

人工智能大语言模型在国际中文写作教学作文评改中的应用研究

莫愁

北京语言大学语言科学与资源学院、语言资源高精尖创新中心, 中国·北京 100083

摘要: 人工智能大语言模型凭借其高效的自然语言处理能力, 为留学生作文评改提供了全新的技术路径, 在提升评改效率、优化反馈质量等方面展现出明显优势。然而, 本文指出大语言模型评改作文也面临评分稳定性不足且高阶语言能力评估性能薄弱、文化适配性欠缺、反馈深度不足、数据安全与伦理风险等问题。通过优化模型训练和提示策略、融合多模型协同评改、加强文化适配性、优化反馈生成机制、加强隐私数据处理能力等路径, 可提升大语言模型在作文评改中的应用效果, 实现技术赋能与教学需求的精准匹配。

关键词: 人工智能; 大语言模型; 国际中文教学; 作文评改

A Study on the Application of Large Language Models in the Evaluation and Revision of Compositions in International Chinese Writing Teaching

Mo Chou

School of Linguistics and Language Resources (Institute of Linguistic Sciences), Beijing Advanced Innovation Center for Language Resources, Beijing Language and Culture University, China Beijing 100083

Abstract: Large language models, with their efficient natural language processing capabilities, offer a novel technological approach for evaluating and revising compositions written by international students. They demonstrate significant advantages in improving evaluation efficiency and optimizing feedback quality. However, this paper points out that using large language models for composition evaluation and revision also faces several challenges, including insufficient scoring stability, weak performance in assessing high-level language skills, lack of cultural adaptability, inadequate feedback depth, and risks related to data security and ethics. By optimizing model training and prompting strategies, integrating multi-model collaborative evaluations, enhancing cultural adaptability, refining feedback generation mechanisms, and strengthening privacy data processing capabilities, the application effectiveness of large language models in composition evaluation and revision can be improved. This will enable a more precise alignment between technological empowerment and teaching needs.

Keywords: Artificial intelligence; Large language models; International Chinese teaching; Composition evaluation and revision

0 引言

写作教学是国际中文教学的重要组成部分, 教师一般在课堂上讲解作文范例, 提供写作脚手架, 并要求学生根据课上所学内容和脚手架在课后完成作文任务。教师通过评改作文来了解课堂的教学效果(许国萍、王一平 2002)。这一模式也能帮助留学生提高写作能力, 学生从反馈者处得到启发, 逐渐培养其自我表达能力(王颖 2007)。然而, 作文评改涉及语言规范、逻辑思维、文化审美等方面的指导, 内容复杂, 耗时较长, 一直面临教师、学生“两头难”的问题。一方面, 由于写作涉及字、词、句、篇章等多方面的语言知识, 教师在点评、纠正学生错误时需要注意各类细节问题, 且需要根据不同学生的特点

进行个性化评价, 导致教师评改任务过重、身心俱疲(翟艳、苏英霞 2022: 414)。另一方面, 由于传统写作教学经常面临课时量不足、师少生多的问题, 因此教师通常在新课集中点评讲解, 这种“写评分离”的方式使得学生无法在写作后第一时间获得评价反馈, 延缓了学生写作能力的发展(马瑞凌、徐娟 2024)。

人工智能技术对教育领域的持续影响使得广大国际中文教师和中文学习者不再满足于传统的教学形态, 亟需各种基于人工智能技术的教学资源 and 模式以满足其多样化、个性化的需求。以 ChatGPT、DeepSeek、豆包等大语言模型(Large Language Model, 以下简称“大模型”)为代表的生成式人工智能技术为语言教学创新提供了广大的发展

空间。大模型强大的语义理解与文本处理能力，能很好地破解国际中文教学中写作评改的痛点，为作文评改提供了智能化解决方案。本文将结合文献梳理大模型在作文评改中的应用优势，并指出这一应用所面临的问题和相应的优化路径，为后续研究提供参考。

1 大模型在留学生作文评改中的应用优势

大模型具备高效处理海量文本的能力，可在数秒内完成单篇作文的多维度评估，大幅提升了评改效率。Han & Li (2024) 指出 ChatGPT 反馈速度快，能有效减轻教师的评阅负担，在写作教学中有极高的应用潜力。刘瑜等 (2024) 采用 ChatGPT 对汉语学习者的日记纠错，验证了大模型修正基础错误（字、词、标点、语法错误）的可行性，肯定了大模型的纠错反馈能力。大模型的规模化处理能力能高效支持考试等批量评阅场景，可快速生成评分结果与反馈报告，为教学活动提供及时支持。同时，大模型“可以突破时空限制为学生提供多样即时反馈”（魏爽、李璐遥 2023），尤其适合师生时区不同步的线上课堂，留学生能在课下利用大模型自行批改作文，满足留学生自主训练、即时纠错的需求，促进其提升自主学习能力。

大模型还能与教师在评改工作中形成优势互补，提升评改质量。Al-Garaady & Mahyoub (2023) 发现 ChatGPT 在识别表层偏误方面优于教师，而教师则更能准确识别内容与语用方面的偏误。李韵等 (2025) 发现教师的单独反馈可能受到“成见效应”影响，使用 ChatGPT 协助评改可以减轻这种影响，且人机协同评改的反馈比传统的教师反馈在内容上更加细致与丰富，能更好地帮助学生维持学习动机。薛嗣媛和周建设 (2024) 指出大模型生成的评语倾向于问题本位，即快速定位学生的写作问题并给予反馈，而教师的评语则倾向于学生本位，偏好给出间接反馈，重视对学生思考能力的培养，因此建议教师以大模型的评语为起点，根据学生需求进行深入指导。

2 大模型在留学生作文评改中的应用问题

虽然大模型有上述应用方面的优势，但在教学实践中大模型的评改反馈仍存在如下问题：

2.1 评分稳定性不足，高阶语言能力评估性能薄弱

尽管大模型在基础语言能力评分中表现较好，其评分结果由于同时受参数设置、提示策略、作文主题与语言难度的影响，评分稳定性仍有待提升。刘瑜等 (2024) 指出虽然大模型具有基本的纠错反馈能力，但也会过度修改甚至越改越错，造成学习者对所学词汇和结构的困惑。此外，由于高阶语言能力（论证能力、创新能力、批判性思维能

力等）的评估依赖主观判断与语境理解，大模型难以精准把握留学生作文中的隐含逻辑与思想深度，在高阶语言能力评估方面存在明显短板，易出现评分偏差。譬如，对于视角独特但表达不够流畅的作文，大模型可能因语言错误而低估其思想价值，从而给出低分；对于观点陈旧但是论证严谨的作文，大模型可能给出过高评分。此外，大模型对修辞手法、文风个性等主观性较强的维度也缺乏有效的评估标准与方法，还需要人工辅助判定。

2.2 文化适配性欠缺，影响评改准确性和公正性

不同文化背景的留学生其写作表达习惯、逻辑思维方式存在显著差异。譬如，美国文化倾向于直截了当表达观点，然而中国文化则偏好含蓄委婉，强调“言有尽而意无穷”。中国传统学术写作允许一定程度的主观情感融入，而西方的学术写作传统则强调客观中立。大模型的预训练语料多以英语等常见语言为主，对使用人数较少的语言训练不足，因此在跨文化表达的适配性方面仍有欠缺，可能将文化表达的差异误判为表达错误。譬如，大模型可能将留学生基于自身文化背景提出的独特观点判定为主题偏离，将符合其文化习惯的段落衔接方式判定为逻辑混乱。这种文化适配性的欠缺会影响评改结果的准确性与公平性，因此教师有必要审核大模型的评改结果，防止大模型的文化误判影响学生的写作积极性。

2.3 反馈缺乏深度，不利于学生反思

现有大模型生成的反馈多数仅指出表面问题，缺乏深入的专业建议。譬如，大模型倾向于给出直接反馈，比如告知某句中某个词需要换成另一个词，却不主动告知换词理由；针对语法错误，仅给出修改结果，却不解释相关语法规则；能指出某段落“逻辑不清晰”，但不说明具体的语篇问题。此外，大模型的反馈存在泛化问题，即不考虑留学生的个体差异而给出内容相似的评语，难以满足学生的个性化学习需求。总体而言，大模型的反馈深度不够，在实用性、针对性方面与教师反馈存在明显差距，尤其在修正复杂错误方面有待优化。

2.4 数据安全问题与伦理风险凸显

留学生作文包含学生的个人经历、思想观点等隐私信息，使用大模型开展评改时，需将作文文本上传至大模型系统，存在隐私数据泄露风险。若模型服务商未建立完善的数据安全保障机制，可能导致留学生隐私信息被滥用或泄露，侵犯其合法权益。同时，大模型评改的伦理风险也不容忽视。过度依赖大模型评改可能导致教师弱化对留学生写作能力的主观判断，忽视个性化教学需求；大模型评

分结果若不准确,可能会打击学生的积极性,甚至影响师生关系。

3 大模型赋能留学生作文评改的优化路径

针对大模型在留学生作文评改工作中存在的上述问题,本文试提出如下优化路径:

3.1 优化大模型训练和提示策略,融合多模型协同评改,提升评分稳定性与精准度

在未来可以通过优化大模型参数来提升大模型评分的稳定性,还可通过优化提示策略,借鉴多种提示方法(薛嗣媛、周建设 2024),设计针对性提示语提升大模型对高阶语言能力评估的精准度。譬如,针对批判性思维能力,可通过提示语引导大模型分析论点合理性、论据有效性、论证逻辑等维度来进行客观全面的评分。

此外,鉴于单一模型的局限性,评改工作还可以融合多模型协同进行。教师可采用多个大模型协同工作的模式,综合各模型的评改结果,通过加权平均等方式生成精准度更高的最终评分。

3.2 加强文化适配性,构建跨文化评改标准

针对大模型文化适配性不足的现状,可以构建留学生作文语料库来进一步训练大模型。收集不同母语背景、不同语言水平的留学生作文语料,内容涵盖日常写作、学术写作、跨文化主题写作等多种类型,通过人工标注多维度评改标签,用于大模型微调。在训练语料中加入具有留学生写作特色的语料,提升大模型对二语表达、跨文化写作的适配性,减少评分偏差。

此外,教师还可以训练大模型的文化认知能力。教师选择跨文化交际理论与案例材料,让大模型学习上述材料,获得不同文化在表达习惯、思维方式与价值观念方面的知识,提升大模型对二语写作的理解能力。同时,邀请跨文化教学专家参与大模型校准,在现有评改标准基础上添加跨文化表达适配性维度,明确不同文化背景下的表达差异与容忍度。譬如,针对不同文化的写作逻辑差异,教师可以制定弹性评估标准,尊重留学生的文化表达习惯。同时,对涉及文化差异的评改结果进行复核,确保评改公平性。

3.3 优化反馈生成机制,提升反馈深度与实用性

根据留学生的语言水平与写作问题类型,教师可以构建“基础-进阶-高阶”分层反馈体系。在基础层面,针对词汇、语法等错误,要求大模型提供修改建议与规则解释;在进阶层面,针对逻辑表达、语篇构架等问题,要求大模型提供具体修改思路与示例参考;在高阶层面,针对观点创新、跨文化表达等技能,要求大模型提供拓展思路

与提升策略。

此外,通过“模型初评+人工复评”的人机协同模式可平衡作文反馈的质量与效率,确保反馈的深度与实用性。譬如,大模型完成语法纠错与基础评分后,教师聚焦逻辑深度、观点创新性等维度进行复评,确保评改的全面性。

3.4 加强隐私数据处理能力,规避数据安全与伦理风险

教师或留学生在上传作文文本前应自觉进行匿名处理或加密处理,防止隐私泄露。此外,教师应选择具备完善数据安全机制的大模型服务商,如有必要可以签订数据保密协议,加强数据安全保障。若条件允许,教师还可搭建本地化大模型评改系统,避免敏感数据上传至第三方平台。同时,有必要对师生进行大模型伦理风险教育,培养其合理使用大模型的习惯。

4 结语

大模型凭借其高效的自然语言处理能力,为留学生作文评改提供了全新的技术路径,在提升评改效率、优化反馈质量等方面展现出明显优势。然而,大模型也面临评分稳定性不足且高阶语言能力评估性能薄弱、文化适配性欠缺、反馈深度不足、数据安全与伦理风险等问题。通过优化大模型训练和提示策略、融合多模型协同评改、加强文化适配性、优化反馈生成机制、加强隐私数据处理能力等路径,可有效提升大模型在作文评改中的应用效果,实现技术赋能与教学需求的精准匹配。未来研究还可探索不同母语背景的留学生与不同大模型评改策略的适配性等问题。随着人工智能技术的不断发展,大模型在留学生作文评改中的应用将更加成熟,为国际中文教育高质量发展注入新动力。

参考文献:

- [1] 李韵,韩继宁,李雨瞳等. 人机双师协同写作反馈在汉语二语教学中的应用及其效果评估[J]. 语言教学与研究, 2025,(02):25-37.
- [2] 刘瑜,谭坦,陈青等. ChatGPT在汉语二语写作中的纠正反馈能力考察[J]. 国际汉语, 2024,(01):83-98+164-165.
- [3] 马瑞祯,徐娟. 国际中文写作智慧教学模式构建与应用探索[J]. 语言教学与研究, 2024,(02):12-23.
- [4] 王颖. 二语写作中的反馈研究述评[J]. 山东大学学报(哲学社会科学版), 2007,(05):48-52.
- [5] 魏爽,李璐遥. 人工智能辅助二语写作反馈研究——以ChatGPT为例[J]. 中国外语, 2023,20(03):33-40.

[6] 许国萍, 王一平. 对外汉语写作教学中的重要一环——谈作文评改的现状和对策[J]. 暨南大学华文学院学报, 2002,(02):35-39.

[7] 薛嗣媛, 周建设. 大语言模型在汉语写作智能评估中的应用研究[J]. 昆明学院学报, 2024,46(02):10-22.

[8] 翟艳, 苏英霞. 汉语技能教学(第2版)[M]. 北京: 北京语言大学出版社, 2022.

[9] Al-Garaady, Jeehaan & Mohammad Mahyoob 2023 ChatGPT's capabilities in spotting and analyzing writing errors experienced by EFL learners. *Arab World English Journal*,

Special Issue on CALL (9) : 3-17.

[10] Han, Jining & Mimi Li 2024 Exploring ChatGPT-supported teacher feedback in the EFL context. *System* 126:103502.

基金项目: 课题项目: 本研究受 2023 年中外语言交流合作中心国际中文教育研究课题一般项目“人工智能‘出题+测评’在国际中文教育领域的应用研究”(项目批准号: 23YH46C)、北京语言大学校级科研项目(中央高校基本科研业务专项资金)“汉语体标记语言知识库资源建设与技术研发”(项目批准号: 23YBB18)资助。