

多模态知识图谱与大模型协同的水利工程风险应对决策推荐

杨阳蕊¹, 潘世峰¹, 刘雪梅^{1,2}, 马威¹, 冯岭¹

(1. 华北水利水电大学 信息工程学院, 河南 郑州 450000;

2. 黄河流域水资源高效利用省部共建协同创新中心, 河南 郑州 450000)

摘要: 随着水利工程发展和信息化需求的增加, 有效处理并利用水利工程巡检数据做出风险应对决策变得更加重要。南水北调工程风险巡检文本涉及诸多影响因素、复杂的相互关系以及多源异构数据。为了充分利用该文本以保证对工程风险处置的及时有效性, 本研究基于巡检文本数据提出一种结合多模态知识图谱与多模态大模型 (MLLM) 的方法, 用于对南水北调工程运行中的风险做出针对性应对决策推荐。首先, 选用更加关注区域位置的开源多模态大模型 VIP-LLaVA 作为对话模型; 其次, 基于水调工程巡检数据, 构建领域多模态知识图谱; 最后, 以水利领域多模态知识图谱作为多模态大模型增强知识, 采用检索与生成协同迭代的方式寻找应对风险的最佳方案。结果表明, 由该方法生成的风险应对决策在 BLEU、ROUGE-L 和 METEOR 评价指标上效果良好, 能够及时对工程中的运行风险做出合适的应对决策推荐。相关研究可以有效提高工程安全维护管理的可靠性, 降低工程风险对供水安全的影响。

关键词: 多模态知识图谱; 多模态大模型; 风险应对; 南水北调工程

中图分类号: TP391.1; X93

文献标识码: A

doi: 10.13243/j.cnki.slxb.20240575

1 研究背景

南水北调项目在解决水资源分配不平衡、促进区域协调发展方面发挥着至关重要的作用。在工程作业日益复杂, 以及系统化程度提高的背景下, 工程运营期潜在风险因素的增加对追求高质量和可持续发展的工程发展提出了挑战。风险处理是风险管理的主要方面, 贯穿于工程稳定运营的整个生命周期。科学、全面地针对各个阶段的关键风险做出决策, 是提高工程安全运行的关键手段。南水北调工程巡检文本中包含大量的有助于工程安全运行的多源异构数据, 充分利用这些数据可为南水北调工程的安全运行做出重要贡献。传统的南水北调工程风险应对决策都是由工程维护人员人工提供。这种人工的方法存在以下局限: (1) 工程维护人员专业水平参差不齐, 在面对各个风险时, 做出的风险应对决策可能不一致; (2) 随着国内水利工程规模的不断扩大及工程复杂性的不断增加, 传统的人工决策方法已然无法满足当下高效、高质量的管理需求。

目前已有一些研究提出了基于知识图谱和深度学习^[1-2]的水利工程智能运维方法。Tang 等^[3]结合深度学习与贝叶斯算法进行设备的故障智能诊断; 刘雪梅等^[4]通过构建水利工程应急方案知识图谱, 实现了风险事件应对方案的生成与推荐。知识图谱 (Knowledge Graph, KG) 作为新的信息管理与组织技术, 是结构化的语义知识库, 以图的形式直观地展示实体 (nodes) 及其关系 (edges), 通过实体-关系-实体三元组的形式, 构成网状的知识结构, 成为了知识结构化存储与管理的有效工具^[5-7], 可对风险应对决策提供结构化、可解释性的知识。然而, 在现实世界中的信息往往是以多模态的形式存在,

收稿日期: 2024-09-08; 网络首发日期: 2025-04-23

网络首发地址: <https://link.cnki.net/urlid/11.1882.TV.20250422.1514.002>

基金项目: 国家自然科学基金项目 (72271091); 河南省高等学校重点科研项目 (25A520006); 河南省科学院科技开放合作项目 (220901008); 华北水利水电大学硕士研究生创新能力提升工程项目 (NCWUYC-202416098)

作者简介: 杨阳蕊 (1982-), 博士, 讲师, 主要从事自然语言处理研究。E-mail: yangyangrui@ncwu.edu.cn

如图像和描述性文字。上述研究进行方案推荐时仅限于单模态数据,未充分利用分布于工程巡检中的多模态数据,从而导致给出的推荐方案准确性较低。因此,充分利用多模态数据对于下游任务至关重要,多模态学习是对多源异构数据的挖掘分析^[8],通过特征提取、模态对齐与模态融合,能够有效融合不同数据模态中的互补信息,提高推荐任务的性能。从符号计算到深度学习,多模态学习的发展备受关注。当下多模态学习的典型工作有维基图像百科(IMGpedia)^[9]和多模态知识图谱(MMKG)^[10]。它们都是通过补充视觉信息知识将传统单模态文本知识扩展成图文多模态知识。然而,经实验发现,在问题超出知识图谱的知识范围后,仅依靠知识图谱作为推荐的后端数据库,仍不能达到较好的推荐效果。鉴于多模态大模型(Multimodal Large Model, MLLM)含有丰富的知识,考虑将两者融合,依此来解决知识溢出问题。现有多模态模型像 LLaVA^[11]、MiniGPT-4^[12]、Flamingo^[13]等可对多模态数据进行分析理解,但上述模型是针对整张图像,在实现特定区域的理解方面存在显著差距,而且现有使用文本坐标或空间编码的解决方法难以满足智能化的视觉提示。

为了解决传统水利工程风险应对决策推荐的局限,保证工程运营期间的稳定,本研究将多模态知识图谱与多模态大模型相结合,提出多模态知识增强大模型的水利工程风险应对决策推荐方法,依此来辅助工程运维部门人员及时解决工程风险。首先使用经过数据微调的 YOLOv8 模型对原始图像进行风险部位识别,随后鉴于 VIP-LLaVA^[14]模型允许将视觉提示标记在原始图像的任意位置处,将经过 YOLOv8 模型智能化识别后的图像传送给 VIP-LLaVA 模型。基于南水北调工程巡检文本并结合专家经验定义领域多模态知识图谱概念模型^[15],并在此基础上采用基于 T5 的水利工程抢险实体和关系联合抽取方法,从非结构化的巡检文本中抽取出结构化三元组知识,以 Neo4j 图数据库^[16]为载体进行水利知识存储,清晰直观呈现出巡检文本中风险知识。在工程风险应对决策推荐方法中,为使得到的推荐方案更加贴合实际,本研究将多模态知识图谱作为增强知识,采用知识检索与多模态模型生成协同方式^[17-18]迭代地寻找应对风险的最佳决策方案。将多模态知识图谱和多模态大模型技术引入到工程巡检风险应对智能化应用中,实现了巡检知识的深度关联和有效利用,对辅助工程运维部门人员,提高南水北调工程运维效率、强化工程风险管理能力具有重要作用。

2 研究方法

为使南水北调工程巡检文本在水利工程风险应对决策推荐方法中发挥重要作用,本研究设置了针对性的方法:首先,选用可着重关注局部特征的开源多模态预训练模型 VIP-LLaVA,让其对风险局部区域有侧重的理解与分析;其次,基于南水北调工程巡检文本构建领域多模态巡检知识图谱,将其作为模型推荐方案的增强知识^[19];最后,结合多模态大模型与水利领域多模态知识图谱上的领域知识^[20],采用检索与生成协同方式迭代式地寻找应对水利工程风险的强相关方案,从而辅助工程运维部门人员进行风险修复。方法的总体架构如图 1 所示。

2.1 VIP-LLaVA 模型 南水北调工程风险巡检图像较为复杂,如果对图像的全局部分进行考察分析,往往存在许多噪声,如果模型能关注图像中的局部风险部位,并针对性地分析风险特征,就能够减少噪声干扰。因此,为使模型着重关注发生风险的位置,并做出有效的应对决策,本研究选取可着重关注局部特征的开源多模态预训练模型 VIP-LLaVA 模型,该模型以 LLaMA^[21]模型为基础,允许在图像的风险位置上针对性标注,从而让模型关注标注提示的位置。例如,南阳管理处,由于持续的降雨,导致闸室的电缆沟里有大量积水,应当采用什么策略解决这个问题。如果没有对积水的电缆沟部分进行标注,模型的回复就是针对整张图像进行分析,可能达不到预期结果,但是,如果事先通过 YOLOv8 模型对图像上积水的电缆沟部分进行智能化标注,然后传送给 VIP-LLaVA 模型,让模型充分分析电缆沟局部积水的部位,着重分析其结构,就能达到符合意图的结果。VIP-LLaVA 架构如图 2 所示。

该架构图中,将输入视觉提示标记到原始图像后,得到图像 X_v ,接着使用预训练的 CLIP^[22]视觉编码器 ViT-L/14 作为视觉特征提取器,提供到图像视觉特征;这些特征被串接并经由归一化层与多层感知机形成视觉标记 H_v ;而文本指令则经由向量嵌入层形成文本标记 H_q 。在进行水利工程风险应

对推荐时，将经局部风险标记后的视觉令牌和文本指令令牌输入到大模型中。让其结合多模态知识图谱，通过知识检索与模型生成协同迭代的方式，让多模态大模型针对风险做出应对决策推荐。所以，为了得到最合适的决策方案，水利多模态知识图谱的构建对于精控大模型尤为重要。

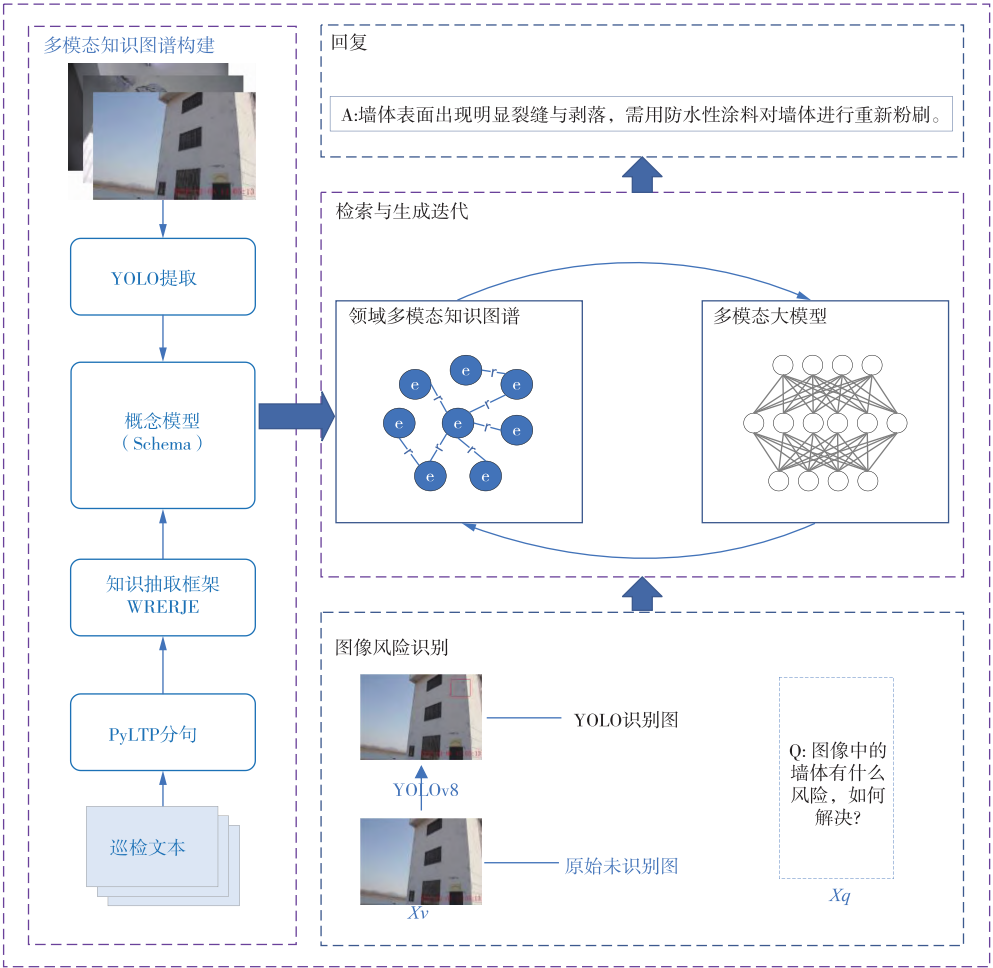


图 1 方法总体框架

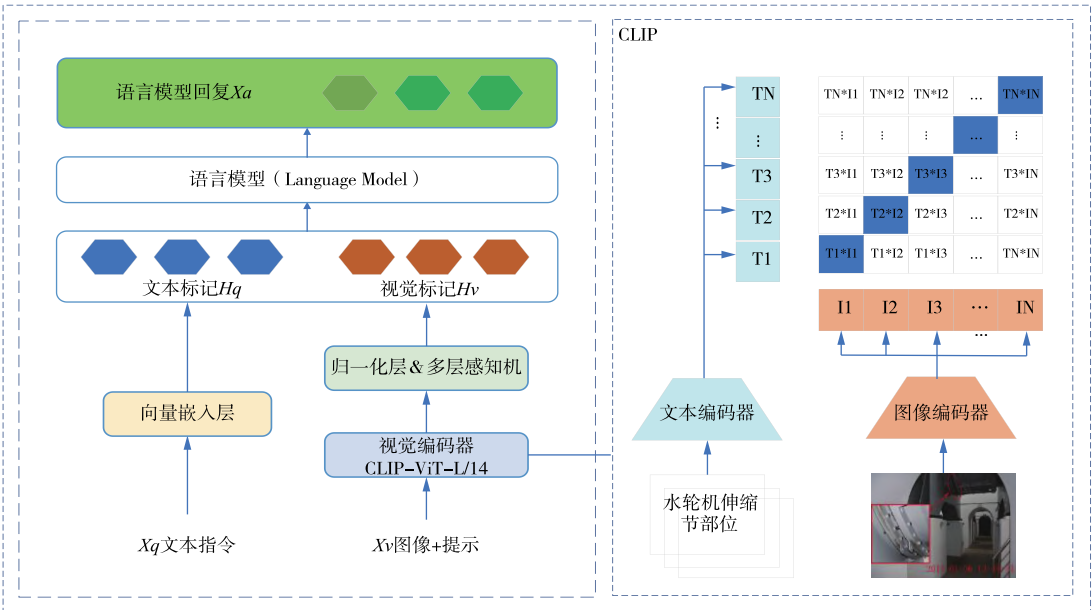


图 2 VIP-LLaVA 架构

2.2 多模态知识图谱构建 水利工程巡检多模态知识图谱是一个以南水北调工程运行风险数据为节点^[23]，以它们之间的各种语义关系为边的结构化语义网络，其建模流程主要包括数据预处理、本体构建、知识抽取、多模态知识融合、知识图谱三元组存储 5 个步骤^[24]。建模结构如图 3 所示。

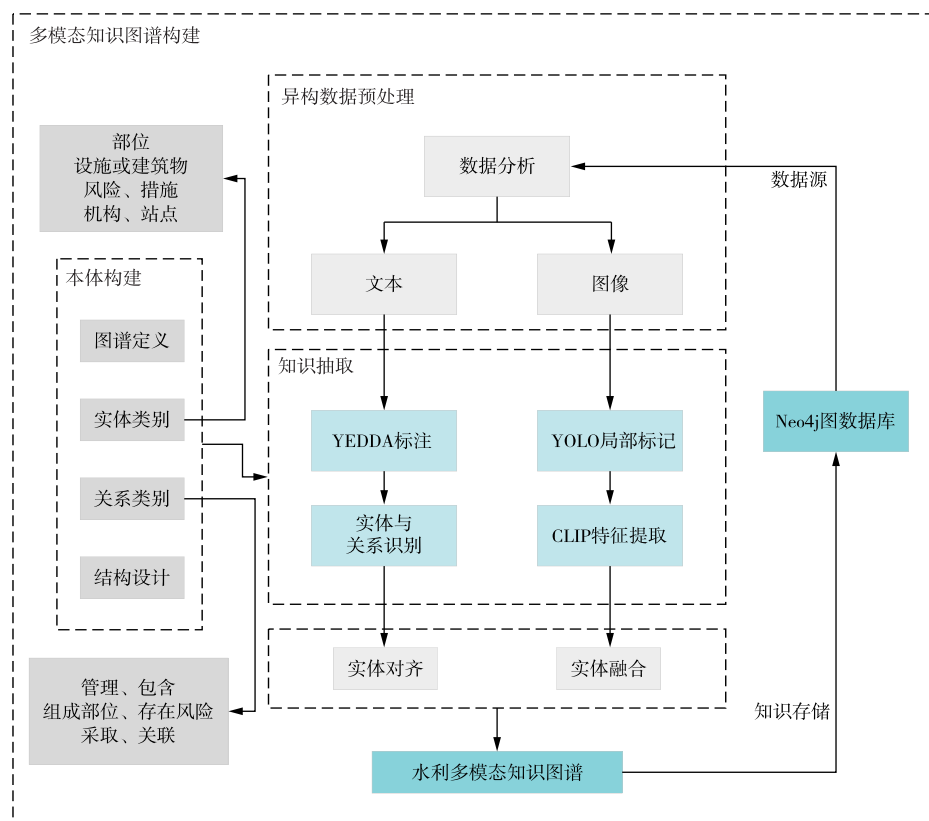


图 3 建模结构图

2.2.1 本体构建 为了规范化水利领域多模态知识图谱中的多模态实体及关系类别，多模态知识图谱的本体结构定义至关重要。从风险的应对决策角度来看，明确具体的风险事件和风险应对决策是有效应对风险、合理做出对应决策以及保证水调工程稳定运行的重要阶段；同时，风险事件图的鲜明特征还可以帮助理解各类风险事件，以保证后续多模态知识图谱与多模态大模型协同迭代做出正确应对风险的判断；风险发生地点的确定利于工程维护资源的有效配置。因此，本研究构建的多模态风险知识图谱本体结构主要包含以下几个方面的概念关系^[15]：一是包含特定风险信息的具体部位风险事件及其鲜明图示；二是具体工程运行风险的应对决策；三是风险事件图的鲜明局部特征；四是风险发生的地点。水利领域多模态知识图谱(MMKG-WC)可表示如下：

$$MMKG-WC = \{E, R, TW, TP\} \quad (1)$$

式中： E 为所有实体集； R 为所有关系集； TW 为文本类水利知识三元组集合； TP 为图像类水利知识三元组集合。每一个三元组可以表示为 $T = \{(s, p, o) \mid T_i \in (TW \cup TP), i = 1, 2, \dots, n\}$ ， s 、 p 、 o 分别代表主语、谓语、宾语。

观察南水北调巡检文本的语义描述，定义风险机构、风险站点、风险设施、风险部位、风险事件、风险事件图和风险应对措施 7 种实体类型，并定义出实体与实体之间存在的 6 种关系类型。

“管理”，即某一工程管理处管理某个站点，则将机构类型与站点类型之间的关系定义为“管理”。

“包含”，即某一工程站点下包含某种设施或建筑物，则将站点类型与设施或建筑物类型之间的关系定义为“包含”。

“组成部位”，即某种组成设施或建筑物的某个部位，则将设施或建筑物类型与部位类型之间的关系定义为“组成部位”。

“存在风险”，即某个部位发现了风险，则将部位类型与风险类型之间的关系定义为“存在风险”。

“采取”，即解决风险应当采取的措施，则将风险类型与措施类型之间的关系定义为“采取”。

“关联(HI)”，即某个风险部位相关联的风险图示，则将部位类型与风险事件图类型之间的关系定义为“关联(HI)”。




各关系类型实例见表 1。

表 1 各关系类型实例

语义关系	语句描述
管理	扬州分公司，洪泽站
包含	蔺家坝泵站，2#水轮机
组成部位	水轮机伸缩节
存在风险	伸缩节下部渗水
采取	伸缩节下部渗水，应当及时更换密封条
关联(HI)	水轮机伸缩节——伸缩节图像(ID 或 URL)

2.2.2 异构数据预处理 由人工记录的南水北调巡检数据中往往存在着影响文本分类的噪声信息，如无关词、重复性词等。因此，对巡检文本进行预处理是构建结构化多模态知识图谱的关键环节，数据预处理影响着多模态知识图谱的质量。表 2 中提供了南水北调工程巡检数据的部分实例。对于文本类数据，通常需要进行文本清洗和去停用词操作，以便于减少文本中的噪声，增强语言模型的整体性能；图像类数据，执行随机旋转、图像缩放等数据增强和归一化处理，对于获取图像的特征至关重要。

表 2 巡检数据实例

机构	站点	风险描述	风险图像
扬州分公司	洪泽站	大沙河闸，启闭机室闸室外墙多处粉刷层起皮、脱落，工程维护人员应当及时用防水涂料进行粉刷。	
徐州分公司	蔺家坝泵站	2#水轮机伸缩节下部渗水，应当检查密封条是否受到氧化，是否老化或破损，如果出现上述问题，应当及时更换。	
保定管理处	易县站	渠道倒虹吸进口上游左岸渐变段渠段处，边坡喷护混凝土破损，坡面土体外露，维护人员应当用混凝土填充渠道边坡破损部位。	

2.2.3 知识抽取 为进一步从多模态信息中抽取相应的知识以构建水利多模态知识图谱，基于前述本体定义，针对文本、图像模态信息分别设计知识抽取方法。

(1) 文本模态知识抽取。在数据进行预处理之后，文本模态知识抽取之前，需要对数据进行标签标注，以确定各个巡检记录文本的具体实体类型。本研究将实体类、关系类与输入语句分别定义为 [node]、[relation]、[text]，以上三类组成部分分别根据定义的实体与关系类型进行标注。其中由 [node] 标注的“风险机构、风险站点、风险设施、风险部位、风险事件、风险事件图和风险应对措施”指实体类型，[relation] 标注的“管理、包含、组成部位、存在风险、采取和关联(HI)”指关系类型，[text] 则表示风险相关文本，基于巡检文本进行实体与关系标注^[25]，用于模型的训练。标注示

例见表 3。

表 3 巡检文本标注示例

名称	标注样式
巡检文本	[text]扬州分公司洪泽站，启闭机室外墙粉刷层脱落，应及时用防水涂料重新粉刷。
扬州分公司	[node]机构
洪泽站	[node]站点
启闭机室	[node]设施或建筑物
外墙	[node]部位
粉刷层脱落	[node]风险事件
及时用防水涂料重新粉刷	[node]风险应对措施

标注数据的质量对巡检知识三元组的确立影响巨大，进而，三元组形成的质量又深层次地影响着多模态知识图谱的构建质量，也会影响后续方法实现中知识的检索。受最近提示工程^[26-27]工作的启发，本研究采取基于 T5 的实体关系联合抽取框架-WRERJE^[28]，使用上表标注的数据对该框架进行微调，使框架向工程巡检文本风险实体与关系抽取方面逐渐靠拢，从而使该框架识别出巡检文本中存在的实体和关系，并转化为三元组形式。联合抽取框架如图 4 所示。WRERJE 首先将输入语句 X 处理为固定提示序列 P ，该过程表示为

$$[p_1, p_2, \dots, p_n] = \text{PG}[x_1, x_2, \dots, x_x] \tag{2}$$

式中：PG 表示基于文本分类模型实现的提示生成器； $X=[x_1, x_2, \dots, x_x]$ 表示输入语句； $[p_1, p_2, \dots, p_n]$ 表示所生成的提示序列 P ；其次由 T5 从该提示序列中同时提取出巡检实体和关系，并转为结构化三元组知识(structured triplet knowledge, S_{ik})序列，该生成过程可表示为

$$[y_1, y_2, \dots, y_y] = \text{ERJE}[p_1, p_2, \dots, p_n] \tag{3}$$

式中：ERJE 表示基于 T5 的实体关系联合抽取器； $[y_1, y_2, \dots, y_y]$ 表示 S_{ik} 序列。

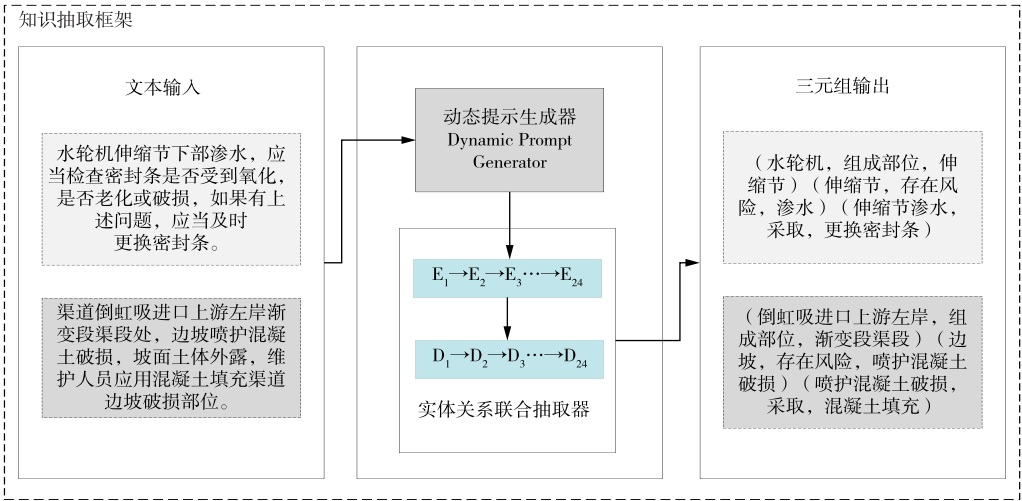


图 4 联合抽取框架

(2)局部风险图像提取。南水北调巡检文本数据中存在大量风险图像数据，能更直观、更形象地可视化巡检文本所描述的有关风险；同时，图像特征对后续多模态知识图谱与多模态语言模型协同迭代地寻找应对风险的强相关方案时具有重要的作用。作为多模态知识图谱的重要组成部分，对于图像模态，本研究借助 CV 处理技术以及人工的方式进行提取，其流程如图 5 所示。

本研究使用经过特征突出数据微调的 YOLOv8^[29]模型作为主干网络，作为南水北调巡检图像的局部风险部位提取器。使用开源工具 LabelImg 对巡检数据中风险图像数据(例如：渗水、开裂、塌陷、

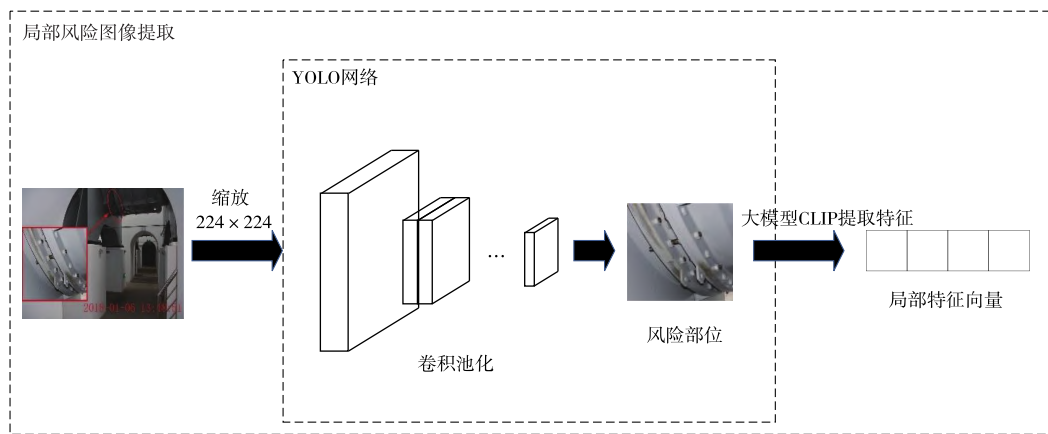


图5 局部风险图像提取框架图

监控损坏、墙皮损坏等)进行人工标记,用于YOLOv8模型的微调,微调后的YOLOv8模型可以将图像当中含有的上述风险进行识别出来,提取到风险局部部位图;最终,风险图像数据都关联于对应的风险实体,并以URL或ID的方式储存在多模态知识图谱中。紧接着在后续检索生成迭代方法中生成风险应对决策的方案时,该局部图像作为检索的重要知识,可经由大模型视觉处理器CLIP进行特征向量提取,用于风险图像特征相似性对比。总之,风险局部部位图的提取在本研究生成最合适决策^[30]方案中尤为重要。

2.2.4 多模态知识融合 通过上述抽取方法从南水北调巡检文本中抽取到的知识可能存在大量同义、重复表述的实体数据,所以需要对上述抽取到的知识进行融合与对齐^[31],融合目的就是将所有相近实体有效融合统一,提高整个多模态知识图谱的质量。本研究知识融合采用Bert^[32]孪生网络框架,通过余弦相似度公式计算相似度来进行实体对齐,公式如下所示。相似度结果大于某一阈值则认为两个表达同属一个实体。

$$\cos(\theta) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|} = \frac{\sum_{i=1}^n u_i v_i}{\sqrt{\sum_{i=1}^n (u_i)^2} \sqrt{\sum_{i=1}^n (v_i)^2}} \quad (4)$$

式中: \mathbf{u} 和 \mathbf{v} 为两个待比较实体向量; $\|\cdot\|$ 为向量取模运算; n 为向量的维度。

2.2.5 知识图谱三元组存储 本研究利用Neo4j图数据库存储知识图谱三元组。文本类实体节点就采用本身内容进行存储,对于图像类数据,采用唯一标识符(如ID或URL)作为实体节点进行存储。多模态风险知识图谱采用三元组知识表示,即“实体-关系-实体”。

多模态知识图谱作为外部知识检索增强的一种方式,其中所包含的文本及图像知识可对多模态大模型的决策生成形成控制的作用,可以让其针对水利工程运行中某一具体风险做出正确的判断,从而让推荐方案更贴合实际、更具解释性。接下来着重介绍如何结合多模态知识图谱与多模态大模型。

2.3 多模态知识图谱与大模型协同的风险决策推荐 MLLM能从大规模语料库当中学习通用知识,并能在多种多模态处理任务中发挥良好的性能。然而,当MLLM运用在特定领域生成任务中,生成的文本常具有“幻觉问题”^[33-34],即为生成的方案看似合理但事实上是错误的内容;另外,尽管MLLM学习到了大量的知识,但学到的知识是基于统计的模式,并不能真正理解那些知识的真实含义,所以在将MLLM直接作为推荐器处理特定领域生成任务时,由于缺乏这方面知识的统计,导致缺乏深度理解,生成的方案也通常不切实际,这表明领域知识对于推荐任务的重要性^[35-37];更重要的,MLLM生成的方案通常可解释性较低,从而限制了其在关键任务中的可用性。针对上述问题,又鉴于KG以明确且结构化的方式存储大量特定领域方面的知识,可用于提升MLLM用于生成任务的性能,所以,提出用领域多模态知识图谱中的领域知识去增强MLLM在领域生成任务方面的性能表现^[38]。知识增强框架如图6所示。

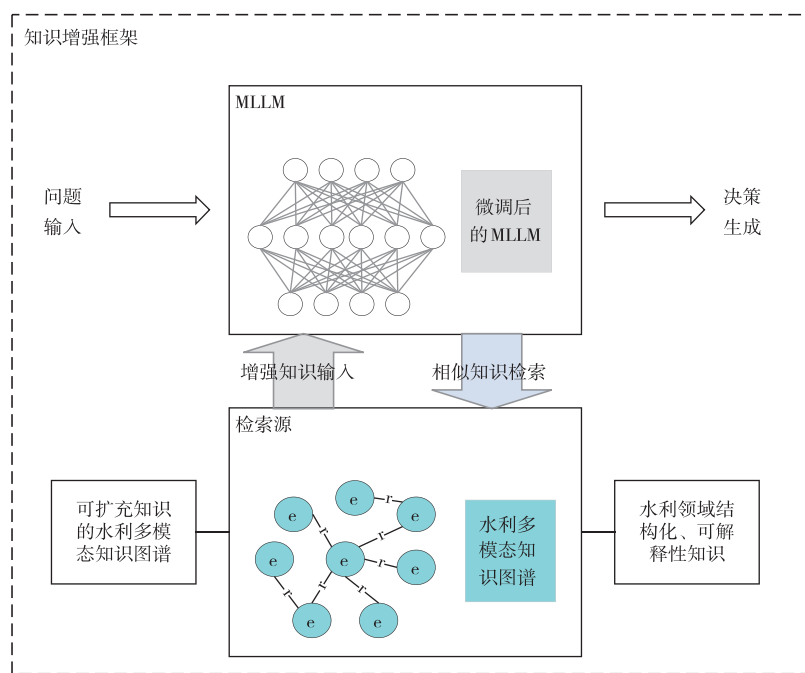


图6 知识增强框架图

在实验过程中，发现一次知识的检索对 MLLM 生成合适方案的有效性较低，针对此缺陷又提出检索与生成迭代协同的方式去寻找最合适的方案^[39-41]。该方法融合大模型的强大自然语言生成能力与知识图谱的检索与推理能力。具体地，当工程运维部门人员将工程现场拍摄图像与文本指令一同输入模型后，首先抓取得到文本中的关键工程部位，用 Cypher 语句针对该部位从知识图谱中查询到相似部位，并将知识图谱中与部位风险相关联的应对措施返还给模型，让模型进行相应推荐方案生成；然后，针对返回的多个相似部位，再次用 Cypher 语句查询知识图谱中与风险部位实体相关联的风险事件图实体，再由多模态大模型视觉处理器最终判断并选择描述当前部位的最佳相关知识。例如，要解决的问题是图像中的水轮机处发生了什么风险，应当采取什么措施解决。首先，针对这个问题，采用文本向量相似度距离方法在水利多模态知识图谱当中寻找与问题中水轮机最相关的 3 个实体，然后将其列出，MLLM 针对检索到的知识做第一轮回复；进而，多模态模型中视觉处理器 CLIP 再根据上述回复，将刚刚列出的实体的相关风险部位图像与当前模型中上传的经 YOLOv8 智能化识别风险位置的图像进行对比，做相应局部风险特征相似性判断，从而确认最相似于当前模型中上传图像局部部位特征的水轮机伸缩节这个实体；最后，MLLM 充分分析上传图像的特定标记位置，并结合前面两轮迭代确定的增强知识^[42]，做出风险应对决策。本研究迭代检索-生成框架的实例如图 7 所示。

3 实验与结果

3.1 文本预处理 本研究主要以南水北调工程巡检知识文本作为主要案例源。在获得相关文本后，对文本进行段落筛选，过滤掉与风险决策知识无关的描述，然后利用自然语言处理工具 PyLTP 对筛选后的段落以句号为分隔符进行分句处理，最终获得 2864 条文本图像对，将 2864 条文本输入到实体与关系抽取器中，获得 5106 个实体对，然后按照 8:2 的比例，利用 Neo4j 将抽取三元组的 80% 以及风险部位对应的图像存放在水利多模态知识图谱当中，另外的 20% 做本研究方法的实验测试数据集，以此证明本研究方法的有效性。

3.2 实验设置 本研究以南水北调工程运营风险应对决策生成任务为应用测试场景，验证提出方法的有效性。主要针对方法的两个模块的有效性以及第一轮检索中列出相关实体的个数设计了以下验证实验，从以下实验设置中得出模块设计的有效性以及多方面考究中需要列出相关实体的最合适数量。

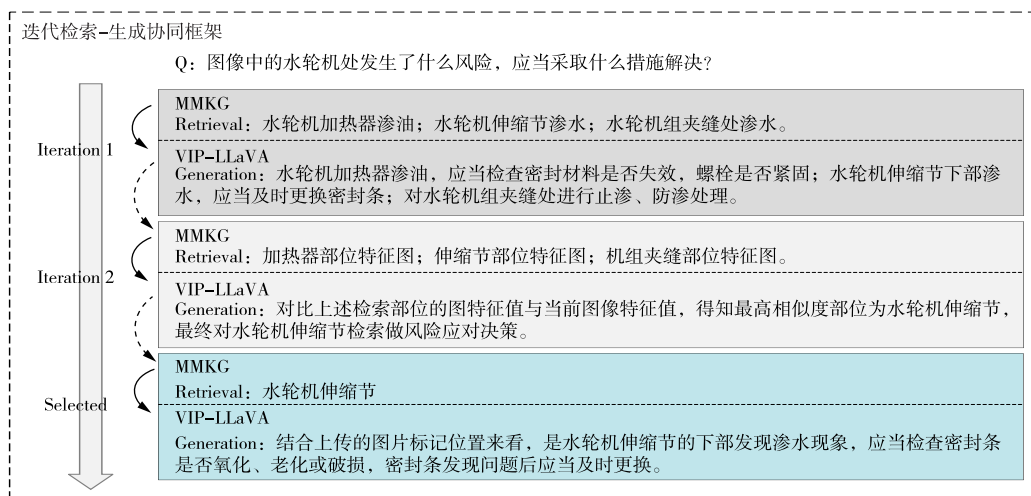


图7 迭代检索-生成框架

3.2.1 子模块领域多模态知识图谱增强的有效性 本研究的一个模块是使用工程巡检多模态知识图谱作为增强知识, 让模型在回答前先进行知识的查询, 以便于针对特定的风险问题进行回复。为了验证这一方法的有效性, 实验设置不采用知识增强, 仅仅使用模型自身的方法 VIP-LLaVA; 以及使用了工程巡检多模态知识图谱作为增强知识的方法 VIP-LLaVA_{KG} (本 VIP-LLaVA_{KG} 使用了两轮次检索生成迭代, 即 VIP-LLaVA_{KG-2})。利用风险图像文本对测试集进行该模块有效性的检验。

3.2.2 子模块检索生成迭代轮次的有效性验证 本研究方法中另一个模块为检索生成协同迭代的方式寻找最合适的方案, 对检索生成协同迭代的轮次次数进行有效性检验。为了验证这一方法的有效性, 实验设置了采用一个轮次的文本知识增强方法 VIP-LLaVA_{KG-1}, 以及两个轮次的文本加图像知识增强方法 VIP-LLaVA_{KG-2}。同样使用风险图像文本对测试集进行该模块的有效性检验。

3.2.3 第一轮次相关实体最合适数量探究 在进行第一轮次的水利知识检索时, 列出相关实体的数量对模型推荐方案的准确性及响应速度有着重要的影响。本研究分别针对这两个方面进行相关实验探究, 分别设置列出实体数量的个数为 2、3、4、5、6。针对不同的数量做准确性与响应速度的分析, 从而找出一个能够平衡准确性与响应速度的数值。此实验的方案准确性仅在指标 BLEU 上进行分析。

3.3 实验结果 关于评价指标, 本研究使用 BLEU、ROUGE-L 和 METEOR 三种^[43], 其中, BLEU@N 表示生成方案占参考方案指定字符串长度的比例。计算公式如下。

$$BLEU@N = BP \cdot \exp \sum_{n=1}^N (w_n \cdot \ln p_n) \quad (5)$$

式中: BP (Brevity Penalty-简短惩罚) 依据生成文本序号进行设置, 用于惩罚生成的句子过短; p_n 为生成文本中第 n 个匹配词的数量与生成文本中连续 n 个词数量的比值, n 由 N 界限; w_n 为每个 $\ln p_n$ 的权重; N 为指定字符串长度。

ROUGE-L 基于最长公共子序列 (LCS), 比较生成方案和参考方案之间的重叠词汇来评估它们的相似性。计算公式如下。

$$ROUGE-L = F_{\beta} = \frac{(1 + \beta^2) \cdot P \cdot R}{\beta^2 \cdot P + R} \quad (6)$$

式中: P 为生成文本和参考文本的 LCS 长度与生成文本长度的比值; R 为生成文本和参考文本的 LCS 长度与参考文本长度的比值; β 为一个可调节的参数, 用来控制 P 与 R 的相对权重。

METEOR 在 BLEU 基础上考虑了同义词、词干、词序等。计算公式如下。

$$METEOR = \frac{10 \cdot P \cdot R}{R + 9 \cdot P} \cdot \left[1 - \left(0.5 \cdot \left(\frac{chunk}{matches} \right)^3 \right) \right] \quad (7)$$

式中: P 为生成文本和参考文本匹配词的数量与生成文本总词数的比值; R 为生成文本和参考文本匹

配词的数量与参考文本总词数的比值；*chunk* 为生成文本和参考文本匹配词的数量；*matches* 为匹配词的数量。

对实验设置小节设置的模块进行验证，模块实验结果见表 4。

表 4 模块实验结果

模型	<i>BLEU@3</i>	<i>ROUGE-L</i>	<i>METEOR</i>
VIP-LLaVA	0.326	0.098	0.232
VIP-LLaVA _{KG-1}	0.503	0.135	0.402
VIP-LLaVA _{KG-2}	0.566	0.169	0.449

由表 4 所示，使用了水利多模态知识图谱增强的 VIP-LLaVA_{KG-2} 方法在三种指标上都要高于未使用知识增强的 VIP-LLaVA 方法，从而验证了实验设置小节中领域多模态知识图谱作为知识增强的有效性；同时，使用了文本加图像知识两个轮次检索生成迭代的 VIP-LLaVA_{KG-2} 方法在三种指标上都要略高于仅使用了文本知识一个轮次检索生成迭代的 VIP-LLaVA_{KG-1} 方法，同样验证了实验设置小节中检索生成迭代轮次的有效性。

由表 5 所示，如果将实体数量设置为 4，较之数量 2、3 在评价指标 *BLEU@3* 上有着明显的提升，较之数量 5、6 在评价指标 *BLEU@3* 上准确性基本没有降低；同时，在响应速度上，实体数量设置为 4 的响应时间适中。最终，综合平衡准确性与响应时间，将实体数量设置为 4 最为合适。

表 5 实体数量分析

数量	<i>BLEU@3</i>	响应时间/s
2	0.412	2.2
3	0.453	3.9
4	0.566	4.7
5	0.569	6.8
6	0.572	7.3

4 结论

南水北调工程运行中的风险知识通常分散于多源异构的巡检文档中，为充分利用其中包含的自然文本和图像两种模态数据，文中以多模态知识图谱集合各种模态数据，以多模态大模型为基础，做南水北调工程风险应对决策推荐，进行巡检知识复用，主要贡献有以下 3 点：（1）首次选用可着重关注局部特征的开源多模态预训练模型 VIP-LLaVA，侧重于工程局部风险区域进行有效理解与分析；（2）基于南水北调工程巡检数据，构建领域多模态知识图谱，对水利领域知识进行明确且格式化管理；（3）通过多模态知识图谱与大模型检索生成协同迭代的方式寻找应对风险的最佳方案。经实验验证了本文方法的有效性，未来将探索更多处理多模态数据的技术，以更好地利用各种模态的数据源，并且将深入研究更新和维护知识图谱的策略，以及其他相关技术和方法，以提高南水北调工程巡检文本的利用率，更好地保证南水北调工程稳定运行，为水利工程行业的发展做出更大的贡献。

参 考 文 献：

[1] 欧阳文字, 叶磊, 顾学志, 等. 深度学习水文预报研究进展综述 II——研究进展及展望[J]. 南水北调与水利科技(中英文), 2022, 20(5): 862-875.

[2] 褚钰, 付景保, 朱九龙. 基于图模型的南水北调中线水源地生态环境多主体治理冲突分析[J]. 南水北调与水利科技(中英文), 2023, 20(6): 1179-1187.

[3] TANG S, ZHU Y, YUAN S. Intelligent fault diagnosis of hydraulic piston pump based on deep learning and Bayesian optimization[J]. ISA transactions, 2022, 129: 555-563.

[4] 刘雪梅, 卢汉康, 李海瑞, 等. 知识驱动的水利工程应急方案智能生成方法—以南水北调中线工程为例[J]. 水利学报, 2023, 54(6): 666-676.

[5] 马恒志, 钱育蓉, 冷洪勇, 等. 知识图谱嵌入研究进展综述[J]. 计算机工程, 2025, 51(2): 18-34.

[6] 张天成, 田雪, 孙相会, 等. 知识图谱嵌入技术研究综述[J]. 软件学报, 2023, 34(1): 277-311.

- [7] 徐有为, 张宏军, 程恺, 等. 知识图谱嵌入研究综述[J]. 计算机工程与应用, 2022, 58(9): 30-50.
- [8] 陈鹏, 李擎, 张德政, 等. 多模态学习方法综述[J]. 工程科学学报, 2020, 42(5): 557-569.
- [9] FERRADA S, BUSTOS B, HOGAN A. IMGpedia: A linked dataset with content-based analysis of wikimedia images[C]//Proc of the 16th International Semantic Web Conference. Berlin, Germany: Springer. 2017.
- [10] LIU Y, LI H, GARCIA-DURAN A, et al. MMKG: Multi-modal knowledge graphs[C]//Proc of the 16th International Semantic Web Conference. Berlin, Germany: Springer. 2019.
- [11] LIU H, LI C, WU Q, et al. Visual instruction tuning[C]//Advances in Neural Information Processing Systems, 2024.
- [12] LI Y, ZHANG Y, WANG C, et al. Mini-gemini: Mining the potential of multi-modality vision language models [EB/OL]. (2024-03-27)[2024-08-25]. <https://arxiv.org/abs/2403.18814>.
- [13] ALAYRAC J B, DONAHUE J, LUC P, et al. Flamingo: a visual language model for few-shot learning[J]. Advances in Neural Information Processing Systems, 2022, 35: 23716-23736.
- [14] CAI M, LIU H, MUSTIKOVELA S K, et al. ViP-LLaVA: Making large multimodal model-s understand arbitrary visual prompts[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.
- [15] 胡志强, 刘鸣飞, 李琦, 等. 基于多源异构数据的风机多模态装配工艺知识图谱建模[J]. 上海交通大学学报, 2024, 58(8): 1249-1263.
- [16] 史政一, 吕君可, 黄弘. 基于 Neo4j 的城市地下管道信息知识图谱构建研究[J]. 中国安全生产科学技术, 2024, 20(6): 5-10.
- [17] CAFFAGNI D, COCCHI F, MORATELLI N, et al. Wiki-LLaVA: Hierarchical retrieval-augmented generation for multimodal LLMs[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.
- [18] FENG Z, FENG X, ZHAO D, et al. Retrieval-generation synergy augmented large language-e models [C]//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. 2024.
- [19] HUANG W, ZHOU G, LAPATA M, et al. Prompting large language models with knowledge graphs for question answering involving long-tail facts[EB/OL]. (2024-05-10)[2024-07-28]. <https://arxiv.org/abs/2405.06524>.
- [20] 冯钧, 杭婷婷, 陈菊, 等. 领域知识图谱研究进展及其在水利领域的应用[J]. 河海大学学报(自然科学版), 2021, 49(1): 26-34.
- [21] TOUVRON H, LAVRIL T, IZACARD G, et al. Llama: Open and efficient foundation language-e models[EB/OL]. (2023-02-27)[2024-08-25]. <https://arxiv.org/abs/2302.13971>.
- [22] RADFORD A, KIM J W, HALLACY C, et al. Learning transferable visual models from natural language supervision[C]//International Conference on Machine Learning. PMLR. 2021.
- [23] 林海香, 胡娜娜, 何乔, 等. 基于建筑信息模型数据驱动的铁路设备运维多模态知识图谱构建[J]. 同济大学学报(自然科学版), 2024, 52(2): 166-173.
- [24] WANG L, LIU X, LIU Y, et al. Multimodal knowledge graph construction for risk identification in water diversion projects[J]. Journal of Hydrology, 2024, 635: 131155.
- [25] 刘雪梅, 程彭圣男, 李海瑞, 等. 基于字词向量的 BiLSTM-CRF 水利工程巡检文本实体识别模型[J]. 华北水利水电大学学报(自然科学版), 2024, 45(3): 9-17.
- [26] 王培冰, 张宁, 张春. 基于 Prompt 的两阶段澄清问题生成方法[J]. 计算机应用研究, 2024, 41(2): 421-425.
- [27] 杨阳蕊, 朱亚萍, 陈思思, 等. 融合群体智能策略的 AI 链在大坝防汛抢险知识推理中的应用[J]. 水利学报, 2023, 54(9): 1122-1132.
- [28] 杨阳蕊, 朱亚萍, 刘雪梅, 等. 水利工程文本中抢险实体和关系的智能分析与提取[J]. 水利学报, 2023, 54(7): 818-828.
- [29] JOCHER G, CHAURASIA A, QIU J. Ultralytics YOLO (Version 8.0.0)[EB/OL]. [2024-11-20]. <https://github.com/ultralytics/ultralytics>.
- [30] 胡志强, 潘鑫瑜, 文思捷, 等. 结合多模态知识图谱与大语言模型的风机装配工艺问答系统[J]. 机械设计, 2023, 40(S2): 20-26.
- [31] 陈囿任, 李勇, 温明, 等. 多模态知识图谱融合技术研究综述[J]. 计算机工程与应用, 2024, 60(13): 36-50.
- [32] 杨阳蕊, 朱亚萍, 刘雪梅, 等. 知识驱动的南水北调工程巡检信息推荐方法[J]. 南水北调与水利科技(中英文), 2024, 22(2): 368-377.

- [33] LIU J, LIU C, ZHOU P, et al. Is chatgpt a good recommender? a preliminary study[EB/OL]. (2023-04-20) [2024-10-16]. <https://arxiv.org/abs/2304.10149>.
- [34] JI Z, LEE N, FRIESKE R, et al. Survey of hallucination in natural language generation[J]. ACM Computing Surveys, 2023, 55(12): 1-38.
- [35] DAI S, SHAO N, ZHAO H, et al. Uncovering chatgpt's capabilities in recommender systems[C]//Proceedings of the 17th ACM Conference on Recommender Systems. 2023.
- [36] KANG W C, NI J, MEHTA N, et al. Do llms understand user preferences? evaluating llms on user rating prediction[EB/OL]. (2023-05-10) [2024-10-15]. <https://arxiv.org/abs/2305.06474>.
- [37] LIN J, DAI X, XI Y, et al. How can recommender systems benefit from large language models: A survey[EB/OL]. (2023-06-09) [2024-10-15]. <https://arxiv.org/abs/2306.05817>.
- [38] SANMARTIN D. KG-RAG: Bridging the Gap Between Knowledge and Creativity[EB/OL]. (2024-05-20) [2024-08-16]. <https://arxiv.org/abs/2405.12035>.
- [39] SHAO Z, GONG Y, SHEN Y, et al. Enhancing retrieval-augmented large language models with iterative retrieval-generation synergy[EB/OL]. (2023-05-24) [2024-08-16]. <https://arxiv.org/abs/2305.15294>.
- [40] YAO S, YU D, ZHAO J, et al. Tree of thoughts: Deliberate problem solving with large language models[C]//Advances in Neural Information Processing Systems, 2024.
- [41] ZHANG Y, WANG X, LIANG J, et al. Chain-of-Knowledge: Integrating knowledge reasoning into large language models by learning from knowledge graphs[EB/OL]. (2024-06-30) [2024-08-13]. <https://arxiv.org/abs/2407.00653>.
- [42] 张殿元, 余传明. 基于知识增强与多模态融合的医疗视觉问答模型[J]. 数据分析与知识发现, 2024, 8(S1): 226-239.
- [43] EVTIKHIEV M, BOGOMOLOV E, SOKOLOV Y, et al. Out of the bleu: how should we assess quality of the code generation models? [J]. Journal of Systems and Software, 2023, 203: 111741.

Multimodal knowledge graph collaborated with large model for decision recommendation of water projects risk response

YANG Yangrui¹, PAN Shifeng¹, LIU Xuemei^{1,2}, MA Wei¹, FENG Ling¹

(1. School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450000, China;

2. Collaborative Innovation Centre for Efficient Utilization of Water Resources, Zhengzhou 450000, China)

Abstract: With the development of water projects and the increasing demand for informatization, it becomes increasingly crucial to effectively process and use inspection data from water project to make risk response decisions. The risk inspection texts of the South-to-North Water Diversion Project involve numerous influencing factors, complex interrelationships, and multi-source heterogeneous data. To effectively utilize the inspection text and ensure the timeliness and effectiveness of the engineering risk response, a method that integrates multimodal knowledge graph with multimodal large model (MLLM) based on inspection text data is proposed, aiming to provide targeted decision-making recommendations for risks in the operation of the South-to-North Water Diversion Project. Firstly, the open-source multimodal large language model VIP-LLaVA, which focuses more on regional locations, is chosen as the dialog model. Secondly, a domain-specific multimodal knowledge graph is constructed based on water project inspection data. Lastly, the multimodal knowledge graph (MMKG) of the water conservancy domain is used as the augmented knowledge of the multimodal large model, and the retrieval and generation of collaborative and iterative approach is adopted to identify the most pertinent solutions and to formulate risk response decisions. The results indicate that the method works well on BLEU, ROUGE-L and METEOR evaluation indexes when generating risk response decisions, and can formulate corresponding response decision recommendations to the operational risks in the project in time. The related research can effectively improve the reliability of engineering safety maintenance management and reduce the impact of engineering risks on water supply safety.

Keywords: multimodal knowledge graph; multimodal large model; risk response; South-to-North Water Diversion Project

(责任编辑: 耿庆斋)