

基于双模型的编程学习辅助系统设计与应用研究

Design and Application of a Dual-Model-Based Programming Learning Assistance System

伍一宋佳, 刘杰, 柏寒青, 卢宇*
北京师范大学教育学部教育技术学院
*luyu@bnu.edu.cn

【摘要】 本研究聚焦生成式人工智能在编程教育领域的应用, 提出并构建了基于双模型的编程学习辅助系统。该系统由基础大模型与代码大模型双引擎驱动, 旨在为学生提供全方位的编程学习支持。基础大模型通过逐步引导学生学习编程知识, 促进深度交互, 培养高阶思维能力; 代码大模型则提供代码生成、补全和注释等核心功能, 助力学生高效完成编程。实验表明, 该系统显著提升学习者的代码撰写质量, 并获得了学习者的认可。此外, 系统在帮助学生理解编程概念、提高学习效率等方面发挥了重要作用, 为生成式人工智能支持编程教育提供了创新型设计和实践范例。

【关键词】 生成式人工智能; 大模型; 编程教育; 学习分析

Abstract: This study focuses on application of generative artificial intelligence in programming education, proposing and constructing a dual-model-based programming learning assistance system. The system is driven by a dual-engine consisting of foundation model and code-specific large model, aiming to provide comprehensive support for learners. The foundation model guides students step-by-step in learning knowledge, facilitates in-depth interaction, and cultivates higher-order thinking skills. The code-specific large model offers core functionalities such as code generation, completion, and annotation to help students efficiently complete program. Experiment demonstrate the system significantly improves learners' code quality and has received widespread recognition. Additionally, the system significantly aids in helping students understand programming concepts and enhancing efficiency, providing a innovation design and practical example of generative AI-supported programming education.

Keywords: Generative artificial intelligence, Foundation model, Programming education, Learning analytics

1. 引言

编程教育作为培养学生计算思维和逻辑思维的重要手段, 近年来受到广泛关注。生成式人工智能作为通过人工智能相关技术, 自动化生成文本、图像、视频、音频等多类型内容的技术(卢宇等, 2023), 在编程领域展现了巨大的潜力。通过使用大模型, 即在预训练中学习大规模语料库并且形成强大基础能力的基础大模型(张熙等, 2023), 可以提高学生编程效率。但是现有编程学习辅助多停留在工具层面, 学生可能会过度依赖大模型生成代码(孙丹等, 2024), 缺乏深入学习思考; 大模型生成的代码不能保证准确性、健壮性和完整性等(徐慧等, 2023)。

为此, 本研究自主研发了基于双模型的编程学习辅助系统。通过基础大模型, 逐步引导学生深入学习编程, 培养高阶思维; 通过代码大模型, 为学生提供编程辅助, 提高代码辅助的准确性与科学性。在双模型驱动的基础上, 为编程学习提供更加专业高效的环境与支撑。

2. 编程学习辅助系统设计与实现

2.1. 系统架构与功能实现

系统架构设计如图 1 所示。在引导学生学习编程知识以及理解问题时，系统流式调用基础大模型 GLM-4⁴。该模型参数量大，拥有较强的语言理解和生成能力，并且能够理解长文本中的上下文关系，回答的准确性较高(Team Glm et al., 2024)。在实现具体代码编程辅助时，系统选用代码大模型 codegeex-4¹，为学生提供纠错建议以及代码注释。该模型支持代码补全生成、代码解释器等全面功能，覆盖了编程的多种场景(北京智谱华章科技有限公司，2022)。

系统前端采用了 HTML、CSS 和 JavaScript 技术进行搭建,后端则选取了 Flask 框架。Flask 作为轻量级 Web 框架，通过定义路由处理客户端请求。其内置开发服务器会自动匹配 URL 路径并执行对应视图函数，返回响应结果。针对大模型生成内容的长文本场景，采用服务器发送事件（SSE）技术实现流式传输，允许服务器主动推送生成内容至客户端。

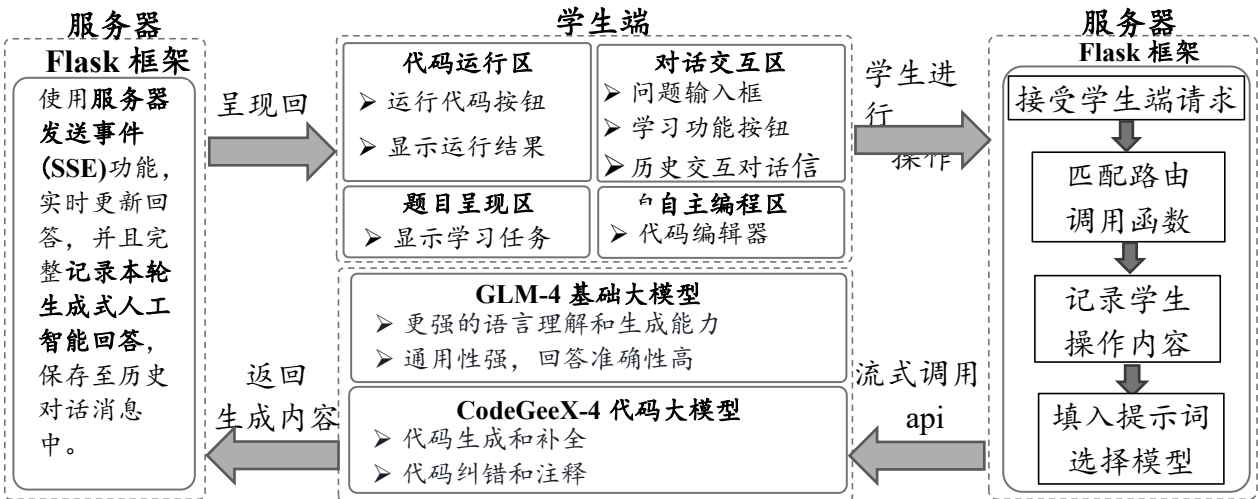


图 1 编程学习辅助系统架构图

基于以上系统架构，本研究设计并实现了以下系统功能。在学生端，学生可根据编程学习任务以及问题，输入疑问或选择系统预设的学习功能按钮，以此与生成式人工智能进行交互，获得个性化学习辅导。同时，学生可在系统的在线代码编辑器中编写和测试代码，以完成学习问题。在服务器端，一旦学生提交问题或点击功能按钮，服务器将自动记录学生的问题，并嵌入适当的提示词，进而调用相关大模型的 API 服务，获得针对性的专业解答，并将其反馈至学生端。在整个过程中，服务器将记录每一轮的交互信息以及操作行为。

2.2. 提示词框架设计

在将大模型运用于编程辅助学习的过程中，为了实现更加准确、科学辅助，本研究构建了一个结构化的提示词框架。该框架由五个核心元素构成，每个元素的定义如表 1 所示。

表 1 提示词元素及其具体含义

提示词元素	具体含义
角色	设置大模型扮演的角色以及听众角色，使其生成更符合场景的回答。
任务	描述大模型需要完成的任务内容、步骤和过程，使其回答更加稳定。
知识参考	提供辅助编程学习所需要的相关知识内容，使大模型回答更加精准。
限制	提供大模型回答需要注意的限定条件，以规范大模型的回答内容。
示例	提供辅助案例，帮助大模型理解任务要求，以获得符合期望的输出。

3. 实验设计

3.1. 案例与实验流程设计

⁴ 智谱 AI 开放平台，<https://bigmodel.cn/>

基于以上系统架构，本研究选取汉诺塔问题的递归算法学习作为应用案例，设计实现了以下功能：(1)知识性学习：利用大模型为学生讲解编程任务所需的核心知识点；(2)解题思路与编写提示：利用大模型为学生提供解题思路和代码框架；(3)代码编写辅助：代码纠错，识别并指出学生代码中的错误点和行数；代码注释，帮助学生提高代码的质量。

同时，本研究依据以上提示词框架，针对不同编程辅助设计提示词，最终构建出编程学习辅助系统，并且设计实验探究其对大学生编程能力的提升效果，以及学生的行为使用模式。

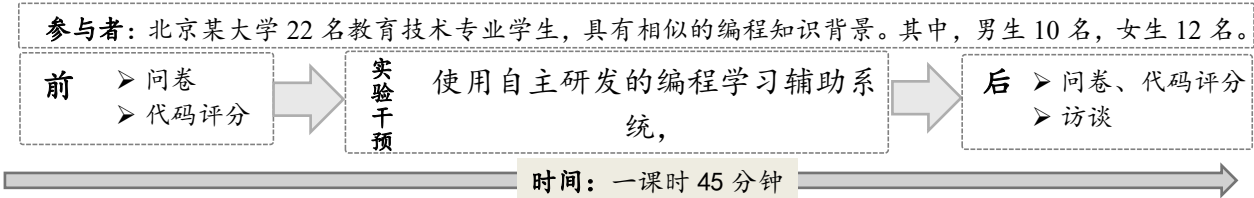


图 2 实验设计流程图

3.2. 研究方法与工具

在前后测中，本研究通过调查问卷的形式收集数据，该问卷包括基本信息部分，5 道递归思想选择题和张屹老师(2024)发布的计算思维量表(包含 15 道题)，以及后测增加的 7 道技术接受度题目。经检验，计算思维量表以及技术接受度量表具有良好的信度。并且本研究开发了代码评分框架，对学生使用前以及使用后提交的代码进行评分，以评价系统的有效性。

本研究采用滞后序列分析法分析学生使用过程。依据商定的评判标准进行编码，共得到 22 组有效样本。取代码评分前 30%为高分组(共 6 人)，评分后 30%为低分组(共 5 人)，生成各组行为转换图，从而对比不同成绩学生学习行为模式。其中操作编码框架如表 2 所示。

表 2 操作编码框架

编码	行为	编码	行为	编码	行为	编码	行为
RC	运行代码	SC	提交代码	LR	学习递归	TU	思路提示
PT	编程提示	EC	代码纠错	CA	代码注释		

并且，本研究在实验结束后采用结构化访谈，对参与实验的人员围绕系统使用体验、问题解决效果和未来使用意愿三个方面展开访谈，随后对访谈内容进行了整理、编码和分析。

4. 实验结果

4.1. 使用前后学生的代码质量分析

对使用前后学生代码评分进行描述性分析和配对样本 t 检验。结果表明，使用前后学生代码评分存在显著性差异($p<0.001$)，使用后较使用前平均值提升 2 分。由此可见，本研究所设计的编程辅助系统在提升学生代码编写质量方面发挥了积极作用。

4.2. 行为模式分析

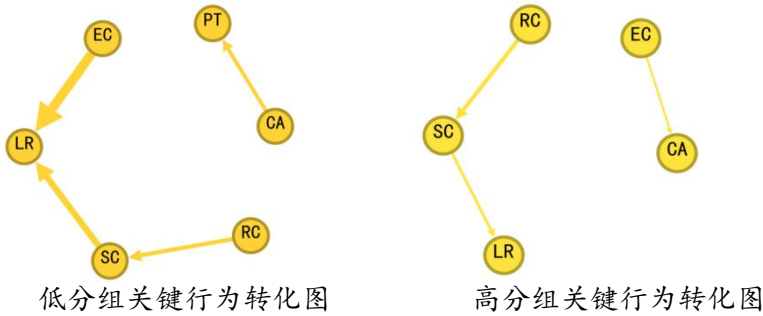


图 3 不同水平学生的关键行为转化图

本研究使用 GSEQ 5.0 和 Gephi 软件获取并绘制高分组与低分组学习者编程学习中的行为序列图。对比两图,研究者发现两组学习者都在独立尝试编写代码后,再选择与系统进行交互,表明都有自主编写代码与独立思考的意识;在“提交代码”之后,两组学习者都倾向于点击“学习递归”。这反映了人工智能辅助编程学习下依然不能忽视基础知识的引导。

但在解决编程问题的方式上两组学习者呈现出差异。高得分、高完成度的学习者更注重利用编程辅助系统的代码纠错与代码注释功能,进行编程的规范和习惯的培养;而低得分、低完成度的学习者则更依赖系统提供的提示与辅助资源来克服学习障碍。

4.3. 使用意愿分析

本研究依据扎根理论对访谈内容进行编码,发现该系统可以辅助学生理解、促进学习,但是仍存在网络、交互设计等问题。并且,对使用后技术接受度问卷进行描述性分析,其中有用性的平均值为 3.52,易用性的平均值为 3.64,可见学生对系统给予较高评价。同时,本次实验系统设置使用意愿的二元选择问卷,其中有效反馈样本数为 10,愿意继续使用占比为 60%。这表明较多学生对当前系统的使用持积极态度,愿意使用该系统辅助自己的编程学习。

通过对以上内容综合分析,得出以下结论:编程辅助系统在编程学习中起到了较为重要的作用,通过提供编程思路、错误提示等,帮助用户提高编程效率,这也是许多学生愿意继续使用该系统的原因。但在用户体验上仍存在问题,例如网络连接时有问题,影响使用体验;注释与编程内容的区分不够清晰,容易混淆,这也可能导致部分用户不愿继续使用。

另外,对问卷结果进行分析可知,学生在使用该系统前后,递归思想知识测试的成绩和计算思维的整体水平没有显著变化,这可能因为编程学习内容局限于单一递归算法案例,缺乏多场景知识迁移训练,并且实验时间过短无法对学生产生显著影响有关。

5. 总结

本研究围绕基于双模型的编程学习辅助系统的构建与应用展开。通过调用基础大模型与学生进行交互和知识性学习辅助,通过调用代码大模型对学生进行代码编写辅助,并且开展了系列实验,验证其有效性。具体结论如下:基于双模型的编程辅助系统在技术实现上可行,有效提升学生撰写代码质量;不同能力水平的学习者在解决编程问题的模式上呈现出差异性,可以提供个性化辅助;系统在帮助学生理解编程概念、提高效率等方面起到较为重要的作用。

另外,本研究也存在不足之处,例如研究选取样本过少,实验时间较短,案例单一,需要持续分析辅助系统对学生编程学习的长期效果。

致谢

本论文受到国家级大学生创新创业训练计划资助(项目编号:202410027030)。

参考文献

- 北京智谱华章科技有限公司.(2022).CodeGeeX 大模型:让 AI 帮你写代码.智谱.
<https://www.zhipuai.cn/news/42>
- 卢宇,余京蕾,陈鹏鹤 & 李沐云.(2023).生成式人工智能的教育应用与展望——以 ChatGPT 系统为例.中国远程教育(04),24-31+51.
- 孙丹,朱城聪,许作栋 & 徐光涛.(2024).基于生成式人工智能的大学生编程学习行为分析研究.电化教育研究(03),113-120.
- 徐慧,鞠小林 & 王皓晨.(2023).大模型下编程教学面临的挑战与应对.计算机教育(11),60-64.

- 张熙,杨小汕 & 徐常胜.(2023).ChatGPT 及生成式人工智能现状及未来发展方向.中国科学基金(05),743-750.
- 张屹, 陈邓康, 付卫东, 刘金芳,林裕如 & 丁双婷.(2024).基于新课标的中小學生计算思维量表构建研究.电化教育研究(03),90-98.
- Team Glm. (2024). ChatGLM: A family of large language models from GLM-130B to GLM-4 all tools. arXiv.Org. <https://arxiv.org/abs/2406.12793>