文章编号：0253-374X(2026)01-0013-09

DOI：10.11908/j.issn.0253-374x.24303

基于协同专家系统的建筑施工大语言模型问答系统

杨　彬[1]，肖鸿儒[1]，高　尚[1]，雷　克[2]，陈文硕[3]，张其林[1]，汪丛军[4]

（1. 同济大学 土木工程学院，上海 200092；2. 中国建筑第八工程局有限公司，上海 200112；
3. 香港科技大学(广州)，广东 广州 511453；4. 中亿丰建设集团股份有限公司，江苏 苏州 215131）

摘要：为解决大型语言模型问答系统在建筑施工场景中存在的生成幻觉与部署成本高的问题，提出了一种基于协同专家机制的大型语言模型施工问答系统。该系统通过共享专家与路由专家的协同工作方式，在保证模型表达能力的同时，显著提升了问答生成的准确性与推理效率，并有效降低了计算开销。此外，设计了一种领域知识库注入的微调策略，在训练阶段引导模型深度学习施工领域专业语义，从而增强其对工程文本的理解能力，确保生成结果更加符合实际工程需求。实验结果表明，在仅激活约1/3模型参数的情况下，所提出系统仍可达到81.1%的生成语义相似度，兼顾了效率与性能，为建筑施工管理提供了一种高效、可靠且具备工程针对性的智能决策支持工具。

关键词：建筑施工；智能建造；问答系统；大语言模型；本地知识库

中图分类号：TP391　　**文献标志码**：A

A Construction Question and Answer System Based on a Collaborative Expert System and Large Language Model

YANG Bin[1], *XIAO Hongru*[1], *GAO Shang*[1], *LEI Ke*[2], *CHEN Wenshuo*[3], *ZHANG Qilin*[1], *WANG Congjun*[4]

(1. College of Civil Engineering, Tongji University, Shanghai 200092, China; 2. China Construction Eighth Engineering Division, Shanghai 200112, China; 3. The Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511453, China; 4. Zhongyifeng Construction Group Co., Ltd., Suzhou 215131, China)

Abstract: To address the issues of hallucinated generation and high deployment costs encountered by large language model (LLM) - based question answering systems in construction scenarios, this paper proposes a construction-oriented question answering system based on a collaborative expert mechanism. The system integrates shared experts and routing experts in a coordinated manner, which significantly improves the accuracy of answer generation and inference efficiency while preserving the model's expressive capacity and reducing computational overhead. In addition, a domain knowledge base - injected fine-tuning strategy is introduced to guide the model to deeply learn professional semantics in the construction domain during training, thereby enhancing its understanding of engineering-related texts and ensuring that the generated responses better align with practical engineering requirements. Experimental results demonstrate that, with only approximately one-third of the model parameters activated, the proposed system achieves a generation semantic similarity of 81.1%, effectively balancing efficiency and performance and providing an efficient, reliable, and construction-specific intelligent decision-support tool for construction management.

Keywords: building construction; intelligent construction; question and answer system; large language model; local knowledge base

建筑施工覆盖场景广泛、工程需求多变，也正由于其动态的非结构化工作场景，增加了项目规范化的难度[1]，阻碍了施工项目的进展。此外，随着施工过程日益复杂，组织协调与工法技术的日趋多样，项目管理中的信息利用率低下，特别是在应对多维度的文档问询和施工场景的动态变化时，显得尤为不足。这种信息障碍严重影响了施工项目的进展效率与科学性[2]。

基于上述问题，部分学者尝试引入自然语言处理技术获取和分析建筑工程非结构化文本，具体体现在事故分析[3]、安全管理[4]、合规性检查[5]等方面。

收稿日期：2024-09-18
基金项目：国家重点研发计划(2022YFC3801702)
第一作者：杨　彬，教授，博士生导师，工学博士，主要研究方向为智能建造。E-mail：yangbin@tongji.edu.cn
通信作者：肖鸿儒，博士生，主要研究方向为智能建造。E-mail：hongru_xiao@tongji.edu.cn

论文拓展介绍

然而,以往的研究受限于语言模型能力,难以显式表达所存储的文本知识,从而不能满足施工场景多维度问询需求,文本知识利用率低下[6]。随着生成式大语言模型的出现[7-9],这类模型在存储和生成自然语言中的结构化知识上展现出了潜力。然而,现有的通用大语言模型在建筑施工领域的应用依然存在瓶颈。首先,通用模型在训练过程中未能充分引入与建筑施工相关的数据和语料,导致在施工领域的问答任务中幻觉现象严重,模型无法准确理解和回答现场问题;其次,目前的问答系统大多基于稠密网络进行微调适配,其庞大的计算开销和推理时间为施工现场使用带来困难[10];最后,基于本地知识库的检索增强策略通常仅在推理阶段添加参考信息,导致模型在训练和推理阶段的知识不一致,削弱了模型对本地知识库的理解和利用能力,限制了其在施工场景中提供精准问答的能力[11]。

为解决以上问题:①构建建筑施工领域本地知识库以及高质量专业问答数据集,提升语言模型在建筑施工领域中的生成能力;②提出基于协同专家系统的生成式大语言模型优化方案,能够在计算开销显著降低的前提下,快速输出更为准确的回答,并适应建筑施工场景中的复杂语义需求;③最后,提出知识注入训练策略,使得模型能够在训练阶段动态学习到本地知识库文档知识,进一步提升模型在实际施工场景中的应用能力。

本系统首次结合土木工程领域知识,从模型架构与训练策略2个维度优化大语言模型,并利用自建的建筑施工数据集进行了多任务验证,旨在实现对建筑施工工程场景的智能问答,为管理人员与专业学习者提供知识支持,以辅助提升施工管理效率。

1 研究背景

由OpenAI开发的生成式预训练变换器(Generative Pre-trained Transformer, GPT),经过海量的非结构化文本预训练,具备了自然语言的理解能力,能够根据学习到的知识流畅生成自然对话语言,为自然语言处理技术带来了巨大的革新[7]。在此基础上,大语言模型在建筑业中的应用也逐渐开展。Zheng和Fischer[12]使用GPT-3.5的API服务,应用于建筑信息模型(BIM)信息检索任务,能够应用自然语言检索BIM模型中的信息。You等[13]使用ChatGPT进行施工顺序规划,驱动机器人处理复杂的施工作业任务并适应环境动态变化。Prieto等[14]使用ChatGPT为施工项目生成计划进度表,并进行了专家主管评测,对生成结果进行主观评估。覃思中等[15]对大语言模型在建筑工程垂域中的应用进行了广泛的测试与探讨,探索了使用本地知识库对大语言模型的增强策略,并在多个建筑问答场景进行测试。丁志坤等[16]针对BIM正向设计,搭建了基于大语言模型高效指令微调的问答系统,能够为设计人员提供可靠的建议参考。

随着大语言模型的出现,其显性输出知识的生成与回复能力为建筑文本处理提供了新的解决方案。通过结合本地知识库和高效指令微调,模型在建筑场景中的幻觉问题得到了初步改善,从而能够对建筑领域问题进行相对可靠的回复,为建筑施工领域的大模型问答系统的搭建提供了技术参考。然而,现有的基于大语言模型问答系统,往往都基于稠密网络,难以兼顾生成准确性与高效响应能力[8]。并且目前的本地知识库增强策略往往只在推理阶段添加相关参考信息,使得模型训练与推理阶段提示不一致,降低了对本地知识库的理解利用能力[11]。

故选取阿里巴巴开源的Qwen1.5-1.8B-Chat[9]为基准模型,并通过一系列增强策略,进一步提升模型针对施工广泛问询与复杂需求的回复能力。

2 施工问答系统设计

本问答系统旨在为建筑施工场景提供可靠的决策知识建议,提高管理人员的信息检索与知识获取效率。考虑建筑施工场景特性,采用一种基于协同专家系统的大语言模型,并结合高效指令微调与本地知识库注入的训练方案,进一步解决大语言模型在建筑施工场景下的幻觉问题,同时提供更快的响应速度,满足多方使用需求。图1为系统总体框架图,本问答系统具体实施方案如下:①收集高质量施工文本数据,构造问答数据与本地施工知识库;②基于新型的路由专家、共享专家系统与协同相融模块,构建施工大语言问答系统;③结合向量相似匹配进行问询数据重组,并利用重组后的提示—问询与标注答案对模型进行LoRA[17]混合高效指令微调;④在推理阶段再次使用文本相似性匹配策略,从而在本地知识库中为相关问询进行知识补充提示,从而得到更加准确的回复。

2.1 施工垂域数据集制备

为使系统能够有效地应用于建筑施工领域,先需要收集高质量知识文档,从而构造本地知识库与

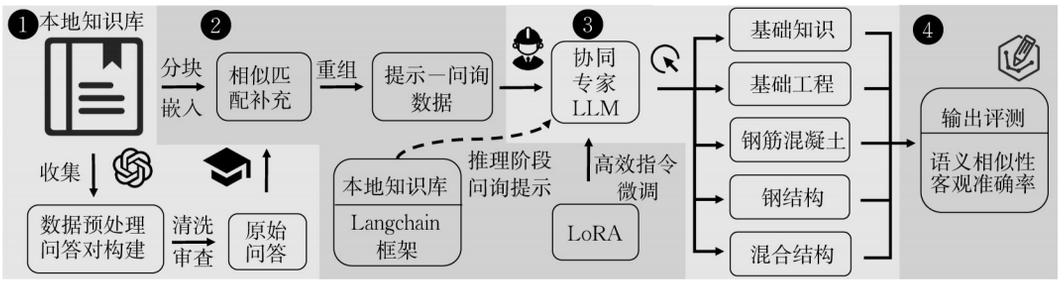


图 1　系统总体框架

Fig. 1　Overall framework of the system

施工专业问答数据。在数据选择中，通过收集建筑施工规范、手册，组织设计文档以及中建八局内部的施工手册技术文档，从中进行问答抽取与筛选。其次，对建筑施工与土木工程施工专业教学过程中的考核题目、问答过程记录、优秀案例解读以及相关建筑施工专栏中的问答信息进行收集，包括施工工艺解读、组织设计步骤、工艺类型选材等。

总体问答类别归纳为五大类，其中，施工基础知识涵盖了施工测量、施工机械与设备以及施工管理等知识；基础施工则包含了地基与桩基施工等知识；钢筋混凝土施工则包括了钢筋工程、混凝土工程、钢筋混凝土结构工程以及预应力工程等知识；混合结构施工则主要覆盖钢—混凝土组合结构工程施工知识。所收集的数据兼顾了行业应用案例与专业课程教学，体现了一线工程师与专业学生在工作与学习中所遇到的代表性问题，能够反映建筑施工领域切实所需的问题类型。除在上述问答数据采集中所涉及的文档资料外，还选择《建筑工程施工》[18]与《土木工程施工》[19]两本专业课教材作为本地知识库专业教材补充数据。为确保数据生成质量，对所获取的数据进行了严格的人工数据清洗，包括文本消歧、错别表达纠正以及转录格式审查等操作。

结合上述数据准备操作，总结出建筑施工领域原始问答 11 652 条，并遵循 alpaca 格式组合成最终指令微调数据。将数据按照 8∶2 的比例划分训练集与测试集，其中，问答数据类型如表 1 所示。

表 1　问答数据统计

Tab. 1　Statistics of question and answer

范围	问答数/条	平均问询子词长度/个	平均答案子词长度/个	问题样例
施工基础知识	2 413	24.2	208.3	钢筋混凝土屋架吊装需进行哪几个阶段的验算？并说明各阶段验算的要点。
基础施工	2 088	21.5	211.4	
钢筋混凝土施工	2 690	26.1	217.7	请叙述单层工业厂房中“分件吊装法”和“综合吊装法”的吊装流程和各自的优缺点？
钢结构施工	2 312	22.2	222.3	
混合结构施工	2 149	20.6	195.4	

2.2　知识注入策略

在利用领域知识库增强策略的研究中，以往的研究[12,15-16,20]选择在推理阶段为模型提供问询相关参考。然而，由于模型在训练阶段未能注入建筑施工领域的知识从而对本地知识库的理解存在不足，给模型的表现带来了隐患。因此，提出了一种知识注入策略。该策略首先通过问询知识相似性匹配策略，对原始问询数据进行知识补充，重构为包含本地知识的提示—问询数据；随后，执行高效指令参数微调操作，将提示中的知识注入到模型参数中，完成知识注入。该策略具体计算式为

$$\begin{aligned} r &= \mathcal{R}(q, K) \\ p &= f(q, r) \end{aligned} \tag{1}$$

式中：q 和 K 分别为原始问询文本与本地知识库；$\mathcal{R}$ 为文本知识相似性检索函数；r 为基于知识库检索得到的相似性参考信息文本语料；f 为提示词构建函数；p 为注入本地知识后重构的提示—问询文本语料。

在本地知识库构建过程中，需要根据问询的相关性，匹配知识库中对应的知识，采取以下策略进行问询知识匹配：①首先对知识库内容进行数据清洗与 markdown 形式格式化存储；②在相关性匹配时，对所存储的内容进行文本分块，以适应模型上下文窗口长度；③通过滑块交叉与文本消歧等策略，确保专业文档的语义连贯性与准确性；④最终，进行语义相似匹配，通过构建参考块嵌入、语义相似匹配过程，完成在本地知识库中检索相关文档知识块，将前

4块参考文本合并入问询,重构Prompt形成提示—问询提示数据,如图2所示。

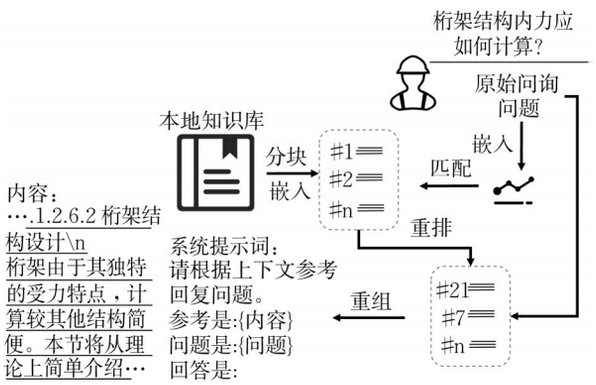


图2 提示—问询的构造流程

Fig. 2 Construction framework for prompt-query

最终,基于上述知识提示数据,通过执行低秩分解高效指令微调方法[17],实现知识注入策略,使模型参数在优化过程中加深对本地知识库的理解,从而进一步提升模型对建筑施工知识的掌握能力。

2.3 协同专家大语言问答系统构建

以GPT为代表的解码器架构生成式大语言模型,通过学习前文相关性预测下一子词的自回归训练方法,能够针对问询提供基于概率的最大可能正确的回复。相较于传统基于稠密网络的模型,基于稀疏网络的生成式语言模型能够带来更少的激活参数开销与更快的响应速度,且基于细粒度专家系统的生成式模型能够进一步提升模型回复的准确性[8,21-22]。基于此,为了更好地提升建筑施工场景广泛问询下的回复准确性,选用新型的路由、共享专家与协同相容模块,针对语言模型在施工场景下的准确回复能力进行增强。整体架构如图3所示,文本通过子词编码、词向量嵌入等模块进入循环解码器层,最终通过线性层与归一化预测出下一子词,计算如式(2):

$$y_{\text{output}}=\mathcal{D}(\mathcal{E}(x)+P)$$
$$y_{t+1}=s_{\text{softmax}}(\mathcal{L}(y_{\text{output}})) \tag{2}$$

式中:x为输入文本序列;$\mathcal{E}$为词向量嵌入函数(embedding);P为位置编码;$\mathcal{D}$为解码器(decoder)映射函数;y_{output}为解码器层的输出表征;$\mathcal{L}$为输出线性层(linear)映射函数;y_{t+1}为下一时刻预测子词的概率分布。

在解码器层中,使用稀疏感知块替代传统的多层感知机,增强稀疏参数调用能力,从而增强模型对范围广阔问询的回复能力,并进一步提升响应速度。在稀疏感知块中,包含路由专家层与共享专家层2个核心模块,从调度与共享2个维度对模型提供增强能力。

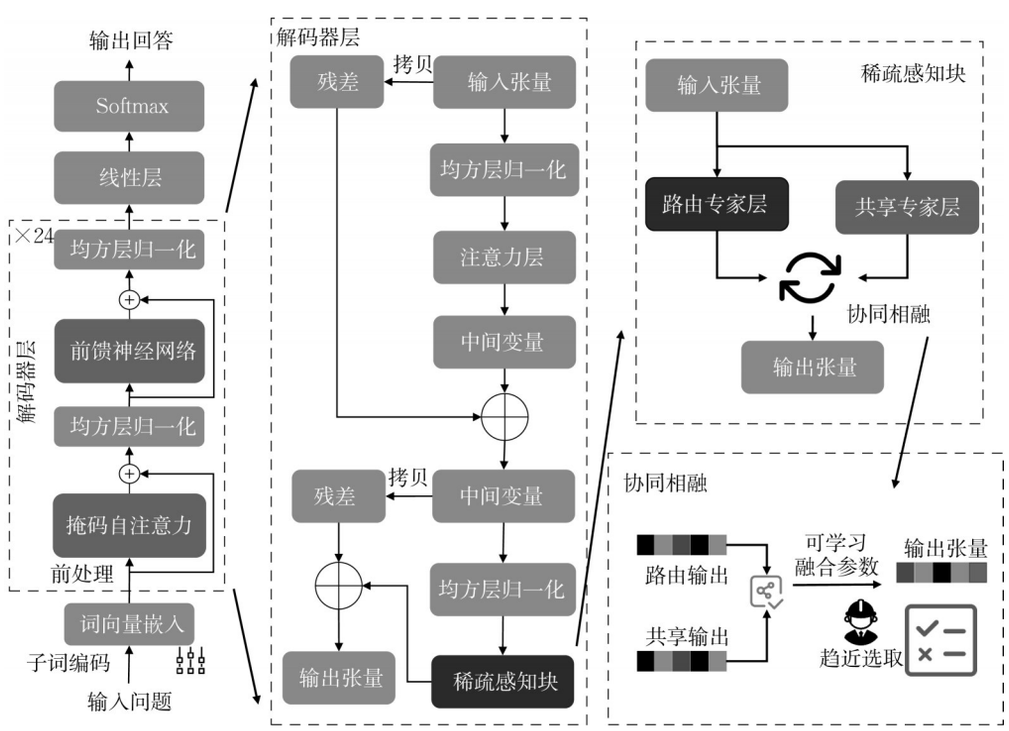


图3 协同专家系统解码层架构

Fig. 3 Architecture of the decoding layer in the collaborative expert system

在路由专家系统中，其问询适应能力尤为突出。鉴于建筑施工领域涵盖了多种结构类型和复杂的工艺操作，因此路由机制使得模型能够灵活地选择和激活适合特定任务的参数，从而高效地处理其擅长的领域。整体处理如图4a所示，式为

$$y_{\text{route}}=\sum_{j=1}^{k}\sigma(\delta_i)_j\cdot\mathcal{E}_{i_j}(x_i) \tag{3}$$

式中：δ_i为通过门控模块计算出的对应专家权重向量；σ为sigmoid归一化激活函数，计算门控模块得到的各个路由专家的权重值；$\mathcal{E}_{i_j}$为门控模型选中的$\mathcal{E}_i$候选专家中的第j层对原始x_i文本隐层进行计算的结果。在每个专家层处理操作中，为提升处理效率，采用专家路由策略，对每个专家层首先进行指派，使得在每次操作中一次性处理分派到指定该专家的隐层状态，并传入相应权重，从而减少系统遍历次数，降低计算复杂性。

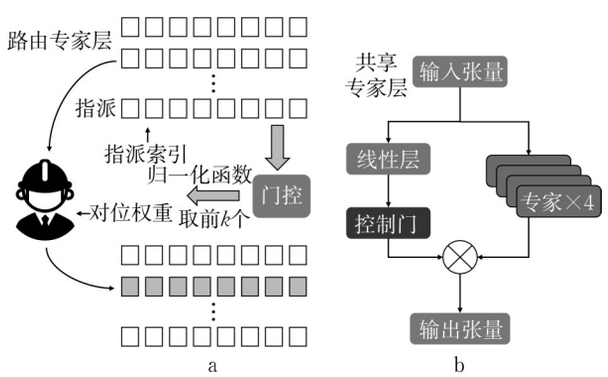


图 4　路由专家与共享专家层架构

Fig. 4　Architecture of the routing expert and shared expert layers

共享专家层的设置旨在确保模型在处理复杂任务时保持平衡与稳定，从而提升整体系统的鲁棒性和可靠性。如图4b所示，与路由策略的多系统不同，共享层中拓展了单一专家层的深度，从而形成4个总被激活的共享专家块。共享专家全部参与计算，由共享门控层生成每个输出的权值。其得到的结果与路由专家输出相加，组成最终的输出张量，从而完成混合专家体系增强。其运算过程为

$$y_{\text{share}}=\sum_{j=1}^{k}\phi(W_{\text{g}}x_i)\odot S_{i_j}x_i \tag{4}$$

式中：W_{g}为共享专家门控模块功能函数，根据输入文本向量动态调节加权参数；$\mathcal{S}_{i_j}$为指定的共享专家层转换函数；ϕ为归一化激活函数。

在得到路由专家和共享专家2项隐层输出后，系统通过协同相融策略实现输出内容的融合。具体来说，协同相融机制包括如下关键步骤：

（1）输出融合。首先，为了控制路由专家和共享专家输出在融合过程中的占比，设置一个可学习调节因子α_{clipped}，动态调整两者的融合权重，确保系统能够适应不同任务的特性。在融合过程中，路由专家的输出y_{route}和共享专家的输出y_{shared}分别被乘以相应的融合参数，随后进行相加。

（2）权重调整策略。在训练初期，α_{clipped}的初始值通过梯度剪裁进行限制，并采用了梯度阶梯下降策略，逐步调整融合权重的梯度更新过程。通过设置平移激活函数，将融合参数的取值范围限定为(0.5,1.5)。确保融合权重不会偏离合理区间，并能根据模型训练过程自适应调整。

整体计算流程如式(5)所示，其中ψ表示为tanh平移激活函数运算式，$\psi\in(0.5,1.5)$，α_{clipped}表示参数梯度剪裁限制的可学习调节因子。

$$y_{\text{output}}=\psi\left(\alpha_{\text{clipped}}\right)\cdot y_{\text{route}}+\left(2-\psi\left(\alpha_{\text{clipped}}\right)\right)\cdot y_{\text{shared}} \tag{5}$$

3　实验结果分析

3.1　生成式文本评测

在生成式文本评测中，使用构造问询数据的评测集进行评测。在评测指标中，考虑到建筑施工场景文本的复杂多样性，应对生成关键信息的准确性与整体信息的一致性进行多维度评测。故选用在自然语言处理领域中常用的最长公共子序列评估指标[23]（Rouge-L）与余弦相似度[24]（cosine similarity）作为评测指标。其中Rouge-L表示为预测与标注之间的最长公共子序列与标注文本长度的比值，能够体现预测与标注之间的细粒度信息准确性；Cosine相似度则先用嵌入模型将预测与标注分别嵌入为文本向量，进而计算两向量之间的夹角余弦值，从而能够体现预测与标注文本之间的整体语义相关性。在推理速度评测中，选取TPS（Tokens Per Second）指标进行评估，其表示模型在生成过程中每秒处理的子词（token）数量。

为综合考量系统回答性能，选用4种主流的大语言模型进行对比实验，模型分别为：Qwen1.5-7B-chat[9]，Qwen1.5-1.8B-chat[9]，Llama3-8B-Instruct[25]与GPT4-Turbo[7]。其中，Llama3为Meta研发的最新开源文本生成式大模型，在多项通用文本生成任务中取得最佳成绩。GPT4-Turbo为OpenAI研发的基于近1 800亿参数训练的强大闭源大语言模型，故本文选择调用其API并接入本地知识库增强。其

余模型的增强策略如表3所示。本实验环境具体配置为:GPU, A6000, 显存48G; CPU Intel(R) Xeon(R) Platinum 8358P 3.4 GHz; python版本3.10; Cuda版本11.8。batch_size设置为2,梯度检查点设置为4。训练过程如图5所示。

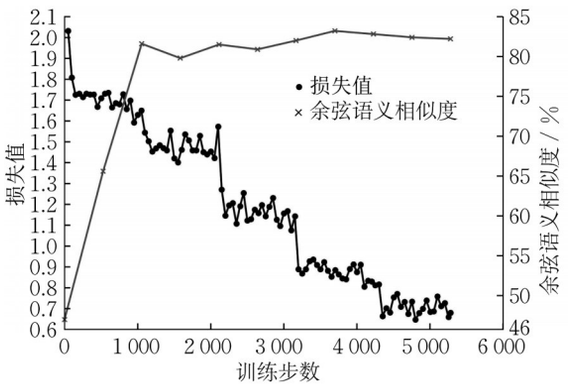


图5 训练过程中损失值与余弦语义相似度变化曲线

Fig. 5 Curves of loss value and cosine semantic similarity during training

由表2评测结果所示,采用知识注入的协同专家增强体系模型,在仅使用1/3激活参数的情况下性能能够优于其他对比模型;并在Rouge-L指标中相较GPT4-Turbo有9.3%的提升;相较于其他本地部署离线模型,具有更快的子词处理速度,能够满足建筑施工场景中多方使用的需求,从而能够提供更加快速准确的建议。

具体生成效果样例如图6所示,在辅助决策问题表现中,图6a为GPT4-Turbo与基准模型训练增强后的生成结果,图6b为本系统生成结果。由图6可见,本系统所提供的回复施工知识性更强。对于问题一,涉及规范条文选择参考,系统能够根据问题回答出相应的条文参考,提供更加准确的实际环境决策推荐,生成内容更加贴合实际生产需求。对于对比模型GPT4-Turbo而言,生成的内容虽与问询相关,但回复内容笼统,建议性不强;对于问题2所示的组合问题,需要施工工艺特性知识,本问答系统回复得更加全面,且回答更加具体,从而为施工决策提供更好的辅助,获得更加准确的建筑施工知识。而基准模型则回复能力较弱,难以提供实质性技术支撑。

表2 评测集评测结果

Tab. 2 Results of test set

模型	激活参数量/B	最长公共子序列占比/%	余弦相似度/%	推理速度/(token·s^{-1})
GPT4-Turbo(本地知识库)	约1 800	34.4	77.2	
Llama3-8B-instruct(微调+本地知识库)	8	37.5	73.4	585
Qwen1.5-7B-chat(微调+本地知识库)	7	39.3	76.3	1 115
Qwen1.5-1.8B-chat(微调+本地知识库)	1.8	26.3	57.7	1 815
Ours(微调+本地知识库+协同专家+知识注入)	2.7	43.7	81.1	1 703

3.2 客观题评测

为验证问答系统在处理建筑施工领域中的逻辑推理与指令遵循能力,从近3年建筑施工专业课程考试题与职业资格真题试卷中抽取除应用在训练集中的单项选择题,建立建筑施工客观题评估测试集。在收集过程中,针对施工基础知识、基础工程施工、钢筋混凝土结构施工、钢结构施工与钢与钢筋混凝土混合结构施工5个领域各遴选30道共150道单项选择题作为评估基准,所选范围覆盖了建筑施工专业课考试范围的核心知识,从而用来验证模型在建筑施工下的推理能力。

在评测过程中,使用few-shot与思维链[26]的方式构建问询prompt,并同样采取与表2相同的对比模型及增强策略。选取回答正确率作为评测指标,即系统回答正确的数量占总问题数量的比例。通过表3可得,本系统在仅有2.7 B激活参数的前提下,在多领域科目中取得了更加优异的表现,并且与使用了相同本地知识库提示增强的GPT4-Turbo相比,平均正确率提升了8%,验证了本系统在建筑施工领域下逻辑推断能力的优越性。相比于基线模型,客观题正确率提升了36%,因此具有更强的推理判断能力。

3.3 消融实验

为探究各个增强模块具体的提升贡献,对其逐一进行了消融实验。具体结果见表4。由消融结果可得,在原版指令微调模型中,使用专家系统增强策略前后,Rouge-l指标由26.3%跃升到36.6%,语义余弦相关性由57.7%上升到73.1%,客观题平均正确率由43.3%上升到71.3%,体现了专家系统增强的有效性。在采用知识注入增强策略后,Rouge-l上升4.9%,客观题平均正确率提升6%,语义余弦相关性提升5.6%。在使用协同相融模块后,通过嵌

图 6　生成效果

Fig. 6　Results of generation

表3　客观题正确率

Tab. 3　Accuracy of objective questions

模型	激活参数量/B	各领域客观题正确率/%					平均正确率/%
		基础知识	基础工程	钢筋混凝土	钢结构	混合结构	
GPT—4—Turbo	约1 800	76.6	70.0	66.7	73.3	70.0	71.3
Llama3—8B—instruct	8	66.7	50.0	70.0	60.0	63.3	62.0
Qwen1.5—7B—chat	7	70.0	73.3	83.3	76.6	83.3	77.3
Qwen1.5—1.8B—chat	1.8	40.0	50.0	36.7	43.3	46.7	43.3
Ours	2.7	76.7	80.0	83.3	73.3	83.3	79.3

入全局、路由隐层信息,为系统带来了2.2%的Rouge-l提升、2.4%的语义相关性提升与2%的客观题正确率提升。进一步说明了系统增强策略的优越性。

表4　消融实验

Tab. 4　Ablation study

指令微调	路由专家	共享专家	知识注入策略	协同相融	激活参数量/B	最长共子序列占比/%	余弦相似度/%	平均正确率/%
√					1.8	26.3	57.7	43.3
√	√				2.3	28.5	61.1	46.0
√	√	√			2.7	36.6	73.1	71.3
√	√	√	√		2.7	41.5	78.7	77.3
√	√	√	√	√	2.7	43.7	81.1	79.3

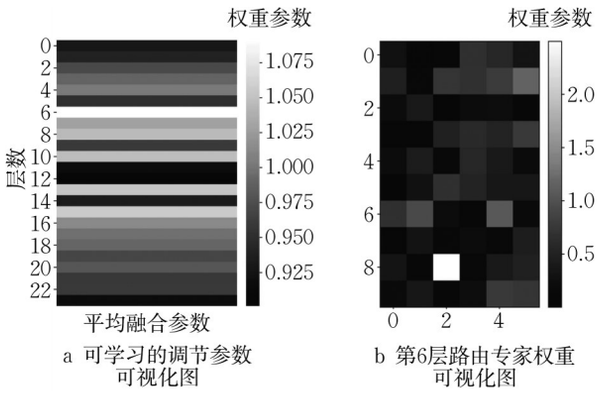
为直观观察协同融合后的系统对路由专家和共享专家层决策的重要性,将每一层的融合参数$\psi(\alpha_{\text{clipped}})$通过求取均值进行可视化展示,具体如图7所示。为了进一步展示不同专家的决策参与度,将图7分为2个子图。

(1)图7a展示了在24层解码层中,每层的融合参数$\psi(\alpha_{\text{clipped}})$大小,该值表示模型在该层中对路由专家的依赖程度,而共享专家的调节系数为2—$\psi(\alpha_{\text{clipped}})$。从图7a中可以看出,模型在前几层逐渐加大对路由专家的依赖,表明这些层中模型更倾向于基于任务的个性化决策。随着层数增加,中间层表现出明显的动态调整,展示了对输入特征的兼容性和灵活性。在后几层,融合参数趋于稳定,这种稳定性有助于模型在深度层次上实现信息的统一整合。

(2)图7b展示了路由专家的权值。由于在第6层中,模型对路由专家的依赖性最强,因此进一步对该层的60个路由专家的权重参数进行了可视化,展示在图7b中。可以观察到,路由专家中的中间部分和末尾部分的决策权重较大,说明在第6层的专家决策过程中,模型更倾向于调用这些特定的路由专家。这种分布模式体现模型在施工场景的多样化问询下具备灵活的专家调度能力,从而应对不同的决策需求。

3.4 人工专家评测

针对建筑施工领域问答系统的表现进行人工评测,从测试集中随机抽取了200条问答数据,并组织专家进行评分。为了确保评估的科学性与权威性,邀请了20位具备建筑施工与人工智能领域知识背景的专家参与打分,其中6位为高年级博士研究生,另外14位均拥有博士学位并具备丰富的行业经验:7位拥有10年及以上的建筑施工工作经验,另7位具备10年以上的施工教学与科研经验。评估维度参考相关研究以及建筑施工特性[16,18-19],分为以下5个维度:技术与工程知识、法律和标准遵循度、回复事实正确性、施工可操作性、问题需求满意度,见表5。这些维度覆盖了建筑施工问答系统的核心功能与使用需求。每个维度基于1—10分进行打分,10分代表最高水平,依次向下递减。将标注答案设置为8分的标准,当模型的回答在所有维度上表现较好,且与参考答案质量相近时,即可获得7~8分的得分水平。最终收集所有专家打分,综合求取平均值,作为人工评测最终得分。



a 可学习的调节参数可视化图 b 第6层路由专家权重可视化图

图7 融合参数可视化

Fig. 7 Visualization of parameters

表5 人工评测

Tab. 5 Human evaluation

模型	各项指标得分					综合得分
	技术与工程知识分值	法规和标准遵循度	回复事实正确性	施工可操作性	问题需求满意度	
GPT—4—Turbo	6.7	6.8	7.3	6.7	6.9	6.9
Llama3—8B—instruct	6.5	6.5	7.1	6.8	7.3	6.8
Qwen1.5—7B—chat	7.0	7.1	7.5	7.3	7.6	7.3
Qwen1.5—1.8B—chat	5.1	5.2	5.4	4.7	5.6	5.2
Ours	7.8	8.0	7.9	7.5	8.2	7.9

通过上述人工评测结果可得,本系统框架在5个维度评分相较于基线模型均有大幅的提升,并且超过了参数量3倍于本模型的Qwen-1.5-7B模型以及使用RAG增强策略的GPT4-Turbo。并且最终综合评分为7.9,在5个维度中均表现出了较好的得分,与人工标注的答案质量相近甚至部分超越,能够满足实际建筑施工场景需求。

4 结论

提出了一项适用于建筑施工领域的生成式问答系统,并提出了基于协同专家系统与知识注入训练策略,通过多维度验证了该系统在建筑施工知识获取与决策辅助等问答任务上的可行性与优越性,具体体现在如下几个方面:

(1)使用协同专家增强策略,使模型在面临广泛问询范围时,学习调用最适合解决当下问询的参数,从而生成更贴切的回答。相比基线模型,带来了12.5%的Rouge-l提升、17.8%的语义余弦相关性提升与30%的客观题平均准确率提升。

(2)为解决使用本地知识库检索所带来的训练与推理问询不匹配问题,采用了知识注入策略,使模型在训练时,通过语义相似性匹配方法,调用知识库中的参考知识与问询重构为问询—提示数据,从而达到在训练时注入垂域施工知识的效果,带来4.9%的Rouge-l提升、5.6%的语义余弦相关性提升与6%的客观题正确率的提升。

(3)为大语言模型在建筑施工垂域中提供了一套离线部署的可能,仅采用2.7 B的激活参数却能够达到最优的回复生成性能与更快的子词吞吐速度,从而能够进一步开拓离线施工大语言模型任务场景,降低信息通讯成本与模型生成成本。

本研究仍有部分不足：由于训练数据有限，在施工场景问答下的准确回复能力仍有待提升；此外，尽管模型在建筑施工领域的回复能力得到了增强，但对通用问题的记忆却有所减弱。因此，在提升模型特定领域表现的同时，如何保持其对通用知识的记忆，将是未来值得进一步研究的方向。

综上，本文验证了在利用协同专家系统与知识注入训练策略增强下，大语言模型问答系统在建筑施工垂域文本生成任务上的可行性与优越性，能够在广泛建筑施工场景中提供可靠的知识检索与决策参考，能够有效帮助施工管理人员进行科学决策，并辅助专业课程教学。本系统已在中建八局施工管理场景中完成测试，在未来将持续探索模型能力增强策略与扩增建筑施工垂域文本数据，从而继续赋能建筑施工管理与专业课程教学。

作者贡献声明：

杨　彬：制定论文思路，论文修改，研究经费支持。

肖鸿儒：制定论文思路，开展实验，论文撰写和修改。

高　尚：协助论文方法优化，论文撰写和修改。

雷　克：测试场景支持，实验进展指导，论文指导。

陈文硕：测试实验，论文修改。

张其林：制定论文思路，论文修改。

汪丛军：指导实验测试，整理测试数据。

参考文献：

[1] PAN Z, YU Y. Learning multi-granular worker intentions from incomplete visual observations for worker-robot collaboration in construction [J]. Automation in Construction, 2024, 158: 105184. DOI:10.1016/j.autcon.2023.105184.

[2] 王煜，邓晖，李晓瑶，等. 自然语言处理技术在建筑工程中的应用研究综述[J]. 图学学报，2020，41(4)：501.
WANG Yu, DENG Hui, LI Xiaoyao, *et al*. A review of natural language processing applications in construction engineering[J]. Journal of Graphics , 2020, 41(4): 501.

[3] 葛恒刚，崔祥民，张友志. 基于文本挖掘的建筑施工坍塌事故成因网络分析[J]. 土木工程与管理学报，2023，40(4)：149.
GE Henggang, CUI Xiangmin, ZHANG Youzhi. Network analysis of construction collapse accidents based on text mining [J]. Journal of Civil Engineering and Management, 2023, 40 (4): 149.

[4] KIM T, CHI S. Accident case retrieval and analyses: using natural language processing in the construction industry [J]. Journal of Construction Engineering and Management, 2019, 145 (3) : 04019004. DOI: 10.1061/(ASCE) CO. 1943-7862.0001625.

[5] XIAO H, ZHUANG J, YANG B, *et al*. Generative knowledge-guided review system for construction disclosure documents[J]. Advanced Engineering Informatics, 2025, 68: 103618. DOI: 10.1016/j.aei.2025.103618.

[6] 赵鑫，窦志成，文继荣. 大语言模型时代下的信息检索研究发展趋势[J]. 中国科学基金，2023，37(5)：786. DOI：10.16262/j.cnki.1000-8217.20231026.007.
ZHAO Xin, DOU Zhicheng, WEN Jirong. Information retrieval research trends in the era of big language modeling[J]. Bulletin of National Natural Science Foundation of China, 2023, 37 (5) : 786. DOI: 10.16262/j. cnki. 1000-8217. 20231026.007.

[7] OPENAI, ACHIAM J, ADLER S, *et al*. GPT- 4 Technical report [J/OL]. (2024-03-04) [2024-05-06]. http://arxiv. org/abs/2303.08774. DOI:10.48550/arXiv.2303.08774.

[8] JIANG A Q, SABLAYROLLES A, ROUX A, *et al*. Mixtral of experts [J/OL]. (2024-01-08) [2024-05-06]. http://arxiv.org/abs/2401.04088. DOI:10.48550/arXiv.2401.04088.

[9] BAI J, BAI S, CHU Y, *et al*. Qwen technical report[J/OL]. (2023-09-28)[2024-05-06]. http://arxiv.org/abs/2309.16609. DOI:10.48550/arXiv.2309.16609.

[10] SAKA A, TAIWO R, SAKA N, *et al*. GPT models in construction industry: opportunities, limitations, and a use case validation[J]. Developments in the Built Environment, 2024, 17: 100300. DOI:10.1016/j.dibe.2023.100300.

[11] SAHOO P, SINGH A K, SAHA S, *et al*. A systematic survey of prompt engineering in large language models: techniques and applications [J/OL]. (2024-02-12) [2024-05-06]. DOI: 10.48550/arXiv.2402.07927.

[12] ZHENG J, FISCHER M. Dynamic prompt-based virtual assistant framework for BIM information search [J]. Automation in Construction, 2023, 155: 105067. DOI: 10.1016/j.autcon.2023.105067.

[13] YOU H, YE Y, ZHOU T, *et al*. Robot-enabled construction assembly with automated sequence planning based on ChatGPT: RoboGPT [J]. Buildings, 2023, 13 (7) : 1772. DOI: 10.3390/buildings13071772.

[14] PRIETO S A, MENGISTE E T, GARCÍA DE SOTO B. Investigating the use of ChatGPT for the scheduling of construction projects[J]. Buildings, 2023, 13(4): 857. DOI: 10.3390/buildings13040857.

[15] 覃思中，郑哲，顾燚，等. 大语言模型在建筑工程中的应用测试与讨论[J]. 工业建筑，2023，53(9)：162. DOI：10.13204/j.gyjzG23081006.
QIN Sizhong, ZHENG Zhe, GU Yan, *et al*. Exploring and discussion on the application of large language models in construction engineering[J]. Industrial Construction, 2023, 53 (9): 162. DOI:10.13204/j.gyjzG23081006.

[16] 丁志坤，李金泽，刘明辉. 基于大语言模型的BIM正向设计问答系统研究[J]. 土木工程与管理学报，2024，41(1)：1. DOI：10.13579/j.cnki.2095-0985.2024.20240046.
DING Zhikun, LI Jinze, Liu Minghui. Research on LLM-based question-answering system for BIM forward design[J].

（下转第 30 页）