

# 基于大语言模型的抽水蓄能电站智能运维 辅助系统构建研究

胡昊<sup>1,2,3</sup>, 许昭一<sup>2</sup>, 崔争艳<sup>1,3</sup>, 张兴奎<sup>2</sup>, 张浩宇<sup>2</sup>

(1. 黄河水利职业技术大学, 河南 开封 475004; 2. 华北水利水电大学, 河南 郑州 450046;

3. 河南省跨流域区域引调水运行与生态安全工程研究中心, 河南 开封 475004)

**摘要:**【目的】针对抽水蓄能电站运维工作复杂、专业知识要求高、传统知识检索效率低下以及通用大语言模型难以准确理解抽蓄运维特有专业术语和复杂工况的问题,为提高运维人员业务水平、运维效率及知识复用效率,满足国家水利电力新质生产力发展要求,构建适用于抽水蓄能电站运维领域的智能辅助系统。【方法】构建基于大语言模型的抽水蓄能电站智能运维辅助系统(PS-AIOM),以本地部署的大语言模型为底座,采用检索增强生成和提示工程技术,通过外挂知识库检索,增强大语言模型在抽水蓄能电站运维领域应用的准确性。为确保运维业务回答的准确性,构建了运维专业知识库;为验证系统的有效性,构建了测试数据集。【结果】①采用综合关键词和语义相似性的方法对PS-AIOM的回答进行评分,既考虑了回答内容的具体匹配程度,也考虑了整体的文本相似性,实现了较为全面的评估。②PS-AIOM对运维业务的回答满意率为83%,综合平均分为0.8138,能够较好地辅助运维人员工作。③PS-AIOM的表现远优于先进的通用大语言模型(模型Z:回答满意率为4%,综合平均分为0.3982;模型T:回答满意率为17%,综合平均分为0.4775)。【结论】通过结合本地化部署的大语言模型底座、检索增强生成技术和专业领域知识库,PS-AIOM系统有效克服了通用大语言模型在专业领域知识匮乏的问题,实现了在抽水蓄能电站运维领域较为准确的知识应用。

**关键词:**抽水蓄能电站;智能运维;大语言模型;检索增强生成

**中图分类号:**TV743;TV738

**文献标识码:**A

**文章编号:**2096-6792(2025)05-0034-09

构建以新能源为主体的新型电力系统是实现碳达峰、碳中和目标的关键举措<sup>[1]</sup>。随着风能、光伏等新能源大规模高比例并网,新型电力系统对调节电源的需求日益凸显。基于我国能源资源禀赋,发展抽水蓄能电站可有效缓解风能和光伏等新能源出力不稳定的问题,对保障电力系统稳定运行具有战略意义<sup>[2]</sup>。2021年9月,国家能源局发布《抽水蓄能中长期发展规划(2021—2035)》,提出构建满足新能源高比例大规模发展需求的现代化抽水蓄能产业<sup>[3]</sup>。2022年3月,国家发改委和能源局联合印发《“十四五”现代能源体系规划》,着重强调要加快推进抽水蓄能电站建设,全面落实全国新一轮抽水蓄能中长期发展规划<sup>[4]</sup>。随着抽水蓄能

电站的大力建设,对具备专业知识和技术的运维人员需求更加迫切。抽水蓄能电站系统复杂且技术密集,运维人员需系统掌握电机学、水力学、电力系统等基础知识,熟练运用电力安全工器具及应急救援技能,深入理解设备运行、维护、安全生产管理全流程,精准执行调度操作与异常事故应急处置,全面掌握电气、机械及辅机设备全生命周期运维技术。在进入抽水蓄能电站工作前,运维人员需经过职业资格培训和技能鉴定,取得相应资格证书后方可从事运维工作<sup>[5]</sup>。由于抽水蓄能电站的运行工况多样、设备结构复杂、枢纽布置独特<sup>[6]</sup>,运维人员需根据业务流程及现场情况做出决策,这些决策高度依赖于运维人员的知识和经验,直接关系到电站

收稿日期/Received: 2025-01-02

修回日期/Revised: 2025-07-08

**基金项目:**国家自然科学基金项目(52079053);河南省中央引导地方科技发展资金项目(Z20241471035);河南省重点研发专项(241111210300);河南省自然科学基金项目(252300420056);河南省高等学校重点科研项目(25B520056)。

**第一作者:**胡昊(1979—),男,教授,博导,博士,从事智慧水利、清洁能源多能互补等方面的研究。E-mail: hh@ncwu.edu.cn。

**通信作者:**崔争艳(1980—),女,讲师,博士,从事自然语言处理、大语言模型等方面的研究。E-mail: cuizy1017@gmail.com。

的安全稳定运行。

目前,抽水蓄能电站的运维研究主要集中在两个方面:一是电站设备数字化智能化,二是电站运维管理。在电站设备数字化智能化方面,通常利用数字孪生技术<sup>[7]</sup>,通过融合水电机组设备、变电设备、水工建筑物等物理实体与数字模型,实现电站的协同运维,为运营管理提供决策支持<sup>[8]</sup>。此外,通过云计算、大数据、物联网等技术,构建以信息运维调度为核心的智慧运维平台,实现自动化智能化运维、本地化远程化维护、全方位监控等功能,智慧运维平台还需构建运维知识图谱,管理运维知识、运维任务、预测模型、故障模型等<sup>[9]</sup>。在运维管理方面,主要研究运行管理模式及运维一体化实践,涵盖生产准备、调试及试运行、人员编制、值班方式等<sup>[10]</sup>。运维一体化以运维班为单位,轮换人员定期参与值守、操作等运维业务,需要大量复合型高素质人才,对运维人员的专业知识、业务知识和岗位技能要求极高<sup>[11]</sup>。当前,运维人员主要通过理论技能培训学习运维业务知识,获取知识效率较低,同时在工作过程中需要查询各种文件资料,工作效率较低。随着生成式人工智能技术的发展,以 ChatGPT 为代表的大语言模型在各个领域的应用潜力凸显。在水利水电行业,通过部署智慧电站领域的垂直大模型,可实现工程建设、生产运维、流域调度与管理决策、安全培训等场景应用<sup>[12]</sup>。在电力行业,通过将大语言模型与电力设备结合,可实现电力设备综合状态评估、面向设备运维的高度智能化人机交互等应用<sup>[13]</sup>。大语言模型时代智能运维的核心在于运维人员与多种智能体之间的交互,通过人机协同,运维人员可以更高效地处理各种运维业务问题,保障运维工作的安全可靠进行<sup>[14]</sup>。

为了满足抽水蓄能电站运维的需求,提高运维人员业务水平,提高运维效率、降低运维成本,同时满足国家对水利新质生产力<sup>[15]</sup>的要求,本文提出基于大语言模型的抽水蓄能电站智能运维辅助系统(Pumped Storage-Artificial Intelligence Operation and Maintenance, PS-AIOM),构建了一个涵盖知识问答、理论知识、实操技能、典型案例、标准规范的运维专业知识库。为了解决大语言模型的幻觉问题,本文采用提示工程(Prompt Engineering)、检索增强生成(Retrieval-Augmented Generation, RAG)技术,以本地部署的大语言模型为底座,通过检索知识库,提高了大语言模型在抽水蓄能电站运维领域应用的准确性。为了验证系统的有效性,本文构建

了一个包含运维业务知识的测试数据集,其中包含专家给出的标准答案,通过综合关键词得分和语义相似性得分对本系统的回答进行评分。

## 1 相关技术方法

### 1.1 大语言模型及微调

人工智能的迅速发展正在形成技术奇点,将为人类社会带来深远影响和无限可能<sup>[16]</sup>,而大语言模型<sup>[17]</sup>作为近年来人工智能技术的重大突破,开启了人工智能应用新时代,引领人工智能技术创新和应用。大语言模型以深度学习和自然语言处理为基础,采用 Transformer 架构<sup>[18]</sup>,通过在大规模无标注文本语料库上自监督学习,获得预训练语言模型<sup>[19]</sup>。预训练语言模型虽然包含了大量的文本知识,能够生成文本补全句子,但缺乏指令理解和上下文理解能力。因此,通常采用有监督微调的方式,在预训练模型的基础上,利用少量高质量的数据集进一步训练得到通用大语言模型。目前开源的 ChatGLM3-6B、Qwen-7B、Llama3-8B、Mistral-7B 等模型均是在预训练模型基础上经过一定指令微调得到的通用大语言模型。通用大语言模型尽管涵盖广泛的知识,但由于缺乏精准的领域知识,无法满足特定应用场景的需求,容易产生幻觉现象<sup>[20]</sup>。为解决大语言模型在垂直领域应用时的不准确性,通常采用的方法有微调、提示工程和检索增强生成等。

相较于预训练模型,微调使用的训练语料和消耗的算力较少。微调是调整大语言模型的参数,针对不同业务场景进行优化。根据参数调整的范围,可分为全量微调和高效微调,全量微调需要调整模型的所有参数,而高效微调仅调整少量参数,使模型能够快速适应具体业务场景。高效微调的方法包括适配器微调(Adapter Tuning)、前缀微调(Prefix Tuning)和低秩适配器微调(Low-Rank Adaptation Tuning, LoRA),其中 LoRA 微调方法在训练参数相近的情况下性能较优,可以在较少的训练参数下使模型达到与全量微调相当的性能<sup>[21]</sup>。

### 1.2 提示工程技术

提示工程通过结构化设计输入指令,定向引导大语言模型输出符合预期的结果,其核心在于语义空间映射与知识路径引导。大语言模型在经过微调后已具备一定的知识,对于简单问题,直接输入问题即可得到相应回答。但对于复杂问题,需要结合多个知识点,按照特定逻辑回答。此时,通过设

计提示词引导大语言模型思考尤为重要。提示词包括执行指令、上下文信息、输入内容、输出格式,提示越具体,大语言模型生成的结果越准确。提示工程技术通过模拟人类思考的过程将问题分解,将解决问题的方法或步骤写入提示词,使大语言模型像人类一样分步骤思考和生成回答。常用的方法包括思维链(Chain-of-Thought, CoT)、自我一致性思维链(Self-Consistency with CoT, CoT-SC)、思维树(Tree of Thought, ToT)<sup>[22]</sup>等。提示工程不仅需要提示技巧,还需要熟悉业务或场景,提示词需要不断地迭代和测试。提示工程是优化大语言模型性能的一种非常有效的方法<sup>[23]</sup>,相比于微调,它侧重挖掘大语言模型已学习的知识,而微调是通过改变模型参数使其学习新的知识。

### 1.3 RAG 技术

检索增强生成(RAG)<sup>[24]</sup>是通过向大语言模型提供外部知识,结合信息检索技术和大语言模型的思维推理能力,对检索到的信息进行总结生成,使模型能够基于外部知识生成更加准确的回答。RAG 技术相当于给大语言模型配置了一本“工具书”,当模型遇到“未学过的”知识时,可以通过检索“工具书”生成回答,避免根据自身知识生成错误回答。

微调和 RAG 是大语言模型在垂直领域应用之前对领域知识的处理。微调通过领域知识训练改变模型参数,使模型“记住”领域知识;RAG 则通过切分领域知识构造向量知识库,使模型“使用”领域知识。提示工程是大语言模型在垂直领域应用时对领域知识的挖掘,它不改变模型参数和领域知识,通过构造具体详细的提示引导模型思考,结合自身知识和外部知识回答问题。微调需要改变模型参数,需要一定的算力支持;RAG 需要构建详细的知识库,需领域专家对知识结构进行分层分布式的划分;提示工程需要领域专家根据经验设计提示模板。在大语言模型应用于垂直领域时,考虑到微调成本和领域文档的私密性,采用提示工程和 RAG 技术是提高模型准确性和可信度的最简单方法,特别适用于知识密集型任务,且便于领域知识的持续更新<sup>[25]</sup>。

## 2 抽水蓄能电站智能运维辅助系统

本文主要研究大语言模型在抽水蓄能电站运维领域的应用。在预训练阶段,由于缺乏高精度的垂直领域知识,通用大语言模型在具体业务场景应

用时存在准确率不高、易产生幻觉等问题。为满足垂直领域业务场景的需要,通过本地部署通用大语言模型,采用外挂领域知识库和构造提示工程相结合的方式,搭建抽水蓄能电站智能运维辅助系统,可更精准地回答运维领域业务问题,辅助运维人员决策,提高工作效率。

### 2.1 系统描述

运维人员在巡视、操作、维护、故障或异常处理过程中,当需要根据现场情况做出决策时,向系统提出问题  $Q$ ,检索模块  $R$  从知识库  $K$  中检索到与问题最相关的前  $k$  个文本片段  $D = \{d_1, d_2, \dots, d_k\}$ ,将问题  $Q$  和文本片段  $D$  通过提示模板  $P$  输入大语言模型 LLM 生成回答  $A$ ,运维人员根据回答  $A$  进行相关操作。系统工作流程可以表示为:

$$D = R(Q, K), \quad (1)$$

$$A = \text{LLM}(P(Q, D)). \quad (2)$$

通过本系统辅助决策,可以提高运维人员的业务水平 and 运维工作质量,保障运维工作的安全进行。

### 2.2 系统框架

本系统主要由 5 个模块组成:知识库构建模块、提示模板构建模块、检索模块、大语言模型生成模块、人机交互模块。系统总体框架如图 1 所示。

1) 知识库构建模块。其主要功能是:根据抽水蓄能电站运维业务分别构建知识问答库、理论知识库、实操技能库、典型案例库、标准规范库 5 个一级知识库,根据专业、设备等知识类别在一级知识库中细分二级知识库。收集抽水蓄能电站运维相关文档资料,分别进行处理后生成向量数据库,按照对应的类别放入二级知识库中。文档资料的处理包括文本切分、向量化存储,模块根据文档结构以及预设的最大文本段长度对文档资料进行切分,选择词嵌入模型进行向量化后存储。

2) 提示模板构建模块。其主要功能是:根据任务要求构建提示模板,用于嵌入检索到的文本片段,引导大语言模型生成回答。提示模板包括任务描述及要求,模块可以根据任务类型分别构建提示模板。

3) 检索模块。其主要功能是:根据业务问题查询相关向量数据库,将业务问题向量化后分别在对应向量数据库中进行检索匹配,得到每个向量数据库中最相似的  $k$  个文本片段,再将所有文本片段拼接起来作为检索结果嵌入到提示模板中对应的位置。模块可以根据设置的  $k$  值选取向量库中检索到的文本片段数量。

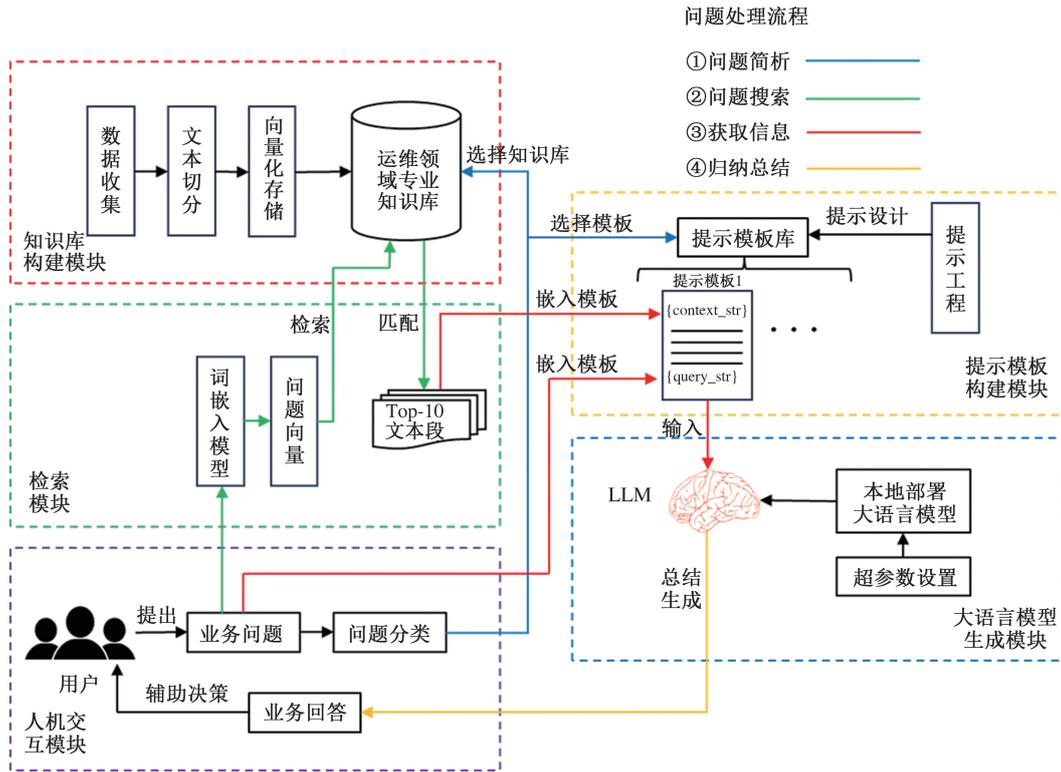


图 1 系统总体框架图

Fig. 1 Overall system framework diagram

4) 大语言模型生成模块。其主要功能是根据提示模板中的检索结果以及任务要求,利用大语言模型进行总结生成。模块包括本地大语言模型部署以及超参数设置。超参数包括模型最大字符数、温度以及采样策略。

5) 人机交互模块。其主要功能:运维人员向系统提问后,根据业务问题选择需要查询的向量数据库,经检索和大语言模型生成后得到业务问题的回答。模块可以手动选择知识库或通过大语言模

型对业务问题进行意图识别自动选择知识库。

### 2.3 系统实现

#### 2.3.1 运维知识库构建

知识库构建包括数据收集、文本切分、向量化存储,构建流程如图 2 所示。

1) 数据收集。收集抽水蓄能电站运维领域专业知识,将专业知识划分为知识问答、理论知识、实操技能、典型案例、标准规范 5 类。知识库具体分类见表 1。

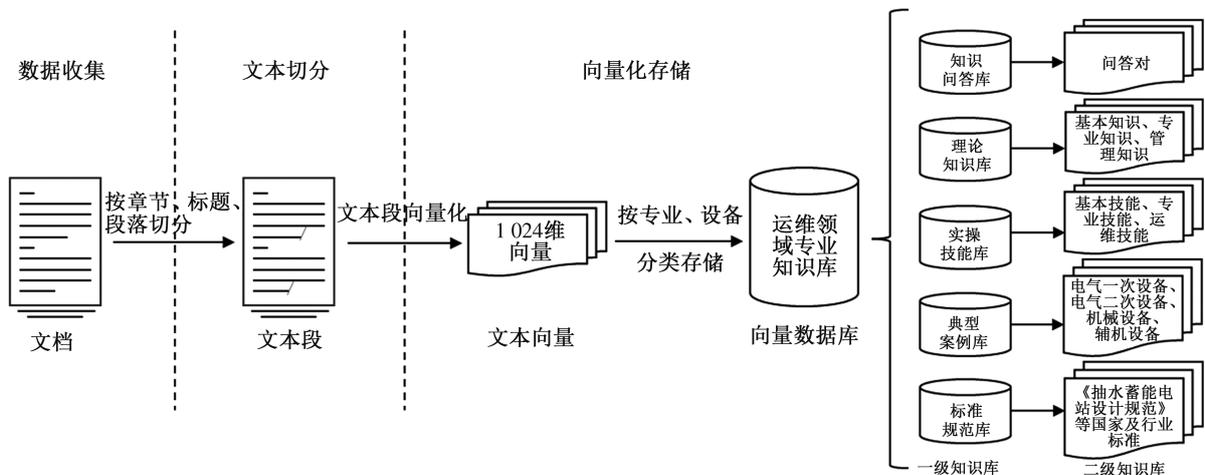


图 2 知识库构建流程

Fig. 2 Knowledge base construction process

表1 知识库划分类别

Tab.1 Classification categories of the knowledge base

类别	知识范围	知识类型
知识问答	运维工作中的常见问题	问答对
理论知识	抽水蓄能电站工作原理、运行方式、水泵水轮机、发电电动机、调速系统、主进水阀、主变压器等设备与设施、设备运维管理及安全管理	自然文本
实操技能	电力安全工器具使用、典型设备操作、调度业务联系、异常及事故应急处理与救护、电气机械辅机设备的运行维护技能	自然文本
典型案例	主变压器、GIS设备、母线、励磁系统、SFC系统、发电电动机本体、轴承、水轮水泵机本体、导水机构、启闭机、闸门本体等设备故障案例	自然文本
标准规范	《抽水蓄能电站设计规范》《国家电网有限公司水电厂重大反事故措施》《国家电网公司十八项电网重大反事故措施》等国家及行业标准规范	自然文本

2) 文本切分。采用 LangChain 框架中的文本加载器 Document\_Loaders 上传各类型数据,采用基于递归的文本分割器 TextSplitter 进行文本分割,分割时按照文本章节、标题、段落构造正则表达式,同时保证文本段长度不超过 1 024 个字符数。由于是按照文本本身语义结构切分,因此文本段前后的重复部分字符数设置为 0。

3) 向量化存储。采用词嵌入模型 Bge-Large-ZH-V1.5 将切分后的文本段进行向量化得到 1 024 维向量。采用 LangChain 框架中的 VectorstoreIndexCreator 将所有文本段向量按段构造索引存入向量数据库 Chroma。

按照以上流程,根据知识类别分别构造不同的向量数据库,使得在检索时能够更精确更快速找到相应的运维知识。

### 2.3.2 提示模板构建

为了使本系统能够严格按照知识库中内容进行回答,需要合理设计提示词模板。本系统提示模板库中的提示词模板示例如图 3 所示,可通过提示模板构建模块对其内容进行设计。

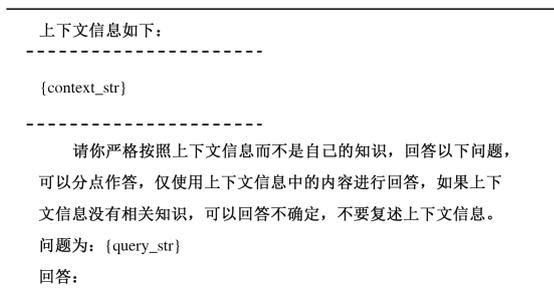


图3 提示词模板示例

Fig.3 Prompt template example

模板中 context\_str 为根据问题检索到的相似文本片段,query\_str 为输入系统的业务问题。

### 2.3.3 RAG 系统搭建

本系统中 RAG 的实现流程包括业务问题分类、业务问题向量化、选择相应的向量数据库检索、将检索结果构建提示、大语言模型总结生成 5 个部分,系统流程如图 4 所示。

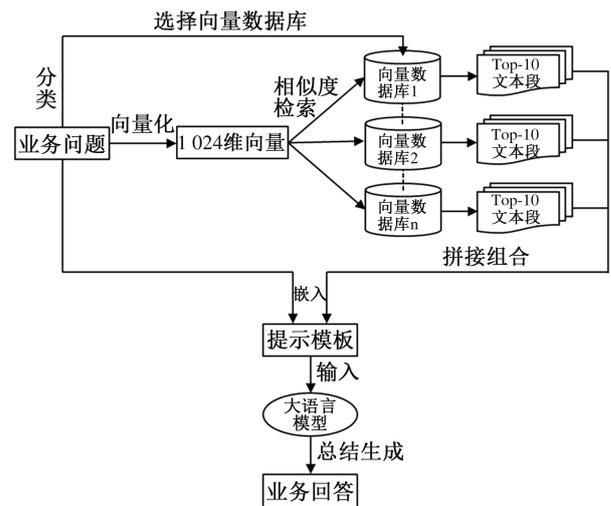


图4 RAG 系统流程

Fig.4 RAG system process flow diagram

1) 业务问题分类。为了更高效精准地选择向量数据库,需要将业务问题按照构建向量数据库时的类别进行划分,采用主动选择和大语言模型识别两种方式。当运维人员知晓是哪一类问题时,自主选择相应的向量数据库;当运维人员不清楚是哪一类问题时,通过大语言模型提取业务问题中的实体与所有类别进行相似度匹配后选择最有可能的向量数据库。

2) 业务问题向量化。采用向量化存储时使用的词嵌入模型将业务问题进行向量化,得到 1 024 维向量。

3) 向量库检索。将业务问题向量与向量数据库中的所有文本段向量分别进行余弦相似度计算,

将计算结果按照从大到小的顺序排列,选取前 10 个最相似的文本片段作为检索结果。

4) 构建提示。将检索到的文本片段以及业务问题替换提示模板中的 context\_str 和 query\_str, 构成完整的提示词作为大语言模型的输入。

5) 总结生成。本系统采用本地部署的 Chat-GLM3-6B 开源模型,选用具有增强长文本理解能力的 128K 预训练模型文件。将包含检索结果和业务问题的提示词输入大语言模型进行总结,生成来源于知识库的业务解决方案。

### 3 试验验证

本试验的环境配置为 Windows10 操作系统,处理器为 Intel(R) Core(TM) i9-11900 @ 2.50 GHz, 内存为 64 GB, 显卡为 NVIDIA Geforce RTX 3090 24G, Python 版本为 3.10, 深度学习框架为 PyTorch 2.2.1+cu121。

#### 3.1 数据集

根据抽水蓄能电站系统的功能结构,构建了一个包含运维知识的测试数据集。测试数据集的具体类别见表 2。

表 2 测试数据集类别

Tab.2 Categories of the test dataset

类别	数量/个	类别	数量/个
基本知识	20	主变压器	6
水泵水轮机	6	发电机出口设备	3
调速系统	3	静止变频器	6
主进水阀	5	机组机械保护	5
油气水系统	14	事故应急处理	2
金属结构	5	抽水蓄能电站规范	20
升压站设备	5	总计	100

#### 3.2 参数设置

本系统运维业务问题的回答严格按照知识库中的知识,仅利用大语言模型的推理能力对检索到的知识片段进行总结生成,不使用大语言模型本身的知识。因此,大语言模型的超参数设置最大字符数为 8 192,温度为 0,采样策略为 0.7。

#### 3.3 评价指标

本系统的评价指标综合考虑生成结果的关键词得分 Score\_k 和语义相似性得分 Score\_s, 满分均为 1 分。评价标准为抽水蓄能电站运维专家对 100 道测试问题进行人工作答给出的标准答案 Answer\_st, 标准答案中给出体现答案主旨的关键词 Answer\_k1、...、Answer\_kn。

假设运维人员提出业务问题 Q, 本系统生成回

答 A, 将语义相似性得分和关键词得分按 4:6 的权重加权平均, 计算系统的综合得分。

$$\text{Score}(A, Q) = 0.4 \times \text{Score}_s(A, Q) + 0.6 \times \text{Score}_k(A, Q), \quad (3)$$

$$\text{Score}_s(A, Q) = f(A, \text{Answer}_{st}), \quad (4)$$

$$\text{Score}_k(A, Q) = f(A, \text{Answer}_{ki}) / (n_k). \quad (5)$$

式中:  $f(A, \text{Answer}_{st})$  是通过将 A 和 Answer\_st 分别进行词嵌入向量化后计算两个向量的余弦相似性;  $f(A, \text{Answer}_{ki})$  是关键词 Answer\_ki 在回答 A 中出现的总次数;  $n_k$  是标准答案中关键词的总数。通过这种加权方式,既考虑了回答内容的具体匹配程度,也考虑了整体的文本相似性,实现了较为全面的评估。

#### 3.4 实验结果及分析

采用本系统对测试数据集中的 100 个问题进行回答, 回答结果的各项得分频数分布如图 5 所示。

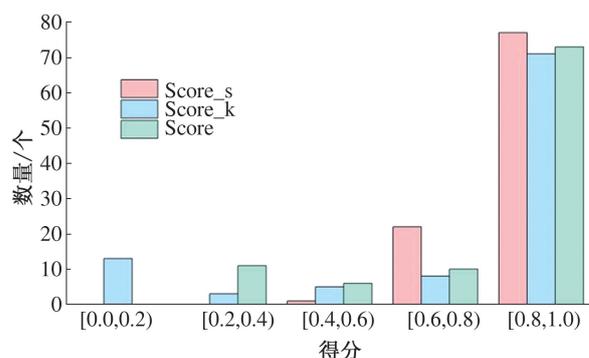


图 5 本系统得分频数分布

Fig.5 Score frequency distribution of the system

由图 5 可以看出,本系统语义相似性得分有 99 个问题回答在 0.6 以上,其中 77 个在 0.8 以上;关键词得分有 79 个在 0.6 以上,其中 71 个在 0.8 以上;综合得分有 83 个在 0.6 以上,其中 73 个在 0.8 以上。由于语言文字的多样性,即使是同一个问题不同的专家也有不同的回答,因此如果按照 0.6 分满意的标准,本系统对运维业务的满意率是 83%,能够辅助运维人员日常工作。

为进一步验证本系统对业务问题回答的效果,采用相同的评价指标对通用大语言模型 Z、T 就同样问题的回答进行评分,得分频数分布分别如图 6 和图 7 所示。由图 6 和图 7 可以看出,由于模型 Z、T 在模型训练或微调阶段缺乏细粒度的专业知识,使得在垂直领域应用时缺乏一定的准确性。在语义相似性方面,模型 Z、T 对 97% 的问题回答得分在 0.6 以上,模型 T 的回答得分在 0.8 以上的数量较模型 Z 多 65%;在关键词方面,模型 Z 对 97% 的问

题回答得分在 0.6 以下,模型 T 对 87% 的问题回答得分在 0.6 以下,它们对关键词的回答效果都不太好。综合考虑语义相似性得分和关键词得分,模型 Z 对运维业务的回答满意率是 4%,模型 T 对运维业务的回答满意率是 17%。

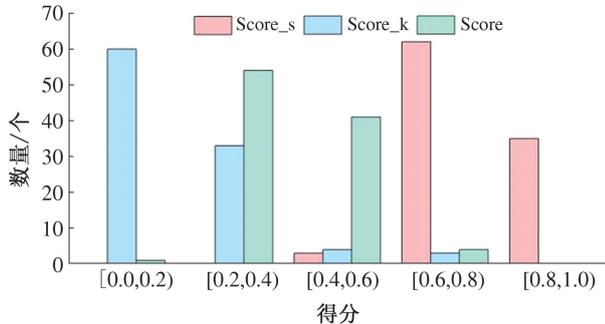


图 6 模型 Z 得分频数分布

Fig. 6 Score frequency distribution of model Z

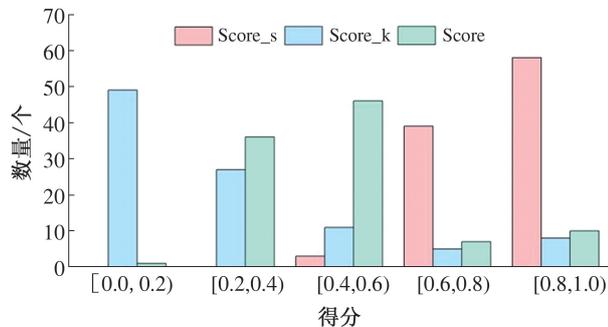


图 7 模型 T 得分频数分布

Fig. 7 Score frequency distribution of model T

本系统 (PS-AIOM) 和模型 Z、T 在语义相似性得分、关键词得分、综合得分的平均分见表 3。

表 3 与最新大语言模型的得分对比

Tab. 3 Score comparison with latest LLM

模型	Ave(Score_s)	Ave(Score_k)	Ave(Score)
模型 Z	0.759 9	0.157 1	0.398 2
模型 T	0.801 4	0.261 6	0.477 5
PS-AIOM	0.849 8	0.789 9	0.813 8

注: Ave(·) 为求平均值函数。

本系统中采用的大语言模型为本地部署的 ChatGLM3-6B 开源模型,相较于千亿参数级模型 Z、T,该模型的参数规模较为有限,因此对于通用知识回答的准确性显著低于最新大语言模型。此外,由于实验设备显卡资源有限以及专业数据集构建工作量较大,本系统未采用微调方式对大模型进行训练。本系统较最新大语言模型在抽水蓄能电站运维领域回答更准确的原因主要是采用了提示工程技术和 RAG 技术,通过构建一个较为全面的运

维知识库和引导大语言模型输出,实现了抽水蓄能电站运维领域较为准确的应用。RAG 的知识库可以在后续运维过程中不断维护改进,根据实际生产需求添加或删除知识库知识,在应用时成本更低。

本系统综合得分平均分为 0.813 8,未达到理想精度主要归因于系统构建与评价两方面的局限性。在系统构建方面,可能是知识库不够全面、检索系统不够准确、本地部署的 ChatGLM3 逻辑推理能力较弱;在系统评价方面,由于自然语言的多样性,同一句话的表述也不一定完全相同,使得语义相似性得分很小几率得到满分。

## 4 结语

为更好地满足国家对水利电力新质生产力的要求,提高抽蓄运维人员业务水平和运维效率,本文主要研究了大语言模型在抽水蓄能电站运维领域的应用。针对通用大语言模型在具体业务场景应用时存在准确率不高、易产生幻觉等问题,提出了基于大语言模型的抽水蓄能电站智能运维辅助系统 PS-AIOM,得出如下结论:

1) 基于抽水蓄能运维国家及行业标准规范、运维业务培训书籍及技术文档等资料构建了运维知识库,将知识库划分为知识问答、理论知识、实操技能、典型案例、标准规范 5 类,使检索范围更精确,为运维领域知识工程提供了数据支持。

2) 通过本地部署大语言模型和外挂运维专业知识库,实现了大语言模型在抽水蓄能运维领域的智能应用。相比先进的通用大语言模型,PS-AIOM 在本地部署参数量仅有 6 B 的大语言模型的基础上,实现了运维垂直领域较为准确的智能应用,同时保证了本地知识的私有性和可维护性。

3) 后续研究将继续优化知识库,使知识范围更全面、知识存储更合理;优化检索流程,使检索更全面、检索结果更准确;优化大模型底座,通过本地部署参数量更大的大语言模型增强对检索结果的总结归纳能力。

## 参 考 文 献

- [1] 张国兴,冯祎琛,聂龔. 能源电力系统绿色转型:政策演进逻辑与实践进路[J]. 兰州大学学报(社会科学版), 2024,52(3):96-108. [ZHANG G X, FENG Y C, NIE Y. Green transformation of energy and power systems: policy evolution logic and practical approaches[J]. Journal of Lanzhou University (Social Sciences), 2024,52(3):96-108.]
- [2] 张静,赵树鑫,宋小洪. 我国抽水蓄能现代化的发展历程、

- 挑战与展望[J]. 中国农村水利水电, 2024(6): 239-243, 251. [ZHANG J, ZHAO F X, SONG X H. The development process, challenges and prospects of china's pump storage modernization[J]. China Rural Water and Hydropower, 2024(6): 239-243, 251. ]
- [3] 国家能源局. 抽水蓄能中长期发展规划(2021—2035年)[EB/OL]. (2021-09-17)[2025-02-16]. [https://zfxxgk.nea.gov.cn/2021-09/17/c\\_1310193456.htm](https://zfxxgk.nea.gov.cn/2021-09/17/c_1310193456.htm). [National Energy Administration. Medium- and long-term development plan for pumped storage (2021-2035)[EB/OL]. (2021-09-17)[2025-02-16]. [https://zfxxgk.nea.gov.cn/2021-09/17/c\\_1310193456.htm](https://zfxxgk.nea.gov.cn/2021-09/17/c_1310193456.htm). ]
- [4] 国家发展改革委员会, 国家能源局. “十四五”现代能源体系规划[EB/OL]. (2022-01-29)[2025-02-16]. [https://www.gov.cn/zhengce/zhengceku/2022-03/23/content\\_5680759.htm](https://www.gov.cn/zhengce/zhengceku/2022-03/23/content_5680759.htm). [National Development and Reform Commission, National Energy Administration. The 14th five-year plan for building a modern energy system[EB/OL]. (2022-01-29)[2025-02-16]. [https://www.gov.cn/zhengce/zhengceku/2022-03/23/content\\_5680759.htm](https://www.gov.cn/zhengce/zhengceku/2022-03/23/content_5680759.htm). ]
- [5] 万正喜, 徐辉, 陈珏, 等. 抽水蓄能电站生产准备现场运维培训实践[C]//中国水力发电工程学会电网调峰与抽水蓄能专业委员会. 抽水蓄能电站工程建设文集 2021. 北京: 中国水利水电出版社, 2021: 483-487. [WAN Z X, XU H, CHEN J, et al. Field operation and maintenance training practice for production readiness of pumped storage power stations[C]//Committee on Grid Peak Shaving and Pumped Storage of the Chinese Society for Hydropower Engineering. Proceedings of the 2021 Conference on Pumped Storage Power Station Engineering. Beijing: China Water & Power Press, 2021: 483-487. ]
- [6] 孙宏. 抽水蓄能电站运行管理实践及运维一体化[J]. 电工材料, 2024(1): 54-56. [SUN H. Integration of operation management and maintenance of pumped storage power stations[J]. Electrical Engineering Materials, 2024(1): 54-56. ]
- [7] 冶涛涛, 蒋云钟, 曹引, 等. 以数字孪生水利为核心的智慧水利标准体系研究[J]. 华北水利水电大学学报(自然科学版), 2023, 44(4): 1-16. [YE Y T, JIANG Y Z, CAO Y, et al. Research on standard system of smart water conservancy with digital twin water conservancy as the core[J]. Journal of North China University of Water Resources and Electric Power (Natural Science Edition), 2023, 44(4): 1-16. ]
- [8] 叶宏, 孙勇, 阎峻, 等. 数字孪生智能抽水蓄能电站研究及其检修应用[J]. 水电能源科学, 2022, 40(6): 201-206. [YE H, SUN Y, YAN J, et al. Research of digital twin intelligent pumped storage station and application of maintenance[J]. Water Resources and Power, 2022, 40(6): 201-206. ]
- [9] 阎峻, 陈洪春, 曹新立, 等. 智慧运维支撑平台对抽水蓄能电站的一体化运维的支撑研究[C]//中国通信学会能源互联网委员会, 中国电力科学研究院有限公司. 2020年中国通信学会能源互联网学术报告论文集. [S. l. : s. n. ], 2020: 349-351. [YAN J, CHEN H C, CAO X L, et al. Research on the support of intelligent operation and maintenance platform for integrated operation and maintenance of pumped storage power stations[C]//Committee on Energy Internet of China Institute of Communications, China Electric Power Research Institute Co., Ltd. Proceedings of the 2020 China Institute of Communications Energy Internet Academic Symposium. [S. l. : s. n. ] 2020: 349-351. ]
- [10] 李翰麟, 林礼清. 抽水蓄能电站运行管理及运维一体化[J]. 福建水力发电, 2018(2): 24-25, 28. [LI H L, LIN L Q. Integration of operation management and maintenance of pumped storage power stations[J]. Fujian Hydropower, 2018(2): 24-25, 28. ]
- [11] 费万堂, 赵利军, 矫镛达, 等. 河北丰宁大型抽水蓄能电站运维模式探讨[J]. 水电与抽水蓄能, 2020, 6(6): 1-4. [FEI W T, ZHAO L J, JIAO R D, et al. Discussion on the operation and maintenance mode of Hebei Fengning large pumped storage power station[J]. Hydropower and Pumped Storage, 2020, 6(6): 1-4. ]
- [12] 金和平, 李德龙, 罗惠恒. AIGC 技术在水电行业的应用前景探索[J]. 水电与抽水蓄能, 2023, 9(6): 6-14, 19. [JIN H P, LI D L, LUO H H. Exploring the application prospects of artificial intelligence generated content technology in the hydropower industry[J]. Hydropower and Pumped Storage, 2023, 9(6): 6-14, 19. ]
- [13] 江秀臣, 臧奕茗, 刘亚东, 等. 电力设备 ChatGPT 类模式与关键技术[J]. 高电压技术, 2023, 49(10): 4033-4045. [JIANG X C, ZANG Y M, LIU Y D, et al. Power equipment ChatGPT-type model and key technologies[J]. High Voltage Engineering, 2023, 49(10): 4033-4045. ]
- [14] 裴丹, 张圣林, 孙永谦, 等. 大语言模型时代的智能运维[J]. 中兴通讯技术, 2024, 30(2): 56-62. [PEI D, ZHANG S L, SUN Y Q, et al. Artificial intelligence for IT operations in era of large language model[J]. ZTE Technology Journal, 2024, 30(2): 56-62. ]
- [15] 左其亭, 秦西, 马军霞. 水利新质生产力: 内涵解读、理论框架与实施路径[J]. 华北水利水电大学学报(自然科学版), 2024, 45(3): 1-8. [ZUO Q T, QIN X, MA J X. New quality productivity of water conservancy: connotation interpretation, theoretical framework and implementation path[J]. Journal of North China University of Water Resources and Electric Power (Natural Science Edition), 2024, 45(3): 1-8. ]
- [16] 陈龙, 刘刚, 戚聿东, 等. 人工智能技术革命: 演进、影响和应对[J]. 国际经济评论, 2024(3): 9-51. [CHEN L, LIU G, QI Y D, et al. AI technology revolution: evolution, impact and countermeasures[J]. International Economic Review, 2024(3): 9-51. ]
- [17] KUMAR P. Large language models (LLMs): survey, technical frameworks, and future challenges[J]. Artificial Intelligence Review, 2024, 57(9): 260.
- [18] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31th International Conference on Neural Information Processing Systems. California: Curran Associates Inc., 2017: 6000-6010.
- [19] KALYAN K S, RAJASEKHARAN A, SANGEETHA S. AM-

- MU; a survey of transformer-based biomedical pretrained language models [J]. *Journal of Biomedical Informatics*, 2022, 126:103982.
- [20] 腾讯研究院. 2024 工业大模型应用报告 [R]. 北京: 腾讯研究院, 2024: 3-5, 11-14. [Tencent Research Institute. 2024 industrial application report of large models [R]. Beijing: Tencent Research Institute, 2024: 3-5, 11-14.]
- [21] SUN Z H, YANG H Q, LIU K, et al. Recent advances in LoRa; a comprehensive survey [J]. *ACM Transactions on Sensor Networks*, 2022, 18(4): 1-44.
- [22] YAO S Y, YU D, ZHAO J, et al. Tree of thoughts: deliberate problem solving with large language models [C]//Proceedings of the 37th International Conference on Neural Information Processing Systems. California: Curran Associates Inc., 2023: 11809-11822.
- [23] LIU P F, YUAN W Z, FU J L, et al. Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing [J]. *ACM Computing Surveys*, 2023, 55(9): 1-35.
- [24] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks [C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. California: Curran Associates Inc., 2020: 9459-9474.
- [25] SIRIWARDHANA S, WEERASEKERA R, WEN E, et al. Improving the domain adaptation of retrieval augmented generation (RAG) models for open domain question answering [J]. *Transactions of the Association for Computational Linguistics*, 2023, 11: 1-17.

## Research on Construction of Intelligent Operation and Maintenance Support System for Pumped Storage Power Stations Based on Large Language Model

HU Hao<sup>1,2,3</sup>, XU Zhaoyi<sup>2</sup>, CUI Zhengyan<sup>1,3</sup>, ZHANG Xingkui<sup>2</sup>, ZHANG Haoyu<sup>2</sup>

(1. Yellow River Conservancy Technical University, Kaifeng 475004, China; 2. North China University of Water Resources and Electric Power, Zhengzhou 450046, China; 3. Henan Engineering Research Center of Project Operation and Ecological Security for Inter-basin Regional Water Diversion Project, Kaifeng 475004, China)

**Abstract:** **[Objective]** To address the complexities of operation and maintenance (O&M) work at pumped storage power stations, the high demand for specialized knowledge, the inefficiency of traditional knowledge retrieval, and the difficulty of general large language models (LLMs) in accurately understanding the unique terminology and complex operating conditions specific to pumped storage O&M, this study aims to enhance the professional competence and O&M efficiency of personnel, improve knowledge reuse efficiency, and meet the national strategic requirements for developing new quality productivity in water conservancy and electric power systems by constructing an intelligent O&M assistance system tailored for pumped storage power stations. **[Methods]** The Pumped Storage-Artificial Intelligence Operation and Maintenance (PS-AIOM) system was constructed based on large language models. Utilizing a locally deployed LLM as the foundation, the system integrated retrieval-augmented generation (RAG) and prompt engineering techniques. Through external knowledge base retrieval, it enhanced the accuracy of LLM applications in pumped storage O&M. To ensure response accuracy for O&M tasks, a dedicated O&M knowledge base was constructed. To validate system effectiveness, a test dataset was created. **[Results]** (1) An evaluation method combining keyword matching and semantic similarity was adopted to score PS-AIOM's responses, considering both the specific matching of content and overall textual similarity for a comprehensive assessment. (2) PS-AIOM achieved a satisfaction rate of 83% for O&M queries, with a comprehensive average score of 0.8138, demonstrating its effectiveness in assisting O&M personnel. (3) PS-AIOM significantly outperformed advanced general LLMs (Model Z: 4% satisfaction rate, 0.3982 average score; Model T: 17% satisfaction rate, 0.4775 average score). **[Conclusion]** By integrating a locally deployed LLM base, RAG technology, and a domain-specific knowledge base, the PS-AIOM system effectively overcomes the limitations of general LLMs regarding professional domain knowledge scarcity. It achieves relatively accurate knowledge application within the O&M domain of pumped storage power stations.

**Keywords:** pumped storage power station; intelligent operation and maintenance; large language model; retrieval-augmented generation

(编辑:陈海涛)