

基于图检索增强生成和少样本学习的美术作品鉴赏

刘天扬¹, 寇思佳², 金旭³, 王文静³, 陆雪松¹

1. 华东师范大学数据科学与工程学院, 上海 200062;
2. 教育部教育技术与资源发展中心(中央电化教育馆), 北京 100032;
3. 北京师范大学附属实验中学, 北京 100032

摘要

随着我国素质教育的不断推进, 美术教育在学科教育中的影响越来越大。美术作品鉴赏是美术教育的重要内容之一, 能够培养学生的美术能力和素养。然而, 优秀美术教师的缺乏和各地区美术教育发展水平的不平衡, 导致众多学生无法受到高质量的美术作品鉴赏教育。在这种情况下, 利用多模态大语言模型辅导学生进行美术作品鉴赏, 成为一种潜在的替代方案。利用多模态大语言模型, 提出一种基于图检索增强生成和少样本学习的方法, 引导模型生成符合高中教学需求的美术作品鉴赏内容。实验结果表明, 相较于对比方法, 该方法能够有效提升多模态大语言模型生成的鉴赏内容的质量。

关键词

图检索增强生成; 美术教育; 多模态大语言模型

中图分类号: TP391.1

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2025061

Art appreciation based on graph retrieval augmented generation and few-shot learning

LIU Tianyang¹, KOU Sijia², JIN Xu³, WANG Wenjing³, LU Xuesong¹

1. School of Data Science and Engineering, East China Normal University, Shanghai 200062, China
2. Center for Educational Technology and Resource Development, Ministry of Education, The People's Republic of China, Beijing 100032, China
3. The Experimental High School Attached to Beijing Normal University, Beijing 100032, China

Abstract

With the continuous advancement of quality education in our country, the influence of aesthetic education in subject education is becoming increasingly important. Appreciation of artworks is one of the important contents of aesthetic education, which can cultivate students' artistic ability and literacy. However, the lack of excellent art teachers and the imbalance of the development level of art education in various regions has led to many students being unable to receive high-quality art appreciation education. In this case, using multimodal large language models to tutor students in art appreciation has become a potential alternative. Using a multimodal large language model, this paper proposes a method based on graph retrieval augmented generation and few-shot learning to guide the model to generate art

appreciation content that meets the needs of high school education. Experimental results show that compared with the comparison methods, this method can effectively improve the quality of the appreciation content generated by the multimodal large language model.

Key words

graph retrieval augmented generation, art education, multimodal large language model

0 引言

随着我国素质教育的不断推进, 美术教育(简称美育)在学校教育中的地位日益凸显^[1]。2024年, 教育部出台《关于全面实施学校美育浸润行动的通知》指导性文件, 标志着我国美育体系建设迈入新阶段。美术作品不仅是艺术的载体, 更是传递文化和价值观的重要媒介。通过美术作品鉴赏教学, 教师能够有效地培养学生的艺术素养和对应能力^[2-3]。近年来, 随着人工智能技术的发展, 大语言模型(large language model, LLM)逐渐成为各类下游应用的主流技术。特别是多模态大语言模型(multimodal large language model, MLLM), 能够整合文本、图像等多种模态的信息源, 在复杂任务中展现出强大的理解和生成能力^[4-5]。现有的多模态大语言模型, 如GPT-4o和Qwen-VL等, 已经能够在一定程度上实现美术作品的智能鉴赏^[6]。但是, 由于这些模型使用了开放式大规模语料进行训练, 其生成的鉴赏内容并不完全符合学生的学习特点和教师的教学目标^[7]。具体来说, MLLM用于美术作品鉴赏存在以下两个主要问题: ①模型生成的鉴赏内容过于冗长, 无法针对不同年龄段的学生生成符合其学习水平的精练的鉴赏文本; ②模型生成的鉴赏内容比较空洞, 不能围绕培养目标, 以教师的口吻来引导学生学习作品鉴赏。例如在图1

中, 中间是一幅待鉴赏的作品, 左边是直接提示Qwen-VL生成的鉴赏文本, 右边则是专业美术老师对作品的鉴赏。可以看到, 左边的鉴赏内容洋洋洒洒, 罗列很多观点, 但是缺乏对核心知识点的讲解; 右边则围绕“线条”这一核心知识点进行鉴赏, 且鉴赏文本简洁精练, 突出要点。从这个例子可以看到, 直接提示MLLM对学生进行美术作品鉴赏的辅导, 与专业教师的水平仍然存在差距。

鉴于上述挑战, 本文围绕美术作品鉴赏智能辅导这一目标, 基于多模态大语言模型的基础内容理解和生成能力, 通过融入外部数据弥合其与专业教师能力的差距, 构建科学合理、能适应不同年龄段学生需求的美术作品鉴赏模型^[8]。该模型不仅能考虑作品本身的美学价值, 并且符合不同学习阶段的教师话语特点, 确保生成的鉴赏内容既能满足学生的认知需求, 又能引导他们逐步建立起正确的审美方式, 最终促进学生美育素质的提高。技术上, 本文提出了一种基于图检索增强生成(graph retrieval augmented generation, GraphRAG)^[9]和少样本学习^[10]的美术鉴赏方法GRAF (graph rag + few-shot), 用于解决多模态大语言模型在高中美术鉴赏场景中的不足。其中, 基于美术知识图谱的图检索增强生成可以使MLLM的鉴赏输出更加符合美术教学的目标, 将专业教师对美术作品的鉴赏作为样本可以使MLLM的鉴赏输出更加精练, 更加贴近专业鉴赏方式。本文的方法主要分为4个阶段: ①构



图1 多模态大语言模型和专业美术老师鉴赏美术作品的区别

建美术作品鉴赏多模态数据集；②检索美术知识图谱中的相关元素；③检索美术作品鉴赏数据集中的示例样本；④结合图检索增强生成和少样本学习构建提示模板，提示 MLLM 生成符合高中美术教学特点和目标的鉴赏内容。

1 相关工作

1.1 图检索增强生成

最近，图检索增强生成^[9]受到了越来越多的关注。传统检索增强生成 (retrieval augmented generation, RAG)^[11]方法基于文本文档，将待检索文档分块，根据给定查询与文档之间的语义相似度检索信息。而 GraphRAG 首先构建图数据库，随后从图数据库中检索与给定查询相关的元素。这些元素可能包括用于生成响应的节点、三元组、路径或子图。与传统 RAG 不同，GraphRAG 通过图结构考虑了文档之间的相关性，从而能够检索到更加准确全面的信息。知识图谱 (knowledge graph, KG)^[12]作为专业知识的一种图存储方式，因其结构化的知识表示，具有较好的可解

释性和精确度，常被用作图检索增强生成的外部知识库，提供文本数据的抽象和摘要，能够显著缩短检索文本的长度，缓解了传统 RAG 输入文本过长的问题。鉴于此，大语言模型和知识图谱的结合近年来受到广泛关注^[13]。Yang 等^[14]提出在各个训练阶段使用 KG 以增强 LLM 的事实推理能力。Luo 等^[15]提出了图推理 (reasoning on graph) 方法，利用 KG 中的关系生成推理路径，增强了 LLM 的推理能力。Sun 等^[16]提出了图思考 (think-on-graph) 方法，将 LLM 视为智能体，以交互方式探索 KG 上的相关实体和关系，并根据检索到的知识进行推理，缓解了 LLM 在知识的可追溯性、及时性和准确性上表现不佳的问题。王敏等^[17]提出了知识图谱和大模型协同增强的推荐系统，结合知识图谱的推理能力和大模型的泛化能力有效提升了推荐系统性能。与上述工作不同，本文结合图检索增强生成和少样本学习，提高了 MLLM 在教育场景下鉴赏美术作品的的能力。

1.2 少样本提示

现有研究表明^[18-19]，大语言模型的性

能不仅依赖于庞大的参数量和昂贵的训练过程，还显著受到采用的提示的影响。这意味着，即使是同一个模型面对同一个任务，由于提示不同，也可能产生截然不同的结果。当模型接收到适当的提示时，能够展现出较好的适应性和灵活性，以应对各种不同的任务需求。因此，如何构建最有效的提示成为大语言模型应用中的一个重要课题，即提示工程。

大量关于提示工程的研究发现，大语言模型具备通过基于零样本（zero-shot）或少样本（few-shot）的上下文学习来完成下游任务的能力^[10,20]。其中，少样本提示（few-shot prompting）是提示工程中的一项经典技术，它通过在提示中提供少量的演示示例来提高大语言模型遵循指令的能力。例如，在情感分析^[21]、文本分类^[22]、机器翻译^[23]、命名实体识别^[24]等多个领域，大语言模型已经被证明了其不需要大量特定任务数据集，仅需在提示中加入少量演示示例，即可激发模型模仿示例完成相应任务的能力，且能达到较高性能。另外，少样本提示的设计不仅涉及对模型工作原理的深入理解，还需要结合具体应用场景的特点，探索出既能激发模型潜能又易于实施的演示示例。

1.3 多模态大语言模型用于美术鉴赏

大语言模型^[25-27]与视觉编码器结合，形成了一种多模态大语言模型^[5]，能够融合图像和文本两个不同模态的数据作为输入，从而处理更复杂的任务。近年来，多模态大语言模型在艺术领域的应用取得了显著进展。Zhang 等^[28]系统地研究了 GPT-4V 作为通用评估器在视觉-语言任务中的表现，涵盖图像到文本生成、文本到图像生成、图像到图像翻译以及多图像

到文本对齐等任务。研究表明，尽管 GPT-4V 在处理复杂艺术图像时能够提供详细的解释性评估，然而，其在视觉清晰度评估和知识密集型任务（如特定艺术风格识别）中仍存在一定的局限性。Hayashi 等^[29]首次定义了艺术品解释生成任务，并从维基百科上构建了艺术品解释数据集 ExpArt。该研究通过引入两个任务设置——基于图像和标题的解释生成以及仅基于图像的解释生成，深入探讨了多模态大语言模型在艺术品理解方面的能力。研究结果表明，尽管多模态大语言模型能够继承其基础语言模型的艺术知识，但在整合视觉和语言信息方面仍存在显著挑战，尤其是在仅依赖视觉信息生成解释时表现更受限。这一发现与 Cetinic^[30]在艺术作品图像描述生成方面的探索形成了呼应，进一步凸显了多模态大语言模型在艺术领域应用的复杂性和局限性。Hayashi 等的工作为后续研究提供了重要的数据集和评估框架，对模型在跨模态知识整合方面的局限性分析也为未来研究指明了方向，例如如何更好地对齐视觉编码器和语言模型的知识表示，以及如何提升模型对艺术史、风格和主题等深层次艺术知识的理解能力。Lee 等^[7]提出 LLaVA-Docent，通过指令微调的方式将多模态大语言模型应用于艺术鉴赏教育领域。该研究创新性地结合视觉编码器与语言模型，构建了一个支持多轮对话的艺术鉴赏辅助系统，提供定制的问题鼓励用户深入欣赏艺术品，支持艺术鉴赏教育。上述工作没有针对美术教育场景进行针对性优化。在教育场景中，对于美术作品的鉴赏更加偏重引导式的教学，鉴赏内容也应该围绕教学目标，尽可能地接近专业教师的鉴赏方式。鉴于此，本文专门面向高中美术教育构建多模态大模型美术作品鉴赏方法。

2 GRAF: 美术作品智能鉴赏

图2展示了GRAF的整体架构。首先,如图2左侧所示,构建一个符合教育场景的高中美术作品鉴赏多模态数据集;随后,如图2“从知识图谱检索能力和素养”部分所示,根据用户提出的鉴赏问题,在高中美术知识图谱中检索知识点、素养等相关内容;接着,如图2“从HAM检索示例样本”部分所示,在高中美术作品鉴赏多模态数据集中检索与用户当前问题关联度最高的图片-文本数据对作为少样本提示的示例样本;然后,根据图2中“MLLM输入”的提示模板,结合知识图谱中检索的相关内容和示例样本,构建多模态大语言模型的提示;最后,将构建好的提示和用户当前的问题一起输入多模态大语言模型,从而得到如图2“MLLM输出”所示的最终鉴赏输出。

2.1 高中美术作品鉴赏多模态数据集构建

为了让MLLM通过少样本学习提高指令遵循的能力,首先需要有一个高质量的数据集,其包含供MLLM学习的样本。然而,目前没有公开的美术作品鉴赏数据集。为此,笔者以国家中小学智慧教育平台上的美术课程视频为数据来源,构建了一个高中美术作品鉴赏多模态数据集HAM (high school art appreciation multimodal dataset)。

国家中小学智慧教育平台的美术课程由国家知名美术教师录制,每节课都与美术教材上的章节对应,会有教师围绕当前课程的教学目标和知识点对名家、教师、学生作品做出鉴赏示例。教师在课程视频中对美术作品的鉴赏大致可分为两种类型:一种是图像长时间停留式的细致分析,这种类型的鉴赏往往与当前课程的教学内容紧密相关,对作品的评价较全面,鉴赏文本也较完整,可以作为高质量数据;一种是图像快速浏览式的简单介绍,这种类型

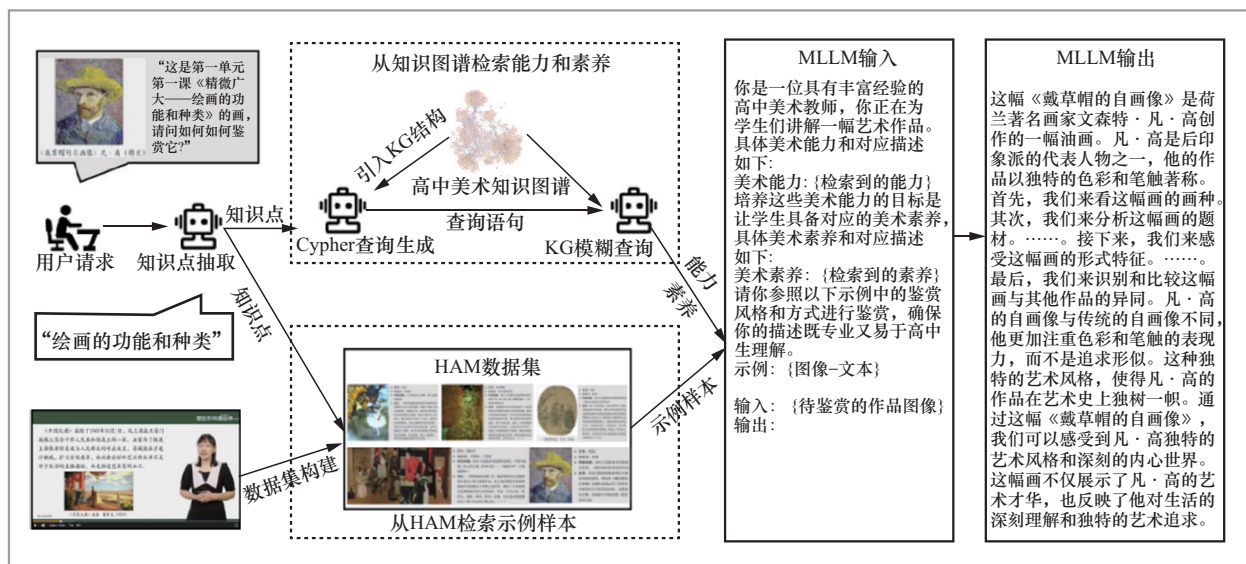


图2 GRAF整体框架

的鉴赏往往出现在对某一教学内容的普及式教学，教师通常会快速展示同一主题或同一类型的多幅美术作品，同时每幅作品仅仅只有一两句对应的鉴赏文本做简单介绍，甚至没有单独介绍的鉴赏文本。基于此观察，笔者观看了所有的高中课程视频，将高中课程视频中的每一幅教师经过长时间停留式的细致分析的作品和教师对其鉴赏的文本抓取下来，构成图像-文本对，最后形成包含 200 个图像-文本对的数据集 HAM。因为教师具有较高的专业素养和教学能力，可以认为 HAM 中的图像数据具有一定的美学价值，其对应的鉴赏文本既能满足学生的认知需求，又能引导他们逐步建立起正确的审美观念，符合美术教育的要求。图 3 展示了 HAM 中的一个作品图像，以及元数据和教师的鉴赏文本。作品的元数据包括该作品的作者、涉及的美



图 3 HAM 数据集样本示例

表1 传统RAG与GraphRAG检索效果的区别		
本文检索的需求	传统RAG	GraphRAG
多跳推理	×	✓
关系推理	×	✓
精准检索结构化信息	×	✓
精简上下文减少噪声	×	✓

术知识点以及数据的网课来源。HAM 中的每个图像-文本对都以相同的形式存储了作品的信息。

2.2 基于高中美术知识图谱的图检索增强生成

检索增强生成 RAG 是一种混合方法，结合了自然语言处理中的检索和生成。通过引入外部知识库，RAG 能够有效缓解 LLM 的幻觉、缺乏特定领域知识和训练语料过时等问题。然而，传统的 RAG 通常在文本数据中进行检索，检索过程依赖文本块之间的相关性，而忽略了知识库中不同实体之间的复杂关系，导致检索精准度不足。此外，传统 RAG 在检索过程中容易引入噪声，来自不同文本块的相似信息可能会导致上下文冗长，LLM 无法抓住重点信息，出现“迷失在中间”^[31]的问题。与传统的 RAG 不同，图检索增强生成 GraphRAG 可以从知识图谱中检索与给定问题相关的结构化信息，同时通过图谱中的关系考虑了信息之间的相互联系，从而能够实现更准确和全面的信息检索。此外，知识图谱提供了文本数据的抽象和总结，有效减少了输入长度，缓解了上下文冗长的问题。表 1 展示了传统 RAG 和 GraphRAG 检索效果的区别。

本文部分笔者来自北京师范大学附属实验中学美术教研组，为国家中小学智慧教育平台美术课程标注了素养、能力和知识点 3 种类型的实体。“素养”类型实体是美育课程期望学生在美术学习中逐步形成的综合性素养，是学生在艺术认知、文化理解、创意实践等素质方面的长期发展目标；“能力”类型实体是学生在美术学习过程中需要掌握的具体能力，是达成素养目标的关键支撑，也是美育课程希望培

养学生具备的具体美术能力；“知识点”类型实体是美育体系中需要学生掌握的具体知识内容，是能力培养与素养形成的基础。每类实体包含以下属性：实体类型、实体名称以及实体描述。其中，素养实体采用三级分类体系，能力实体采用两级分类体系，知识点实体采用五级分类体系。不同级别的同类型实体之间存在从属关系，例如一级素养“图像识读”包含二级素养“感受图像的形式特征”，同级别的知识点类型实体之间存在并列关系，如“水彩画”和“水粉画”属于同一级别的知识点。除此之外，知识点与能力和素养之间存在关联关系，每个知识点可关联多项能力和素养。不同的知识点下关联的能力和素养所需要的掌握程度也有所不同，由低到高可分为了解、理解、感悟和探索。基于上述实体的属性及其关联关系，笔者构建了高中美术知识图谱，图谱中不同类型的节点即对应的实体，该图谱清晰

地呈现了素养、能力和知识点之间的层次结构与关联关系，将此图谱作为外部知识库。本文使用 Neo4j 图数据库存储上述知识图谱。Neo4j 支持图数据的模糊查询，这与本文的应用场景十分匹配。构建的知识图谱有素养、能力和知识点 3 种类型的节点，同一类型的节点内部存在层级关系。其中，每个素养和能力都有具体描述，每个知识点都有其对应的素养和能力。鉴于高中美术教育的目标是培养学生综合运用美术学科及跨学科知识和技能解决问题的能力，形成高中生必备的美术学科核心素养，上述知识图谱能够将当前鉴赏作品涉及的知识点与其相关的素养和能力关联起来，从而确定当前鉴赏作品的教学目标。美术知识图谱中的部分节点与关系可视化如图 4 所示，其中红色节点为素养节点，橙色节点为知识点节点，紫色节点为能力节点。如图 4 所示，该素养节点存储有该素养的名称、具体描述以及具体

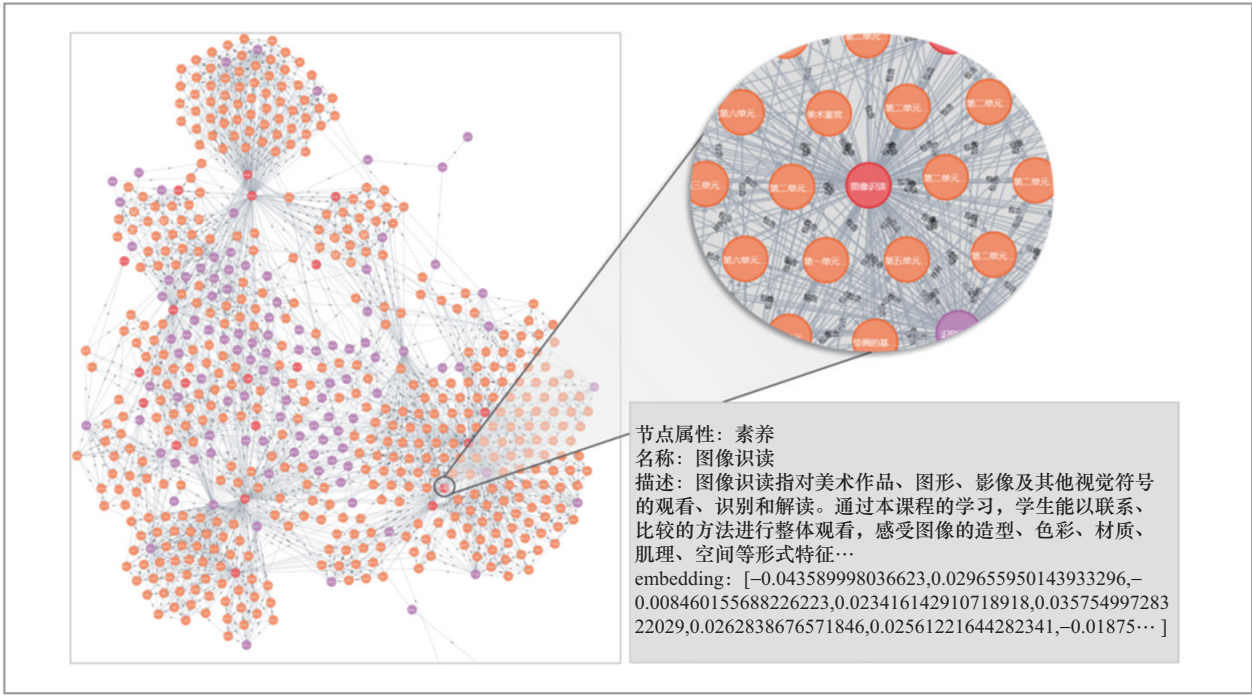


图4 知识图谱部分可视化

描述的向量化嵌入。每个节点都以相同的形式存储了对应实体的信息，存在关联的节点之间有边相连。

基于上述知识图谱，笔者采用 Graph-RAG 为用户请求检索相关内容，作为对请求的上下文补充。在美术作品鉴赏的应用场景下，用户请求被分为两部分：图像内容和文本内容。在知识图谱中检索与文本内容最相关的 k 个知识点，再根据这 k 个知识点检索相关的能力和素养，得到用于辅助 MLLM 输出的外部知识。这一过程由 3 个大语言模型智能体来实现：知识点抽取智能体、查询语句智能体和查询执行智能体。首先，用户请求的文本内容通常是一段自然语言，例如“请你从{知识点}的角度分析鉴赏这幅画”，直接根据文本内容难以在知识图谱中执行检索。因此，设置知识点抽取智能体为知识点抽取专家，从用户当前请求的文本内容中抽取知识点。其次，知识点抽取智能体抽取得到的知识点并不一定与知识图谱中的节点直接匹配，需要通过模糊查询检索知识图谱中的相似知识点。因此，设置查询语句智能体，根

据知识点抽取智能体返回的知识点，生成基于 Cypher 的模糊查询语句。Cypher 是一种声明性的图形查询语言，可以实现图数据库的查询、更新和管理，并且可以通过正则表达式实现模糊查询。利用 Cypher，笔者通过正则表达式匹配的方式从图谱中模糊查询相关知识点，以及与知识点相关的能力和素养。最后，设置查询执行智能体，在 Neo4j 图数据库中执行 Cypher 查询语句，从知识图谱中检索出与用户请求中的知识点相似度最高的 k 个知识点。图 5 展示了一个模糊查询的示例，以用户请求文本“这是第二单元第三课《象外之境——中国传统山水画》里展示的画，怎么鉴赏？”为例，知识点抽取智能体会从中抽取“中国画”和“山水画”两个知识点，查询语句智能体会根据知识点生成如图 5 所示的 Cypher 查询语句。接下来，查询执行智能体会根据该查询语句查找图数据库中 with “中国画”和“山水画”相关的前 k 个知识点，并最终返回与这些知识点相关的能力和素养的具体描述。在实验中，设置 $k=5$ 。上述过程如图 2 中间上半部分所示。

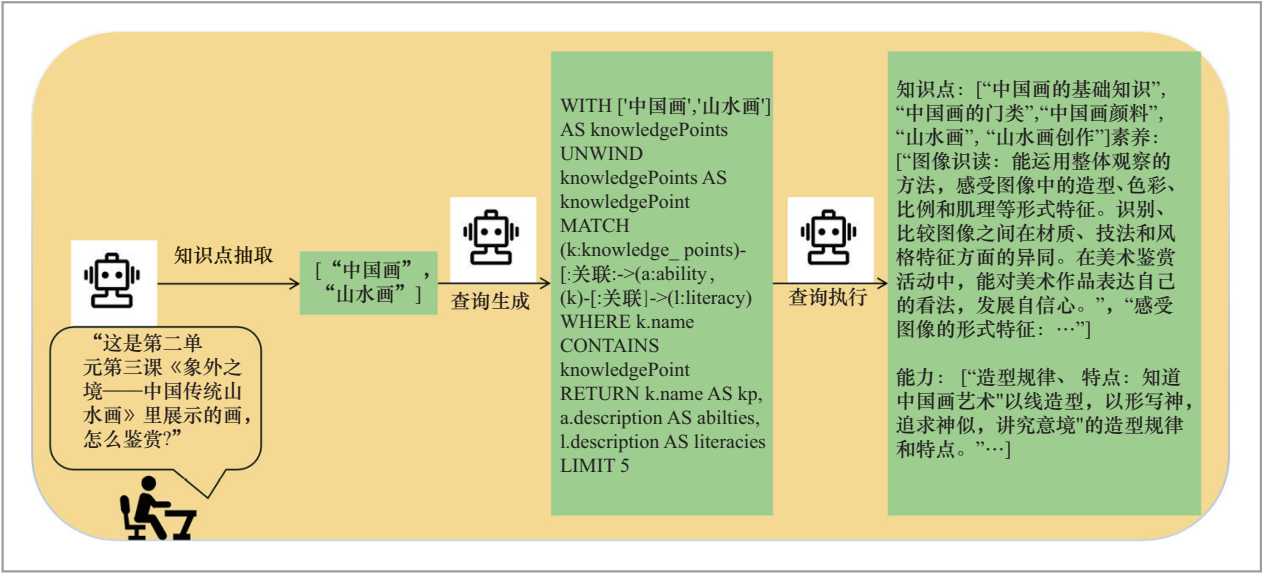


图5 模糊查询示例

2.3 少样本学习

直接提示 MLLM 对美术作品进行鉴赏，可能会生成冗长、空洞的鉴赏文本，与高中美术教育的目标不符。鉴于此，采用少样本学习的方式，从 HAM 中选取与用户请求相关的图像-文本对，作为鉴赏示例。少样本学习通过在提示中提供演示示例来启用上下文学习，引导大语言模型参考演示示例，生成更符合任务需求的响应。

如图 2 中间下半部分所示，用知识点抽取智能体从用户当前请求中得到知识点，然后基于知识点在 HAM 中检索参考示例。由于知识抽取智能体抽取到的知识点在 HAM 的数据对中并不一定存在，笔者采用近似计算的方式检索参考示例。具体来说，计算 HAM 中存在的知识点与用户请求中的知识点之间的语义相似度，并使用相似度最高的知识点对应的图像-文本对作为当前用户请求的参考示例。示例样本中专业教师对图像内容的鉴赏文本可以规范 MLLM 生成鉴赏的内容和风格。

2.4 美术作品鉴赏提示模板构建

经过上述步骤，可以检索出知识图谱中与当前用户请求最相关的高中美术能力和素养，以及 HAM 中与当前用户请求最相关的参考示例。使用这些检索结果构建美术作品鉴赏提示模板。如图 6 所示，首先是任务描述，告知 MLLM 其角色和目标；随后，告知 MLLM 可以参考的、与用户请求有关的美术能力和素养；接着，向 MLLM 展示一个图像-文本示例，用于规范鉴赏文本；最后，给出用户当前请求中的作品图像，要求 MLLM 进行鉴赏。通过这种方式，笔者将用户的原始请求转化为更精细的提示，以期

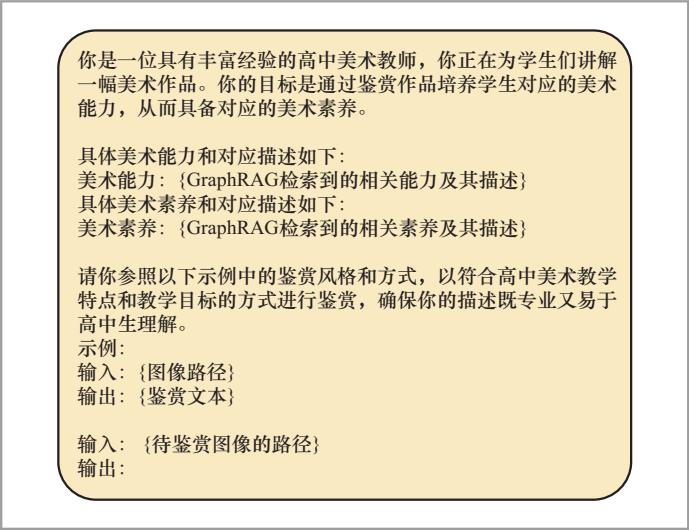


图 6 完整的美术作品鉴赏提示模板

MLLM 生成的鉴赏文本更加符合美术教学的需求。

3 实验结果与分析

3.1 数据集和模型

笔者从 HAM 中随机挑选了 165 张图像作为测评数据，其余 35 张作为供少样本学习挑选的示例。使用 Qwen2.5 为每张图像生成鉴赏请求文本，模拟用户输入并构建 3 个智能体。使用 Qwen-VL-Max 作为美术作品鉴赏模型。Qwen-VL-Max 是阿里巴巴发布的一个多模态大模型，可以完成图片理解相关任务，是 Qwen-VL 多模态大模型系列的最强模型。为了避免模型输出的随机性，设置 temperature=0。

3.2 对比方法

将 GRAF 与 3 种基准方法进行对比。据笔者调研，现有相关研究工作（如

ExpArt^[29]和 LLaVa-docent^[7]) 都是简单基于提示的方法完成美术鉴赏任务, 其中, LLaVa-docent 是微调之后再提示。基于此, 笔者选择的第一个对比方法就是简单基于提示的方法, 即从图6的提示模板中, 去除能力、素养、示例, 仅保留任务描述, 称之为 SimplePrompt。其次, 去除少样本示例, 仅保留任务描述和检索到的能力与素养, 称之为 GraphRAG。最后, 去除检索到的能力和素养, 仅保留任务描述和少样本提示, 称之为 FewShot。GraphRAG 和 FewShot 同时作为另外的两个消融实验, 分别验证两个模块在 GRAF 方法中的重要性。

3.3 评价指标

采用自动评价指标和人工评价指标来评价模型生成的鉴赏文本。自动评价指标包括 ROUGE-L^[32]和 Bertscore^[33], 分别用来衡量生成的文本在语义和流畅性方面与参考文本的匹配程度。ROUGE-L 通过计算生成的鉴赏文本和教师鉴赏文本之间的最长公共子序列的长度来评估。Bertscore 使用嵌入模型来衡量生成的鉴赏文本和教师鉴赏文本的相似度。两个指标均是数值越高, 效果越好。

此外, 笔者设计了人工评价指标。从4个维度考虑高中美术作品的专业鉴赏: ①客观技术解析, 包括技法、构图、色彩等视觉元素的分析; ②主观内涵解读, 包括情感传递、符号隐喻、创作意图推断等; ③人类偏好验证, 包括与教师教学文本的吻合度、是否符合对应年龄段学生的学习特点及理解能力; ④教学目标实现, 包括知识点覆盖、术语规范性、教学引导性等。基于以上维度, 设计4项可量化人工评价指标, 如下:

- 模型生成的鉴赏文本的可读性, 对应鉴赏文本对技术解析的表述清晰度;
- 模型生成的鉴赏文本与图像内容的相关性, 对应鉴赏文本对内涵解读的客观依据;
- 模型生成的鉴赏文本与教师鉴赏文本的吻合度, 对应鉴赏文本的人类偏好验证;
- 模型生成的鉴赏文本对相关美术素养/能力/知识点的覆盖度 (衡量是否符合教学目标), 对应鉴赏文本对教学目标的实现程度。

对于每一个指标, 定义了0~3的分数, 分值越高, 代表鉴赏效果越好。笔者邀请美术专业的3位研究生按照上述要求打分, 并计算每个指标的平均值。评价指标和分数定义见表2。

3.4 实验结果

实验结果见表3。从自动评价指标看到, GRAF 的表现略低于 FewShot。这是因为在课堂上教师并不会完全围绕能力和素养来鉴赏作品, 语言的自由度更高。在 FewShot 的基础上加入能力和素养信息后, GRAF 导致模型的输出内容与教师的鉴赏内容有更多差别。不过, 自动评价指标仅比较模型输出和标准输出的文字差异, 并不能从语义角度反映模型的性能。

从人工评价指标可以看出, GRAF 在可读性和相关性上表现并非最佳。GRAF 的可读性略低于 FewShot, 这可能是因为要求模型根据培养目标中的能力和素养生成鉴赏文本, 增加了任务难度, 导致可读性下降。但是与 SimplePrompt 和 GraphRAG 相比, GRAF 的可读性仍有较大提升。在相关性上, 4种方法表现相近, GRAF 略低于 GraphRAG, 但没有显著差

表2 人工评价指标				
分数	3	2	1	0
可读性	回答语句通顺,话语精确不冗余,没有无关信息	回答语句较为通顺,可读性较好	回答语句具有一定可读性	回答语句出现乱码,没有可读性
相关性	鉴赏的内容完全符合图片内容,不仅能描述图像中的显性内容,还能关联图像中的隐性特征	鉴赏的内容基本符合图片内容,能够描述图像中客观存在的显性内容	鉴赏的内容少量与图片内容相关	鉴赏的内容与图片内容完全不相关,甚至出现幻觉等误导内容
指标定义				
吻合度	鉴赏的内容与教师赏析内容高度吻合,基础要素、图像分析以及对图像的延伸解读都能匹配	鉴赏的内容与教师赏析内容基本吻合,基础要素与部分分析匹配	鉴赏的内容与教师赏析内容不太吻合,只有基础要素匹配	鉴赏的内容与教师赏析内容完全无关
覆盖度	鉴赏的内容覆盖相关美术素养/能力/知识点,同时高度关联并系统化分析,完全符合教学目标	鉴赏的内容覆盖相关美术素养/能力/知识点,同时简单关联,很好地匹配教学目标	鉴赏的内容提及相关美术素养/能力/知识点,一定程度匹配教学目标	鉴赏的内容与相关美术素养/能力/知识点完全没有,不匹配教学目标

表3 生成鉴赏文本的自动评价和人工评价结果						
方法	ROUGE-L	Bertscore	可读性	相关性	吻合度	覆盖度
SimplePrompt	0.87	0.79	2.35	2.54	2.08	1.48
GraphRAG	0.84	0.79	2.30	2.58	2.52	2.83
FewShot	1.05	0.86	2.88	2.53	2.55	2.20
GRAF(本文方法)	0.93	0.85	2.65	2.56	2.62	2.95

别。这主要得益于多模态大语言模型本身的图像理解能力。值得注意的是, GRAF在吻合度和覆盖度上,均超过了对比模型,取得了最佳表现。这证明结合 GraphRAG和少样本学习,能够让多模态大语言模型输出更符合高中美术教育应用场景的作品鉴赏文本。

3.5 案例分析

3.5.1 方法对比案例

图7展示了4种方法生成的《韩熙载夜宴图》的鉴赏文本。笔者将能力与素养相关的文字标注成红色。可以看到, SimplePrompt 生成的鉴赏内容冗长且空

洞, 仅有“色彩与线条”的一小段与高中美术能力与素养培养要求有关。GraphRAG虽然引入了多个素养与能力鉴赏点, 但是鉴赏文本比较生硬简单, 没有像教师那样用生动的语言来描述作品。FewShot模仿了教师的鉴赏口吻, 但对能力和素养的描述比较松散和口语化。这是因为笔者构建的数据集来自课堂上的教师语录。最后, 本文提出的 GRAF既能有效模仿教师的鉴赏口吻, 同时对与能力和素养相关语句的组织更加精练和完整。

3.5.2 偏序对比案例

图8展示了用 GRAF 评价美术素描考试作品的表现。作品来源为河北省某年美



这幅画作为《韩熙载夜宴图》，是中国五代十国时期南唐画家顾闳中所绘的一幅长卷人物画。它描绘了南唐大臣韩熙载在家中设宴的情景，展现了当时贵族阶层的生活风貌。

一、画面分析

- 1) 构图与布局：画面分为四个部分，每个部分都展示了不同的场景和人物活动。
- 2) 色彩与线条：色彩以淡雅为主，使用了大量浅色和柔和的色调，营造出一种宁静而优雅的氛围。线条细腻流畅，人物的衣纹和表情刻画得非常生动，显示出画家高超的技艺。
- 3) 人物与服饰：人物形象栩栩如生，神态各异，表现出不同的情绪和性格特征。服饰华丽，细节丰富，反映了当时贵族阶层的生活水平和审美情趣。
- 4) 背景与环境：背景简洁，主要以室内环境为主，通过家具和装饰物来表现空间感。室内布置精致，反映出主人的品味和地位。

二、艺术价值

- 1) 历史价值：这幅画不仅是一幅艺术作品，也是研究五代十国时期社会生活的重要史料。
- 2) 艺术风格：体现了中国绘画中的“写意”风格，注重人物神态和情感的表达，而非单纯的形体描绘。
- 3) 文化内涵：通过描绘贵族阶层的生活，反映了当时的社会风气和文化习俗。

三、教学目标

- 1) 观察与分析：引导学生仔细观察画面，分析构图、色彩、线条等艺术元素。
- 2) 历史背景：介绍画作的历史背景，帮助学生理解其文化意义。
- 3) 艺术鉴赏：培养学生的艺术鉴赏能力，学会从不同角度欣赏和评价艺术作品。

通过这幅画的学习，希望同学们能够更好地理解和欣赏中国传统绘画的魅力，同时也能从中感受到古代文化的深厚底蕴。

SimplePrompt

《韩熙载夜宴图》是一幅非常著名的古代中国绘画作品，它通过细腻的笔触和生动的场景，展现了五代时期南唐名臣韩熙载的一次豪华宴会。这幅画由南唐画家顾闳中创作，采用了长卷式的构图，分为五个部分，分别是听乐、观舞、休息、轻吹和送别，完整地呈现了宴会的整个过程。在第一段听乐的场面中，我们可以看到韩熙载和他的宾客们正全神贯注地欣赏琵琶演奏。每个人的表情都非常专注，仿佛被音乐深深吸引。画家巧妙地将所有人的视线和注意力都集中在琵琶女的演奏上，营造出一种紧张而安静的氛围。尽管画面中的人物众多，动作各异，但布局合理，线条工整，色彩绚丽，层次分明，给人一种和谐统一的感觉。这幅画不仅展示了南唐官宦阶层的奢华生活，还通过主人公韩熙载的表情，揭示了他的内心世界。韩熙载虽然表面上享受着宴会的欢乐，但他的眼神中透露出一丝忧郁和矛盾，似乎在思考着什么。这种对人物心理的细腻刻画，使得这幅画不仅仅是一幅描绘宴会场面的作品，更是一幅反映人性和社会现实的艺术杰作。

FewShot

这幅画作是一幅典型的中国画，属于人物画的一种。它描绘了一个古代家庭聚会的场景，人物众多，细节丰富。我们可以看到画面中有男女老少，他们围坐在一起，有的在弹琴，有的在交谈，还有的在侍奉。这种场景通常被称为“风俗画”，因为它反映了当时社会的生活习俗和人际关系。

一、美术能力：

区分题材

- 1) 题材：这幅画属于“风俗画”，因为它展示了人们日常生活的场景。
- 2) 画种：这是中国画，使用了传统的毛笔和墨色。

识别了解

- 1) 画种：中国画，使用毛笔和墨色。
- 2) 题材内容：风俗画，展示日常生活场景。
- 3) 特征：人物众多，细节丰富，色彩柔和，线条流畅。

二、美术素养

分析美的丰富性，形成发现美、欣赏美的审美判断能力

- 1) 形式美原理：这幅画运用了平衡、对比、和谐等形式美原理，使得画面看起来非常和谐。
- 2) 审美感受：通过观察人物的表情和动作，我们可以感受到画家对生活的热爱和对人物情感的细腻描绘。

图像识读

- 1) 整体观看：通过整体观看，我们可以感受到画面的和谐与生动。
- 2) 识别与解读：通过识别人物的服饰和环境，我们可以推测出这可能是一个古代贵族家庭的聚会。
- 3) 图像价值：这幅画不仅展示了古代的生活习俗，还体现了画家高超的技艺技巧和深厚的文化底蕴。

通过这幅画的鉴赏，我们不仅可以了解中国古代的生活习俗，还可以提高我们的审美能力和图像识读能力。希望同学们在欣赏这幅画作的同时，也能感受到其中蕴含的美和文化价值。

GraphRAG

《韩熙载夜宴图》是一幅表现现实生活题材的长卷式工笔人物画作品，是五代时期的宫廷画家顾闳中受南唐后主李煜之命所作，描绘了五代南唐名臣韩熙载开宴行乐的场面。作品一共分五段，即听乐、观舞、休息、轻吹、送别，以长卷式的构图表现了整个宴会的完整过程。第一段听乐画面中，韩熙载与到访的宾客正聚精会神地倾听琵琶演奏，众人屏气凝神，沉醉其中。画家将全场上每个人的视线和精神都集中于琵琶女的演奏动作上，紧凑的节奏与现场的宁静氛围相吻合。虽然人物繁多，动态各异，场景复杂，却安排得宾主得当，疏密有致，线条工整细腻，遒劲流畅，运用了高古游丝描、铁线描等描线法，绚丽清雅，层次丰富。画面整体注重客观叙事的情节性，真实反映了南唐官宦青年的生活面貌。主人公韩熙载表情的刻画，又透露出其失意矛盾的内心世界。

GRAF

图7 4种方法生成的《韩熙载夜宴图》的鉴赏文本

术类质检模拟联考的考生作品。联考专业阅卷组将作品分为了10个档次，第1档最好，第10档最差。为了更好地体现GRAF鉴赏作品的区分能力，选择同一个考题（临摹女生人像）的3幅作品，分别为第1档、第5档和第9档，由此构成偏序集。图8中3幅作品从左到右依次为第1档、第5档和第9档。将专业阅卷组的分档要求作为GRAF方法的少样本示例，作品下方为GRAF生成的对应鉴赏文本。将模型对作

品的正面评价用绿色字体高亮，对作品的负面评价用红色字体高亮。可以看到，模型从鉴赏素描作品的专业角度，如透视、形体比例、黑白灰关系等出发，对3幅作品做出评价。对于第1档的作品，GRAF鉴赏文本以绿色高亮的正面评价为主。对于第5档的作品，GRAF鉴赏文本在肯定作品优点的同时指出了画面中存在的不足，正面评价与负面评价同时存在。对于第9档的作品，GRAF鉴赏文本着重分析了作品



图8 GRAF在偏序集上的实验结果

存在的问题，以红色高亮的负面评价为主。可以看到，GRAF 鉴赏结果整体上符合专业阅卷组的分档结果，能够从素描作品评价的各个维度分析作品的优缺点，同时分析内容符合人类评价偏好。

4 结束语

围绕美术作品鉴赏智能辅导这一目标，本文提出了一种结合图检索增强生成和少样本学习的高中美术作品智能鉴赏方法 GRAF。该方法的核心思想是从美术知识图谱中检索相关能力和素养，并从少样本示例中学习专业教师对美术作品的评价方式，从而使多模态大语言模型的鉴赏输出符合教育场景的需求。相比于一般的图像理解任务，美术作品具有以下特性：技法多样性（不同艺术形式具有独特的技法体系）、文化语境

依赖性（作品内涵与历史背景、流派演变、文化符号密切相关）、审美主观性（艺术鉴赏需结合专业理论与个人审美经验）。本文针对以上美术作品的专有特性对方法进行了改进和优化，包含三级语义的美术知识图谱构建、基于美术知识图谱的 GraphRAG、针对美术作品鉴赏的提示模板、针对美术作品的人工评估方式等。其中，三级语义的美术知识图谱构建为美术作品鉴赏提供的技法和文化语境等外部知识来源；基于美术知识图谱的 GraphRAG 检索知识图谱中相关的外部知识以提供作品对应的技法与文化语境，从而提升模型生成文本的质量；针对美术作品鉴赏的提示模板让多模态大模型的输出更符合“教育场景下的美术作品鉴赏”这一下游任务。

实验结果表明，GRAF 在多项关键指标上表现最佳。在未来，笔者将进一步探索生成更符合教育场景需求的美术作品鉴

赏文本的方法，并将这些方法应用到更多学段的美术教育中。

参考文献:

- [1] 梁兰兰. 美育在中小学教育教学中的实施——评《中小学生美育教育知识》[J]. 教育理论与实践, 2023, 43(11): 2.
LIANG L L. The implementation of aesthetic education in primary and secondary school teaching—a review of “Knowledge of Aesthetic Education for Primary and Secondary School Students” [J]. Theory and Practice of Education, 2023, 43(11): 2.
- [2] 吴鸿昌. 浅析高校美术欣赏教学实践对实施素质教育意义[J]. 教育与职业, 2010(26): 118–119.
WU H C. On the significance of art appreciation teaching practice in colleges and universities to the implementation of quality education[J]. Education and Vocation, 2010(26): 118–119.
- [3] 夏岩. 新课改背景下美术鉴赏课的有效性教学研究[J]. 大舞台, 2012(12): 194–195.
XIA Y. Research on effective teaching of art appreciation course under the background of new curriculum reform[J]. Dawutai, 2012(12): 194–195.
- [4] WU J Y, GAN W S, CHEN Z F, et al. Multimodal large language models: a survey[C]//Proceedings of the 2023 IEEE International Conference on Big Data. Piscataway: IEEE Press, 2023: 2247–2256.
- [5] BAI J Z, BAI S, YANG S S, et al. Qwen-VL: a versatile vision-language model for understanding, localization, text reading, and beyond[EB]. arXiv preprint, 2023, arXiv: 2308.12966.
- [6] RACHABATUNI P K, PRINCIPI F, MAZZANTI P, et al. Context-aware chatbot using MLLMs for Cultural Heritage[C]//Proceedings of the 15th ACM Multimedia Systems Conference. [S. l.: s. n.], 2024: 459–463.
- [7] LEE U, JEON M, LEE Y, et al. LLaVA-docent: Instruction tuning with multi-modal large language model to support art appreciation education[J]. Computers and Education: Artificial Intelligence, 2024, 7: 100297.
- [8] 谢赠生. 核心素养背景下高中美术鉴赏教学策略探讨[J]. 中国教育学刊, 2020(S1): 130–131.
XIE Z S. Discussion on teaching strategies of art appreciation in senior high school under the background of core literacy[J]. Journal of the Chinese Society of Education, 2020(S1): 130–131.
- [9] EDGE D, TRINH H, CHENG N, et al. From local to global: a graph RAG approach to query-focused summarization [EB]. arXiv preprint, 2024, arXiv: 2404.16130.
- [10] BROWN T B, MANN B, RYDER N, et al. Language models are few-shot learners[C]//Advances in Neural Information Processing Systems. [S. l.: s. n.], 2020: 1877–1901.
- [11] LEWIS P, PEREZ E, PIKTUS A, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks[EB]. arXiv preprint, 2020, arXiv: 2005.11401.
- [12] HOGAN A, BLOMQUIST E, COCHEZ M, et al. Knowledge graphs[J]. ACM Computing Surveys, 2022, 54(4): 1–37.
- [13] JIN B W, LIU G, HAN C, et al. Large language models on graphs: a comprehensive survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(12): 8622–8642.
- [14] YANG L Y, CHEN H Y, LI Z, et al. Give us the facts: enhancing large language models with knowledge graphs for fact-aware language modeling[EB]. arXiv preprint, 2023, arXiv: 2306.11489.
- [15] LUO L H, LI Y F, HAFFARI G, et al. Reasoning on graphs: faithful and interpretable large language model reasoning [EB]. arXiv preprint, 2023, arXiv:

- 2310.01061.
- [16] SUN J S, XU C J, TANG L M Y, et al. Think-on-Graph: deep and responsible reasoning of large language model on knowledge graph[C]//Proceedings of the 12th International Conference on Learning Representations. [S.l.:s.n.], 2024.
- [17] 王敏, 高晓影, 汪诗蕊, 等. 基于知识图谱和大语言模型增强的推荐系统研究[J]. 大数据, 2025, 11(2): 29-46.
- WANG M, GAO X Y, WANG S R, et al. Research on recommendation system based on knowledge graph and large language model enhancement[J]. Big Data Research, 2025, 11(2): 29-46.
- [18] LU Y, BARTOLO M, MOORE A, et al. Fantastically Ordered Prompts and Where to Find Them: Overcoming Few-Shot Prompt Order Sensitivity[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. [S.l.:s.n.], 2022: 8086-8098.
- [19] ZHAO T Z, WALLACE E, FENG S, et al. Calibrate before use: improving few-shot performance of language models [EB]. arXiv preprint, 2021, arXiv: 2102.09690.
- [20] RADFORD A, WU J, CHILD R, et al. Language models are unsupervised multitask learners[J]. OpenAI Blog, 2019, 1(8): 9.
- [21] WANG Z, XIE Q, XIA R. A simple yet effective framework for few-shot aspect-based sentiment analysis[C]//Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. [S.l.:s.n.], 2023: 1765-1770.
- [22] MIN S, LEWIS M, HAJISHIRZI H, et al. Noisy channel language model prompting for few-shot text classification[C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. [S.l.:s.n.], 2022: 5316-5330.
- [23] GARCIA X, BANSAL Y, CHERRY C, et al. The unreasonable effectiveness of few-shot learning for machine translation[C]//Proceedings of the International Conference on Machine Learning. [S.l.:s.n.], 2023: 10867-10878.
- [24] LIU J, FEI H, LI F, et al. TKDP: threefold knowledge-enriched deep prompt tuning for few-shot named entity recognition[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(11): 6397-6409.
- [25] TOUVRON H, MARTIN L, STONE K, et al. Llama 2: open foundation and fine-tuned chat models[EB]. arXiv preprint, 2023, arXiv:2307.09288.
- [26] BAI J Z, BAI S, CHU Y F, et al. Qwen technical report[EB]. arXiv preprint, 2023, arXiv: 2309.16609.
- [27] ACHIAM J, ADLER S, AGARWAL S, et al. Gpt-4 technical report[EB]. arXiv preprint, 2023, arXiv:2303.08774.
- [28] ZHANG X L, LU Y J, WANG W Z, et al. GPT-4V(ision) as a generalist evaluator for vision-language tasks[EB]. arXiv preprint, 2023, arXiv: 2311.01361.
- [29] HAYASHI K, SAKAI Y, KAMIGAITO H, et al. Artwork explanation in large-scale vision language models[EB]. arXiv preprint, 2024, arXiv: 2403.00068.
- [30] CETINIC E. Iconographic image captioning for artworks[M]//Pattern Recognition. ICPR International Workshops and Challenges. Cham: Springer, 2021: 502-516.
- [31] LIU N F, LIN K, HEWITT J, et al. Lost in the middle: how language models use long contexts[J]. Transactions of the Association for Computational Linguistics, 2024, 12: 157-173.
- [32] LIN C Y. ROUGE: a package for automatic evaluation of summaries[C]//Proceedings of the Workshop on Text Summarization Branches Out. [S.l.:s.n.], 2004: 74-81.
- [33] ZHANG T Y, KISHORE V, WU F, et al. BERTScore: evaluating text generation with BERT[EB]. arXiv preprint, 2019, arXiv: 1904.09675.

作者简介



刘天扬（2001- ），男，华东师范大学数据科学与工程学院硕士生，主要研究方向为教育领域的大模型应用。



寇思佳（1992- ），女，教育部教育技术与资源发展中心（中央电化教育馆）助理研究员，主要研究方向为教育大数据与知识图谱。



金旭（1982- ），女，北京师范大学附属实验中学教师，主要研究方向为中学美术教育。



王文静（1987- ），女，北京师范大学附属实验中学教师，主要研究方向为中学美术教育。



陆雪松（1985- ），男，华东师范大学数据科学与工程学院副教授，主要研究方向为数据驱动的计算教育学。

收稿日期: 2024-10-31
通信作者: 陆雪松, xslu@dase.ecnu.edu.cn
基金项目: 国家重点研发计划项目(No.2023YFC3341200)
Foundation Item: The National Key Research and Development Program of China (No.2023YFC3341200)