

李姝颖, 周申蓓, 徐琪. 基于人工智能技术的抽水蓄能电站预选址研究[J]. 水利水电技术(中英文), 2025, 56(6): 199-213. DOI: 10.13928/j.cnki.wrahe.2025.06.017

LI Shuying, ZHOU Shenbei, XU Qi. Research on pre-site selection of pumped storage power stations based on artificial intelligence [J]. Water Resources and Hydropower Engineering, 2025, 56(6): 199-213. DOI: 10.13928/j.cnki.wrahe.2025.06.017

基于人工智能技术的抽水蓄能电站预选址研究

李姝颖, 周申蓓, 徐琪

(河海大学 商学院, 江苏 南京 211100)

摘要:【目的】为解决目前抽水蓄能电站预选址大多依靠人工比选, 存在耗时费力、自动化程度低等问题, 提出一种将多模态大模型应用于抽水蓄能电站预选址的方法。【方法】基于抽水蓄能电站选址规范, 构建站址评价体系, 结合模糊综合评价法计算各站址综合得分, 作为电站标签。通过设计特定提示引导 GPT 模型生成与遥感图像相关的指令微调数据。在此基础上, 采用提示工程(prompt)与 LoRA 微调技术对多模态大语言模型 LLaVA 进行训练, 并将模型应用于安徽绩溪抽水蓄能电站预选址工程, 对模型实施系统评估。【结果】结果显示: 模型对安徽绩溪抽水蓄能电站工程水文、地形、经济等方面的关键指标评分准确, 得出其综合得分为 84.4 分, 符合理想站址标准; 在 1091 个样本的测试集上进行验证时, 模型能识别出 74.1% 的理想站址样本和 82.4% 的非理想站址样本; 微调后的 LLaVA 模型的 AUC 值为 0.822, 相比 Qwen-VL-Chat、InternLM-XComposer-VL、VisualGLM 和 InstructBLIP 模型分别高出了 0.106、0.152、0.205 和 0.207。【结论】结果表明: 利用上述方法进行微调的 LLaVA 模型在站址分类方面的精准率、召回率和误检率相较于多模态通用模型有显著提升, 并在实际的选址中展现出良好的站址评分能力, 具有较高的推广应用价值。通过对 LLaVA 模型进行领域微调和应用, 充分展示了多模态大模型在提高选址工作效率和自动化程度中的独特作用, 可为抽水蓄能行业的智能化转型提供有力支撑。

关键词: 抽水蓄能电站; LLaVA 模型; 预选址; 指令微调; 影响因素

DOI: 10.13928/j.cnki.wrahe.2025.06.017

开放科学(资源服务)标志码(OSID):

中图分类号: TV743

文献标志码: A

文章编号: 1000-0860(2025)06-0199-15



听语音 | 聊科研

Research on pre-site selection of pumped storage power stations based on artificial intelligence

LI Shuying, ZHOU Shenbei, XU Qi

(Business School, Hohai University, Nanjing 211100, Jiangsu, China)

Abstract: [Objective] Current pre-site selection of pumped storage power stations heavily relies on manual comparison and selection, which suffers from time-consuming processes and low-automation levels. To address these issues, a method integrating

收稿日期: 2025-03-13; 修回日期: 2025-04-06; 录用日期: 2025-04-07; 网络出版日期: 2025-04-15

基金项目: 国家重点研发计划项目(2019YFC0409000); 教育部人文社会科学研究规划基金项目(23YJAZH225); 江苏省高等学校大学生创新创业训练项目(202410294250Y)

作者简介: 李姝颖(2004—), 女, 本科生, 主要从事人工智能研究。E-mail: 3024458669@qq.com

通信作者: 周申蓓(1974—), 男, 教授, 博士研究生导师, 博士, 主要从事水资源及技术创新管理研究。E-mail: shenbei@hhu.edu.cn

©Editorial Department of Water Resources and Hydropower Engineering. This is an open access article under the CC BY-NC-ND license.

multimodal large models into the pre-site selection of pumped storage power stations is proposed. [Methods] Based on site selection criteria for pumped storage power stations, an evaluation system for potential sites was established. The fuzzy comprehensive evaluation method was employed to calculate an overall score for each site, which served as the station label. Then, specific prompts were designed to guide the GPT model in generating prompt fine-tuning data associated with remote sensing images. Based on this, prompt engineering and Low-Rank Adaptation (LoRA) fine-tuning techniques were used to train the multimodal large language model LLaVA. Subsequently, the trained model was applied to the pre-site selection of the Jixi Pumped Storage Power Station in Anhui Province, followed by a systematic evaluation of the model performance. [Results] The result showed that the model accurately scored for key indicators such as hydrology, topography, and economic factors for the Jixi Pumped Storage Power Station, yielding a comprehensive score of 84.4 that met the criteria for an ideal site. When validated on a test set of 1 091 samples, the model successfully identified 74.1% of ideal site samples and 82.4% of non-ideal site samples. The fine-tuned LLaVA model achieved an Area Under the Curve (AUC) value of 0.822, outperforming Qwen-VL-Chat, InternLM-XComposer-VL, VisualGLM, and InstructBLIP models by 0.106, 0.152, 0.205, and 0.207, respectively. [Conclusion] The findings indicate that the LLaVA model fine-tuned by the proposed method achieves significant improvements in accuracy, recall, and false detection rates for site classification compared to general-purpose multimodal models. Additionally, it demonstrates excellent site evaluation in practical applications, showing high potential for broader application. The domain-specific fine-tuning and application of the LLaVA model effectively highlight the unique advantages of multimodal large models in improving the efficiency and automation level of site selection, providing robust support for the intelligent transformation of the pumped storage industry.

Keywords: pumped storage power station; LLaVA Model; pre-site selection; prompt fine-tuning; influencing factors

0 引言

为助力实现“双碳”目标,我国正积极推动能源结构的转型升级。作为目前电力系统中最成熟的大规模储能技术,抽水蓄能电站具有调峰填谷、调频调相、事故备用和黑启动等多项功能,能够有效应对能源转型带来的调峰难题和储能需求^[1]。抽水蓄能电站通过上下两座水库以及水轮机和泵机系统的相互作用,在电力需求低谷时段利用多余电力将水从下库抽至上库进行储能;在电力需求高峰时段释放上库的水来驱动水轮机发电,实现能量的释放^[2]。电站的选址是决定其运行效率、经济性和环境影响的关键环节,具有重要意义^[3]。合理的位置布局能确保上下水库具备充足的水位差,有效提升抽水蓄能电站的蓄能能力和发电效率^[4]。同时,恰当的坝址位置有助于简化工程设计,降低建设成本,规避地震、滑坡等自然灾害的潜在风险,为电站的长期、稳定和高效运行提供有力保障。

目前,国内外学者对抽水蓄能电站选址的研究主要集中在选址影响因素和选址方法两个方面。在选址影响因素方面,学界普遍认为可以从水文地质特征、工程技术特征、经济社会特征和生态环境特征四个维度出发,构建站址评价体系^[5-8]。在选址方法方面,现有选址方法大致可概括为以下3类:(1)实地勘测法。通过现场访问和调查的方式,直接获取目标区域

的地理、环境、资源、基础设施等信息,用于选址决策^[9]。(2)基于地理信息系统(Geographic Information System, GIS)的空间分析法。借助GIS技术整合多源地理空间数据,并通过缓冲区分析、叠加分析、坡度分析等空间分析方法对各类因素进行识别和量化评估^[10]。进一步地,宋云丽等^[11]利用ArcPy站点包调用ArcGIS的空间分析工具,综合考虑库容、坝长及副坝数量等因素,通过编程实现地理空间数据的批量处理,提高了选址工作的自动化程度与准确性。(3)基于遥感和激光雷达技术的选址方法。利用无人机搭载激光雷达、RGB相机获取地面的三维地形数据和地形表面的RGB图像,结合抽水蓄能电站上水库与下水库的位置关系,确定抽水蓄能电站的候选站点^[12]。上述方法在当前的选址工作中得到了大规模应用,但存在人力成本高、易遗漏部分优质站点、对数据质量要求较高、自动化程度低等问题,难以适应新形势下抽蓄选点工作对“高精度、高效率”的要求。

随着人工智能和大数据技术的发展,以GPT-4为代表的多模态大模型通过“多模态预训练+多模态任务调优+多模态提示生成”的技术路线^[13],推动融合图文音视频等多种模态数据的感知、推理与决策一体化,促进人工智能迈向新高度^[14]。与之前的大语言模型相比,多模态大模型在处理复杂数据结构和多源数据融合方面表现卓越,能够挖掘出隐藏在不同模

态数据之间的深层次关联和模式^[15]。近年来,学者们对多模态大模型进行了广泛的研究,通过融入多模态预训练、多任务学习和强化学习等策略,使得模型不仅能够理解图像和文本,还能够进行复杂的推理任务,推动多模态推理向智能推理的方向发展,形成了多种在不同任务和领域中各具特色的大模型^[16-17]。例如,MiniGPT-4模型实现了图像和文本之间的深度语义对齐,在多任务的协同学习中表现出色^[18]。XU等^[19]提出的mPLUG-Ow模型优化了推理路径和生成能力,在某些动态的决策任务中表现出色。LI等^[20]提出的LLaVA模型不仅依赖传统的图像-文本对进行训练,还结合细粒度的视觉-语言预训练和多任务学习框架,能够更精细地捕捉图像与文本间的深层次关系,在处理图像描述生成和回答复杂视觉问题时展现出明显的优势。

以多模态大模型为基座开发垂直专用大模型,为各专业领域的创新变革注入了新动能,在教育、医疗、文化遗产保护、情感分析等领域被广泛应用^[21]。然而当前国内外将多模态大模型应用于抽水蓄能电站选址却鲜有研究。鉴于LLaVA模型在细粒度视觉-语言交互和视觉推理方面具有显著优势,本文借助LLaVA多模态大语言模型,结合抽水蓄能电站的卫星遥感图像,以及区域水文地质、地理地形、社会经济、生态环境等数据,提出一个综合权衡多种关键因素的电站预选址模型,以期缩短选址的工作周期、提升选址的准确性和效率,推动抽水蓄能行业高质量发展提供可行性建议。

本文研究框架如下:首先,参照国家行业标准及前人研究成果,构建抽水蓄能电站预选址评价体系,运用博弈论组合赋权法、模糊综合评价法得出全国范围内153个电站(包括已建、在建、核准的抽水蓄能电站)的综合得分,将该得分作为各电站的标签。其次,通过设计特定的提示(prompt)来引导GPT模型以问答的形式生成与图像相关的指令数据,避免人工生成数据耗时且效率低下的问题。然后,利用已选址电站以及不适合建电站区域的相关数据,对多模态大语言模型LLaVA进行训练,并引入LoRA微调,增强模型对电站预选址任务的理解能力。最后,通过比较模型在不同类型的提示工程(prompt)下的精准率、召回率和 F_1 值,并将微调后的LLaVA模型与其他多模态大模型进行对比,结合安徽绩溪抽水蓄能电站的具体案例,进一步验证LLaVA模型在抽水蓄能电站预选址上的效能。本文的研究框架如图1所示。

1 抽水蓄能电站选址综合评价体系构建

本文构建了涵盖水文、地理、地形、经济、环境等5个因素、10个评价指标的评价体系,运用层次分析法与CRITIC(Criteria Importance Through Intercriteria Correlation)方法相结合的主客观赋权法,引入博弈论计算评价指标综合权重,实现主观权重和客观权重的统一^[2],并在此基础上采用模糊综合评价法对各电站进行量化打分,将每个电站的站址可行性评分作为电站的标签。

1.1 指标选取

参照国家发布的NB/T 35009—2024《抽水蓄能电站选点规划编制规范》^[22],综合考虑影响抽水蓄能站址选择的刚性要求和模糊约束,从水文条件、地形条件、地理条件、经济效益和环境效应5个维度,选取了10项评价指标,构建抽水蓄能电站预选址评价体系(见表1)。具体指标选取和测量方式如下。

1.1.1 水文条件

本文用多年平均径流量和多年平均泥沙含量来衡量站址的水文条件。抽水蓄能电站的稳定运行依赖于充足的水量以满足抽水和发电的循环用水需求。若多年平均径流量过低,可能导致电站无法满足抽水需求,进而影响发电效率;若流量过高,则会增加防洪和调度的复杂性。此外,泥沙含量是影响电站运行寿命和调节能力的重要因素^[23]。抽水蓄能电站的机组兼具水轮机和水泵功能,运行过程中频繁启停,泥沙含量过高易导致泥沙颗粒对机组部件造成磨损,缩短机组寿命。同时,过多的泥沙淤积会占据水库的调节库容,减少水库的有效蓄水空间,进而降低电站的调节能力^[24]。因此,多年平均径流量和多年平均泥沙含量是保障抽水蓄能电站长期稳定运行、优化防洪调度的关键因素。

1.1.2 地形条件

本文选用水头和距高比作为衡量指标。水头是指上下水库的高度差,水头的大小直接影响到电站在储能与释能过程中的能量转换效率,进而决定了该地建设抽水蓄能电站能够提供的电能资源^[11]。而距高比作为衡量电站布局合理性的关键参数,是指抽水蓄能电站上、下水库水平距离与垂直高度的比值,其数值大小对电站的工程投资、水头损失以及电站布局有着重要影响。具体而言,距高比较小通常意味着上、下水库的水平距离较短,能够减少水流输送过程中的能量损失,提高能量转换效率。然而,若距高比过小,则可能导致电站布局的空间不够充分,不利于输水

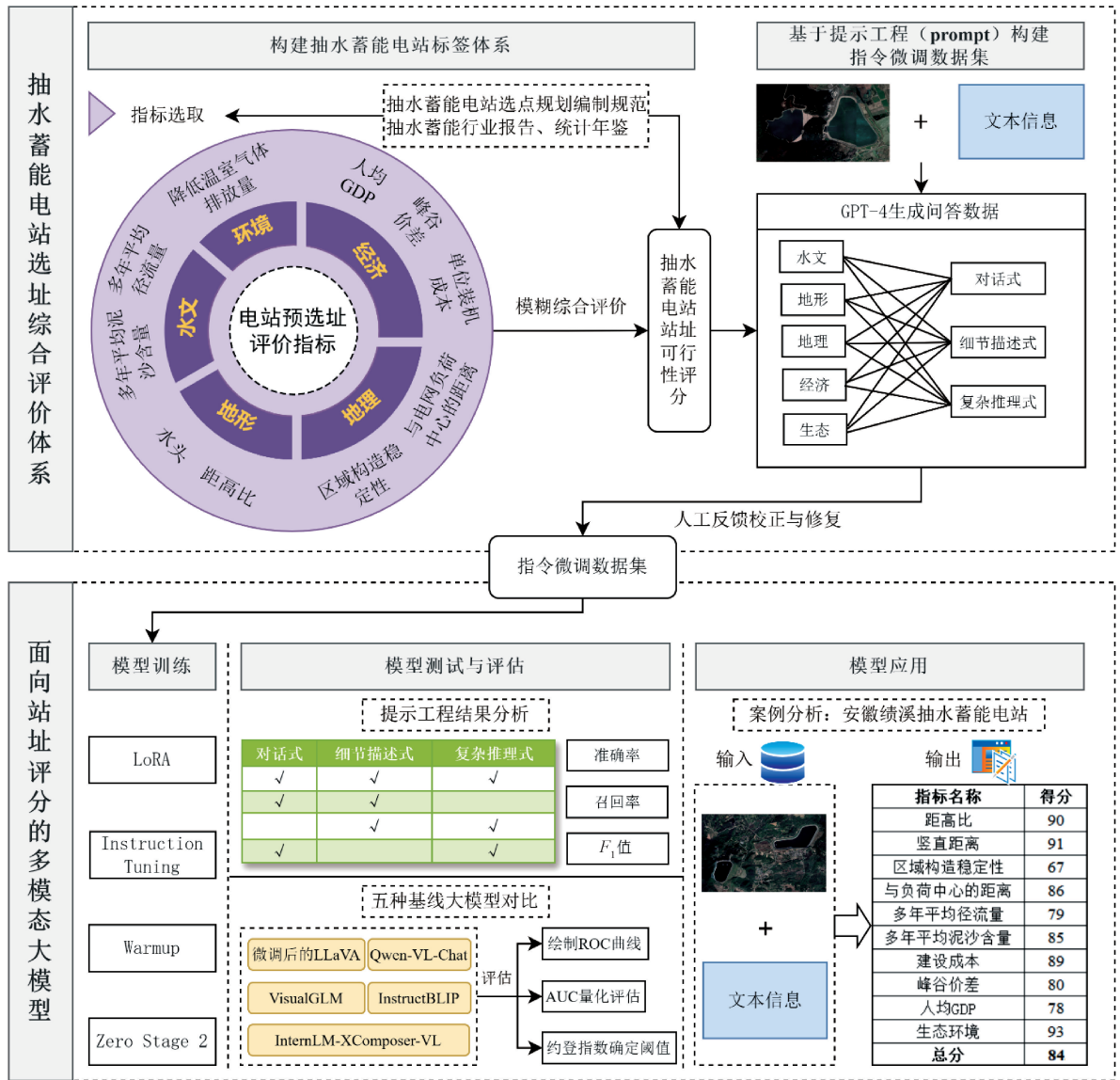


图 1 研究思路框架

Fig. 1 Research framework

表 1 抽水蓄能电站选址综合评价体系

Table 1 Comprehensive evaluation system for site selection of pumped storage power stations

一级指标	二级指标	指标量化
水文条件	多年平均径流量 多年平均泥沙含量	多年时间尺度上, 通过水体某一固定断面的平均流量 多年时间尺度上, 水体中泥沙含量的平均值
地形条件	水头 距高比	上下水库的高度差 上下水库水平距离与垂直高度的比值
地理条件	区域构造稳定性 与电网负荷中心的距离	量化为稳定性好、稳定性较差、稳定性差三个级别 电站与最近的地级市之间的距离
经济效益	抽蓄单位装机成本 峰谷价差	每单位装机容量所需的投资成本 电力系统中高峰时段电价与低谷时段电价之间的差值
环境效应	地区人均 GDP 降低温室气体排放量	国内生产总值(GDP)与该地区常住人口总数的比值 电站建设替代火电发电所减少的二氧化碳排放量

系统和地下主厂房的合理布置^[25]。相反, 若距高比过大, 则意味着上水库与下水库之间的水平距离较

长, 这将导致输水道延长, 引发工程量和投资的增加。

1.1.3 地理条件

本文选用区域构造稳定性、与电网负荷中心的距离作为衡量指标。区域构造稳定性评价是水利工程规划、坝址选择和工程设计的一项基本工作,其通过识别潜在的地质风险,如地震、断层活动、滑坡和泥石流等,有效降低工程实施的风险和投资成本。水利水电工程对区域构造稳定性提出了三分法方案,将区域构造稳定性分为三级,分别是稳定性好、稳定性较差、稳定性差^[26]。电网负荷中心是指电力需求集中、电能消费量大的地区,选择其附近区域作为电站站址,可以有效缩短输电线路的长度、降低输电损耗,提高电力系统的整体效率。

1.1.4 经济效益

抽水蓄能电站建设的经济效益主要体现在提供电力调节、储能和负荷平衡等服务所带来的经济回报和价值。本文分别以地区人均国内生产总值(Gross Domestic Product, GDP)、峰谷价差和抽蓄单位装机容量三个指标加以测度。地区人均GDP是衡量一个地区经济实力和居民生活水平的重要指标。一般而言,地区人均GDP越高,表明该地区经济发展水平越高,对抽水蓄能电站的建设和运营的经济支撑能力越强。峰谷价差是电力市场供需关系的直接体现,若某一区域的峰谷价差较大,则表明该地区电力供需的波动性更为显著,更需要抽水蓄能电站等灵活性资源以保障电网稳定运行,抽水蓄能电站在此区域的建设将更具现实价值和经济意义^[27]。此外,由于不同地区的地理条件和资源禀赋存在差异,抽水蓄能电站的投资成本也会因地区而异。为了更清晰地评估不同装机容量下的成本分布,本文采用抽蓄单位装机容量来衡量抽水蓄能电站的投资成本。

1.1.5 环境效应

本文以降低温室气体排放量来衡量电站建设的环境效应。抽水蓄能电站通过在电力需求低谷期以及新

能源发电量较大的时段进行抽水,吸收系统内过剩的新能源电力;在电力需求高峰期和新能源发电量较小的时段,释放先前储存的电能,替代火电出力,以降低系统碳排放量。同时,由于不同地区的发电结构不同,火电占比也不尽相同,平均排放因子能够准确地反映该地区电力系统的碳排放水平^[27]。因此,本文基于替代效应的思想,采用“抽水蓄能电站年发电量×省级电力平均二氧化碳排放因子”来估算抽水蓄能电站替代火电发电所减少的二氧化碳排放量。

1.2 确定评价指标权重

电站选址是一个涉及多因素、多层次、多准则的复杂决策问题。在权重赋值过程中,主观赋权法能够体现决策者对评价指标的主观偏好和重视程度,但客观性相对较差,而客观赋权法基于客观数据,能有效减少主观判断带来的偏差^[28]。为了充分发挥主观赋权法和客观赋权法各自的优势,本文将层次分析法与CRITIC方法相结合,以提高赋权的科学性和合理性。针对二者之间可能存在的冲突或偏差,博弈论理论可以有效弥补这一不足,通过对两种方法所确定的指标权重进行优化组合,使“主观权重”和“客观权重”达到“纳什均衡”状态,进而形成最优权重值^[29]。

本研究邀请了10位行业专家(涵盖设计、咨询、科研以及投资方),对各指标的相对重要程度进行主观赋值,并基于国家能源局、水电水利规划设计总院、中国水力发电工程学会发布的相关指标数据,利用CRITIC法计算得到客观权重,结合博弈论理论进行组合赋权。选址评价指标的权重结果如表2所列。

1.3 模糊综合评价法建立电站综合得分体系

模糊综合评价方法是一种基于模糊数学理论的多因素决策分析方法,广泛应用于复杂系统的评价与决策领域。该方法通过引入模糊集合的概念,综合考虑多个评价因素,并利用模糊数学对评价结果进行量化处理,从而有效应对评价过程中存在的模糊性和不

表2 站址评价指标权重结果

Table 2 Weights of evaluation indicators for site selection

准则层	主观权重	客观权重	组合系数	组合权重	指标层	主观权重	客观权重	组合系数	组合权重	综合权重
水文	0.257	0.200	$\alpha_1^* = 0.6636$ $\alpha_2^* = 0.3364$	0.238	多年平均径流量	0.750	0.420	$\alpha_1^* = 0.7194$	0.657	0.156
					多年平均泥沙含量	0.250	0.580	$\alpha_2^* = 0.2806$	0.343	0.082
地形	0.206	0.230		0.214	水头	0.500	0.474	$\alpha_1^* = 0.0000$	0.474	0.101
					距高比	0.500	0.526	$\alpha_2^* = 1.0000$	0.526	0.113
地理	0.203	0.240		0.216	区域构造稳定性	0.667	0.416	$\alpha_1^* = 0.6462$	0.578	0.125
					与电网负荷中心距离	0.333	0.585	$\alpha_2^* = 0.3538$	0.422	0.091
经济	0.137	0.150		0.141	抽蓄单位装机容量	0.500	0.296	$\alpha_1^* = 0.7368$	0.446	0.063
					峰谷价差	0.333	0.309	$\alpha_2^* = 0.2632$	0.327	0.046
					地区人均GDP	0.167	0.395		0.227	0.032
环境	0.197	0.180		0.191	降低温室气体排放量	1.000	1.000	$\alpha_1^* = 1.0000$	1.000	0.191

确定性^[30-32]。具体流程如下。

1.3.1 建立综合评价因素集

因素集涵盖了所有与评价目标相关的关键因素, 通常用 U 表示。本研究中, 因素集包括上文所述的所有与电站选址相关的因素, 具体为

$$U = \{u_1, u_2, \dots, u_n\} \quad (1)$$

1.3.2 确定评语集

评语集是指对评价对象进行综合评价时所采用的一组定性描述或等级划分, 用于对评价结果进行归类 and 描述。本文建立的评语集为 $V = \{\text{优秀, 良好, 一般, 较差, 差}\}$, 为了计算综合评价的具体分值, 将评价集量化为 $V = \{90 \sim 100, 80 \sim 89, 70 \sim 79, 60 \sim 69, \leq 60\}$ 。

1.3.3 构建模糊综合评价矩阵

模糊综合评价矩阵中的每个元素表示某个评价因素对某个评语的隶属度, 本文对多个专家的意见进行汇总和平均处理, 构造模糊评价矩阵 R , 其公式为

$$R = (\overline{r_1}, \overline{r_2}, \dots, \overline{r_n})^T \quad (2)$$

$$\overline{r_i} = \sum_{m=1}^k r_{im}/k \quad (3)$$

式中, k 为专家数量; r_{im} 为第 m 个专家对第 i 个因素的评价, $m \in (1, k)$; $\overline{r_i}$ 为第 i 个因素的综合得分, $i \in (1, n)$ 。

1.3.4 进行模糊综合评价运算

若已知因素集的权向量 $W = (w_1, w_2, \dots, w_n)$, 根据构建的模糊综合评价矩阵和确定的权重向量, 计算出各电站的综合评价结果。具体计算公式为

$$S = WR = (w_1, w_2, \dots, w_n)(\overline{r_1}, \overline{r_2}, \dots, \overline{r_n})^T \quad (4)$$

式中, S 为电站的综合评价得分。

2 面向站址评分体系的多模态模型

LLaMA 作为一个强大的开源语言模型, 已经成为许多多模态大模型的语言模型基础。LAmigo、FROMAGE 和 LLaVA 等多模态模型均使用 LLaMA 作为语言模型核心, 并通过大规模的无监督数据进行预训练, 具有出色的零样本学习和任务转移能力。然而, LAmigo、FROMAGE 等多模态大模型主要依赖图像文本对进行训练, 并未专门使用视觉-语言指令数据进行调整, 在处理多模态任务时的表现不如处理单一语言任务时出色^[33]。相比之下, LLaVA 模型通过结合视觉编码器和语言模型, 创新性地引入视觉-

语言指令数据进行微调, 实现了对图像和文本信息的深度理解和融合, 在处理复杂地理空间数据和多因素综合评估方面具有显著优势^[33]。因此, 本文采用 LLaVA 模型进行研究。

2.1 构建高质量的指令微调数据集

本文在全国范围内采集了已选址电站、不适合建电站区域的相关数据(包含图像数据和文本信息), 形成多组图像-文本对。在图像数据方面, 采集了相关区域的卫星遥感图像, 以提取水文、地理和地形特征; 在文本信息方面, 收集了大量数据集条目, 每条记录涵盖电站与最近地级市的距离、建设成本、所在地区峰谷价差、地区人均 GDP 和降低温室气体排放量等基本信息, 从而对电站的经济效益和环境效应进行评估。

鉴于现有图像-文本对数量相对有限, 且人工创建此类数据的过程较为耗时, 本文基于提示工程(prompt)填充任务的背景信息, 借助 GPT-4 生成与图像相关的问答数据(Question-Answering, QA)。问题由多模态数据、具体的站址评价指标构成, 回答为对站址某一维度的分析评价和打分。电站选址分析问答内容涵盖水文、地形、地理、经济和环境条件五个维度。为避免生成的数据缺乏多样性和深入的推理, 本文采用“渐进式”引导的方式, 针对每个维度, 生成了三种类型的指令跟随数据, 包括对话式、细节描述式和复杂推理式。对话式数据生成围绕图像内容展开, 注重互动性和自然对话的流畅性, 涵盖对象类型、数量、位置以及对象间的相对位置等, 模拟真实世界中的沟通交流情景。细节描述式数据生成旨在提供对图像内容全面而丰富的描述, 引导模型捕捉图像的重要细节。复杂推理式则要求模型不仅要识别图像内容, 还要进行深入的逻辑推理。此外, 为了减少 GPT-4 生成的问答样本中可能出现的脏数据, 如图像误解和问题重复等, 最后的数据还需要经过人工反馈校正与修复, 从而提供可微调、精确的指令数据集。指令微调数据示例如图 2 所示。

2.2 基于 LLaVA 大模型的站址评分问答

LLaVA 模型通过将视觉编码器和大型语言模型(Large Language Model, LLM)相结合, 利用“双流”Transformer 结构分别处理图像和文本输入, 在多模态数据联合处理方面展现了高度灵活性^[34]。本研究利用 LLaVA 模型中的语义嵌入技术, 将图像的水文、地形特征转化为统一的向量表示, 并与经过多级编码的涵盖地理、经济和环境等文本信息的向量进行拼接, 形成一个综合表征抽水蓄能电站选址情景的特征


	<p>文本信息: 这是响水涧抽水蓄能电站, 位于安徽省芜湖市弋江区峨桥镇, 距电网负荷中心芜湖市中心58km, 装机容量1000MW, 年发电量17.62亿kWh, 建设成本3.8元/W, 可减少二氧化碳排放12.4662万吨/a。该地区2023年人均GDP为76830元, 峰谷价差0.8759元。</p>
<p>类型1. 对话式</p>	
<p>Question: 图中的区域地形条件怎么样?</p> <p>Answer: 图中水库周围的地形呈现明显起伏, 该地区可能是丘陵或山地。</p>	
<p>类型2. 细节描述式</p>	
<p>Question: 请详细描述一下上图的内容。</p> <p>Answer: 这张图是一幅卫星地图, 图中有两个大型水库, 形状不规则, 可能是自然形成的湖泊或人工建造的水库, 水库的水面呈现出深蓝色, 显示出水体较深, 水库的边缘由混凝土或土坝围成, 表现出明显的人工痕迹。水库周围的植被覆盖较为密集, 说明该地区可能具有较好的土壤条件和适宜的气候, 有利于植被的生长。在水库的一侧可以看到一些建筑物和设施, 可能是水库的管理站、泵站或其他相关建筑。水库的右侧可以看到一些农田, 这些农田分布在地形较为平坦的区域, 说明该地区也有一定的农业活动。</p>	
<p>类型3. 复杂推理式</p>	
<p>Question: 依据上述描述, 请对该区域地形条件进行打分。</p> <p>Answer: 水库周围地形有明显起伏, 有助于自然地汇集水流, 形成水库, 适合建设抽水蓄能电站, 该地区地形条件评分为80分。</p>	

图2 指令微调数据示例

Fig. 2 Examples of prompt fine-tuning data

向量。在此基础上, 将构建的综合特征向量输入到LLM模型中, 以抽水蓄能电站的综合得分作为目标变量, 从而构建一个评估区域建站可行性的抽水蓄能电站预选址模型。上述分析框架如图3所示, 下面将对其中的关键流程进行详细介绍。

2.2.1 视觉流

在视觉信息处理方面, 模型采用基于ViT-L/14架构的CLIP视觉编码器, 以提取抽水蓄能电站选址相关的视觉特征^[35]。具体而言, 遥感影像首先被输入CLIP视觉编码器, 模型通过视觉Transformer处理图像数据, 将输入图像分割为一系列小块(patch), 并通过多层Transformer网络提取丰富的视觉特征, 包括地形形态、周围水资源分布、地质条件、植被覆盖等关键信息, 进而将视觉特征转换为高维嵌入表示。接着, 应用一个可训练的投影矩阵, 将图像特征转换成语言嵌入标记, 使得抽水蓄能电站选址相关的水文、地形信息能够与语言模型中的词嵌入向量具有相同的维度, 从而将电站图像中关键的物理信息提取出来, 便于后续与文本信息的融合。计算公式为

$$\mathbf{Z}_v = g(\mathbf{X}_v) \quad (5)$$

$$\mathbf{E}_{Image} = \mathbf{W} \cdot \mathbf{Z}_v \quad (6)$$

式中, \mathbf{X}_v 为输入的遥感影像; \mathbf{Z}_v 为视觉特征向量, 通过函数 g 从 \mathbf{X}_v 中提取; \mathbf{E}_{Image} 为遥感图像的视觉语言嵌入标记; $\mathbf{W} \in \mathbf{R}^{d_{Model} \times d_{Image}}$ 为用于将视觉特征向量 \mathbf{Z}_v 转换为语义嵌入向量的可训练权重矩阵。

2.2.2 文本流

在文本流方面, LLaVA采用预训练的LLaMA语言模型处理文本信息^[33]。LLaMA模型首先对输入文本 X_q 进行分词处理(Tokenization), 将其分解为更小的语义单元token, 以便模型能够细粒度理解站址描述的关键信息。然后, 模型利用其预训练的参数, 结合token嵌入(Token Embeddings)、段嵌入(Segment Embeddings)、位置嵌入(Position Embeddings), 将语义单元token编码为对站址评分有参考意义的高维语义嵌入 H_q , 并在每个站址文本信息的前后分别添加特殊的<s>和</s> token, <s>作为起始标记, 使得模型在生成或分析文本时能够识别完整的输入信息, </s>用于标记一个站址文本信息的结束。为深入理解各个站址方案的可行性, LLaMA模型依托自注意

力机制(Self-Attention)和 Transformer 架构,在语义层面捕捉站址描述中的关键因素及其相对重要性。自注意力机制能够动态计算不同要素的相关性,量化水文条件、地质稳定性、输电距离等因素对选址决策的影响,确保模型对选址文本的理解符合实际工程需求。例如,水文条件的可靠性会影响可供调度水量的稳定性,地质条件决定了工程施工的可行性和安全性,而输电距离则与电站运营成本密切相关等。语义单元 token 编码的计算公式为

$$\mathbf{E}_{Token}(T) = \mathbf{W}^M(\mathbf{W}^T T) \quad (7)$$

$$\mathbf{E}_{Segment}(T) = \mathbf{W}^M(\mathbf{W}^S T) \quad (8)$$

$$\mathbf{E}_{Position}(T) = \mathbf{W}^M(\mathbf{W}^P T) \quad (9)$$

$$\mathbf{E}_{Text} = \mathbf{E}_{Token} + \mathbf{E}_{Segment} + \mathbf{E}_{Position} \quad (10)$$

式中, \mathbf{E}_{Token} 为每个 token 的唯一嵌入; $\mathbf{E}_{Segment}$ 为不同的站址进行区分; $\mathbf{E}_{Position}$ 为编码序列中每个标记的位置; \mathbf{E}_{Text} 为文本信息的语义嵌入向量; $\mathbf{W}^M \in \mathbf{R}^{d_e \times d_{model}}$, $\mathbf{W}^T \in \mathbf{R}^{1 \times d_e}$, $\mathbf{W}^S \in \mathbf{R}^{1 \times d_N}$, $\mathbf{W}^P \in \mathbf{R}^{1 \times d_{max}}$ 为用于嵌入函数的可训练权重矩阵。

2.2.3 视觉流和文本流的融合

LLaVA 在中间层引入多模态融合模块,将遥感影像中的地区地形、坡度特征与区域规划文件、环境影响评估报告、调查数据等文本信息进行拼接,形成包含电站选址相关地理、环境、水文等多维特征的完整语义表示。计算公式为

$$\mathbf{E}_{Concat} = \mathbf{E}_{Image} \oplus \mathbf{E}_{Text} \quad (11)$$

式中, \mathbf{E}_{Concat} 为拼接后的完整语义嵌入向量; \mathbf{E}_{Image} 为遥感图像的视觉语言嵌入标记, \mathbf{E}_{Text} 为文本信息的语义嵌入向量。

2.2.4 LLaVA 模型的多模态理解与站址评分

LLaVA 模型将拼接后的文本和图像 token 序列作为输入,通过线性变换将文本特征映射为查询向量(Query),并将图像特征映射为键向量(Key)和值向量(Value)。以“水头”这一指标为例,给定文本描述:“绩溪抽水蓄能电站上水库的正常蓄水位设定为 959.50 m,下水库为 339.30 m,上下水库之间的高差为 620.20 m,为电站提供了充足的水头差,保证了发电和抽水的效率”,模型会将“水头”转化为查询向量。同时,根据图像中的河流、湖泊等视觉特征,模型为每个图像区域生成键向量(K)和值向量(V)。接着,LLaVA 通过点积计算文本查询向量与图像区域的键向量之间的相似度。在与“水头”指标密切相关的图像区域,文本描述与视觉特征之间的相似度较高,从而使得模型能够识别出与“水头”相关

的图像区域,实现文本与图像的有效匹配和关联。

在评分问答过程中,用户输入指令“请对该区域进行建站的适宜性评分”后,LLaVA 模型会对指令进行解析,识别任务目标及其相关要求,从文本中提取与建站适宜性相关的特征(如水文、地形、地质等),并将这些文本信息与图像中相应区域的视觉特征进行匹配与推理。基于跨模态推理的结果,LLaVA 模型依据预设的评分标准,从多个维度(如水资源条件、地形适宜性、环境影响等)对各项评分指标进行量化评估,最终生成综合评分。

2.3 训练策略

2.3.1 LLaVA 模型预训练阶段

第一阶段的训练目标是将图像特征 \mathbf{E}_{Image} 与预训练的 LLM 词嵌入 \mathbf{E}_{Text} 对齐^[34]。在这一阶段中,模型冻结视觉编码器和 LLM 的权重,只训练投影矩阵 \mathbf{W} ,从而使投影模块能够在实现最优对齐性能的同时显著降低训练成本^[35]。具体而言,模型先从 CC3M 数据集中筛选出 595K 图像-文本对,并将其转换为指令跟随数据。其次,对于每张图像,随机采样一个问题作为语言指令,采用单轮对话的形式,要求模型简要描述图像内容,预测的目标答案是图像的原始描述文本。

2.3.2 针对电站评分任务的微调阶段

在端到端微调阶段,冻结视觉编码器,对投影层 \mathbf{W} 和语言模型 LLM 进行训练,以针对电站评分任务进行优化。实验中,使用构建的高质量指令微调数据集对 LLaVA 模型进行微调,采用多轮问答和单轮问答相结合的方式,重点训练模型在抽水蓄能电站选址领域的多模态理解能力和问答能力。

其中,指令微调数据集通过结构化的 JSON 格式进行表示,包含以下字段。

(1) id: 数据记录的唯一标识符。

(2) image: 遥感图像的路径,用于提供视觉信息。

(3) text: 用于描述或补充选址任务的文本信息。

(4) conversations: 包含一系列对话的数组,每个对话包含表用户提问(指令)和模型回答。

用户指令(value): 在每轮对话中,用户提出的三种任务指令被作为 value 字段提供给 LLaVA 模型,引导模型对电站选址的不同维度进行多层次的分析和推理,使得模型能够在多种选址场景中生成多样化且富有逻辑的回答。

模型回答(value): 将此前 GPT-4 根据用户指令生成并经过人工修正的回答数据填充在 value 字段

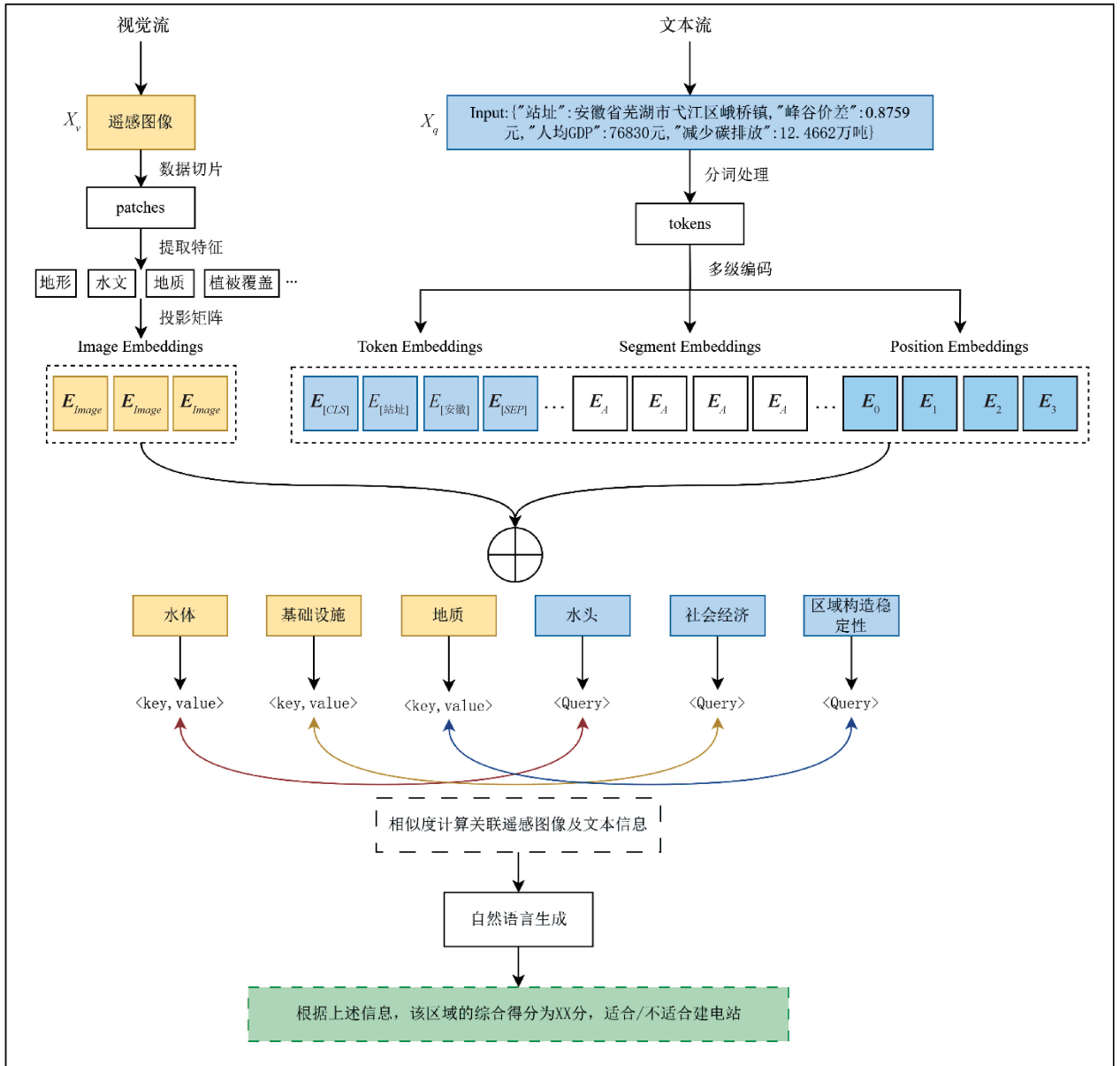


图3 面向站址评分体系的多模态模型架构

Fig. 3 Multimodal model architecture for site evaluation system

中,具体包含分析性回答、描述性回答和推理性回答三类回答数据。

3 试验与分析

3.1 数据预处理

在数据预处理阶段,研究对原始数据集进行了以下操作:(1)数据清洗:对文本数据进行去噪、填补缺失值、标准化数据格式、删除重复记录等基本的清洗操作,提高数据质量。(2)图像翻转:以0.5的概率分别对图像进行随机水平、垂直翻转,增加数据的多样性。(3)图像旋转:在 $-20^{\circ} \sim 20^{\circ}$ 之间随机旋转图像,增强模型对不同角度的适应性。(4)图像裁剪

和缩放:模拟图像在不同视角、不同尺度下的视觉效果,提升模型对输入图像的泛化能力。(5)图像颜色调整:随机调整图像的亮度、对比度、饱和度和色调。

3.2 试验内容与超参数设置

本文采用消费级服务器进行实验,具体环境配置如下:系统,64位 Ubuntu 16.04 LTS; CPU, Intel (R) Xeon (R) Gold 6132 CPU @ 2.6 GHz×45; 内存,128 GB RAM; GPU, NVIDIA A800; Python 版本,3.10.12; Cuda 版本,12.4; Pytorch 版本,2.1.2。

根据前文提出的站址指令数据集生成方法,针对现有图像-文本对进行数据生成。为获得足够的问答数据,每个问题会生成5~8种语义表述相近的指令

数据, 最终共生成了 7200 条指令数据, 用于微调预训练的 LLaVA 模型, 以提高其在抽水蓄能电站选址领域的性能和准确性。训练参数配置如表 3 所列。

表 3 训练参数配置

Table 3 Configuration of training parameters

模型参数	参数值
训练批次大小	40
训练周期	4
学习率	2×10^{-4}
学习率调度类型	Cosine
权重衰减系数	0.01
预热比例	0.03
多模态投影器学习率	2×10^{-5}
多模态视觉选择层	-2
LoRA rank	128
LoRA alpha	256

由表 3 可知, 本文采用较高的基础学习率($2e-4$), 结合 Cosine 学习率调度策略, 确保模型在训练初期快速收敛, 并在训练后期降低学习率以增强模型的稳定性。同时, 本实验引入了 Warmup 策略, 有助于模型在训练初期逐渐增加学习率, 避免因初始学习率设置过高而导致训练不稳定。

在多模态适配设计方面, 试验采用了基于 CLIP 模型的视觉编码器, 以实现视觉特征抽取, 且设置多模态视觉选择层(mm_vision_select_layer)为-2, 指定模型使用视觉编码器输出的倒数第二层作为图像特征, 优化图像特征的嵌入效果。与此同时, 通过加载预训练的多模态适配器 mm_mlp_adapter, 确保多模态融合模块在任务中的表现最优。

为提高模型的训练效率, 本实验选用 DeepSpeed 作为深度学习优化库, 利用 ZeRO Stage 2 作为核心优化技术。ZeRO Stage 2 通过将数据分片存储在不同的 GPU 上, 使得每个 GPU 只需存储其负责的部分数据, 从而大幅降低了单个 GPU 的显存占用, 实现了参数的高效分配。在此基础上, 本文对 DeepSpeed 的优化配置文件进行了定制化调整, 保证在复杂模型任务下的内存和计算资源分配更加合理。此外, 为进一步优化内存使用, 采用了混合精度训练和梯度检查点技术。混合精度训练利用 16 位浮点运算显著减少了显存占用, 而梯度检查点则允许模型在特定点保存梯度, 使得更深层的模型结构可以在有限的显存资源下进行训练。

在模型微调过程中, 启用了 LoRA 技术, 并对其低秩矩阵维度和缩放因子进行了配置。将低秩矩阵分解维度设为 128, 用于控制模型参数的压缩程度, 使

模型在保留适度特征的同时, 能够更高效地进行参数更新。同时, LoRA 缩放因子设为 256, 放大低秩矩阵权重, 使模型在更新时对目标任务特定信息具有更高的灵敏度。模型微调过程中损失(Loss)的变化趋势如图 4 所示。

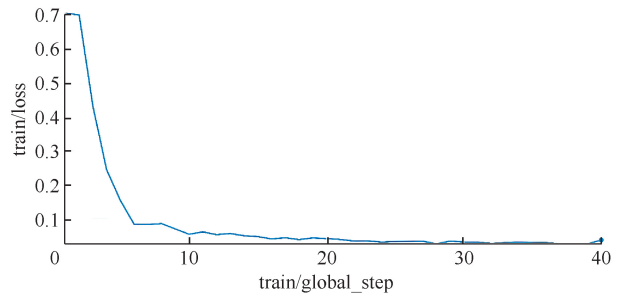


图 4 模型微调 Loss 损失

Fig. 4 Loss curve of model fine-tuning

3.3 结果分析

利用上述方法及训练集对大模型进行训练, 为了验证大模型的有效性及其精度, 继而采用测试集进行测试。测试集包含不适合建电站的区域 562 个, 适合建电站的区域 529 个。模型对每个测试集样本的预测结果是一个分数, 该分数在 0~100 范围内。为提高模型输出的可比性、可解释性和一致性, 研究将分数校准到 0~1 之间, 分数越小, 代表该站址是一个理想站址的概率越低, 反之, 分数越大, 代表该站址是一个理想站址的概率越高。

3.3.1 提示工程(prompt)试验结果分析

为深入分析各类 prompt 在模型指令微调中的作用, 本文进一步评估了缺失任意一种 prompt 对模型性能的影响。在站址评分效果测度上, 采用精准率、召回率和 F_1 值作为评价指标, 具体计算方式为

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (14)$$

式中, TP 为模型正确预测为理想站址的数量; FP 为模型错误预测为理想站址的数量; FN 为模型错误预测为非理想站址的数量; $Precision$ 为对站址的预测精度; $Recall$ 为对理想站址的召回率; F_1 值为精准率和召回率的调和平均值。

试验结果如表 4 所列, 以下为具体分析。

(1) 对话式+细节描述式+复杂推理式: LLaVA 模型在此组合下的精准率为 82.4%, 召回率为 74.1%,

F_1 值为 78.0%，表明模型在进行电站选址任务时，能够有效结合不同维度的任务要求，既保持了较高的精准率，又提升了召回率。其原因在于实验对模型进行了多维度训练，对话式 prompt 帮助模型能够通过与用户的互动深入理解任务，细节描述式 prompt 则增强了模型对电站选址各项指标以及图像的精准把握，而复杂推理式 prompt 使得模型能够进行深度推理，综合分析电站选址中的潜在因素。因此，模型能够在准确性与全面性之间取得平衡。

表 4 不同 prompt 下模型性能对比

Table 4 Comparison of model performance under different prompts

prompt 类型			精准率	召回率	F_1 值
对话式	细节描述式	复杂推理式			
✓	✓	✓	82.4	74.1	78.0
✓	✓		65.6	60.2	62.8
✓		✓	69.8	62.7	66.1
	✓	✓	75.5	68.2	71.7

(2) 对话式+细节描述式：该组合下的 F_1 值较“对话式 + 细节描述式 + 复杂推理式”组合下降了 15.2%，且召回率较低，主要原因在于其侧重通过基本信息提取和细节描述来理解任务，未能引入推理机制，限制了模型的高维度信息综合分析能力，导致在面对具有多重因素交织的电站选址任务时，模型无法从整体上抓住任务的复杂性，对理想站址的识别较为局限。

(3) 对话式+复杂推理式：该组合的各项指标值相对适中，但相比于“对话式 + 细节描述式 + 复杂推理式”组合，召回率下降了 11.4%、 F_1 值降低了 11.9%。本研究将其原因归为缺乏对细节的全面理解。尽管复杂推理式 prompt 增强了模型的推理能力，使其能够更好地进行图像文本匹配，但在没有细节描述式 prompt 的支持下，模型可能会忽视某些关键细节，如植被的覆盖情况、基础设施的布局等，导致未能全面把握电站选址任务中的各个维度，进而影响对理想站址的预测。

(4) 细节描述式+复杂推理式：该组合的 F_1 值达到了 71.7%，精准率和召回率的差距较小，表现出较为稳定的综合性能。究其原因，是因为细节描述式与复杂推理式 prompt 互为补充，细节描述式使得模型能深入理解电站选址的具体要求和特定参数，复杂推理式帮助模型在多个维度之间建立关联，进行更深层次的推理，二者结合使得模型具备了从微观到宏观的分析能力。然而，该组合缺乏对话式 prompt 的训

练，更侧重于静态的信息描述和推理，动态调整和应变能力较弱。因此，在处理高度复杂和多变的任务时，灵活性不及“对话式 + 细节描述式 + 复杂推理式”组合。

3.3.2 模型与基线方法的试验结果对比分析

为全面衡量经过微调的 LLaVA 模型的效能，本研究对比测试了 4 种多模态大模型在电站选址任务上的性能表现，包括 Qwen-VL-Chat、InternLM-XComposer-VL、VisualGLM 和 InstructBLIP 模型。研究在对模型进行站址分类效果评定的时候，采用理想站址的召回率和非理想站址的误检率作为评测指标，同时，为五种模型在测试集上的表现绘制 ROC (Receiver Operating Characteristic) 曲线(见图 5)，并计算对应的 AUC 值(Area Under ROC Curve)以量化评估模型的性能(见表 5)。

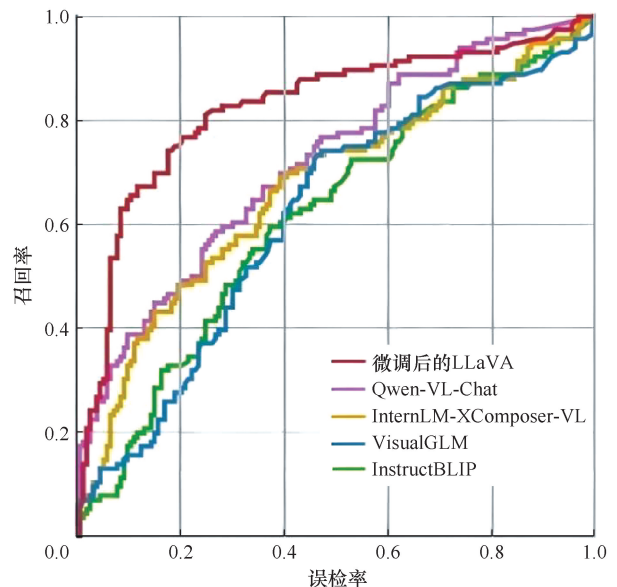


图 5 各模型在测试集上的 ROC 曲线

Fig. 5 Comparison of ROC curves of different models on test set

表 5 多模态模型 AUC 汇总表

Table 5 Summary of AUC for multimodal models

模型	AUC	标准误差	P 值	95%置信区间
微调后的 LLaVA	0.822	0.014	0.000**	0.795~0.849
Qwen-VL-Chat	0.716	0.016	0.000**	0.686~0.747
InternLM-XComposer-VL	0.670	0.017	0.000**	0.637~0.703
VisualGLM	0.617	0.017	0.000**	0.583~0.651
InstructBLIP	0.615	0.017	0.000**	0.581~0.649

注： $P < 0.05$ ；* 表示结果在统计上具有显著性； $P < 0.01$ ；**表示结果在统计上具有高度显著性。

从 ROC 曲线的整体趋势来看，所有模型的曲线均位于对角线之上，表明五个模型均具备一定的站址

分类能力, 但不同模型之间存在显著差异。具体而言, 微调后的 LLaVA 模型的曲线位于其他模型之上, 表现出较高的召回率和较低的误检率, 说明 LLaVA 模型能够根据遥感图像和文本信息有效识别适宜的电站选址区域, 并排除不适宜的区域, 在电站选址任务中具有较为优越的性能。相较之下, Qwen-VL-Chat 模型的误检率较高, 在电站选址过程中存在一定的错误分类风险, 可能将不符合建设条件的区域误判为适合电站建设的地点。InternLM-XComposer-VL、VisualGLM 和 InstructBLIP 模型的整体表现相对较弱, 且其 ROC 曲线明显低于 LLaVA 和 Qwen-VL-Chat, 尤其是在误检率低于 0.3 的情况下, 这些模型的召回率增长较为缓慢, 表明在相同的误检率下, 这些模型的正确分类能力相对有限。

结合表 5, 可以进一步分析各模型性能的差异。一般而言, AUC 值越高, 表示模型的预测能力越强。根据 MAO 等^[36]提出的 AUC 判断标准, AUC 值在 0.5 至 0.7 之间表示模型的站址预测能力较弱, 介于 0.7 至 0.9 则表示模型具有较好的预测能力, 而超过 0.9 则表明模型具有出色的预测能力。微调后的 LLaVA 模型 AUC 值为 0.82, 说明本文提出的框架在结合图像和文本进行抽水蓄能电站预选址时, 具有较高的准确性。相比之下, Qwen-VL-Chat、VisualGLM 和 InstructBLIP 的 AUC 值均低于 0.7, 在电站站址分类任务中的表现较差。所有模型的 p 值均小于 0.01, 说明实验结果具有统计学显著性^[36]。根据实验结果, 可以得出结论: 微调后的 LLaVA 模型优于 VisualGLM 等多模态通用模型, 原因在于通用模型缺乏针对相关领域的知识训练, 导致对电站数据的编码能力较弱。而微调后的 LLaVA 模型经过大规模数据集预训练, 并结合领域知识进行微调, 能更有效地学习和理解抽水蓄能电站选址的相关知识。

为确定微调后的 LLaVA 模型在站址分类中的最佳阈值, 本研究采用了约登指数 (Youden's Index) 作为评价标准。该指数用于评估二分类模型的整体性能, 通过平衡召回率和特异度, 衡量分类器对正负样本的整体预测能力, 从而帮助确定使模型分类性能最佳的阈值。当约登指数达到最大值时, 相应的阈值即为最佳界值^[37]。计算公式为

$$J = Precision - Recall \quad (15)$$

式中, J 为约登指数; $Precision$ 为精准率; $Recall$ 为召回率。

通过计算不同阈值下的约登指数, 发现当阈值为 0.57 时, 约登指数达到最大值, 此时模型的分类性

能最优。因此, 在定量评估中, 选择 0.57 作为最终的决策阈值, 预测值高于该阈值的样本被分类为理想站址, 低于该阈值的样本则为非理想站址。在此阈值下, 模型能够识别出 74.1% 的理想站址样本和 82.4% 的非理想站址样本。


3.3.3 案例分析

为了更加充分地验证 LLaVA 模型在电站选址任务中的有效性, 本小节通过案例分析对微调后的 LLaVA 模型的实验结果进行了深入探讨。以安徽绩溪抽水蓄能电站为例, 重点分析了模型在实际应用中的站址评分预测能力。模型在选址任务中对各影响因素的预测结果如表 6 所列, 涵盖了水文、地形、地理、社会经济以及环境效应等多个方面。

首先, 模型能够准确识别遥感图像中的关键特征。在分析该地区的水头与距高比时, 模型成功识别出该区域地形高差显著, 坡度变化较大, 上下水库的高度差明显。因此, 模型认为地区距高比和水头条件较好, 分别给出了 90 分和 91 分。这一结果表明, 微调后的 LLaVA 模型在处理遥感图像方面表现出了较高的准确性与可靠性。其次, 模型在文本信息处理方面同样展现出良好的能力。例如, 在分析文本“电站装机容量 1 800 MW, 年平均发电量 40.20 亿 kWh, 年发电利用小时 2 233 h, 抽蓄单位装机成本 5.49 元/W”时, 模型成功捕捉到电站装机容量较大、年发电小时数充足、经济性与运营能力较强等信息, 针对抽蓄单位装机成本给出了 89 分; 根据文本“电站在促进清洁能源消纳方面效果十分显著, 相当于每年减少电网煤炭消耗量约 21.6 万 t, 减少排放二氧化碳约 47.5 万 t”, 模型认识到该电站的环境效益显著, 为“生态效益”赋予了 93 分的评分。此外, 模型能够有效地关联遥感影像和文本信息, 进行综合分析和深入推理。例如, 模型将图像中的两片水体与文本中提到的“上水库位于登源河的北支流, 下水库位于登源河的北支流”进行关联, 将图像中上下水库的地势差异和文本中描述的“上水库和下水库的高差为 620 m”相匹配, 进一步验证水头和距高比的评分; 同时, 模型结合图像中的道路和村落与文本信息“距离绩溪县城约 29 km”, 进一步确认该区域的交通便捷性和经济条件, 并为“负荷中心距离”这一指标赋予了 86 分。通过综合分析各项指标, 模型最终得出电站综合得分为 84.4 分, 大于上一节中提出的阈值 57 分, 表明该区域适合建设电站。该结果进一步验证了本研究所提出的模型在电站选址任务中的潜力。

表 6 案例分析

Table 6 Case analysis

文本信息	遥感图像	指标	预测分数
安徽绩溪抽水蓄能电站位于安徽省绩溪县伏岭镇境内, 距绩溪县城公路里程约 29 km, 地理位置优越。电站装机容量 1 800 MW, 年平均发电量 40.20 亿 kWh, 年发电利用小时 2 233 h, 建设成本 5.49 元/W。电站在促进清洁能源消纳方面效果十分显著, 相当于每年减少电网煤炭消耗量约 21.6 万 t, 减少排放二氧化碳约 47.5 万 t。该地区 2023 年人均 GDP 为 76 830 元, 峰谷价差为 0.875 9 元。绩溪抽水蓄能电站上水库的正常蓄水位设定为 959.50 m, 下水库为 339.30 m, 上下水库之间的高差为 620.20 m, 为电站提供了充足的水头差, 保证了发电和抽水的效率。上水库位于登源河的北支流、近东西向展布的赤石坑沟内, 由一座钢筋混凝土面板堆石坝及库周山岭围成。下水库位于登源河的北支流、赤石坑沟口的上岭前、下岭前村所在的山涧盆地, 库坝坝型为混凝土面板堆石坝。登原河年平均径流量 1 亿~3 亿 m ³ , 多年平均泥沙含量 0.1~0.3 kg/m ³ 。		多年平均径流量	79
		多年平均泥沙含量	85
		距高比	90
		水头	91
		区域构造稳定性	67
		与负荷中心距离	86
		抽蓄单位装机容量	89
		峰谷价差	80
		地区人均 GDP	78
		生态效益	93
		总分	84

尽管如此, 模型在某些隐含信息的推断上仍存在不足。例如, 针对遥感图像和文本中提到的“上水库由一座钢筋混凝土面板堆石坝及库周山岭围成, 下水库库坝坝型为混凝土面板堆石坝”, 模型未能对“区域构造稳定性”进行准确预测, 导致该指标的评分偏低。这可能是由于图像分辨率或影像特性限制, 模型难以有效识别图像中的地质特征, 未能充分将图像中的地理信息与文本中提到的地质描述关联起来, 从而影响了区域稳定性的准确判断。总体而言, 微调后的 LLaVA 模型在跨模态推理和多维度综合分析方面表现出色, 但在处理低分辨率影像和识别复杂地质特征方面的能力仍有待提升。

4 结论及展望

针对传统抽水蓄能电站选址方案存在耗时费力、自动化程度低、人为主观影响较大等问题, 构建一个抽水蓄能电站预选址领域的 LLaVA 模型, 取得了如下成果。

(1) 从水文、地形、地理、经济和环境五个方面出发, 建立了抽水蓄能电站选址评价体系, 提出一种基于博弈论组合赋权的抽水蓄能电站选址评价方法, 并应用于全国范围内的 153 个抽水蓄能电站, 得出各站址的综合得分。

(2) 针对抽水蓄能电站选址领域可利用的图像和文本数量少、质量低, 手动构建数据代价高的问题, 提出了一种多维度、递进式的站址指令数据集生成方法。该方法通过指导 GPT-4 根据实际的遥感图像和文本信息, 从不同维度生成三种类型的指令跟随数据, 丰富了问答数据的多样性, 提供了构建抽水蓄能电站选址领域模型所需的高质量微调数据集。

(3) 针对通用 LLaVA 大模型领域知识不足的问题, 首先, 使用面向电站选址任务的指令微调数据对大模型进行指令微调, 以增强模型对特定任务的理解能力。其次, 利用 LoRA 技术进行领域有监督微调, 构建了抽水蓄能电站预选址领域模型。

(4) 经过预训练和指令微调的 LLaVA 模型在站址分类方面的精准率、召回率和误检率要显著优于 Qwen-VL-Chat、InternLM-XComposer-VL、VisualGLM 和 InstructBLIP 等多模态通用模型, 表明 LLaVA 模型在抽水蓄能电站预选址任务中具有较高的适应性。

本文提出了一种基于 LLaVA 大模型的抽水蓄能电站预选址方法, 可为水利水电工程的智能化转型提供新思路。未来研究中, 可以考虑集成地方政策、气候等更多的影响因子, 进一步提高抽水蓄能电站预选址的科学性、准确性和实用性; 此外, 还可扩展模型参数量, 增加模型数据集规模, 并引入强化学习方法以进一步提升模型性能, 推动抽水蓄能行业的数字化转型和高质量发展。

参考文献 (References):

- [1] 李国英. 进一步全面深化水利改革 为推动水利高质量发展、保障我国水安全作出新的贡献: 在 2025 年全国水利工作会议上的讲话[J]. 水利发展研究, 2025, 25(1): 1-12.
LI G Y. Further comprehensively deepen water conservancy reform to promote high-quality development of water conservancy and safeguard national water security: A speech at the 2025 national water conservancy work conference [J]. Water Resources Development Research, 2025, 25(1): 1-12.
- [2] 李国英. 为以中国式现代化全面推进强国建设、民族复兴伟业提供有力的水安全保障[J]. 水利发展研究, 2024, 24(3): 1-3.

- LI G Y. Improved water security for China's efforts to build itself into a stronger country and rejuvenate the Chinese nation on all fronts by pursuing Chinese modernization[J]. *Water Resources Development Research*, 2024, 24(3): 1-3.
- [3] 吴丰昌. 我国水体污染控制与治理成效、科技支撑与展望[J]. *水利发展研究*, 2023, 23(12): 1-8.
- WU F C. Effectiveness, scientific and technological support, and prospects for water pollution control and management in China[J]. *Water Resources Development Research*, 2023, 23(12): 1-8.
- [4] 费香泽, 顾克, 刘佳龙, 等. 基于卫星遥感地形数据的抽水蓄能电站上下水库选址方法研究[J]. *水电能源科学*, 2023, 41(2): 79-82.
- FEI X Z, GU K, LIU J L, et al. Study on site selection of upper and lower reservoirs of pumped storage power station based on satellite remote sensing terrain data[J]. *Water Resources and Power*, 2023, 41(2): 79-82.
- [5] 邓子昂, 张继勋, 常飞, 等. 基于博弈论组合赋权的抽水蓄能电站选址评价[J]. *水力发电*, 2024, 50(8): 22-27.
- DENG Z A, ZHANG J X, CHANG F, et al. Evaluation of pumped-storage power station site selection based on game theory-combinatorial weighting[J]. *Water Power*, 2024, 50(8): 22-27.
- [6] NZOTCHA U, KENFACK J, BLANCHE M M. Integrated multi-criteria decision making methodology for pumped hydro-energy storage plant site selection from a sustainable development perspective with an application[J]. *Renewable and Sustainable Energy Reviews*, 2019, 112: 930-947.
- [7] 张涛, 左双英, 沈春勇, 等. 岩溶洼地抽水蓄能建库适宜性评价体系及目标优选[J]. *中国岩溶*, 2023, 42(6): 1161-1172.
- ZHANG T, ZUO S Y, SHEN C Y, et al. Evaluation system for the suitability of reservoir construction for pumped storage and optimization of its site selection in karst depressions[J]. *Carsologica Sinica*, 2023, 42(6): 1161-1172.
- [8] 任岩, 侯尚辰. 基于多能互补的抽水蓄能电站站址选择的研究[J]. *水电与抽水蓄能*, 2021, 7(6): 37-39.
- REN Y, HOU S C. Study on site selection of pumped storage power station based on multiple complementarity [J]. *Hydropower and Pumped Storage*, 2021, 7(6): 37-39.
- [9] 王玉威, 徐子桥. 微动在广西田东县抽水蓄能电站大坝选址勘查中的应用研究[J]. *地下水*, 2023, 45(3): 293-295.
- WANG Y W, XU Z Q. Study on application of fretting in dam site selection exploration of pumped storage power station in Tiandong County, Guangxi[J]. *Ground Water*, 2023, 45(3): 293-295.
- [10] LU B, STOCKS M, BLAKERS A, et al. Geographic information system algorithms to locate prospective sites for pumped hydro energy storage[J]. *Applied Energy*, 2018, 222: 300-312.
- [11] 宋云丽, 严云籍, 翟林博, 等. 抽水蓄能电站建设的地理要素分析及GIS选址[J]. *云南水力发电*, 2022, 38(4): 131-134.
- SONG Y L, YAN Y J, ZHAI L B, et al. Geographic element analysis and GIS location selection for pumped storage power station construction[J]. *Yunnan Water Power*, 2022, 38(4): 131-134.
- [12] THIRUNAVUKARASU A J, TING D S J, ELANGO VAN K, et al. Large language models in medicine[J]. *Nature Medicine*, 2023, 29(8): 1930-1940.
- [13] SZOT A, MAZOURE B, AGRAWAL H, et al. Grounding multimodal large language models in actions[J]. *Advances in Neural Information Processing Systems*, 2024, 37: 20198-20224.
- [14] 刘静, 郭龙腾. GPT-4 对多模态大模型在多模态理解、生成、交互上的启发[J]. *中国科学基金*, 2023, 37(5): 793-802.
- LIU J, GUO L T. Inspiration of GPT-4 on multimodal foundation models in multimodal understanding, generation, and interaction [J]. *Bulletin of National Natural Science Foundation of China*, 2023, 37(5): 793-802.
- [15] KOH J Y, FRIED D, SALAKHUTDINOV R R. Generating images with multimodal language models[J]. *Advances in Neural Information Processing Systems*, 2023, 36: 21487-21506.
- [16] GANDHI A, ADHVARYU K, PORIA S, et al. Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions [J]. *Information Fusion*, 2023, 91: 424-444.
- [17] GU T, ZHOU Z, HUANG K, et al. Mllmgard: A multi-dimensional safety evaluation suite for multimodal large language models [J]. *Advances in Neural Information Processing Systems*, 2024, 37: 7256-7295.
- [18] ALSABBAGH A R, MANSOUR T, AL-KHARABSHEH M, et al. MiniMedGPT: Efficient large vision-language model for medical visual question answering [J]. *Pattern Recognition Letters*, 2025, 189: 8-16.
- [19] XU X, WANG T, YANG Y, et al. Cross-modal attention with semantic consistence for image-text matching[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(12): 5412-5425.
- [20] LI C, WONG C, ZHANG S, et al. Llava-med: Training a large language-and-vision assistant for biomedicine in one day [J]. *Advances in Neural Information Processing Systems*, 2023, 36: 28541-28564.
- [21] 罗江华, 张玉柳. 多模态大模型驱动的学科知识图谱进化及教育应用[J]. *现代教育技术*, 2023, 33(12): 76-88.
- LUO J H, ZHANG Y L. Evolution and educational application of discipline knowledge graph driven by multimodal large model [J]. *Modern Educational Technology*, 2023, 33(12): 76-88.
- [22] 国家能源局. 抽水蓄能中长期发展规划(2021—2035年)[Z]. 北京: 国家能源局, 2021.
- National Energy Administration. *Medium and Long term Development Plan for Pumped Storage (2021—2035)* [Z]. Beijing: National Energy Administration, 2021.
- [23] 彭煜民, 刘德旭, 王雪林, 等. 计及水库运行约束的极端气象场景下抽蓄电站优化调度研究[J]. *水利学报*, 2023, 54(11): 1298-1308.
- PENG Y M, LIU D X, WANG X L, et al. Optimization scheduling of pumped-storage power stations under extreme weather scenarios considering reservoir operation constraints[J]. *Journal of Hydraulic Engineering*, 2023, 54(11): 1298-1308.
- [24] 李更丰, 孙少华, 别朝红, 等. 面向新型电力系统弹性提升的储能优化配置与灵活调度研究综述[J]. *高电压技术*, 2023, 49

- (10): 4084-4095.
- LI G F, SUN S H, BIE Z H, et al. Review on optimal configuration and flexible scheduling research of energy storage for resilience improvement of new power system [J]. *High Voltage Engineering*, 2023, 49(10): 4084-4095.
- [25] GÜNEN M A. A comprehensive framework based on GIS-AHP for the installation of solar PV farms in Kahramanmaraş, Turkey [J]. *Renewable Energy*, 2021, 178: 212-225.
- [26] 王惠卿, 谭成轩, 丰成君, 等. 京津冀协同发展区活动构造与地壳稳定性[J]. *地质通报*, 2022, 41(8): 1322-1341.
- WANG H Q, TAN C X, FENG C J, et al. Activity structure and crustal stability in Beijing-Tianjin-Hebei collaborative development zone[J]. *Geological Bulletin of China*, 2022, 41(8): 1322-1341.
- [27] 徐三敏, 张弓, 王放, 等. 基于 CCER 规则的抽水蓄能碳减排计算方法[J]. *中国电力*, 2024, 57(1): 175-182.
- XU S M, ZHANG G, WANG F, et al. Carbon emission reduction calculation method for pumped storage based on CCER rules [J]. *Electric Power*, 2024, 57(1): 175-182.
- [28] SAHA P, PAUL B. Identification of potential strategic sites for city planning based on water quality through GIS-AHP-integrated model [J]. *Environmental Science and Pollution Research International*, 2021, 28(18): 23073-23086.
- [29] 任楠波, 刘宏权, 潘增辉, 等. 基于 DPSIR-博弈论组合赋权 TOPSIS 模型的河北省水资源保障度演进评价[J]. *南水北调与水利科技(中英文)*, 2023, 21(5): 873-885.
- REN N B, LIU H Q, PAN Z H, et al. Evaluation on the evolution of water resources security degree in Hebei Province based on DPSIR-game theory combined weighting TOPSIS model [J]. *South-to-North Water Transfers and Water Science & Technology*, 2023, 21(5): 873-885.
- [30] 王春燕, 乔娟. 改进的层次分析法及模糊综合评价法在病险水库除险加固治理效果评价中的应用[J]. *水电能源科学*, 2019, 37(10): 64-67.
- WANG C Y, QIAO J. Application of improved analytic hierarchy process and fuzzy comprehensive evaluation in evaluation of treatment effect of disease-dangerous reservoir [J]. *Water Resources and Power*, 2019, 37(10): 64-67.
- [31] 张修宇, 李颖博, 闫倩倩, 等. 河南省农村供水工程标准化管理水平综合评价[J]. *人民黄河*, 2023, 45(8): 134-137.
- ZHANG X Y, LI Y B, YAN Q Q, et al. Comprehensive evaluation of standardization management level of rural water supply projects in Henan Province [J]. *Yellow River*, 2023, 45(8): 134-137.
- [32] 何武全, 刘群昌, 娄宗科, 等. 管道输水灌溉工程质量模糊综合评价方法研究[J]. *人民黄河*, 2017, 39(1): 145-148.
- HE W Q, LIU Q C, LOU Z K, et al. Fuzzy comprehensive evaluation for engineering quality of pressure pipeline irrigation [J]. *Yellow River*, 2017, 39(1): 145-148.
- [33] LIANG Y, QIN G, SUN M, et al. Multi-modal interactive attention and dual progressive decoding network for RGB-D/T salient object detection [J]. *Neurocomputing*, 2022, 490: 132-145.
- [34] LIU Z, HUANG Y, WANG Z, et al. Visual instruction tuning for multi-modal large language models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024, 46(3): 1234-1248.
- [35] WANG J, LI M, LUO H. et al. Power-llava: Large language and vision assistant for power transmission line inspection [J]. *IEEE International Conference on Image Processing*, 2024, 12024(1): 963-969.
- [36] MAO Y, HAO Y, CAO X, et al. Boosting GNN-based link prediction via PU-AUC optimization [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2025, 37(4): 1635-1649.
- [37] CLAUS B O M, DE SMEDT D, DE COCK P A. Therapeutic drug monitoring versus Bayesian AUC-based dosing for vancomycin in routine practice: A cost-benefit analysis [J]. *Journal of Antimicrobial Chemotherapy*, 2025, 80(3): 857-867.

(责任编辑 王海锋)