

文章编号:1672-3961(2026)01-0097-08

DOI:10.6040/j.issn.1672-3961.0.2025.190

基于检索增强生成和智能体的建筑材料碳排放单位换算问答模型

阎俏^{1,2}, 焦飞³, 严毅^{1,2*}, 杜向华⁴, 刘鹏程¹

(1. 山东建筑大学信息与电气工程学院, 山东 济南 250101; 2. 山东省智慧建筑与建筑节能重点实验室(山东建筑大学), 山东 济南 250101; 3. 中国铁塔股份有限公司菏泽市分公司, 山东 菏泽 274000; 4. 山东建筑大学计算机与人工智能学院, 山东 济南 250101)

摘要:为解决建筑材料生产及运输阶段碳排放计算时建筑材料计量单位与碳排放因子单位不匹配的问题,提出一种基于检索增强生成(retrieval-augmented generation, RAG)和智能体(Agent)的建筑材料碳排放单位换算问答模型。通过解析典型材料换算步骤构建本地知识库,设计RAG模块,为换算提供步骤参考;开发可调用计算工具的Agent,执行换算过程中的数学运算;设计提示词模板并接入大语言模型,实现基于本地知识库的文本问答。试验结果表明,所提模型能够准确回答建材的单位换算问题,支持Web端与本地控制台交互,实现单位换算结果及推理步骤的可视化。

关键词:建筑材料碳排放;单位换算;检索增强生成;智能体;问答模型

中图分类号:TU5 **文献标志码:**A

引用格式:阎俏,焦飞,严毅,等. 基于检索增强生成和智能体的建筑材料碳排放单位换算问答模型[J]. 山东大学学报(工学版), 2026, 56(1): 97-104.

YAN Qiao, JIAO Fei, YAN Yi, et al. Question-answering model for building material carbon emissions unit conversion based on retrieval-augmented generation and Agent[J]. Journal of Shandong University (Engineering Science), 2026, 56(1):97-104.

Question-answering model for building material carbon emissions unit conversion based on retrieval-augmented generation and Agent

YAN Qiao^{1,2}, JIAO Fei³, YAN Yi^{1,2*}, DU Xianghua⁴, LIU Pengcheng¹

(1. School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan 250101, Shandong, China; 2. Shandong Key Laboratory of Smart Buildings and Energy Efficiency, Shandong Jianzhu University, Jinan 250101, Shandong, China; 3. Heze Branch of China Tower Co., Ltd., Heze 274000, Shandong, China; 4. School of Computer and Artificial Intelligence, Shandong Jianzhu University, Jinan 250101, Shandong, China)

Abstract: To solve mismatching between the measurement units of building materials and the units of carbon emission factors during the calculation of carbon emissions in the phase of building materials production and transportation, a question-answering model for building material carbon emissions unit conversion based on retrieval-augmented generation (RAG) and Agent was proposed. A local knowledge database was constructed by analyzing typical material conversion processes, and a RAG module was designed to provide step-by-step references for the conversions. An Agent capable of calling the calculation tool was developed to perform the mathematical operations required in the conversion process. The prompt templates were designed and integrated with a large language model to answer the question based on local knowledge database. The experimental results showed that the proposed model could accurately answer the unit conversion questions of building materials, and realized visualization of the unit conversion results and reasoning steps displayed on the Web interface and local console.

Keywords: building material carbon emissions; unit conversion; retrieval-augmented generation; Agent; question-answering model

收稿日期: 2025-09-05

基金项目: 国家自然科学基金青年科学基金资助项目(52007109)

第一作者简介: 阎俏(1976—), 女, 山东荣成人, 教授, 硕士生导师, 博士, 主要研究方向为建筑智能化与能效管理、人工智能技术应用。

E-mail: yanqiao@sdjzu.edu.cn

* 通信作者简介: 严毅(1987—), 男, 山东济南人, 副教授, 硕士生导师, 博士, 主要研究方向为新能源、储能、建筑节能。

E-mail: yanyi19@sdjzu.edu.cn

0 引言

建筑领域的节能降碳对我国达成“双碳”目标至关重要。精确测算碳排放是评估建筑碳排放状况的有效依据。根据文献[1],可将建筑全生命周期划分为建材生产及运输、建造及拆除、运行3个阶段。2021年数据显示,建材生产及运输阶段的碳排放约占建筑全生命周期碳排放总量的51.9%^[2]。因此,精准计算建材生产及运输阶段的碳排放是推动建筑全生命周期节能减排的关键。

碳排放因子法是广泛应用于建材生产阶段的碳排放计算方法。碳排放量为建材数量与每种建材对应的碳排放因子(简称为碳因子)的乘积。现有技术能够通过文本数据处理方式为输入的不同建筑材料匹配正确的碳因子^[3],但存在建材计量单位与碳因子单位的不匹配问题。

生成式人工智能技术的发展为解决语言类的复杂问题提供解决途径。大语言模型(large language model, LLM)(如生成式预训练变换器(generative pre-trained Transformer, GPT)、星火^[4]、Qwen^[5]和 DeepSeek^[6]等)代表自然语言处理领域的前沿技术。这些模型通过大规模语料库的预训练,能够高效理解和生成自然语言文本。在通用问答领域,尽管许多 LLM 已展现出接近人类水平的表现,但仍存在一些不足:一方面,LLM 有时会产生不可控的“幻觉”,导致输出内容与现实脱节或逻辑不连贯^[7];另一方面,LLM 的训练数据往往相对封闭且单一,在特定专业领域的表现不佳。

针对上述问题,主流解决方案包括微调和检索增强生成(retrieval-augmented generation, RAG)等技术。微调通过在特定领域数据上进一步训练模型,只需调整少量参数即可达到与全量训练相近的效果,但计算成本较高,难以应对频繁的知识更新^[8]。RAG 结合检索与生成技术,预先检索权威知识库,筛选可信信息作为生成依据,减少误导性内容,确保答案与用户问题的高度匹配^[9]。文献[10]提出自适应 RAG,在多项开放式知识问答任务中均表现出色,体现出 RAG 在调用外部知识库的优势。然而,RAG 受知识库范围的限制,难以处理超出知识库的问题,无法执行复杂数值计算、访问实时信息或调用外部程序等任务,影响落地应用的效果。人工智能体(artificial intelligence Agent, AI Agent)技术可以通过调用计算器、搜索引擎、应用程序接

口(application programming interface, API)等工具,赋予 LLM 更强的推理与执行能力,使 LLM 突破静态知识库限制,实现动态信息处理与复杂任务的逐步解决^[11]。针对天然气阀室泄漏问题,文献[12]利用 RAG 技术构建基于相关泄漏文档的本地知识库,辅助 LLM 进行知识问答和推理,设计泄漏检测智能体(Agent),集成泄漏检测、知识推理及处置方案生成等功能,实现智能化泄漏管理。RAG、AI Agent 和 LLM 作为人工智能领域的三大核心技术,各自发挥不同的作用,通过协同工作,不断拓展新的应用场景,为各种应用和服务的开发提供更多可能性。

本研究构建一种基于 RAG 和 Agent 的建筑材料碳排放单位换算问答模型。通过解析典型材料的单位换算步骤,完成本地知识库搭建,为后续换算提供计算步骤参考;构建 RAG 组件,完善提示词设计并接入 LLM,使 LLM 能够基于本地知识库实现文本问答功能,但无法提供正确的数学计算结果;构建能够调用计算工具的 Agent,完成模型搭建。当用户向模型提问与单位换算相关的问题时,模型可依托本地知识库中的换算步骤,调用计算工具进行完整的数学计算。通过创建 FastAPI 应用,模型可在 Web 端与本地控制台展示换算过程。

1 问题的提出

在实际工程中,建筑材料清单的记录格式缺乏统一标准。由于建筑材料种类较多,不同的供应商或项目阶段可能采用不同的计量单位。因此,在利用碳排放因子法进行建筑材料碳排放计算时,会出现建筑材料工程量清单(简称为工程清单)中记录的建筑材料计量单位与碳因子数据库中的单位不匹配问题。工程清单中建筑材料的计量单位通常为 m^3 、 m^2 、kg、个等,碳因子的单位大多为 $kgCO_2e/t$ ^[13]。此时,往往依赖人工查阅材料密度、体积等参数,或根据材料名称中标注的特定型号进行单位换算,使建筑材料的计量单位与碳因子的单位一致,再计算碳排放量。某厂商碳排放云计量平台如图 1 所示,平台中不锈钢管的数量以 m 为单位,碳因子单位为 $kgCO_2e/t$ 。用户需线下自行查询换算关系,手动将不锈钢管的数量单位从 m 换算为 t ,再将换算后的结果填入换算系数字段中,系统才能进行后续计算。这一过程不仅增加计算复杂度,也影响计算效率。

序号	材料名称	规格型号	材料类别	二级分类	数量	单位	质量换算系数 (t/unit)	碳因子换算系数	换算系数	碳因子单位	kgCO ₂ e/t
9	不锈钢	U形卡	钢铁金属类	钢材	233.235	个	-	-	-	t	2.350
10	不锈钢栏杆	-	钢铁金属类	钢材	1946.653	m	-	-	-	t	2.350
11	不锈钢法兰	φ59	钢铁金属类	钢材	5.219 8	只	-	-	-	t	2.350
12	不锈钢焊丝	-	钢铁金属类	钢材	1.148 7	kg	-	-	-	t	2.350
13	不锈钢管	φ32×1.5	钢铁金属类	钢材	51.492 1	m	-	-	-	t	2.050

图 1 某厂商碳排放云计量平台的单位不匹配问题

Fig.1 Unitmismatch issue in the carbon emission cloud measurement platform of a manufacturer

2 建筑材料碳排放单位换算问答模型

2.1 模型构建

本研究模型主要由 RAG 和 Agent 两大模块构成,模型结构如图 2 所示。

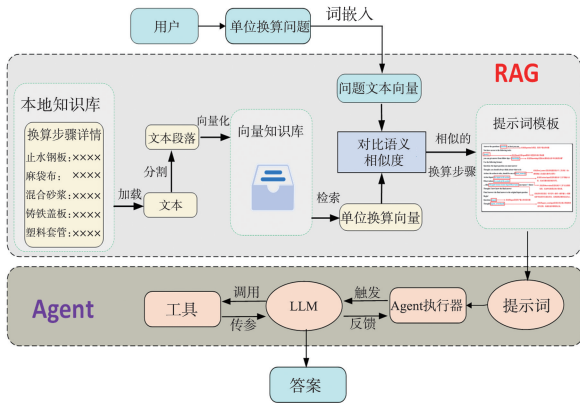


图 2 单位换算问答模型结构

Fig.2 Architecture of the unit conversion question-answering model

(1) RAG 模块:基于换算步骤详情构建本地知识库,完成文档加载;利用文本切分器将加载的文本分割成文本段落,通过词嵌入模型对段落进行向量化处理,形成向量知识库;对于输入的文本问题,使用同样的词嵌入模型进行文本向量化处理;在已有的词向量库中搜索与问题向量最相似的若干条文本,将文本整合进预设的提示词模板内,生成提示词发送至 Agent。

(2) Agent 模块:Agent 执行器接收到提示词后,触发 LLM 解答问题;LLM 根据提示词内容分析出问题的参数信息与要调用的工具,通过工具计算得到结果;LLM 根据计算结果完成提示词的汇总整理,得到包含计算结果的完整解答;将完整解答反馈

给 Agent 执行器并输出。

2.2 本地知识库与 RAG 的构建

2.2.1 本地知识库的构建

根据建材清单和碳因子之间的单位差异,确定典型建材的单位换算步骤和相应的数学表达式,形成包含单位换算步骤专业知识的本地知识库。利用生成式语言模型,得到其他材料的换算步骤和相应的数学表达式,用于扩充本地知识库内容。

比对工程清单和碳因子数据库中相同材料的单位,找出材料类型相同而单位不同的情况,将换算方式归纳为以下 4 种类型:m 向 t 换算、m² 向 t 换算、m³ 向 t 换算、个或套向 t 换算。

对每种换算类型构建“换算步骤详情”的语言描述,具体构建步骤如图 3 所示。提取工程清单中的材料名称、单位和数量,利用设定变量代替固定数值,形成通顺的语言描述;根据建材形状提取关键尺寸参数,增加建材密度的描述;根据碳因子单位、建材形状及数量,给出由尺寸参数和密度信息组成的单位换算计算公式。本地知识库中的“换算步骤详情”描述如表 1 所示,每种换算类型以一种材料举例。

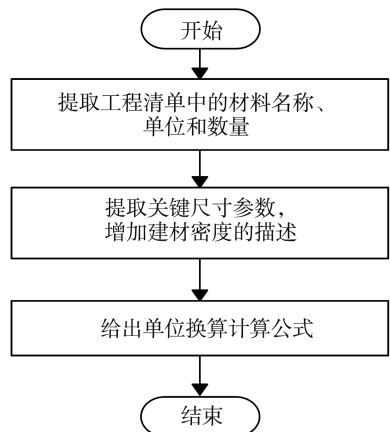


图 3 “换算步骤详情”的构建步骤

Fig.3 Construction steps of "conversion step details"

表1 “换算步骤详情”描述
Table 1 Description of "conversion step details"

换算类型	材料名称	工程清单中的材料单位	数据库中的碳因子单位	换算步骤详情
m 向 t 换算	止水钢板 $\delta 3 \times 400$	m	tCO ₂ e/t	止水钢板(规格为 $\delta x \times y$)的长为 A ,详细参数如下:厚为 x (单位 mm),将 x 的单位换算为 m 为 $0.001 \times x$;宽为 y (单位 mm),将 y 的单位换算为 m 为 $0.001 \times y$;钢的密度为 7.85 t/m^3 。换算步骤为 $0.001 \times x \times 0.001 \times y \times A \times 7.85$
m ² 向 t 换算	麻袋布	m ²	tCO ₂ e/t	麻袋布的面积 B ,详细参数如下:密度为 1.33 t/m^3 ,厚为 0.0024 m 。换算步骤为 $B \times 0.0024 \times 1.33$
m ³ 向 t 换算	混合砂浆	m ³	tCO ₂ e/t	混合砂浆的体积 C ,详细参数如下:密度为 1.8 t/m^3 。换算步骤为 $C \times 1.8$
个或套向 t 换算	铸铁盖板 600×500	个	tCO ₂ e/t	铸铁盖板(规格为 $a \times b$)的数量 D ,详细参数如下:长为 a (单位 mm),将 a 的单位换算为 m 为 $0.001 \times a$;宽为 b (单位 mm),将 b 的单位换算为 m 为 $0.001 \times b$;密度为 7.86 t/m^3 ;厚为 0.01 m 。换算步骤为 $0.001 \times a \times 0.001 \times b \times 0.01 \times 7.86 \times D$

注:止水钢板 $\delta 3 \times 400$ 是指厚为 3 mm、宽为 400 mm 的止水钢板,铸铁盖板 600×500 是指长为 600 mm、宽为 500 mm 的铸铁盖板。

本地知识库的构建统一了计算规则,通过明确换算步骤,确保不同系统对同一材料计算逻辑一致,减少人为换算出现的错误。后期可以通过更新文本格式文档适配新单位或新材料,实现动态更新。

2.2.2 RAG 的构建

RAG 框架主要由检索器和生成器两部分组成。检索器负责在本地知识库中查找与用户问题相关的信息,生成器则基于检索结果生成最终答案。在检索器中,采用词嵌入模型对本地知识库中的文本进行向量化处理,建立向量知识库,对用户输入的问题进行同源词嵌入表示,通过向量相似度检索,从向量知识库中匹配最相关的知识段落,完成 RAG 的构建。

2.3 利用 Agent 技术调用计算工具

本研究解决的单位换算问题本质是一个数学计算问题。RAG 技术通过检索相关数学公式、定理和解题步骤,给 LLM 提供解题思路,提高解决数学问题的准确性和效率。然而,RAG 本身并不执行数学计算。因此,利用 Agent 可以使 LLM 具备规划思考、记忆及使用工具函数的能力,最终能够自主完成给定任务。

本研究开发的 Agent 工具为一段数值计算程序,将文本中出现的阿拉伯数字指定为浮点数,执行本地知识库中数学表达式的数值计算并返回计算结果。

在设置工具之前,定义一个装饰器 `clean_tool_input`,用于清理传递给工具函数的输入。该装饰器通过去除输入问题文本字符串中的多余空白字符并仅保留第一行,确保输入的一致性和规范性。装饰器应用于 `calculate` 函数。该函数利用 `NumExpr`

库中的 `numexpr.evaluate` 函数计算输入的数学表达式并返回结果,保证数学表达式字符串的安全解析与高效运算。装饰器将 `calculate` 函数注册为一个可用工具,使 Agent 执行器能够调用该工具。

将提示词输入 Agent 模块后,Agent 执行器根据提示词内容触发 LLM 进行推理思考,调用相关工具解决问题。模型通过反复推理和自我检查验证答案的准确性。若发现错误,Agent 会对回答进行修正并重新生成输出。Agent 解决问题的流程如图 4 所示。

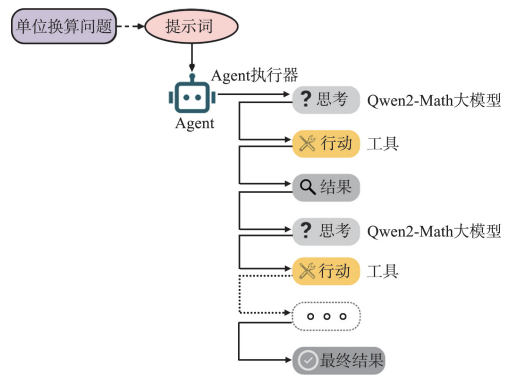


图4 Agent 解决问题流程
Fig.4 Problem-solving process of the Agent

2.4 完善提示词

在 RAG 模块中,当查询并返回与问题文本相似的本地知识库中相关文本后,输入问题和返回文本将同时传递至提示词模板。提示词模板通过编辑和整合这些输入,生成适当的提示词以引导后续任务的执行。提示词工程可视作一种指令或上下文引导^[14],主要目标是明确告知 LLM 所需执行的任务类型及期望模型以何种形式对输入数据作出

响应。提示词工程的部分代码及说明如图 5 所示。

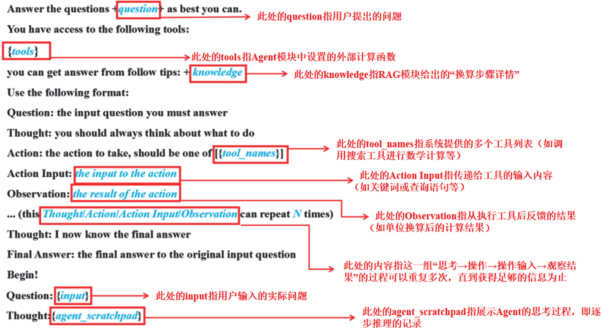


图 5 提示词工程的部分代码及说明

Fig.5 Part of the code and description of the prompt

提示词工程采用模板化方法构建一个多轮推理框架。首先明确问题,然后要求 LLM 基于本地知识库进行初步推理,并在必要时调用 Agent 工具,通过工具的输出不断调整推理过程,直至得出最终答案。这种结构化的提示词设计使模型能够逐步推理,选择合适的工具执行任务,并根据反馈结果更新思考过程,最终收敛至正确答案。

本研究构建 RAG 与 Agent 协同的问答模型。RAG 依托本地知识库与词嵌入检索提供知识支撑; Agent 模块以数值计算程序为核心工具弥补 RAG 的计算短板,通过提示词工程构建多轮推理框架,形成完整逻辑体系。本研究将通过试验验证模型在建筑材料碳排放单位换算任务中的有效性。

3 试验与分析

3.1 数据集

本研究使用的数据集为某建筑工程的建筑材料清单,共包含 1 700 种建筑材料,换算方式涵盖 m 向 t 换算、m² 向 t 换算、m³ 向 t 换算、数量(个或套)向 t 换算等多个维度。

3.2 试验设置

本试验环境配置如下:GPU 为 A4000,显存为 16 GB,Python 版本为 3.10,LangChain-Chatchat 版本为 0.2.6,分词器为 ChineseRecursiveTextSplitter,嵌入模型为 m3e-base@ cuda。

3.3 大语言模型选取

当前常见的大语言模型包括多款主流产品,各模型的预训练规模与能力特点各有不同。Baichuan-7B 是一款支持中英双语、基于约 1.2 万亿 tokens 的预训练大语言模型,但本身不具备对话能力^[15]。ChatGLM2-6B-32k 模型(基于 1.4 万亿 tokens 预训练)在逻辑推理和长上下文处理方面相对较弱。Llama-3.1-70B 模型(基于 15.6 万亿 tokens

预训练)和 DeepSeekMath-Base-7B 模型(基于约 1.4 万亿 tokens 预训练)均为业内领先的开源模型。Qwen2-Math 模型(基于 18 万亿 tokens 预训练)在数学相关任务中展现出显著优势,其数学推理能力、公式处理精度及复杂问题求解效率均优于同类模型,在 MATH、GSM8K 等权威数学基准测试中取得当前开源模型的最高成绩^[16]。与 Qwen2-Math 相比,Llama-3.1-70B 和 DeepSeekMath-Base-7B 在基准测试中的表现欠佳。因此,本研究选用 Qwen2-Math 作为基准大语言模型。

3.4 试验对比与分析

为了验证本研究模型的有效性,设置 3 种方法开展对比试验:方法 1 为无 RAG 参与,仅由 LLM 独立回答的 LLM 对话模式;方法 2 为 LLM 结合 RAG 回答的知识库对话模式;方法 3 为本研究构建的基于 RAG 和 Agent 的知识库及智能体对话模式。

3.4.1 方法 1、2 的试验结果分析

前 2 种方法利用 LangChain 框架开发一个基于 LLM 的应用^[17],通过外接多种市面上主流的大语言模型 API,实现单位换算问题的问答。试验测试星火、Qwen2 和文心一言等模型的 API,回答结果在内容上较为相近。以星火大模型 API 作为对话模型,方法 1 和方法 2 的回答结果示例如图 6、7 所示。2 种方法在该大模型下的回答篇幅较为简短,便于进行直观对比。



图 6 方法 1 的回答结果
Fig.6 Answer results of the method 1



图 7 方法 2 的回答结果
Fig.7 Answer results of the method 2

由图6中的回答结果可以看出,由于LLM缺乏相关专业知识,无法提供正确的计算结果。由图7中的回答结果可以看出,LLM参考本地知识库中的相关信息,使回答符合正确的计算步骤,但最终结果仍然错误,因为LLM采用自回归预测机制,擅长语义理解,但难以处理数字运算的位值对齐、进位等复杂逻辑。

3.4.2 方法3的试验结果分析

本研究构建模型可通过创建FastAPI应用,在Web端与本地控制台查看模型的完整运行结果。请求initdb端口对本地知识库进行处理,将结果加载进向量库,生成向量文件;请求query端口,向模型输入问题,得到最终答案。例如,要求模型将30 m²麻袋布换算为麻袋布质量的回答结果如图8所示。



图8 方法3的试验结果

Fig.8 Experimental results of the method 3

首先,模型接收输入的单位换算问题,根据问题内容检索向量知识库中对应的换算步骤详情;随后,模型将问题和换算步骤信息填入提示词模板,生成本次问题的最终提示词。Agent根据生成的提示词信息,逐步执行推理过程,得出计算逻辑为“面积乘厚度乘密度”。调用工具calculate进行数学计算,提取计算信息作为输入,经过工具计算得到初步结果为0.9972。Agent进行自我检查,验证结果的正确性。意识到计算错误后,Agent重复推理过程,直到最终得出正确结果为0.09576,将结果展示于FastAPI的Web端。上述推理过程如图9所示。

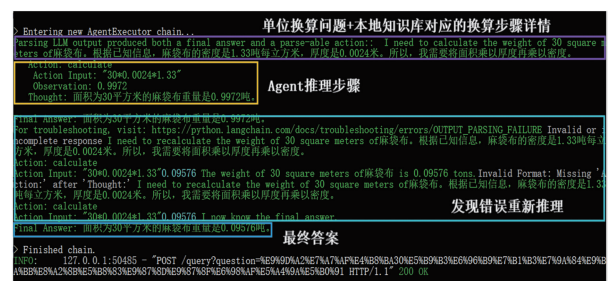


图9 控制台的推理过程

Fig.9 Reasoning process on the console

在模型的RAG模块执行过程中,利用向量数据库对单位换算问题进行相似度搜索时,模型会根据问题的内容,从数据库中检索出与换算问题词向量相似度最高的换算步骤详情。例如,当要求模型将长2 m、宽3 mm、厚400 mm的止水钢板换算为止水钢板质量时,模型回答问题的情况如图10所示。为了更清楚地展示推理过程,相似度最高的换算步骤详情如图11所示。



图10 回答问题情况

Fig.10 Performance of question answering

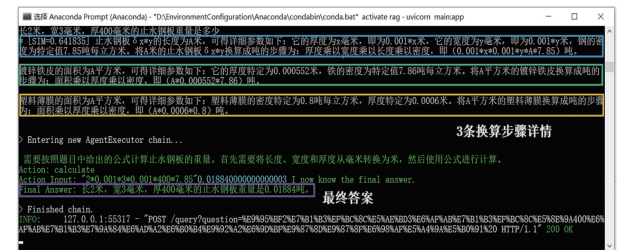


图11 相似度检索结果与推理过程

Fig.11 Results of similarity retrieval and reasoning process

3.4.3 不同任务下3种方法的对比分析

为验证所提模型的有效性,本研究设定3个复杂度由低到高的任务,分别为单步换算、多步换算和动态换算,测试3种方法的表现。

单步换算是针对基础量纲间的直接转换任务,聚焦单一转换关系的执行。例如将kg换算为t。

多步换算是针对需要经过多环节转换的复合计算任务,强调分步逻辑链的完整执行。以钢材碳排放计算为例,先通过RAG获取钢材单位质量碳因子,再完成钢材从kg到kgCO₂e的第一级转换,最后基于1 t = 1 000 kg的关系完成从kgCO₂e到tCO₂e的第二级转换,形成“钢材质量(kg)→碳排放总量(kgCO₂e)→碳排放总量(tCO₂e)”的完整逻辑链。

动态换算针对建材单位量纲与碳因子量纲不匹配的复杂推理任务,突出参数适配与动态计算能力。以麻袋布碳排放计算为例,根据RAG匹配最近似的换算步骤详情,利用Agent进行反复推理计算,直至完成从m²到t的量纲转换,得到与碳因子

适配的单位。该过程需同时处理固定知识库数据与动态输入参数的融合计算。

3 种方法面向不同任务的测试结果如表 2 所示,测试问题数量为 200 个,其中准确度表示模型回答正确的问题数量占测试问题数量的百分比。由表 2 可以看出:在单步换算任务中,因换算关系明确,3 种方法回答的准确度均为 100%;在多步换算任务中,方法 1 仅依赖 LLM 内置逻辑,易出现步骤遗漏,且碳因子等专业数据易偏离实际,准确度最低,方法 2 虽可通过 RAG 获取各环节所需数据,但需人工明确分步指令,缺乏自动串联逻辑链的能力,方法 3 的 Agent 能够将任务自动拆解为“数据调用→分步计算→结果整合”环节,基于 RAG 获取的精准数据自动串联多步逻辑,无需人工干预即可完成完整计算,步骤完整性和结果准确性均最优;在动态换算任务中,方法 1 无法识别动态参数与碳因子的关联关系,直接使用固定值计算,准确度较低,方法 2 可通过 RAG 获取基础碳因子和厚度影响规律,但需人工输入参数计算修正系数,无法自动完成动态校准,方法 3 的 Agent 能够自动识别动态参数与核心指标的关联规则,基于 RAG 获取基础数据和修正规律,自动计算校准系数并完成动态推理,结果最贴合实际场景需求。

表 2 3 种方法面向不同任务的准确度
Table 2 Accuracy of three methods for different tasks

方法	准确度/%		
	单步换算	多步换算	动态换算
方法 1	100	80	64
方法 2	100	90	80
方法 3	100	98	92

4 结论

本研究提出一种基于 RAG 和 Agent 的单位换算问答模型,通过设计本地知识库给出换算规则,完善提示词工程,搭建 RAG 模块,利用 Agent 的工具调用能力完成数学计算,解决建筑材料的计量单位与碳因子单位不匹配的问题。模型的运行结果通过 FastAPI 应用展示,用户可在 Web 端提问并查看模型回答,同时控制台展示问题的完整推理过程。

当前模型的本地知识库主要依赖人工手动构建,尽管通过大类划分策略降低了部分工作量,但整体构建效率和单位覆盖范围仍有不足。未来研究可考虑采集国标文档、学术公式、工程记录等多源数据,通过实体识别、关系抽取等自然语言处理技术抽取关键信息,构建结构化换算数据集;针对

不同复杂度场景,分别采用预训练语言模型微调或领域大语言模型生成固定系数或带场景约束的动态换算规则,拓展单位换算覆盖范围,提升模型泛化能力,使模型能迁移至更多领域,增强实用价值与推广潜力。

参考文献:

- [1] 中华人民共和国住房和城乡建设部. 建筑碳排放计算标准: GB/T 51366—2019[S]. 北京: 中国建筑工业出版社, 2019: 2.
- [2] 中国建筑节能协会, 重庆大学城乡建设与发展研究院. 中国建筑能耗与碳排放研究报告(2023 年)[J]. 建筑, 2024(2): 46-59.
- [3] YAN Q, JIAO F, PENG W. Building materials classification model based on text data enhancement and semantic feature extraction[J]. Buildings, 2024, 14: 1859.
- [4] 钱力, 张智雄, 伍大勇, 等. 科技文献大模型: 方法、框架与应用[J]. 中国图书馆学报, 2024, 50(6): 45-58. QIAN Li, ZHANG Zhixiong, WU Dayong, et al. The large language model for scientific literature: method, framework, and application[J]. Journal of Library Science in China, 2024, 50(6): 45-58.
- [5] 汪伦, 艾斯卡尔·艾木都拉, 张华平, 等. 基于大语言模型的开源情报摘要生成研究[J]. 情报理论与实践, 2025, 48(5): 43-48. WANG Lun, HAMDULLA Askar, ZHANG Huaping, et al. Research on open source intelligence summary generation based on large language models[J]. Information Studies: Theory & Application, 2025, 48(5): 43-48.
- [6] 赵阳. 从 DeepSeek 探讨大语言模型在建筑及能源行业的应用趋势和技术方法[J]. 建筑节能(中英文), 2025, 53(2): 1-6. ZHAO Yang. Application trends and technical methods of LLM in the construction industry and energy industry from DeepSeek[J]. Journal of Building Energy Efficiency, 2025, 53(2): 1-6.
- [7] DU Y L, LI S, TORRALBA A, et al. Improving factuality and reasoning in language models through multiagent debate[C]//Proceedings of the 41st International Conference on Machine Learning. Vienna, Austria: ACM, 2024: 11733-11763.
- [8] LU J Y, YU L, LI X J, et al. LLaMA-reviewer: advancing code review automation with large language models through parameter-efficient fine-tuning[C]//2023 IEEE 34th International Symposium on Software Reliability Engineering (ISSRE). Florence, Italy: IEEE, 2023: 647-658.
- [9] 刘雪颖, 云静, 李博, 等. 基于大型语言模型的检索增

- 强生成综述[J]. 计算机工程与应用, 2025, 61(13): 1-25.
- LIU Xueying, YUN Jing, LI Bo, et al. Survey of retrieval-augmented generation based on large language models[J]. Computer Engineering and Applications, 2025, 61(13): 1-25.
- [10] HINDI M, MOHAMMED L, MAAZ O, et al. Enhancing the precision and interpretability of retrieval-augmented generation (RAG) in legal technology: a survey[J]. IEEE Access, 2025, 13: 46171-46189.
- [11] WANG L, MA C, FENG X Y, et al. A survey on large language model based autonomous agents[J]. Frontiers of Computer Science, 2024, 18(6): 186345.
- [12] WEI Q, SUN H J, XU Y, et al. Exploring the application of large language models based AI Agents in leakage detection of natural gas valve chambers[J]. Energies, 2024, 17(22): 5633.
- [13] 中国产品全生命周期温室气体排放系数库[DB/OL]. (2025-05-12)[2025-09-04]. <http://lca.cityghg.com>
- [14] 沈思, 冯暑阳, 吴娜, 等. 融合大语言模型的政策文本检索增强生成研究[J]. 数据分析与知识发现, 2025, 9(9): 37-48.
- SHEN Si, FENG Shuyang, WU Na, et al. Retrieval-augmented generation of policy texts based on large language models[J]. Data Analysis and Knowledge Discovery, 2025, 9(9): 37-48.
- [15] 齐俊, 曲睿婷, 教传铭, 等. 基于知识图谱增强大语言模型双碳领域服务[J]. 计算机与现代化, 2024(9): 8-14.
- QI Jun, QU Ruiting, JIAO Chuanming, et al. Enhanced big language model dual carbon domain services based on knowledge graph[J]. Computer and Modernization, 2024(9): 8-14.
- [16] Qwen Team. Qwen2-Math, 新一代数学模型[EB/OL]. (2024-08-08)[2025-09-04]. <https://qwenlm.github.io/zh/blog/qwen2-math/>
- [17] 张天鸿, 王晓玲, 余红玲, 等. 基于大语言模型的灌浆工程知识服务系统[J]. 水利学报, 2025, 56(1): 130-142.
- ZHANG Tianhong, WANG Xiaoling, YU Hongling, et al. Grouting works knowledge service system based on large language model[J]. Journal of Hydraulic Engineering, 2025, 56(1): 130-142.

(编辑:孙亚彤)

(上接第96页)

- [24] HÜNGER K J. The contribution of quartz and the role of aluminum for understanding the AAR with greywacke [J]. Cement and Concrete Research, 2007, 37(8): 1193-1205.
- [25] 龚青南, 王德辉. 不同离子对混凝土碱硅酸反应影响的研究进展[J]. 材料导报, 2024, 38(2): 89-103.
- GONG Qingnan, WANG Dehui. Research progress on the effects of different ions on alkali-silica reaction of concrete[J]. Materials Reports, 2024, 38(2): 89-103.
- [26] 武汉地质学院矿物教研室. 结晶学及矿物学:上册[M]. 北京:地质出版社, 1979: 91-95.
- [27] 魏丽丽, 胡明玉, 郑江, 等. 陶瓷抛光渣对水泥基材料碱集料反应的影响[J]. 硅酸盐学报, 2018, 46(11): 1568-1574.
- WEI Lili, HU Mingyu, ZHENG Jiang, et al. Effect of ceramic polishing powder on alkali-aggregate reaction of cement-based materials[J]. Journal of the Chinese Ceramic Society, 2018, 46(11): 1568-1574.
- [28] LOTHENBACH B, BERNARD E, MÄDER U. Zeolite formation in the presence of cement hydrates and albite [J]. Physics and Chemistry of the Earth, Parts A/B/C, 2017, 99: 77-94.
- [29] 董伟, 周梦虎, 王雪松, 等. 碳化-冻融作用对风积沙混凝土氯离子传输的影响[J]. 山东大学学报(工学版), 2024, 54(1): 123-130.
- DONG Wei, ZHOU Menghu, WANG Xuesong, et al. Effect of carbonation freeze-thaw on chloride ion transport in aeolian sand concrete[J]. Journal of Shandong University(Engineering Science), 2024, 54(1): 123-130.

(编辑:郭少华)