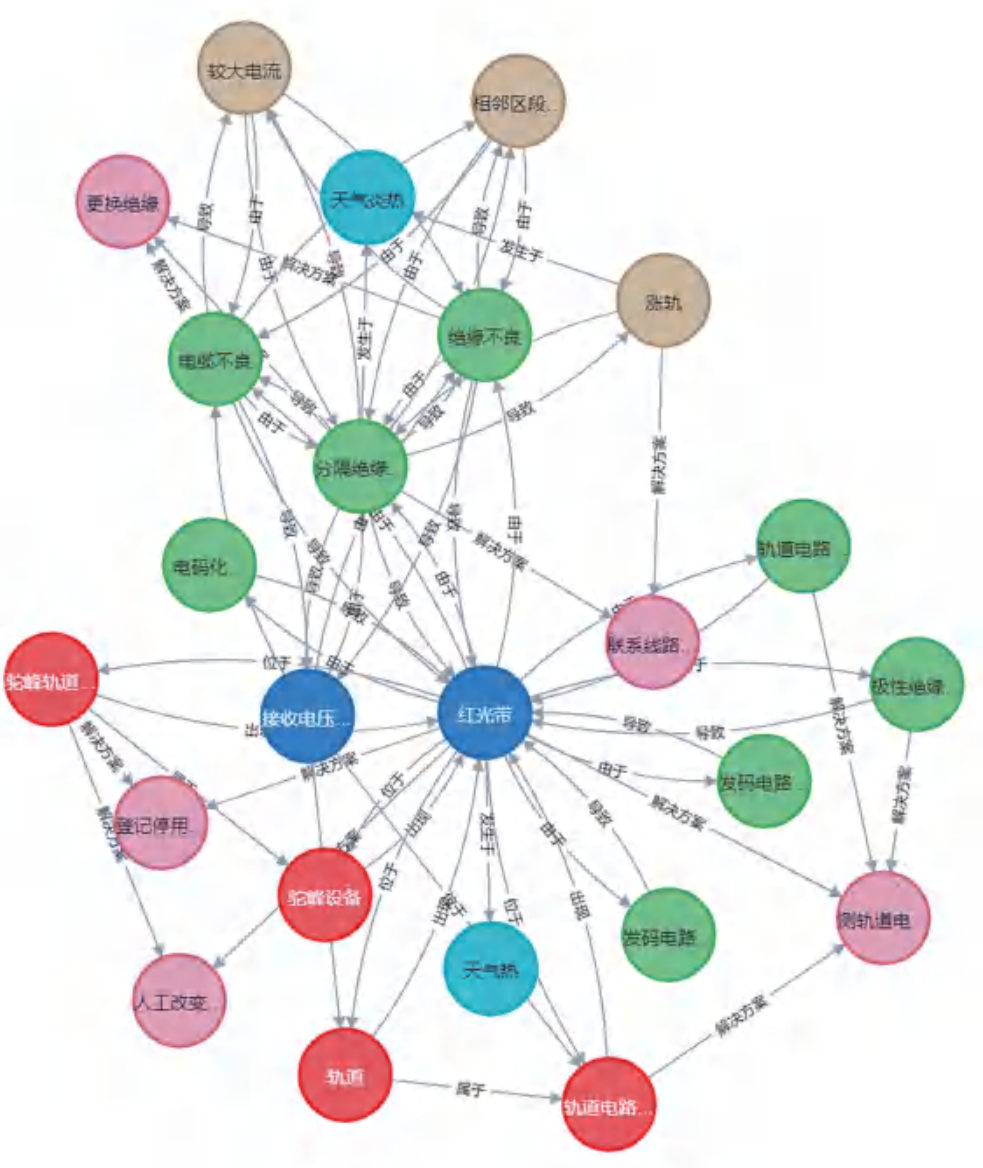


图 8　高速铁路信号设备故障领域故障图谱

Fig. 8　Fault spectrum of high-speed railway signal equipment fault field

高铁信号设备常见故障现象如图 9 所示。通过 Neo4j 检索可知，高铁信号设备常见故障现象主要有驼峰场信号故障、无轨面电流、室外电机不转、CIR 提示注册失败、图像暗或黑屏等。其中，驼峰场信号故障关系展示如图 10 所示。分析指出，驼峰场信号故障出现于驼峰自动化控制系统，故障原因可能包括监控 A 机未启动、有断路或者短路、电缆电气特性差异、接受不到工控机传输的信息等，会导致

减速顶模块亮红灯等。依据故障现象,可以定位故障设备、追溯故障原因、了解故障影响,以做好故障防范、运营维护与应急防控。其他类型数据的关系查询、关联展示亦与之类似。



图 9 高铁信号设备常见故障现象

Fig. 9 Common fault phenomena of high-speed railway signal equipment



图 10 驼峰场信号故障关系展示

Fig. 10 Display of signal fault relationship of hump field

5 结束语

本文设计提出了一种基于 LLMs 的高铁信号设备故障文本数据自动标注方法,明确界定了自动标注任务及其标注规范,设计提示学习策略,实现了文本的自动标注。接着,构建了基于 LoRA 指令微调的高铁信号设备故障文本信息抽取模型,以 Qwen-14B 为基础进行进一步优化。实验结果显示,该模型信息抽取精确度为 66.06%,召回率为 85.99%,*F*1 分数为 74.64%,显示出对复杂三元组的高精度识别能力,并在三元组发现方面表现优异,证明了多 LoRA 微调大模型在故障领域信息抽取任务中的有效性。在此基础上,本文将模型输出结果导入 Neo4j 图数据库中,总计 6 903 个实体,19 215 条关系,实现了知识图谱的可视化展示与应用,有效帮助高铁从业人员迅速定位故障及其原因,为提升铁路安全运营与保障水平提供技术支撑。

参考文献

[1] 刘辉. 中国高速铁路的创新与发展[J]. 领导科学论坛, 2018(12):42-62.

[2] 董兴芝. 面向智能高铁安全保障的知识图谱构建及应用关键技术研究[D]. 北京:中国铁道科学研究院, 2022.

[3] 李新琴. 高速铁路安全文本大数据分析方法研究[D]. 北京:中国铁道科学研究院, 2020.

[4] 左冀铭. 基于铁路故障文本数据的知识图谱建模方法研究[D]. 北京:北京交通大学, 2022.

[5] WEI J, TAY Y, BOMMASANI R, et al. Emergent abilities of Large Language Models[J]. arXiv preprint arXiv, 2206.07682, 2022.

[6] ZHAO W, ZHOU K, LI J, et al. A survey of large language models[J]. arXiv preprint arXiv, 2303.18223, 2023.

[7] ELHAMMADI S, LAKSHMANAN L V S, NG R, et al. A High Precision Pipeline for Financial Knowledge Graph Construction [C]//Proceedings of the 28th International Conference on Computational Linguistics. ICCL, 2020: 967-977.

[8] LI Linfeng, WANG Peng, YAN Jun, et al. Real-world data medical knowledge graph: Construction and applications [J]. Artificial Intelligence in Medicine, 2020, 103: 101817.

[9] 李瑞琪, 李佳, 韦莎, 等. 知识图谱产业应用及标准化需求研究[J]. 信息技术与标准化, 2021(5): 11-16.

[10] LIU Jiaying, REN Jing, ZHENG Wenqing, et al. Web of scholars: A scholar knowledge graph[J]. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'20). New York: ACM, 2020: 2153-2156.

[11] WANG Xiang, HE Xiangnan, CAO Yixin, et al. KGAT: Knowledge graph attention network for recommendation[J]. arXiv preprint arXiv, 1905.07854, 2019.

[12] BAO J, DUAN N, YAN Z, et al. Constraint-based question answering with knowledge graph [EB/OL]. (2024-03-20). https://aclanthology.org/C16-1236/.

[13] 杨连报. 铁路事故故障文本大数据分析关键技术研究及应用[D]. 北京:中国铁道科学研究院, 2018.

[14] 林海香, 赵正祥, 卢冉, 等. 面向高速铁路道岔设备故障处置的知识图谱构建与应用[J]. 铁道学报, 2024, 46(7): 73-80.

[15] 纪贵阳, 王裴岩, 余卓. 面向工艺规范文本的大语言模型知识注入方法研究[J]. 计算机科学与探索, 2024, 18(9): 2361-2369.

[16] 李佳沂, 黄瑞章, 陈艳平, 等. 结合提示学习和 Qwen 大语言模型的裁判文书摘要方法[J]. 清华大学学报(自然科学版),

2024，64（12）：2007-2018.
[17]孙熠衡，刘茂福. 基于知识提示微调的标书信息抽取方法[J]. 计算机应用，2025，45（4）：1169-1176.
[18]杨燕，叶枫，许栋，等. 融合大语言模型和提示学习的数字孪生水利知识图谱构建[J]. 计算机应用，2025，45（3）：785-793.
[19]SEKINE S. NYU：Description of the Japanese Ne system used for Met-2[EB/OL]. （1998-04-29）. https://aclanthology. org/M98-1019. pdf.
[20]AMIT S. Introducing the knowledge graph[R]. Official Blog of Google，2012.
[21]ZHONG Lingfeng，WU Jia，LI Qian，et al. A comprehensive survey on automatic knowledge graph construction[J]. ACM Computing Surveys，2024，56（4）：94.
[22]刘峤，李杨，段宏，等. 知识图谱构建技术综述[J]. 计算机研究与发展，2016，53(3)：582-600.
[23]LIU Pengfei，YUAN Weizhe，FU Jinlan，et al. Pre-train，prompt，and predict：A systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys，2023，55(9)：35.
[24]黄俊涛. 科技领域知识图谱构建技术研究[D]. 北京：北方工业大学，2024.
[25]PETRONI F，ROCKT A，SCHEL T，et al. Language models as knowledge bases？[J]. arXiv preprint arXiv，1909. 01066，2019.
[26]LESTER B，AL-RFOU R，CONSTANT N. The power of scale for parameter-efficient prompt tuning[J]. arXiv preprint arXiv，2104. 08691，2021.