

第 16 届大学生服务外包创新创业大赛 A15 赛题 参赛作品

“第十章”项目概要 介绍

基于动态 RAG 和大模型的算术与数学问题解答系统

十方算阁 T2403770

2025-4-6

目录

前言	2
一、 目标回顾	2
二、 系统功能简介	2
三、 创新点说明	3
1. 深度融合大模型与动态 RAG 技术	3
2. 构建垂直领域的知识库	3
3. 注重用户体验和交互设计	4
4. 支持图片识别与自定义知识库	5
四、 特色综述	5
五、 开发工具与技术	5
六、 应用对象与应用环境	6
1. 应用对象	6
2. 应用场景	6
七、 结语	6

前言

代数作为数学学科的重要基石，在高科技人才培养中占据核心地位。然而，传统的代数教学模式和解题软件长期面临着智能化、个性化不足的挑战。党的十八大以来，人工智能被提升到战略高度，实现教育智能化转型已成为必然趋势。

本项目融合大模型与动态检索增强生成技术开发了“第十章”智能代数解答系统，旨在成为中小学生的“AI 代数导师”，通过大模型强大的语义理解和推理能力，为学生提供更精准、更深入、更个性化的代数学习辅导。

下文将对本项目进行概要介绍。

一、目标回顾

本次项目目标在于系统化整理代数领域的数据资源，构建结构化知识库，并依托本地部署大语言模型与检索增强生成技术开发智能问答系统，其服务定位于高中及以下学生的“AI 代数导师”，解决当今学生代数学习过程中遇到问题时，得不到及时、精准、个性化的服务问题。具体目标如下：

1. 前期调研^[1]。
2. 构建知识库^[2]。
3. 本地部署大模型，通过 RAG 技术，搭建网页端对话系统。该系统需要：
 - a) 支持至少两轮的问答交互；
 - b) 具有友好的交互界面，问答响应时间不超过 10 秒；
 - c) 获取相关问题列表，帮助用户深入研究；
 - d) 检索参考文件，帮助用户查阅额外信息或支持材料；
 - e) 允许上传图片，回答图片中的问题；
 - f) 允许上传用户自定义的知识库文件；

二、系统功能简介

本组完成了上述目标，开发的“第十章”系统主要提供两项服务：

代数领域问答服务可以解答包括但不限于代数题目与解法示例、代数概念、数学家信息、数学典籍信息等用户问题；

AI 赋能的代数领域学习服务可以根据用户填写的知识点表单生成代数课程大纲与教案，后台人员将根据 AI 教案来录制线上微型课程；

1 明确核心用户群体及其需求，分析行业形势、主要竞品与市场机会，探索技术可行性与法律合规性，初步思考商业模式与落地场景，并形成调研报告

2 通过网络、教材和其他资源整合和收集相关的算术与数学知识数据，包括但不限于基础运算、公式推导、常见定理、解题技巧等信息，数据量不少于 2000 条。

此外，系统还可用于帮助老师写教案、向古典数学爱好者介绍古代数学典籍与数学成就、向代数研究者介绍高难度代数问题等。

三、创新点说明

1. 深度融合大模型与动态 RAG 技术

该系统并非简单地调用大模型或使用传统的 RAG 技术，而是使用推理引导的动态 RAG 技术^[3]，与传统 RAG 相比增加了“推理用户意图”的环节。通过大模型先推理用户意图，再基于结果检索相关知识，可以更精准地定位所需知识，降低检索噪声，降低用户自然语言导致的不准确性^[4]，提升后续生成质量，大大提高整个系统的可靠性。

传统RAG与推理引导的动态RAG流程对比图

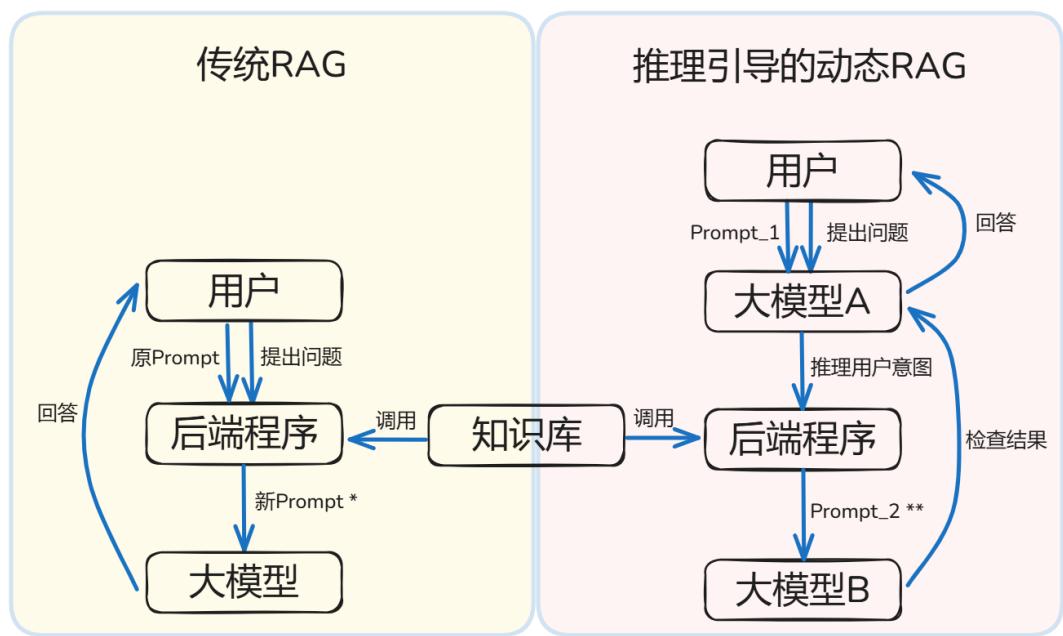


图 1 传统 RAG 与推理引导的动态 RAG 流程对比图

2. 构建垂直领域的知识库

系统并非通用型 AI 助手，而是专注于代数学习领域，构建专业的代数知识库，并针对知识问答与日常交流做 Prompt 区分，提升了系统在代数领域的专业性和精

3 具体的动态 RAG 技术发展及其与传统 RAG 技术的比较见项目详细方案-5.3 动态 RAG 技术解释。

4 关于用户自然语言导致不确定性，在项目详细方案 - 4.3.1 深度融合大模型与动态 RAG 技术 中有举例。

准度。

通用型AI助手与专业型AI助手训练数据对比示意图（简化版）

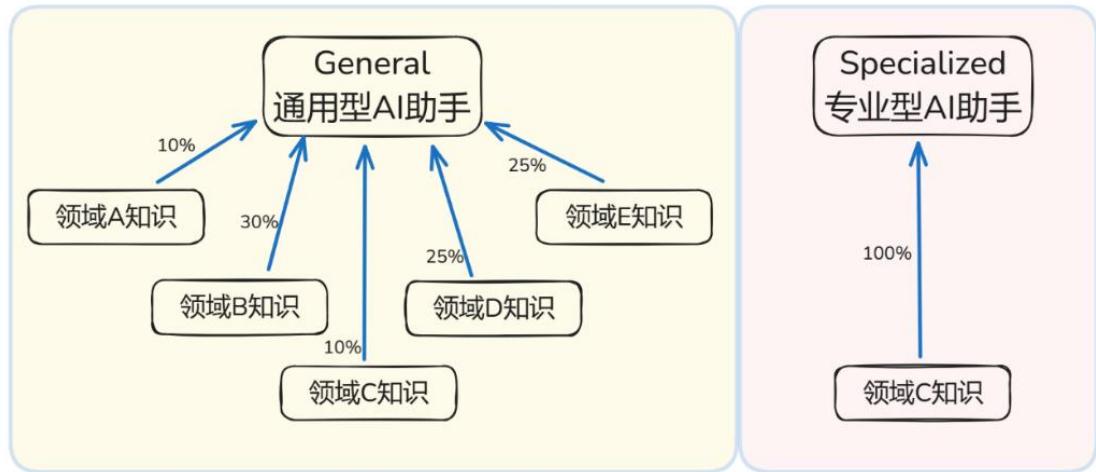


图 2 通用型 AI 助手与专业型 AI 助手训练数据对比示意图（简化版）

3.注重用户体验和交互设计

本系统具有简洁友好的用户界面，采用流式数据传输技术，提升用户体验。

传统输出与流式输出流程对比图

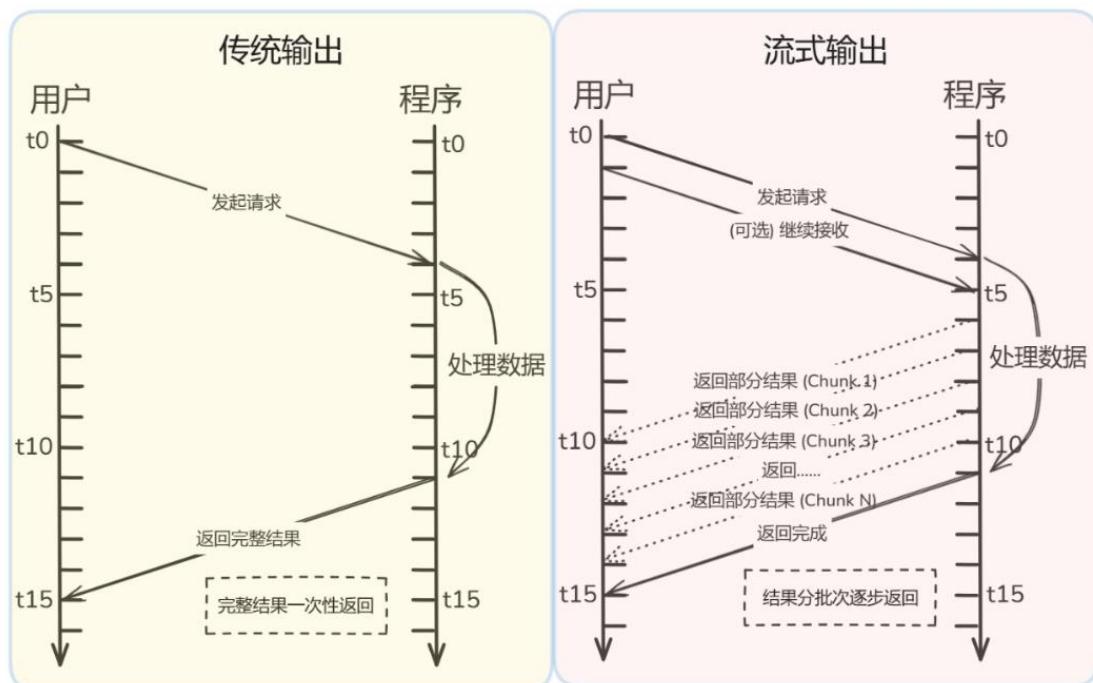


图 3 传统输出与流式输出流程对比图

4. 支持图片识别与自定义知识库

本系统支持用户直接上传代数题目的图片与上传自己的知识库文件^[5]。

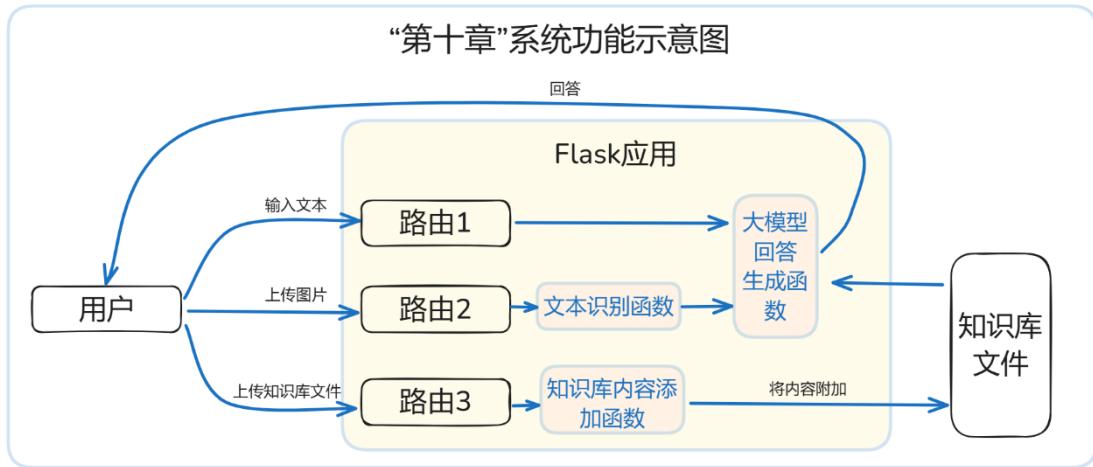


图 4 “第十章” 系统功能示意图

四、特色综述

技术上，“第十章”深度融合大模型与推理引导的动态 RAG 技术，并构建代数领域的垂直数据库，克服撞库与传统 RAG 技术的检索精度不足、容易出现知识错误或逻辑混乱的问题。在用来测试的 100 道高中及以下代数题目中，“第十章”正确率达到 87%，用以对比的无 RAG 技术的 QWen2.5-7B 仅有 77%，而响应时间仅慢不到 20%^[6]。

功能上，“第十章”不仅支持用户输入来与大模型交互，而且允许用户上传图片，与自己的知识库文档，获得更好的交互体验与个性化解答。

此外，本项目小组并非只进行了系统产品搭建，而是系统进行了用户画像分析、市场与竞品分析、用户付费意愿调研、小范围测试与用户满意度调研等一系列工作，并提供了项目时间规划、工作安排与工作方法等经过团队实践的内容，力求为企业提供具有实践价值的先验验证方案^[7]。

五、开发工具与技术

工具：VS Code、ollama、各 Python 库^[8]

语言：

前端：原生 Web 技术（HTML5、CSS3 和 JavaScript）

后端：Python 轻量级框架 Flask

5 有格式要求的.csv 文件。

6 同等算力条件下测试。

7 详见项目详细方案-2 项目规划与工作安排回顾。

8 Python 库及版本可见附件-项目源文件-START- requirements.txt.

技术：

1. 系统架构：Browser-Server；
2. 大模型：本地部署 Qwen2.5-1.5B-Instruct-GGUF:Q8_0 与 Qwen2.5-7B-Instruct-GGUF:Q8_0；
3. 核心：“推理引导”的动态 RAG 技术；
4. 数据搜集：人工搜集与爬虫结合；
5. 提取关键字：利用大模型，提取知识库条目关键字^[9]，用以更好地转为向量。
6. 数据处理与文本向量化：使用 jieba 库等进行数据处理，使用开源预训练模型 Sentence-BERT (all-MiniLM-L6-v2) 进行文本向量化；
7. 数据存储：FAISS 开源向量数据库；
8. 结果输出：通过流式数据传输技术实时、分段地输出至用户交互界面；
9. 渲染：本地部署 MathJax 与 Markedjs 实现 LaTeX 公式和 Markdown 的显示；
10. 安全防护：使用 DOMPurify 防止 XSS 攻击；
11. 图片处理：使用 base-64 编码传输图片数据，在后端进行文本识别；
12. 文件上传：在后端接收用户文件并附在原知识库文件之后。

六、应用对象与应用环境

1. 应用对象

面向高中及以下的代数学习者。

学习代数的过程伴随着晦涩复杂的公式和解题思路，然而传统搜题软件只提供正确答案，无法进行更深层的交流以深化理解。根据团队的调查结果^[10]，学生群体已经开始意识到并接受 AI 在解答思路的清晰度、更智能的互动体验等优势，但目前市面上^[11]多位通用 AI 助手，没有专为该领域构建的专业 AI。而“第十章”系统具有及时、专业、个性化、交互性强的特点，以用户的“AI 代数导师”为定位解决这一痛点。

2. 应用场景

家庭作业辅导、复习、课堂互动、教案编写辅助等。

七、结语

“第十章”智能代数解答系统验证了大模型与动态 RAG 技术在代数教育领域的

9 该程序见附件-5_提取关键字程序.py.

10 具体结果见项目详细方案-3 项目前期调研与分析。

11 具体市场及竞品分析见项目详细方案-3 项目前期调研与分析。

应用潜力，为构建智能化、个性化、高效化的代数学习平台提供了有效的解决方案。随着人工智能技术的不断进步和应用深入，“第十章”类似系统将会在推动代数教育智能化转型、提升学生代数学习效率和兴趣方面发挥越来越重要的作用，最终迈向智能化代数教育的新时代。