文章编号: 1671-251X(2025)02-0076-09 DOI: 10.13272/j.issn.1671-251x.2024080031

基于大语言模型的矿山事故知识图谱构建

张朋杨[1,2], 生龙[1,2], 王巍[1,2], 魏忠诚[1,2], 赵继军[1,2]

(1. 河北工程大学 信息与电气工程学院, 河北 邯郸 056038;
2. 河北工程大学 河北省安防信息感知与处理重点实验室, 河北 邯郸 056038)

摘要: 现有矿山领域知识图谱构建方法在预训练阶段需要大量人工标注的高质量监督数据, 人力成本高且效率低。大语言模型(LLM)可在少量人工标注的高质量数据下显著提高信息抽取的质量且效率较高, 然而 LLM 结合 Prompt 的方法会产生灾难性遗忘问题。针对上述问题, 将图结构信息嵌入到 Prompt 模板中, 提出了图结构 Prompt, 通过在 LLM 上嵌入图结构 Prompt, 实现基于 LLM 的矿山事故知识图谱高质量构建。首先, 收集煤矿安全生产网公开的矿山事故报告并进行格式修正、冗余信息剔除等预处理。其次, 利用 LLM 挖掘矿山事故报告文本中蕴含的知识, 对矿山事故报告文本中的实体及实体间关系进行 K-means 聚类, 完成矿山事故本体构建。然后, 依据构建的本体进行少量数据标注, 标注数据用于 LLM 的学习与微调。最后, 采用嵌入图结构 Prompt 的 LLM 进行信息抽取, 实例化实体关系三元组, 从而构建矿山事故知识图谱。实验结果表明: 在实体抽取和关系抽取任务中, LLM 的表现优于通用信息抽取(UIE)模型, 且嵌入图结构 Prompt 的 LLM 在精确率、召回率、F_1 值方面均高于未嵌入图结构 Prompt 的 LLM。

关键词: 矿山事故; 知识图谱; 大语言模型; 图结构 Prompt; 本体构建; 信息抽取

中图分类号: TD67 文献标志码: A

Construction of a mine accident knowledge graph based on Large Language Models

ZHANG Pengyang[1,2], SHENG Long[1,2], WANG Wei[1,2], WEI Zhongcheng[1,2], ZHAO Jijun[1,2]

(1. School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, China;
2. Hebei Provincial Key Laboratory of Security Information Perception and Processing, Hebei University of Engineering, Handan 056038, China)

Abstract: Current methods for constructing knowledge graphs in the field of mining require a large amount of manually labeled high-quality supervised data during the pre-training stage, resulting in high labor costs and low efficiency. Large Language Models (LLMs) can significantly improve the quality and efficiency of information extraction with only a small amount of manually labeled high-quality data. However, the prompt-based approach in LLMs suffers from catastrophic forgetting. To address this issue, graph-structured information was embedded into the prompt template and a Graph-Structured Prompt was proposed. By integrating this prompt into the LLM, high-quality construction of a mine accident knowledge graph based on the LLM was achieved. First, publicly available mine accident reports were collected from the Coal Mine Safety Production Network and preprocessed through formatting corrections and redundant information removal. Next, the LLM was utilized to extract knowledge embedded in the accident reports and K-means clustering was used to classify entities and

收稿日期: 2024-08-14; **修回日期**: 2025-02-28; **责任编辑**: 盛男。

基金项目: 国家自然科学基金资助项目(61802107); 河北省高等学校科学技术研究项目(ZD2020171); 河北省省级科技计划资助项目(22567624H)。

作者简介: 张朋杨(1998—), 男, 河北邯郸人, 硕士研究生, 主要研究方向为自然语言处理、知识图谱, E-mail: zhangpy996@163.com。通信作者: 生龙(1982—), 男, 河北邯郸人, 副教授, 博士, 主要研究方向为自然语言处理、人工智能与城市公共安全, E-mail:shenglong@hebeu.edu.cn。

引用格式: 张朋杨, 生龙, 王巍, 等. 基于大语言模型的矿山事故知识图谱构建[J]. 工矿自动化, 2025, 51(2): 76-83, 105.
ZHANG Pengyang, SHENG Long, WANG Wei, et al. Construction of a mine accident knowledge graph based on Large Language Models[J]. Journal of Mine Automation, 2025, 51(2): 76-83, 105.

扫码移动阅读

relationships, thereby completing the construction of the mine accident ontology. Then, a small amount of data were labeled based on the ontology, which was used for LLM training and fine-tuning. Finally, the LLM embedded with the Graph-Structured Prompt was employed for information extraction, instantiating entity-relation triples to construct the mine accident knowledge graph. Experimental results showed that LLMs outperformed the Universal Information Extraction (UIE) model in entity and relationship extraction tasks. Moreover, the LLM embedded with the Graph-Structured Prompt achieved higher precision, recall, and F1 scores compared to those without it.

Key words: mine accident; knowledge graph; Large Language Model; Graph-Structured Prompt; ontology construction; information extraction

0　引言

知识图谱是结构化的语义网络知识库，其以三元组的形式结构化表示客观世界中存在的概念、实体及其关联关系[1]。在矿山领域，大量的事故信息通常以报告文本的形式存在，结构化程度低，难以实现事故信息的数据挖掘及知识推理。构建矿山事故知识图谱可有效整合报告文本中事故概述、经过及原因中离散的实体及实体间关系，将矿山事故中事故地点、类型、原因等关键因素及其之间的关系以三元组的形式进行存储，提高矿山事故信息的结构化程度，从而实现对事故信息的数据挖掘及知识推理，为矿山风险识别与预防、应急响应与决策支持、事故分析与原因追溯、事故预防措施制订等一系列矿山智能化安全管理系统建设提供数据支撑[2]。

在矿山领域知识图谱构建中，郭晓黎等[3]对煤矿安全事故的种类及类间关系进行分析，建立了煤矿安全事件本体，为构建煤矿安全事件知识图谱提供了理论指导。潘理虎等[4]提出了一种基于七步法、METHONTOLOGY 法的本体构建方法，采用知识存储映射算法将煤矿领域本体映射到 Neo4j 图数据库中，完成了煤矿领域知识图谱的构建。李蓓等[5]基于煤矿灾害事件概念语义分类和煤矿灾害事件描述属性，构建了煤矿灾害事件本体，为构建煤矿灾害知识图谱提供了理论借鉴。曹现刚等[6]采用预训练的 Lattice−LSTM 模型进行实体识别，采用基于弱监督学习的 Bootstrapping 方法进行关系抽取，完成了煤矿设备维护知识图谱的构建。王忠强等[7]针对智慧矿山领域的知识要素，提出了基于依存句法分析的实体抽取方法，并根据语句结构特点，设计了依存句法树结构，构建了智慧矿山知识图谱。韩一搏等[8]采用联合编码器将收集到的综采设备数据转换为向量表示，在解码时采用预训练的 Lattice−LSTM 模型，完成了综采设备实体识别，实现了煤矿综采设备知识图谱构建。现有矿山领域知识图谱构建多采用基于预训练模型的方法，该方法在预训练阶段需要大量人工标注的高质量监督数据[9]，而标注高质量的监督数据需要投入大量人力资源，并且效率较低。

近年来，大语言模型（Large Language Model，LLM）在自然语言理解、学习和表达上取得重大突破，LLM 可在少量人工标注的高质量数据下显著提高信息抽取的质量且效率较高，广泛应用于各领域的信息抽取任务[10-12]。M. Agrawal 等[13]证明了 LLM 在没有针对专业领域进行训练的情况下，仍可在零样本和少样本的医疗文本信息抽取任务中表现良好。S. Wadhwa 等[14]证明了 LLM 可高质量地完成少样本新闻信息抽取。冯钧等[15]证明了 LLM 在未针对水利调度领域文本进行训练的情况下，可在少样本的调度文本中实现高质量信息抽取。因此，将 LLM 应用于零样本和少样本的矿山事故信息抽取任务，从而构建矿山事故知识图谱是可行的。

随着 LLM 的不断发展，Prompt 已经成为自然语言处理领域的一种前沿方法，为 LLM 的使用提供了一种更有效和更具成本效益的方法[16]。然而，LLM 结合 Prompt 的方法会产生灾难性遗忘问题[17]，致使模型原始理解上下文能力丧失，难以处理蕴含复杂关系的信息抽取任务。图结构信息可增强模型对实体间复杂关系的理解能力，提高实体抽取和关系抽取的准确率。Li Lei 等[18]提出了一种基于上下文感知图结构的图卷积网络来进行事件检测任务，提高了模型理解语义上下文信息的能力。Zhang Qianjin 等[19]将实体间的隐式图结构信息融入知识图谱嵌入模型，在关系预测任务上实现了性能提升，增强了模型对上下文的理解能力。因此，本文将图结构信息嵌入到 Prompt 模板中，提出了图结构 Prompt，通过在 LLM 上嵌入图结构 Prompt，提升矿山事故知识图谱的构建质量。首先，对收集到的矿山事故报告进行预处理得到原始语料。其次，按照相关文件要求，使用 LLM 对矿山事故报告文本中的事故信息进行 K−means 聚类分析，挖掘事故信息中的实体及实体

间关系，完成事故本体构建。然后，将矿山事故报告文本中蕴含的图结构信息嵌入到 Prompt 模板中，进行矿山事故实体及关系的信息抽取，实例化实体关系三元组。最后，根据抽取到的实体关系三元组构建知识图谱。

1　基于 LLM 的矿山事故知识图谱构建

本文采用自顶向下的方式构建矿山事故知识图谱，流程如图 1 所示。知识图谱涵盖模式层和数据层[20]。模式层在数据层之上，主要通过本体来规范数据层中的一系列事实表达；数据层主要由一系列事实三元组组成，知识以事实为单位进行存储。通过网络爬虫技术，收集煤矿安全生产网公开的矿山事故报告，经过预处理得到原始语料，使用 LLM 对事故报告中的名词、名词短语及动词进行批量化抽取。在模式层中，实体集由事故报告中的名词、名词短语组成，关系集由事故报告中的动词组成。通过 LLM 对实体集和关系集中的元素进行聚类分析，同时结合《矿山生产安全事故报告和调查处理办法》《生产安全事故报告和调查处理条例》《煤矿安全生产条例》中要求事故报告应包含的内容，构建矿山事故本体。本体构建完成后，对原始语料进行少量的人工标注，标注数据用于 LLM 的学习与微调。按照本体中的概念定义设计信息抽取模板。在数据层中，将矿山事故报告中不同文本中实体及实体间关系的图结构信息嵌入到信息抽取模板中，使用 LLM 进行实体及关系抽取，得到矿山事故文本中的实体关系三元组，完成数据的实例化。

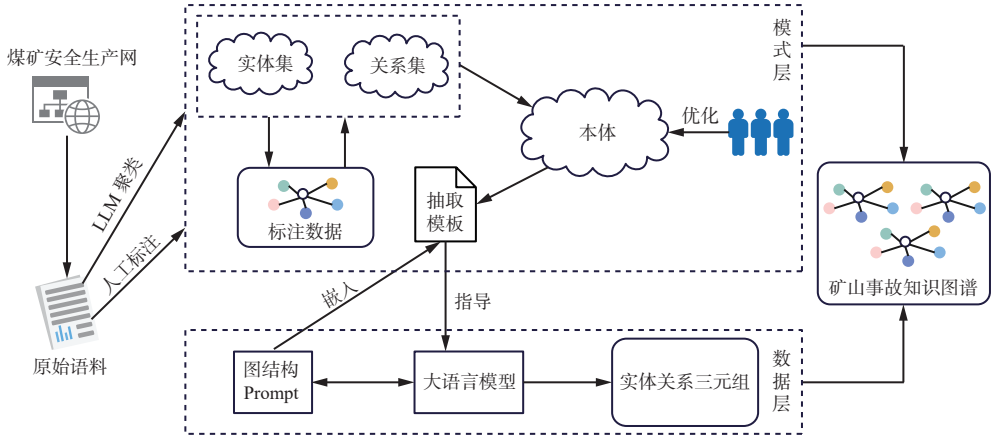


图 1　矿山事故知识图谱构建流程

Fig. 1　Construction process of mine accident knowledge graph

1.1　模式层构建

模式层是知识图谱的概念模型和逻辑基础，可借助本体定义的规则和公理对数据层进行规范约束[3]。对矿山事故报告文本分析可知，该报告文本中蕴含丰富的实体对象和关系。使用 LLM 并结合煤矿生产文件、煤矿设备文件和安全防治文件对矿山事故报告文本进行了实体关系挖掘、聚类和总结归纳。

矿山事故报告按照结构可划分为事故概述、事故原因、事故单位情况和事故发生经过。实体关系挖掘过程如图 2 所示。首先，本文利用 LLM 按事故报告结构分批获取矿山事故报告文本中的所有名词及名词短语，同时，提示 LLM 采用粗粒度分词标准。例如，事故原因文本为“事故直接原因：工作面放炮崩歪单体液压支柱，工人在空顶情况下违章打设支柱，冒落的岩石砸倒支柱，支柱砸伤其头部致死。”，采用粗粒度分词标准后的分词结果为“事故/直接原因/：/工作面放炮崩歪单体液压支柱/，/工人/在/空顶情况下/违章打设支柱/，/冒落的岩石/砸倒支柱/，/支柱/砸伤其头部致死/。”。采用粗粒度分词标准可以保留事故原因的语义完整性，有助于模型理解上下文，减少分词歧义。其次，获取事故报告中的所有名词及名词短语后，通过 LLM 对所有名词及名词短语进行 K−means 聚类。如将具体名词“单体液压支柱”“风镐”“液压枪”等聚类在一起，并进一步映射为“设备”标签；将“运输事故”“顶板事故”“水害事故”等聚类在一起，并映射为“事故类型”标签；将“2 号采煤工作面”“硐室”“106 号—115 号液压支架间”等聚类在一起，并映射为“地点”标签；将“2023 年 6 月 8 日 6 时许”“60 万 t/a”“未打设临时支护”等分类为其他标签。得到聚类数据后，将同标签的名词及名词短语放入同一集合中，采用 Dice 系数对聚类后的每个标签集合进行相似性度量，即两两比较集合中文本元素的重复度。Dice 系数越接近 1，表示 2 个集合越相似。如果相似，则重复上述步骤进行进一步聚类，否则根据集合中的元素并结合事故文本特征进行标签映射。最后，得到事故核心、机构、事故原因、设备、事件、人员和证照 7 类实体。

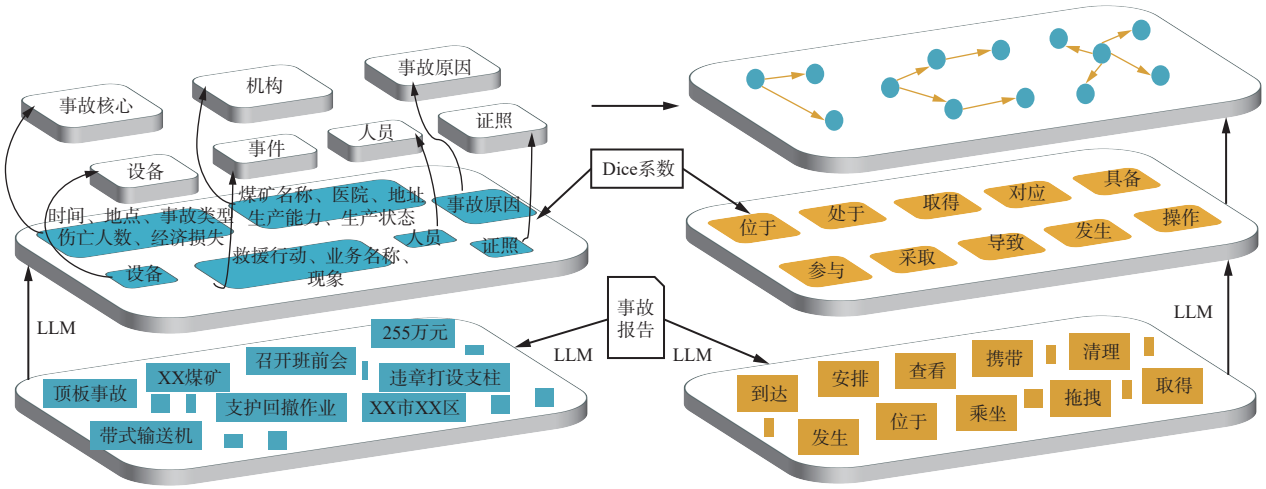


图 2　实体关系挖掘过程

Fig. 2　Entity-relationship mining process

在对语料中的关系进行挖掘时，首先，利用 LLM 按事故报告结构分批获取矿山事故报告文本中的所有动词。然后使用 LLM 对获取到的所有动词进行 K−means 聚类，同样使用 Dice 系数对聚类后的动词集合进行相似性度量，结合行业实际情况进行调整。最后获得位于、处于、取得、对应、具备、参与、采取、导致、发生、操作 10 种关系。

此外，在对训练数据中少量样本进行数据标注时，为提高人工标注的效率，提升实体辨识度，对前文所述 7 类实体中的事故核心、机构、事件和证照 4 类实体进行了细分，细分后的实体及实体间关系如图 3 所示。将事故核心实体细分为时间、地点、事故类型、死亡人数、受伤人数和经济损失，将机构实体细分为煤矿名称、地址、生产能力和生产状态，将事件实体细分为业务名称、救援行动和现象，将证照实体细分为证照编号和证照有效期。最终得到矿山事故领域实体及实体间关系。

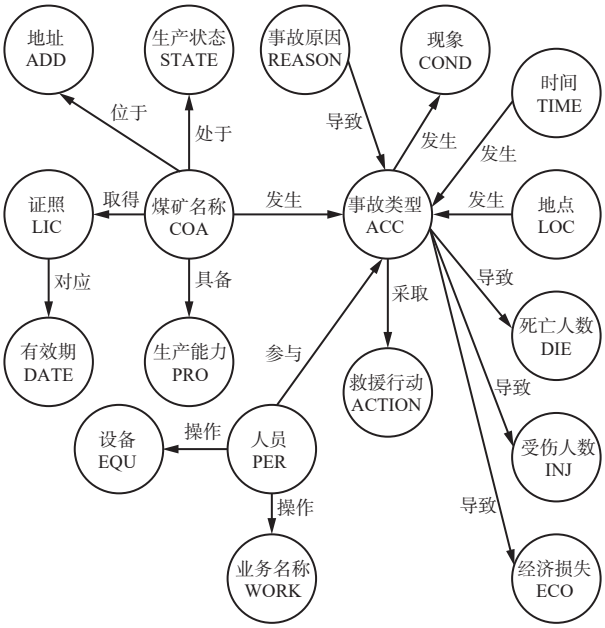


图 3　实体及实体间关系

Fig. 3　Entities and relationships between entities

1.2　数据层构建

在数据层中，知识以“实体−关系−实体”或“实体−属性−属性值”的三元组形式存在。根据模式层中对实体及实体间关系的定义，对事故文本进行信息抽取，构建矿山事故知识图谱的数据层。

根据矿山事故报告文本中实体及实体间关系结构，可将图结构信息分为 3 类。事故概述文本和事故原因文本的图结构信息相同。以事故概述文本的图结构信息（图 4）为例，按照矿山事故本体中实体及实体间关系，该文本中 XX 煤矿为起始节点，其余节点为终止节点。起始节点与各个终止节点之间存在发生、导致等不同的关系，并且节点之间只有一对多的图结构信息，在对事故概述文本进行信息抽取时，可定义该部分文本的 Prompt 模板，将各节点之间的

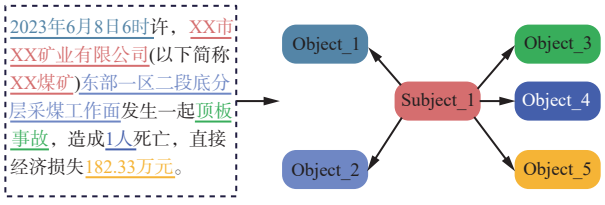


图 4　事故概述文本的图结构信息

Fig. 4　Graph structure information of accident overview text

关系和图结构信息嵌入 Prompt 模板。

事故单位情况文本的图结构信息如图 5 所示。按照矿山事故本体中实体及实体间关系，该文本中 XX 煤矿为起始节点，证号为中间节点，证照有效期为终止节点。起始节点和终止节点通过中间节点进行连接，各个节点之间存在取得、对应等不同的关系，并且起始节点与中间节点之间只存在一对多的

关系，中间节点和终止节点之间存在一对一的关系，在对事故单位证照情况进行信息抽取时，可固定该部分文本的 Prompt 模板，将各节点之间的关系和图结构信息嵌入 Prompt 模板。

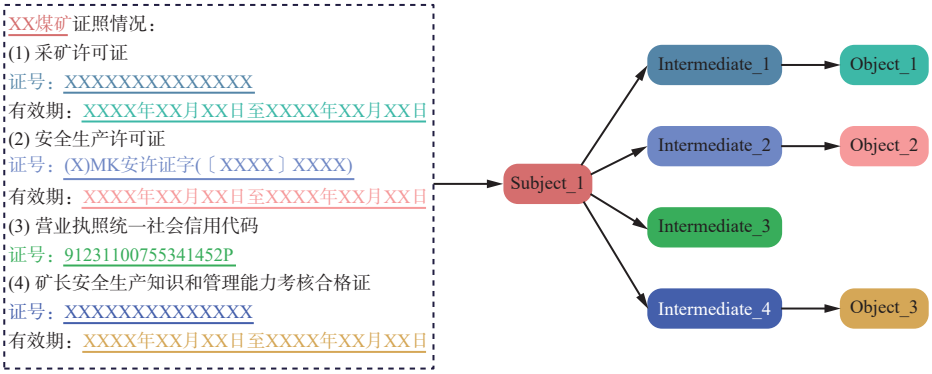


图 5　事故单位情况文本的图结构信息

Fig. 5　Graph structure information of accident unit situation text

事故发生经过文本的图结构信息如图 6 所示。在该文本中，事件是按照时间顺序发生的，各个时间点中都有人员参与，且参与其中的人员都在进行相应活动，如操作设备、进行业务施工、采取救援措施等。因此，按照矿山事故本体中实体及实体间关系，将该文本中时间点作为起始节点，人员作为中间节点，具体业务名称作为终止节点。起始节点和终止节点通过中间节点进行连接，各个节点之间存在参与、操作、对应等不同的关系，并且起始节点与中间节点之间存在一对多的关系，中间节点与终止节点之间存在一对一和一对多的关系，在对事故发生经过文本进行信息抽取时，可固定该部分文本的 Prompt 模板，将事故发生经过文本中各节点之间关系和图结构信息嵌入 Prompt 模板。

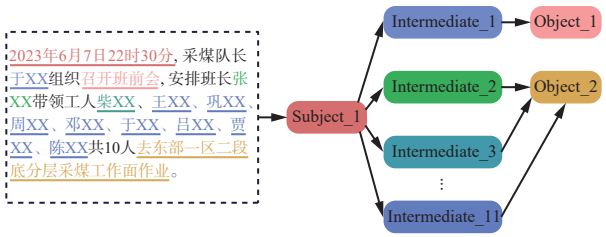


图 6　事故发生经过文本的图结构信息

Fig. 6　Graph structure information of accident occurrence text

根据矿山事故报告文本各部分内容的差异性，对事故概述、事故原因、事故单位情况和事故发生经过进行信息抽取时采用不同的 Prompt 模板。信息抽取过程如图 7(a)所示，在 Prompt 模板中嵌入原始语料中实体之间的图结构信息，将嵌入图结构信息的 Prompt 模板和待抽取文本输入 LLM，使用人工标注的训练集数据指导 LLM 进行矿山事故中实体及实体间复杂关系的学习，对模型参数进行微调，使 LLM 在当前对话中保持对该任务的信息抽取能力。具体信息抽取案例如图 7(b)所示，模板中的 triples 表示当前待抽取文本中所包含的三元组，target 表示嵌入图结构信息后的文本，Subject_X 标签表示起始节点，Object_X 标签表示终止节点，target_text 表示待抽取文本的内容，ner2ent 表示待抽取文本中所包含的实体节点与标签的对应关系。

在信息抽取时，按矿山事故报告结构对原始语料进行划分，将嵌入图结构信息的 Prompt 模板和待抽取文本输入 LLM 进行批量化信息抽取，最终 LLM 输出抽取到的实体关系三元组。

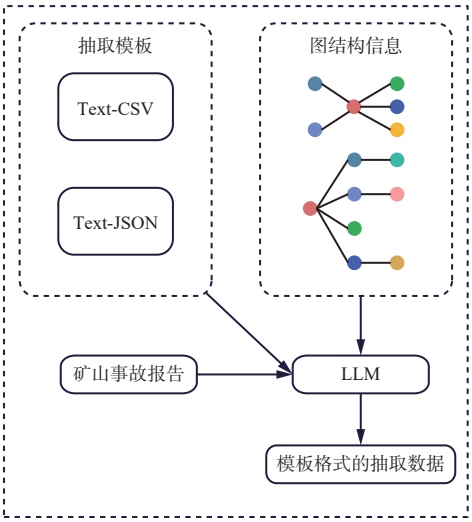
2　实验验证

为验证本文方法的可行性和有效性，开展实验验证。用于实验验证的 LLM 包括 GPT−3.5，GLM_4，ERNIE−4.0 及 Qwen−7B−chat，将 LLM 的信息抽取结果与通用信息抽取(Universal Information Extraction，UIE)模型[21]的信息抽取结果进行对比。

2.1　数据集构建

在矿山事故信息抽取任务中，目前尚无公开的数据集，因此需要自行构建数据集。本文收集的数据来源于煤矿安全生产网，通过网络爬虫获取原始语料文本，选取 7 类矿山事故，共包含 2 532 个矿山事故报告文本，人工标注 253 个矿山事故报告文本，将标注后的数据按照 7∶3 的比例划分为训练集和测试集。

通过网络爬虫获取到的原始矿山事故报告文本存在实体关系和专业词汇复杂及实体嵌套等问题，使得本体构建变得困难，且直接对原始语料进行信息抽取并不能得到高质量的抽取结果。此外，收集到的原始矿山事故报告存在诸多冗余信息和格式混乱数据，无法将其直接用于信息抽取任务。为改善上述问题，需要对数据进行预处理，以提高语料库的



（a）信息抽取过程

模板：{"triples":[["XXXX","发生","XXXX"],["XXXX","发生","XXXX"],["XXXX","发生","XXXX"],["XXXX","导致","XXXX"],["XXXX","造成","XXXX"],["XXXX","造成","XXXX"]],"target": "Object_1 许，XX市XX矿业有限公司(简称 Subject_1)Object_2 发生一起 Object_3，事故造成 Object_4，依据《企业职工伤亡事故经济损失统计标准》(GB 6721—1986)和有关规定统计，直接经济损失 Object_5，属Object_6 事故。","target_text": "2023年6月8日6时许，XX市X矿业有限公司(简称XX煤矿)东部一区二段底分层采煤工作面发生一起顶板事故，事故造成1人死亡，依据《企业职工伤亡事故经济损失统计标准》(GB 6721—1986)和有关规定统计，直接经济损失182.33万元，属一般事故。"

"ner2ent":

{"Subject_1":"XXXX","Object_1":"XXXX","Object_2":"XXXX","Object_3":"XXXX","Object_4": "XXXX","Object_5": "XXXX","Object_6": "XXXX"}}输入：2023年6月8日6时许，XX市X矿业有限公司(简称XX煤矿)东部一区二段底分层采煤工作面发生一起顶板事故，事故造成1人死亡，依据《企业职工伤亡事故经济损失统计标准》(GB 6721—1986)和有关规定统计，直接经济损失182.33万元，属一般事故。

提示：1. “ner2ent”部分：Subject_1，Object_1等中填入"XX煤矿","2023年6月8日6时","东部一区二段底分层采煤工作面","顶板事故","1人死亡","182.33万元","一般"中最合适的一个。输出为json格式。2. 将target键对应值的内容用ner2ent中的键进行替换，用空格与相邻文本进行分隔…

（b）信息抽取案例

图 7　信息抽取过程及案例

Fig. 7　Information extraction process and case example

构建质量。

根据国家矿山安全监察局《关于印发〈矿山安全生产事故报告和调查处理办法〉的通知》(矿安〔2023〕7 号)第十条要求，对采集到的原始矿山事故报告进行预处理，如图 8 所示。首先，对原始语料进行数据清洗，修正格式混乱的数据，同时对报告内容进行精简，删除矿山事故报告中的冗余信息，去除事故责任追究与处理建议等与本体构建无关信息，保留事故发生单位概况，事故发生的时间、地点、事故类别，事故的简要经过，事故已经造成伤亡人数、涉险人数、失踪人数和初步估计的直接经济损失等必要内容。然后，进行实体对齐，例如针对 XX 市 XX 区 XX 煤业有限公司(以下简称“XX 煤业”)，统一使用简称之后的煤矿名称。最后，统一矿山事故报告结构，将矿山事故报告保留的内容进一步精炼为事故概述、事故原因、事故单位情况和事故发生经过 4 个部分内容。

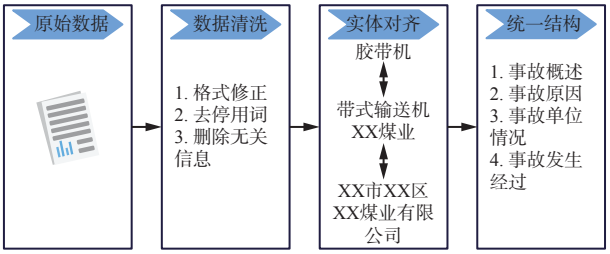


图 8　数据预处理流程

Fig. 8　Data preprocessing process

2.2　信息抽取结果

分别采用 GPT−3.5, GLM_4, ERNIE−4.0 及 Qwen−7B−chat 这 4 个 LLM 与 UIE 模型对矿山事故报告中的实体和关系进行抽取，结果见表 1。

表 1　UIE 模型与 LLM 在信息抽取任务上的对比结果

Table 1　Comparison results of Universal Information Extraction(UIE) model and Large Language Model(LLM) in information extraction tasks

模型	实体抽取			关系抽取		
	精确率	召回率	F_1	精确率	召回率	F_1
UIE	0.894	0.827	0.859	0.713	0.627	0.667
GPT−3.5	0.893	0.847	0.870	0.887	0.904	0.895
GLM_4	0.956	0.850	0.901	0.910	0.885	0.898
ERNIE−4.0	0.752	0.836	0.792	0.788	0.817	0.802
Qwen−7B−chat	0.883	0.855	0.869	0.862	0.881	0.871

由表 1 可知：在实体抽取任务中，UIE 模型表现稳定但整体略差于 LLM；在关系抽取任务中，LLM 表现显著优于 UIE 模型。这是因为 UIE 模型依赖于预定义的结构化模式，难以灵活处理多样化的关系类型；而 LLM 凭借强大的上下文理解能力、生成式框架及对大规模预训练数据的深度学习能力，能够更好地捕捉语义关联和隐含关系，此外，LLM 在处理动态和多样化任务时表现出更强的泛化能力，能够更准确地构建实体之间的关系，从而在实体抽取和关系抽取任务中取得更好的效果。

在 GPT−3.5, GLM_4, ERNIE−4.0 和 Qwen−7B−chat 上开展嵌入图结构 Prompt 和未嵌入图结构 Prompt 的对比实验，分别对测试集数据进行实体抽取和关系抽取，结果见表 2。

由表 2 可知，在 LLM 中嵌入图结构 Prompt 后的信息抽取结果明显优于未嵌入图结构 Prompt。未嵌入图结构 Prompt 的 LLM 虽能捕捉一定的语义信息，

表 2　LLM 嵌入图结构 Prompt 前后在信息抽取任务上的对比结果

Table 2　Comparison results of information extraction tasks before and after LLM embedded with Graph-Structured Prompt

模型		实体抽取			关系抽取		
		精确率	召回率	F_1	精确率	召回率	F_1
GPT−3.5	未嵌入图结构Prompt	0.775	0.835	0.804	0.803	0.791	0.797
	嵌入图结构Prompt	0.893	0.847	0.870	0.887	0.904	0.895
GLM_4	未嵌入图结构Prompt	0.831	0.679	0.793	0.785	0.794	0.789
	嵌入图结构Prompt	0.956	0.850	0.901	0.910	0.885	0.898
ERNIE−4.0	未嵌入图结构Prompt	0.673	0.731	0.701	0.695	0.683	0.689
	嵌入图结构Prompt	0.752	0.836	0.792	0.788	0.817	0.802
Qwen−7B−chat	未嵌入图结构Prompt	0.761	0.748	0.754	0.792	0.731	0.760
	嵌入图结构Prompt	0.883	0.855	0.869	0.862	0.881	0.871

但在精确率和召回率上存在局限性，尤其在处理复杂图结构数据时，难以充分利用节点和边之间的关系信息。而嵌入图结构 Prompt 可帮助 LLM 更好地理解图中节点和边之间的关系，并将图结构信息保留至低维空间表征中，提升捕捉实体间复杂关系的能力。

2.3　知识图谱构建结果

利用嵌入图结构 Prompt 的 LLM 从矿山事故报告中抽取事故概述、事故原因、事故单位情况和事故发生经过所包含的实体及实体间关系信息，生成矿山事故知识图谱三元组，并将其存储在 Neo4j 图数据库中，从而构建矿山事故知识图谱。

使用 Cypher 语句可对 Neo4j 图数据库中的矿山事故进行查询。以顶板事故为例，查询某一煤矿发生的顶板事故，该顶板事故的事故概述、事故原因、事故单位情况和事故发生经过所涵盖的实体关系三元组构成的知识图谱如图 9 所示。

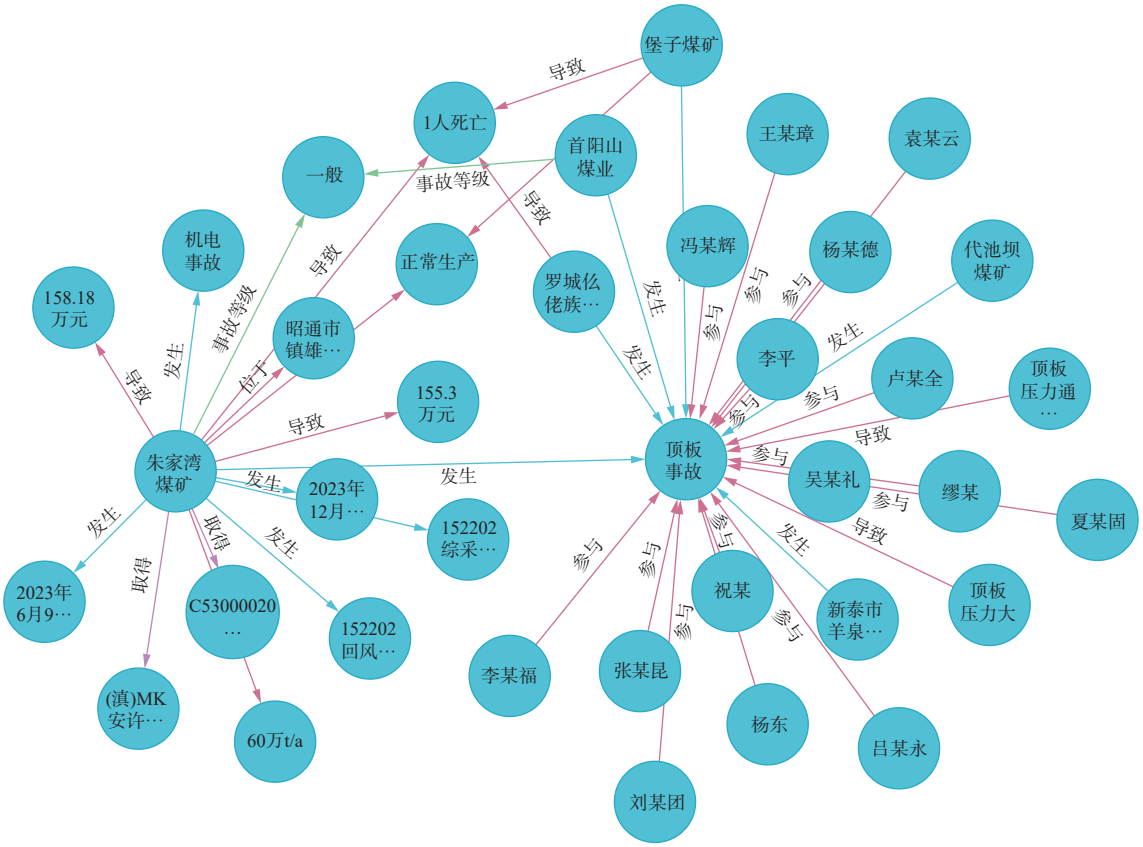


图 9　顶板事故知识图谱

Fig. 9　Knowledge graph of roof accident

3　结论

1）通过 LLM 对矿山事故报告文本中的名词、名词短语及动词进行 K−means 聚类分析，使用 Dice 系数对聚类后的集合进行相似性度量，并结合煤矿领域相关规范性文件，可快速、高效地完成煤矿领域事故本体构建，生成矿山事故知识图谱三元组，实现矿山事故信息的结构化表示。

2）在 LLM 上嵌入图结构 Prompt，提升了 LLM 实体抽取和关系抽取的准确率，从而在少量的标注数据下快速实现矿山事故知识图谱的高质量构建。

3）由于数据来源于煤矿安全生产网的矿山事故报告，文本结构相对固定，文本类型相对单一。在未来的研究中，可提高数据源的多样性，进一步完善矿山事故知识图谱，探索在矿山事故原因分析、救援策略决断、防范措施制订和事故报告自动生成等场景下的应用。

参考文献（References）：

[1] JI Shaoxiong，PAN Shirui，CAMBRIA E，et al. A survey on knowledge graphs：representation，acquisition，and applications[J]. IEEE Transactions on Neural Networks and Learning Systems，2022，33(2)：494-514.

[2] 王国法，任怀伟，赵国瑞，等. 智能化煤矿数据模型及复杂巨系统耦合技术体系[J]. 煤炭学报，2022，47(1)：61-74.
WANG Guofa，REN Huaiwei，ZHAO Guorui，et al. Digital model and giant system coupling technology system of smart coal mine[J]. Journal of China Coal Society，2022，47(1)：61-74.

[3] 郭晓黎，王宇，刘瑞祥. 面向煤矿安全事件本体模型研究与应用[J]. 中国煤炭，2014，40(12)：113-116.
GUO Xiaoli，WANG Yu，LIU Ruixiang. Research and application of event ontology model of coal mine accidents[J]. China Coal，2014，40(12)：113-116.

[4] 潘理虎，张佳宇，张英俊，等. 煤矿领域知识图谱构建[J]. 计算机应用与软件，2019，36(8)：47-54，59.
PAN Lihu，ZHANG Jiayu，ZHANG Yingjun，et al. Construction of knowledge graph in coal mine domain[J]. Computer Applications and Software，2019，36(8)：47-54，59.

[5] 李蓓，王鹏，杨政，等. 基于多层次语义约束的煤矿灾害事件本体模型构建[J]. 陕西煤炭，2024，43(4)：146-149.
LI Bei，WANG Peng，YANG Zheng，et al. Disaster event ontology model building of coal mine based on multi-level semantic constraints[J]. Shaanxi Coal，2024，43(4)：146-149.

[6] 曹现刚，张梦园，雷卓，等. 煤矿装备维护知识图谱构建及应用[J]. 工矿自动化，2021，47(3)：41-45.
CAO Xiangang，ZHANG Mengyuan，LEI Zhuo，et al. Construction and application of knowledge graph for coal mine equipment maintenance[J]. Industry and Mine Automation，2021，47(3)：41-45.

[7] 王忠强，宋建鑫，余数三，等. 基于依存句法分析的智慧矿山知识图谱构建方法[J]. 矿业研究与开发，2023，43(10)：232-240.
WANG Zhongqiang，SONG Jianxin，YU Shusan，et al. A method of constructing knowledge graph of intelligent mines based on dependency syntax analysis[J]. Mining Research and Development，2023，43(10)：232-240.

[8] 韩一搏，董立红，叶鸥. 基于联合编码的煤矿综采设备知识图谱构建[J]. 工矿自动化，2024，50(4)：84-93.
HAN Yibo，DONG Lihong，YE Ou. Construction of knowledge graph for fully mechanized coal mining equipment based on joint coding[J]. Journal of Mine Automation，2024，50(4)：84-93.

[9] ZHONG Lingfeng，WU Jia，LI Qian，et al. A comprehensive survey on automatic knowledge graph construction[J]. ACM Computing Surveys，2023，56(4)：1-62.

[10] DAGDELEN J，DUNN A，LEE S，et al. Structured information extraction from scientific text with large language models[J]. Nature Communications，2024，15(1). DOI：10.1038/s41467-024-45563-x.

[11] HU Yan，CHEN Qingyu，DU Jingcheng，et al. Improving large language models for clinical named entity recognition *via* prompt engineering[J]. Journal of the American Medical Informatics Association，2024，31(9)：1812-1820.

[12] REMADI A，EL HAGE K，HOBEIKA Y，et al. To prompt or not to prompt：navigating the use of large language models for integrating and modeling heterogeneous data[J]. Data & Knowledge Engineering，2024，152. DOI：10.1016/J.DATAK.2024.102313.

[13] AGRAWAL M，HEGSELMANN S，LANG H，et al. Large language models are few-shot clinical information extractors[EB/OL]. [2024-07-25]. https://arxiv.org/abs/2205.12689v2.

[14] WADHWA S，AMIR S，WALLACE B C. Revisiting relation extraction in the era of large language models[EB/OL]. [2024-07-25]. https://doi.org/10.48550/arXiv.2305.05003.

[15] 冯钧，畅阳红，陆佳民，等. 基于大语言模型的水工程调度知识图谱的构建与应用[J]. 计算机科学与探索，2024，18(6)：1637-1647.
FENG Jun，CHANG Yanghong，LU Jiamin，et al. Construction and application of knowledge graph for water engineering scheduling based on large language model[J]. Journal of Frontiers of Computer Science and Technology，2024，18(6)：1637-1647.

（下转第 105 页）

40(11)：3172-3177.

[9] 卢万杰，付华，赵洪瑞. 基于深度学习算法的矿用巡检机器人设备识别[J]. 工程设计学报，2019，26(5)：527-533.
LU Wanjie，FU Hua，ZHAO Hongrui. Equipment recognition of mining patrol robot based on deep learning algorithm[J]. Chinese Journal of Engineering Design，2019，26(5)：527-533.

[10] 倪旺旺. 基于音频的矿井提升机异常检测方法研究及应用[D]. 淮南：安徽理工大学，2023.
NI Wangwang. Research and application of anomaly detection method for mine hoists based on audio [D]. Huainan：Anhui University of Science and Technology，2023.

[11] 曾锃，张震，缪巍巍，等. 基于卷积神经网络的放电声音故障检测[J]. 电子器件，2024，47(1)：176-181.
ZENG Zeng，ZHANG Zhen，MIAO Weiwei，et al. Fault detection of discharge sound based on convolutional neural network[J]. Chinese Journal of Electron Devices，2024，47(1)：176-181.

[12] 卢安琪. 基于集成学习与注意力机制的泵机设备异常声音检测方法研究[D]. 合肥：合肥学院，2023.
LU Anqi. Research on abnormal sound detection method for pump equipment based on ensemble learning and attention mechanism[D]. Hefei：Hefei University，2023.

[13] 翟洪婷，张庆锐，卞若晨，等. 基于图聚类的电力设备异常声音检测方法[J]. 南京理工大学学报，2022，46(3)：270-276.
ZHAI Hongting，ZHANG Qingrui，BIAN Ruochen，et al. Abnormal sound detection method of power equipment based on graph clustering[J]. Journal of Nanjing University of Science and Technology，2022，46(3)：270-276.

[14] 宁永安. 堆栈自编码器下机电故障信号多尺度滤波方法研究[J]. 自动化仪表，2024，45(8)：42-46，51.
NING Yong'an. Research on Multi-scale filtering method of electromechanical fault signals under stack self-encoder[J]. Process Automation Instrumentation，2024，45(8)：42-46，51.

[15] 郝洪涛，倪凡凡，陈亮，等. 远程带式输送机托辊故障巡检方法[J]. 煤矿机械，2018，39(11)：133-135.
HAO Hongtao，NI Fanfan，CHEN Liang，et al. Investigation of inspection method on roller of remote belt conveyor[J]. Coal Mine Machinery，2018，39(11)：133-135.

[16] 黄光球，赵梦娜，陆秋琴. 融合时域卷积网络和深度自编码器的 VOCs 数据异常检测[J]. 安全与环境学报，2023，23(10)：3749-3759.
HUANG Guangqiu，ZHAO Mengna，LU Qiuqin. VOCs data anomaly detection based on time-domain convolutional network and depth self-encoder[J]. Journal of Safety and Environment，2023，23(10)：3749-3759.

[17] 李斌，朱杰，马志贤. WebRTC 中一种基于 DNN 的噪声抑制算法的研究[J]. 信息技术，2019，43(5)：1-5.
LI Bin，ZHU Jie，MA Zhixian. Research on noise suppression algorithm based on DNN in WebRTC[J]. Information Technology，2019，43(5)：1-5.

[18] 陶瀚宇，陈换过，彭程程，等. 基于 MFCC−IMFCC 混合倒谱的托辊轴承故障诊断[J]. 机电工程，2024，41(7)：1215-1222.
TAO Hanyu，CHEN Huanguo，PENG Chengcheng，et al. Fault diagnosis of idler bearings based on MFCC-IMFCC hybrid cepstral coefficients[J]. Journal of Mechanical & Electrical Engineering，2024，41(7)：1215-1222.

[19] 史爱武，马淑然. 基于多特征融合与插值卷积自编码器的机械异常声音检测研究[J]. 软件导刊，2025，24(2)：40-47.
SHI Aiwu，MA Shuran. Research on detection of mechanical abnormal sounds based on multifeature fusion and interpolation convolutional neural auto-encoder[J]. Software Guide，2025，24(2)：40-47.

[20] TAX D M J，DUIN R P W. Support vector data description[J]. Machine Learning，2004，54：45-66.

[21] VIROLI C，MCLACHLAN G J. Deep Gaussian mixture models[J]. Statistics and Computing，2019，29(1)：43-51.

(上接第 83 页)

[16] WANG Jiaqi，SHI Enze，YU Sigang，et al. Prompt engineering for healthcare：methodologies and applications[EB/OL]. [2024-07-25]. https://arxiv.org/abs/2304.14670?context=cs.

[17] TONEVA M，SORDONI A，DES COMBES R T，et al. An empirical study of example forgetting during deep neural network learning[EB/OL]. [2024-07-25]. https://arxiv.org/abs/1812.05159.

[18] LI Lei，JIN Li，ZHANG Zequn，et al. Graph convolution over multiple latent context-aware graph structures for event detection[J]. IEEE Access，2020，8：171435-171446.

[19] ZHANG Qianjin，WANG Ronggui，YANG Juan，et al. Structural context-based knowledge graph embedding for link prediction[J]. Neurocomputing，2022，470：109-120.

[20] 张吉祥，张祥森，武长旭，等. 知识图谱构建技术综述[J]. 计算机工程，2022，48(3)：23-37.
ZHANG Jixiang，ZHANG Xiangsen，WU Changxu，et al. Survey of knowledge graph construction techniques[J]. Computer Engineering，2022，48(3)：23-37.

[21] LU Yaojie，LIU Qing，DAI Dai，et al. Unified structure generation for universal information extraction[C]. The 60th Annual Meeting of the Association for Computational Linguistics，Dublin，2022：5755-5772.