doi:10.3772/j.issn.2095-915x.2025.03.002

基于大语言模型多模态融合驱动的图书分类优化方法研究

开放科学（资源服务）标识码（OSID）

潘雪峰[1] 陈洁[2] 王超[1]

1. 辽宁工业大学图书馆 锦州 121000
2. 桂林理工大学图书馆 桂林 541000

摘要：[目的 / 意义] 探索基于多模态数据的大语言模型图书分类方法，解决传统图书分类方法存在的瓶颈。[方法 / 过程] 构建单模态数据集和多模态数据集，通过小样本学习策略构建提示语，对比测试 GPT-4 和 DeepSeek 模型图书分类结果。建立人工编目团队参照系，进行 4 次实验，每次实验通过 5 组对照量化模型效能。[结果 / 结论] 单模态分类中，DeepSeek 平均准确率（80.69%）显著高于 GPT-4（64.20%），提示语信息增加使 DeepSeek 准确率提升，GPT-4 因信息冗余出现性能波动。多模态场景下，DeepSeek 在精确率（94.5%）、召回率（84.0%）及 F1 值（88.9%）上全面优于 GPT-4，资深编目员与 DeepSeek 协同准确率达 99.0% 以上。大语言模型通过多模态协同表征与动态知识蒸馏机制，显著优化图书分类效能。

关键词：单模态；多模态；大语言模型；图书分类

中图分类号：G35

Research on Book Classification Optimization Method Driven by Multimodal Fusion of Large Language Models

PAN Xuefeng[1] CHEN Jie[2] WANG Chao[1]

1. Liaoning University of Technology Library, Jinzhou 121000, China;
2. Guilin University of Technology Library, Guilin 541000, China

Abstract: [Objective/Significance] To explore a large language model book classification method based on multimodal data and solve the bottleneck of traditional book classification methods. [Methods/Processes] A unimodal dataset and a multimodal dataset were constructed. Prompt language was constructed using a small sample learning strategy, and the book classification results of GPT-4 and DeepSeek models were compared and tested. Establishing a reference frame for the manual cataloging team, conduct 4 experiments, and quantify the effectiveness of the model through 5 control groups in each experiment. [Results/Conclusions]

基金项目 2023 年度教育部人文社会科学研究规划基金“‘人工智能驱动的科学研究’对科研数据基础设施建设的影响机制研究”（23YJA870002）。

作者简介 潘雪峰（1990-），硕士，馆员，主要研究方向为智能计算与图书馆信息化建设；陈洁（1981-），博士，副研究馆员，主要研究方向为信息资源建设与信息服务，E-mail：441751938@qq.com；王超（1983-），硕士，副研究馆员，主要研究方向为智能计算与图书馆信息化建设。

引用格式 潘雪峰 , 陈洁 , 王超 . 基于大语言模型多模态融合驱动的图书分类优化方法研究 [J]. 情报工程 , 2025, 11(3): 15-26.

doi:10.3772/j.issn.2095-915x.2025.03.002

In unimodal classification, the average accuracy of DeepSeek (80.69%) was significantly higher than that of GPT-4 (64.20%). The increase in prompt information improved the accuracy of DeepSeek, while GPT-4 experienced performance fluctuations due to information redundancy. In multimodal scenarios, DeepSeek outperforms GPT-4 in accuracy (94.5%), recall (84.0%), and F1 score (88.9%), with a collaborative accuracy rate of over 99.0% achieved by experienced catalogers and DeepSeek. The large language model significantly optimizes the efficiency of book classification through multimodal collaborative representation and dynamic knowledge distillation mechanism.

Keywords: Unimodal; Multimodal; Large Language Model; Book Classification

引言

图书分类作为知识组织的核心环节，深刻影响图书馆服务效能与学术资源利用效率。传统基于图书内容的分类方法通过人工标注与文本元数据（单模态）的融合，构建了层级化知识管理体系，解决了信息组织与检索难题。基于中图法等树状分类框架，将知识划分为逻辑清晰的大类与子类，形成清晰的知识体系。但在数字时代传统图书分类方法面临四维瓶颈[1]：第一，单模态语义割裂，数字图书内容形态已从纯文本扩展至多模态（如封面设计、图表布局），传统方法仅能解析结构化文本，无法捕捉视觉元素的语义价值。例如，儿童绘本的封面色彩直观反映适读年龄，科技专著的信息图表与文本形成互补关系，仅依赖文本分类将导致关键信息丢失。第二，领域知识理解浅层化，专业图书（如法律、医学类）中术语（如“善意取得”“ICD 编码”）[2]需结合上下文解析，传统规则引擎难以识别领域特异性，易引发误分类。例如，法律文本中“管辖”的语义需动态解析地域、层级等维度信息。第三，分类体系动态适配滞后，新兴交叉领域（如“量子计算伦理”）缺乏预定义标签，人工扩展层级结构成本高、响应慢。第四，人工图书分类耗时且易疲劳。基于多模态的图书自动分类主要解决传统单模态（如图书名称、作者信息、摘要）图书分类的信息局限性与准确性不足问题。通过融合文本、图像、版式等多维特征，克服单一文本分析对封面设计、图表内容等非结构化信息识别不足的缺陷，提升复杂主题、跨领域书籍的分类精度。同时自动化处理海量数据，降低人工标注成本，解决分类效率与扩展性难题，适应数字时代图书形态多样化的发展需求。大语言模型（LLMs）为多模态协同分析提供了创新范式，通过多模态语义对齐（如封面地域符号与文本关键词共现识别旅游图册类别）、领域自适应推理（基于提示工程解决术语消歧），以及动态知识蒸馏（自监督学习扩展分类标签），显著提升分类效能。本研究提出“提示语引导的多模态融合架构”，聚焦两大科学问题——如何实现文本、视觉与结构信息的互补融合？如何量化分类结果可信度并溯源错误？研究成果将为图书馆智能化转型提供方法论支撑。

1 相关研究

1.1 传统分类方法的技术局限

早期图书分类研究主要依赖人工规则与统

计学模型（如 LSA 潜在语义分析 [3]），通过关键词—类别映射实现分类。此类方法虽具备可解释性，但存在显著缺陷：第一，特征工程需人工设计规则与权重，难以适配多模态非结构化数据，导致处理效率低下且成本高昂；第二，语义鸿沟问题突出，传统模型无法捕捉文本隐含语义（如讽刺、隐喻）及跨模态关联（如图表与正文的协同表达），不利于提升分类的深度与准确性。

1.2 深度学习驱动的分类优化

神经网络的发展呈现多元化发展趋势。早期探索中，Kim[4] 构建了基于卷积神经网络的文档预测模型，开创性地将图像处理技术引入文本分类领域；Chang 等 [5] 进一步拓展应用场景，通过融合 CNN 与图像特征实现电子图书分类；Ting 等 [6] 则系统评估了朴素贝叶斯算法在图书文档分类中的效能边界。随着技术演进，以 CNN、RNN 为代表的深度学习通过自动特征提取机制显著优化了分类性能，典型应用包括文档主题预测模型 [7] 以及儿童读物的层级多标签分类等 [8]。然而当前技术体系仍存在显著局限：首先，多模态信息融合机制尚不完善，文本语义与视觉特征的跨模态对齐仍缺乏有效实现方案 [9]；其次，模型在专业领域或小样本场景下易出现泛化能力不足与过拟合现象；更为关键的是，现有范式普遍将分类过程与结果验证割裂，缺乏预测可信度的自评估机制，导致错误分类溯源困难。此外，高计算成本、参数调优复杂等技术门槛也制约着非专业用户的应用体验提升，这些问题共同构成了当前图书分类智能化转型的主要障碍。

1.3 大语言模型推动图书分类技术革新

大语言模型通过重构技术框架与实验验证推动了图书智能分类领域发展。当前研究热点显示，基于预训练—微调范式的大语言模型正突破传统方法的技术边界 [10]，以 GPT-4、Claude、PaLM 及 DeepSeek 为代表的模型体系，依托超大规模参数（>100B）与跨领域知识蒸馏机制，展现出语义关联解析、逻辑推理及少样本学习等突破性能力。相较于传统分类方法，其技术优势形成三个创新维度 [11-12]：（1）多模态协同表征机制通过跨模态预训练融合文本语义、图像特征及版面编码，构建多维联合表征空间，有效捕获学术专著中公式—插图语义关联等深层特征；（2）利用提示工程和思维链技术，可以实现知识在不同分类体系间的快速迁移，并依托内置知识图谱解决专业术语消歧与长尾类别识别难题；（3）生成式自验证框架突破传统分类—验证分离模式，建立端到端的可信度评估体系。

1.4 提示语优化驱动多模态图书分类效能提升

提示语作为人机协同的关键枢纽，其设计质量直接影响模型性能表现。如陈萍等 [13] 提出的“CIFE”框架系统探讨图书馆提示语工程的作用和价值，推动图书馆领域提示语设计专业化进程，实验数据显示优化后的提示模板可以提升分类准确率。针对图书信息的多元特征，当前研究已形成多模态解析体系 [14-16]：文本维度涵盖书目数据与全文内容特征；视觉维度解析封面设计（CLIP 模型特征提取）与版式布局（YOLOv7 检测）；时空维度则融合出版年代

doi:10.3772/j.issn.2095-915x.2025.03.002

时序编码与地域特征 GIS 映射。

本文提出“提示语 + 多模态数据 + 大语言模型”的新范式，探索大语言模型在图书自动分类中的有效性和实用性，以单模态（文本模态）和多模态（视觉模态）数据为切入点，采用小样本学习策略构建提示语，应用 GPT-4 模型 [17] 和 DeepSeek 模型测试图书自动分类的影响因素。

2 研究方法

2.1 数据来源

本文共收集两部分数据用于研究：（1）单模态数据。研究者收集了某大学图书馆 2023 年的图书采购信息，总计 16492 条记录。由于该校以工科为主，多领域协同发展，尤其在人工智能、信息技术等前沿领域的资源投入力度较大，因此在数据分布上“自动化技术”“计算机技术”等相关数据占比较多，达 15.63%。该校也对经管学科和法学领域较为重视，其购置资源分别占比 14.31% 和 8.73%，但均未超过 16%。同时该校兼顾人文素养培养，因此文学类书目较多。其余类别较为均衡，整体数据分布较为合理，如表 1 所示。由于图书购置过程常常出现副本等现象，因此经过数据预处理（去除缺失数据、重复数据及异常数据），最终筛选出 10994 条有效数据用于研究。每条有效数据包含图书名称、作者信息及摘要 3 个关键组成部分。（2）依据（1）的数据分布，等比例选取 360 组图书封面图片用于多模态数据研究。

表 1 不同类别图书占比情况

类别	占比（%）	类别	占比（%）	类别	占比（%）
A	1.11	O	4.27	TK	0.49
B	2.39	P	0.23	TM	2.80
C	1.97	Q	0.09	TN	3.02
D	7.79	R	0.86	TP	15.63
E	0.05	S	0.17	TQ	1.29
F	14.31	T	0.20	TS	0.99
G	3.88	TB	2.26	TU	8.73
H	3.25	TD	0.03	TV	0.03
I	7.86	TE	0.05	U	3.20
J	3.96	TF	0.02	V	0.06
K	2.36	TG	1.23	X	3.05
N	0.45	TH	1.51	Z	0.25

2.2 研究方法

2.2.1 单模态数据分析

本文基于大语言模型对单模态数据预测图书分类情况进行研究 [18]，采用小样本学习策略构建提示语信息 [19-20]。

提示语具体构成如下：

P1：{ 提示语 + 图书名称 }

P2：{ 提示语 + 图书名称 + 作者信息 }

P3：{ 提示语 + 图书名称 + 作者信息 + 摘要 }

将 P1、P2、P3 输入 GPT-4 模型和 DeepSeek 模型中，获得图书分类结果，将结果与人工分类结果进行对比，判断正确与否，提示语示例如表 2 所示。

表 2 单模态数据分析提示语示例

提示语信息	小样本学习策略		
你是某大学的图书馆员，请根据下面信息对图书进行分类，如《机器人动力学控制》，图书分类号：TP24	P1 题目：趣学算法	P2 题目：趣学算法 作者：陈小玉	P3 题目：趣学算法 作者：陈小玉 摘要：本书是用轻松有趣的方法学习算法的入门指南。按照算法策略分为 8 章。

2.2.2 多模态数据分析

（1）GPT-4 图书分类工作原理

应用 GPT-4 模型进行图书分类的核心逻辑是通过多模态协作处理，结合封面图像的 OCR 文字提取与视觉元素分析，输入 GPT-4 模型生成复合语义特征，最终映射到标准分类体系。

具体流程如下：将图书封面图片输入模型后，首先通过 OCR 提取文字，分析视觉特征进行多模态解析；然后构建语义 Prompt 对 GPT-4 模型进行多轮对话；GPT-4 模型采用分类推理策略，该部分进行两个操作（A、B），A：直接输出分类建议，通过置信度对分类结果进行评估，满足置信度预设 t 时自动生成索书号，完成分类。不满足 t 时进行人工复核，馆员修正系统，进行反馈学习；B：重新生成追问指令，人工补充描述，再次对话 GPT-4。GPT-4 图书分类工作原理如图 1 所示。

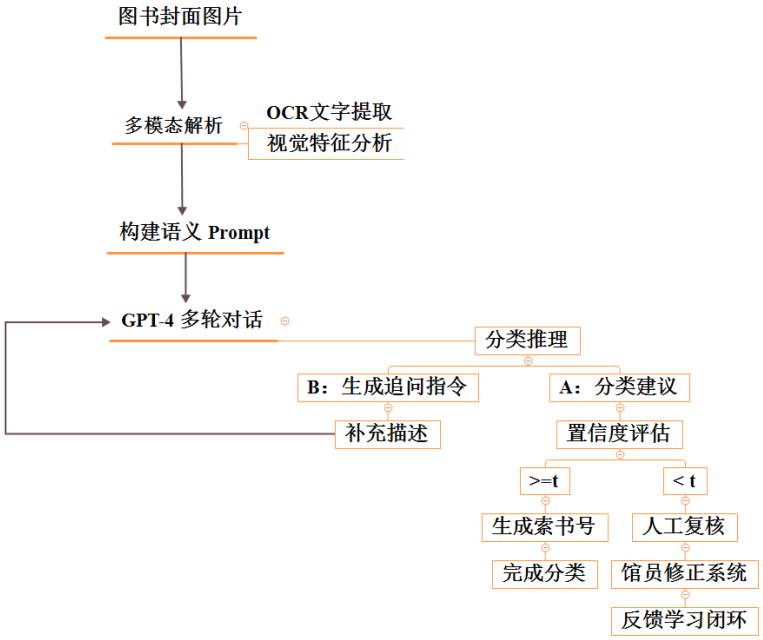

图 1 GPT-4 图书分类工作原理

（2）DeepSeek 图书分类工作原理

应用 DeepSeek 进行图书分类的核心逻辑是通过中文 OCR 增强引擎提取封面关键信息，结合 DeepSeek 的垂直领域知识库进行语义推理，实现图像转化为文本进而分类的端到端映射。

具体流程如下：图书封面图片输入模型后，首先进行预处理，对图像进行增强（如去噪和纠偏处理）；然后分割图像区域（如分割标题区和图案区域），该部分共进行两部分操作（A、B），A：通过视觉语义编码生成图案特征描述文字，B：通过 DeepSeek 模型内置的中文优化 OCR 引擎提取图片中的标题、作者、出版社等信息，将 A 与 B 结合构建分类模型 Prompt；DeepSeek 依据 Prompt 进行推理，推理过程采用多级分类决策方法。依据该方法输出中图法编码和主题标签。采用置信度矩阵对输出结果进行评估。满足预设的置信度 t 的结果自动入库，并更新领域知识库，模型继续增强训练；对不满足预设的置信度 t 的结果启用智能追问，请求补充信息，如人工输入摘要、目录等，并重新进入 DeepSeek 推理环节。DeepSeek 图书分类工作原理如图 2 所示。

doi:10.3772/j.issn.2095-915x.2025.03.002

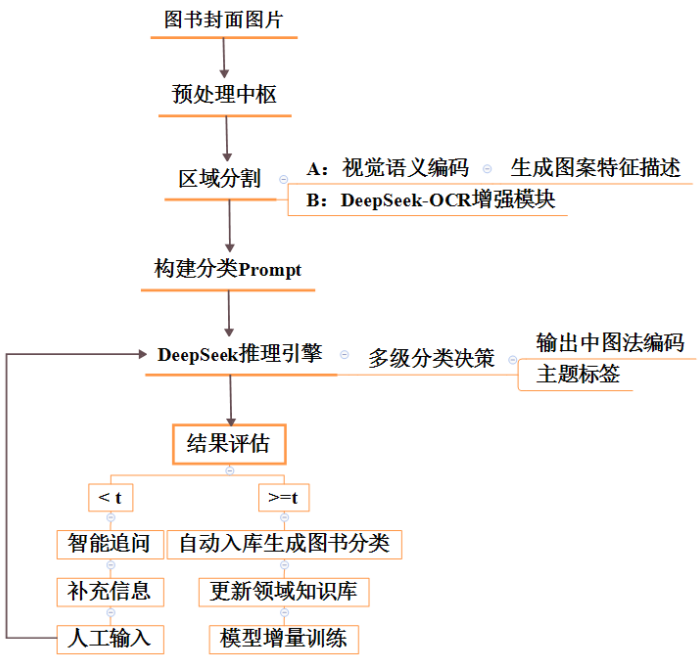

图 2　DeepSeek 图书分类工作原理

本文构建了一个涵盖 360 组图书封面图片的视觉—文本混合数据库[21]，每项数据单元包含图书的基础元数据（题目、作者）及复合型特征信息（封面视觉设计元素及补充性文本描述）。实验设计中，将带有特定引导指令的图片数据分别输入到 GPT-4 和 DeepSeek 模型框架中，如表 3 所示，系统记录模型的文本输出，并与专家标注的分类标签进行定量比对，旨在探讨不同方法在图书分类任务中的表现与差异。

表 3　多模态数据分析提示语（P4）示例

你是一名图书馆员，依据下面图书封面信息，根据《中图法》（第五版）对图书进行分类，例如：I242.47，10 字以内回复


注：单模态三种提示语分别为 P1、P2、P3，多模态的提示语顺次为 P4

2.2.3　研究过程

本文基于单模态和多模态两种方式，通过大语言模型对图书分类进行预测研究。单模态着重研究提示语对于图书分类的有效性；多模态探究大语言模型在图书自动分类方向上的可用性。为了更有效地验证结论，本文进行了实验验证，共进行 4 次实验，比较不同条件下图书分类的时效性和准确率等。

3　实证验证

3.1　单模态数据分析结果

基于小样本学习策略，设置 P1、P2、P3 三种提示语结构，采用 API 接口循环调用 GPT-4 模型和 DeepSeek 模型，模型预测结果与人工分类结果比较如图 3 所示。图中深灰色区域表示未返回有效预测值的数据量，灰白色区域为

预测正确量，灰色区域为预测错误量。左侧为GPT-4 模型，右侧为 DeepSeek 模型，圆环由里向外分别表示提示语为 P1~P3 情况下两种模型图书分类的情况。

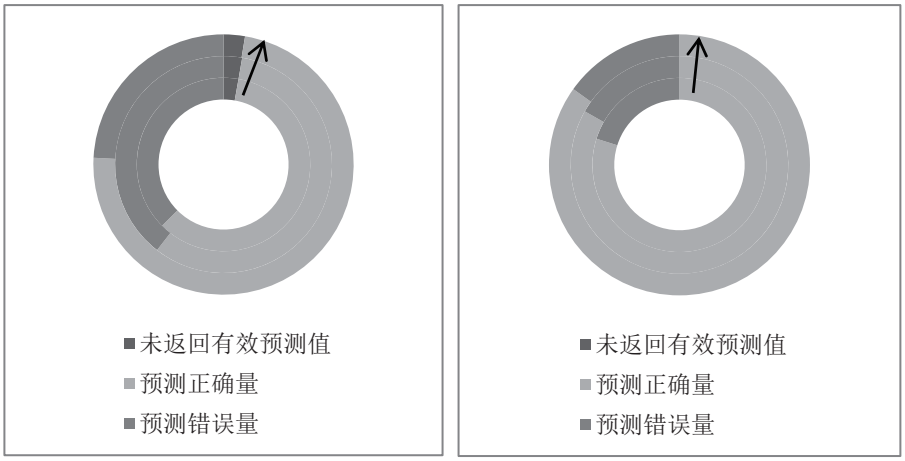

图 3　单模态数据下大语言模型预测情况

从图中发现 GPT-4 模型未能返回有效值的比率达 2.75%（P1~P3 均值），而 DeepSeek 模型均有返回值，说明 DeepSeek 模型稳定性较强；GPT-4 预测正确率均值为 64.20%（P1：6570/10944；P2：6500/10944；P3：8009/10944），DeepSeek 为 80.69%（P1：8102/10944；P2：9104/10944；P3：9288/10944），并且随着提示语信息的增加，两种模型的预测准确率变化稍有不同。GPT-4 在 P2 的准确率较 P1 降低了 0.639%，说明提示语冗余可能导致信息过载。然而，随着提示语信息的增加，DeepSeek 的预测准确率呈不断上升趋势。为了更加清晰地表述两种大模型基于 3 种提示语进行图书分类的结果，本文进行了样例测试，测试结果如表 4 所示。

表 4　GPT-4 与 DeepSeek 基于 P1~P3 提示语图书样例测试分类结果

P1：《汽车系统优化与最优控制》
P2：《汽车系统优化与最优控制》（奥）哈拉尔·瓦希尔 等 著
P3：《汽车系统优化与最优控制》（奥）哈拉尔·瓦希尔 等 著 本书结合优化理论与汽车工程实践，探讨汽车系统设计与控制的实际应用……

大语言模型	P1 分类结果	P2 分类结果	P3 分类结果
GPT-4	O221 或 TP273	TP273	O221 或 U46
DeepSeek	U46 或 U461.6	U461.6 或 TP273	U461.6

《汽车系统优化与最优控制》一书依据中图法第五版分类号应为 U461.6。从表 4 中发现，提示语为 P1 时 GPT-4 模型的分类结果为 O221 或 TP273。当提示语信息增加（P2）时，分类结果仍为 TP237，图书分类结果并不准确；当提示语为 P3 时，分类结果（U46）接近正确分类，出现上述问题的主要原因在于 GPT-4 模型在图书分类过程中过度关注“优化”与“控制”关键词，忽略“汽车系统”的领域归属，导致将“汽车工程类”图书错分为“自动控制类”图书。并且作者信息对于 GPT-4 模型预测图书分类效果不明显。

doi:10.3772/j.issn.2095-915x.2025.03.002

DeepSeek 模型通过书名高频词统计（如“汽车系统”）匹配领域库，优先锁定工程类而非纯数学类（P1），将图书分为 U46 或 U461.6；调用作者—研究领域映射表，识别该作者长期从事汽车工程研究（P2），将图书分为 U461.6 或 TP273；利用摘要关键词权重分析“汽车工程实践”>“优化理论”，基于中图法规则引擎（U 类优先于 O/TP 类）（P3），将图书分为 U461.6。随着提示语信息的增加，DeepSeek 模型预测图书分类结果更加准确。

DeepSeek 通过领域知识固化与结构化规则引擎，在工程类图书分类中实现了更高鲁棒性，而 GPT-4 的通用性使其在信息不全时更易偏离目标分类。

3.2 多模态数据分析结果

本文针对单模态模式下两种模型预测性能均未突破 85% 阈值的现状，转向多模态图书分类方法探索。实验选取 360 组多模态数据，通过提示语 P4 模式分别输入 GPT-4 与 DeepSeek 模型，将预测结果与专家标注对比，综合评估准确率、精确率、召回率及 F1 值 [22]，关键指标如表 5 所示。DeepSeek 的精确率（0.945）高于 GPT-4（0.913），表明 DeepSeek 对图书正确分类判定的严格性更强，这种特性在珍本古籍分类等容错率极低的场景中尤为重要，如将《永乐大典》残卷误标为普通史料可能引发学术争议。DeepSeek 的召回率（0.840）较 GPT-4（0.788）提升 6.6%，这对新兴交叉领域图书的覆盖率提升具有实质价值，避免知识孤岛现象，如《物理学的哲学》需同时归入哲学与物理学。两模型 F1 值差异（0.847 vs. 0.889）折射出策略分歧：GPT-4 以牺牲漏检率为代价换取高确信度，而 DeepSeek 通过混合专家架构的语义解析优势 [23]，在维持超高精确率的同时降低漏检率，更适合处理多标签细粒度分类，如“明清小说”类图书需关联文学、历史、社会学等多个类目。

分析两种模型图书分类工作原理不难发现，两种模型均基于 Transformer 架构，遵循“输入解析→特征提取→分类映射→置信评估”的基础流程并且支持人工修正数据的再训练闭环。不同之处在于 GPT-4 模型为通用多模态大模型；支持多种语言，特别是英文；以云端 API 调用为主；知识更新截至训练时间点；DeepSeek 模型为垂直领域优化模型；更适合中文场景尤其是古籍、方言方面；支持本地化轻量部署；可动态加载领域知识库。本文应用数据多为中文图书数据，因此 DeepSeek 模型的预测效果更佳。

表 5 多模态数据下大语言模型预测情况

模型	准确率（A）	精确率（P）	召回率（R）	F1 值
GPT-4	0.799	0.913	0.788	0.847
DeepSeek	0.905	0.945	0.840	0.889

4 图书分类实验、偏差解构与协同矫正研究

4.1 单/多模态图书分类效能实验分析

为了验证单模态数据情况下提示语对于图书分类的有效性，以及多模态数据情况下大语言模型在图书自动分类方向上的可用性，本文进行了实验验证。本研究从某高校文献数据库随机抽取 2000 册已完成标准编目的馆藏图书，

并组建包含 9 名不同专业资历编目员的编目团队（含 3 名具有 1 年以上实务经验的实习编目员、3 名具备 5 年编目工作经历的专业编目员及 3 名拥有 10 年以上分类实践经验的资深编目员），构建多层次人工分类参照系。两种模态共进行 4 次实验。结合 GPT-4 和 DeepSeek 模型对图书分类进行测试，并记录分类时间、分类错误数和准确率，实验结果如表 6 所示。

表 6　实验结果

数据模态	实验过程	测试对象	分类时间（分钟）	分类错误数量（本）	准确率（A）
单模态	实验 1：将 500 本图书名称和提示语（P1）提供给测试对象，获得图书分类结果。	GPT-4 模型	20	37	0.926
		DeepSeek 模型	28	26	0.948
		实习编目员	117	57	0.886
		专业编目员	92	25	0.950
		资深编目员	81	17	0.966
	实验 2：将 500 本图书名称、作者信息和提示语（P2）提供给测试对象，获得图书分类结果。	GPT-4 模型	18	39	0.922
		DeepSeek 模型	29	25	0.950
		实习编目员	112	49	0.902
		专业编目员	87	23	0.954
		资深编目员	79	16	0.968
	实验 3：将 500 本图书名称、作者信息、摘要和提示语（P3）提供给测试对象，获得图书分类结果。	GPT-4 模型	15	26	0.948
		DeepSeek 模型	34	19	0.962
		实习编目员	109	42	0.916
		专业编目员	76	19	0.962
		资深编目员	64	13	0.974
多模态	实验 4：将 500 本图书封面图片和 P4 提供给测试对象，获得图书分类结果。	GPT-4 模型	18	7	0.986
		DeepSeek 模型	23	5	0.990
		实习编目员	100	17	0.966
		专业编目员	71	6	0.988
		资深编目员	60	4	0.992

本文通过对比单模态与多模态实验数据下 5 种测试对象进行图书分类的时间、错误数量和准确率，揭示了不同信息维度对图书分类效能的影响规律。以单模态实验为例，当测试对象仅获取图书名称和提示语（实验 1）时，GPT-4 模型较其他 4 种测试对象耗时最少，资深编目员准确率（0.966）最高。相较于仅处理书名，增加作者信息（实验 2）后，GPT-4 模型的准确率微降 0.4 个百分点，源于部分作者跨领域创作引发误判，如将科普作家撰写的科幻小说《星际漫游》误归入科技类。而在引入摘要信息（实验 3）后，模型通过内容概要增强语义理解，错误量显著减少，印证了补充文本信息对 AI 分类决策的正向作用。实习编目员在实验 1 中错误量达 57 本，主要集中于学科交叉领域，如将《计算语言学导论》错误标注为“语言学”而

doi:10.3772/j.issn.2095-915x.2025.03.002

非“计算机科学”。但随着信息维度扩展，其准确率从 88.6% 提升至 91.6%，专业编目员和资深编目员的准确率分别增长 0.012 和 0.008，说明辅助信息的认知补偿效应。

多模态实验呈现突破性进展，封面图像的引入使分类准确率跃升。以《中国古建筑构造》为例，单模态下 GPT-4 仅凭书名有 7.4% 概率误判为“历史地理”类，但结合封面中的榫卯结构图解后准确归入“建筑工程”类，错误率降至 1.4%。值得注意的是，得益于对中文出版特征的深度学习，DeepSeek 模型在多模态场景下展现出独特优势，5 本错误量低于 GPT-4 的 7 本，并且其准备率（0.990）逼近资深编目员（0.992）水平。

4.2 模型认知偏差深层解构

针对大语言模型的误判现象，第一在于语义理解的偏差性。自然语言的多义特性导致模型在文本特征提取时易产生认知偏移，尤其在处理“政治法律”与“历史学”交叉领域时更为显著。以图书《探源》的分类为例，DeepSeek 模型对“政治”和“哲学”的语义权重分配失衡，将本应归入 D61 类别的图书错误标注为 B0，这种错误源于模型对复合型政治文本的深层语义解构能力不足，大语言模型构建时应引入领域知识图谱，强化政治哲学交叉概念的向量表征。第二是训练数据分布失衡问题，如 GPT-4 在处理《新时代 新思想》时，因训练集中 D6 类样本的稀疏性及标注噪声，导致该著作被误分为 D25，这凸显出现有语料库在专业领域细分类目的构建缺陷。大语言模型构建时应建立主动学习 [24] 的语料动态扩充机制，优先采集中图法 4 级以上细分类目数据。第三是术语消歧机制失效，如误将属于 TN410.2（电子器件制造）的《一板成功》错分至 F832.51（股票市场）。由于大模型机械关联“板”字与金融概念，忽视“电路板”专业语义，反映了模型在“板”字的多义消解中存在认知盲区。大语言模型需构建跨学科术语消歧引擎，引入动态语境窗口。第四是分类体系更新滞后，主要体现在大模型面对某些复杂图书数据时无法给出准确的分类意见。如大模型误将《数学领域中的发明心理学》归入 B804.4（创新性思维），未识别 O1-0（数学理论）类目，暴露出复杂图书分类解析缺陷。大语言模型构建时应开发细粒度决策树算法，对复杂类目建立多层判别规则。

4.3 人工分类误差溯源与人机协同矫正体系构建

人工分类的误差溯源研究显示，认知框架的局限性是主要诱因。分类标准的理解偏差直接导致系统性错误，如《压缩空气泡沫灭火技术及应用》本属于 TU998.1 建筑科学类图书，却被误归入 TU53 有机材料、建筑化工材料类，这种错误映射反映出实习编目员对中图法层级结构的掌握缺陷。大语言模型训练中嵌入分类法知识图谱增强模块以矫正认知锚点偏移会避免上述问题。进一步观察发现，跨学科著作的分类困境具有普遍性，《陕西文化》在地理风貌与艺术传承的双重属性下，专业编目员因缺乏多维度分析框架而难以确定主类目归属。构建大语言模型时应注重学科关联度量化矩阵，如当艺术传承与地理风貌权重差值 <0.2

时触发资深编目员人工复审机制会对模型确定主类目归属有所帮助。实验数据还揭示生理性误差的显著影响，连续工作 120 分钟后，资深编目员在分类《风控》一书时将 F830.5 误录为 F830.7，证实了注意力衰减与分类准确率的负相关性。

基于误差产生机理的深度解析，构建人机协同的智能分类系统具有实践必要性。如自动化系统对 500 本图书的初级分类仅需 18~34 分钟，同时保留资深编目员对争议案例的终审权。在《量子力学哲学导论》的分类实践中，这种协作模式成功规避了单一模态的认知局限：AI 系统通过摘要分析建议标注为 B08 类别，资深编目员结合封面设计的量子符号元素最终确认为 N02 交叉学科类。这种双向校验机制使整体错误率显著下降，且处理效率有所提升。验证了智能辅助系统在文献组织中的增效价值。值得注意的是，系统需建立动态学习机制，将人工复核产生的修正数据实时反哺模型训练，从而形成分类精度持续优化的闭环体系。

5 结语

本文针对传统图书分类方法的局限性，以多模态数据为切入点探索大语言模型的创新应用。实验表明，多模态数据较单模态具有显著预测优势，但模型性能并不与数据规模呈绝对正相关，不同架构模型的信息处理能力存在显著差异。通过引入小样本学习策略优化模型提示机制，可有效提升分类准确率。未来研究将重点推进模型轻量化与知识嵌入技术的协同发展，借助 DeepSeek 等模型的本地化部署优势，结合多模态融合技术，在保障数据安全的基础上构建动态适配系统，实现出版机构知识库的实时分类更新。

参 考 文 献

[1] 胡泽文 , 任萍 , 崔静静 . 基于机器学习模型的科技论文潜在 “ 精品 ” 识别研究 [J]. 情报学报 , 2023, 42(2): 189-202.

[2] 伍祥林 , 刘煜民 . 深度学习技术在 ICD 智能编码中的应用综述 [J]. 现代医药卫生 , 2025, 41(1): 192-195.

[3] 包翔 , 刘桂锋 . 一种基于概率潜在语义分析的专利主题标引方法研究 [J]. 情报工程 , 2020, 6(3): 15-24.

[4] KIM Y. Convolutional neural networks for sentence classification[C]// Association for computational linguistics. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha: Association for Computational Linguistics, 2014: 1746-1751.

[5] CHANG T P, CHEN H M, CHEN J Q. A deep learning co-training framework for e-book classification[C]// Chin-Yi University of Technology. In 2020 International Symposium on Computer, Consumer and Control (IS3C). Taiwan: IEEE Power and Energy Society, 2020: 376-379.

[6] TING S L, IP W H, TSANG A H. Is Naive Bayes a good classifier for document classification[J]. International Journal of Software Engineering and Its Applications, 2011(3): 37-46.

[7] 田保军 , 刘爽 , 房建东 . 融合主题信息和卷积神经网络的混合推荐算法 [J]. 计算机应用 , 2020, 40(7): 1901-1907.

[8] 成全 , 董佳 . 面向分级阅读的儿童读物层级多标签分类研究 [J]. 数据分析与知识发现 , 2023, 7(7): 156-169.

[9] 吕璐成 , 韩涛 , 周健 , 等 . 基于深度学习的中文专利自动分类方法研究 [J]. 图书情报工作 , 2020, 64(10): 75-85.

[10] 戎璐 . 面向图书自动分类的大语言模型提示学习研究 [J]. 图书馆学研究 , 2024(1): 86-103.

doi:10.3772/j.issn.2095-915x.2025.03.002

[11] 闫晓慧 , 杨杰 , 杨文霞 , 等 . 学者合作网络外部影响力与学术成就的关系研究 [J]. 图书情报工作 , 2023, 69(2): 96-107.
[12] 韩佳艺 , 刘建伟 , 陈德华 , 等 . 深度长尾学习研究综述 [J]. 自动化学报 , 2025, 51(5): 985-1020.
[13] 陈萍 , 朱军华 . 图书馆提示语工程：价值、设计及实现 [J]. 图书馆工作与研究 , 2024(9): 62-70.
[14] 符荣鑫 , 杨小华 . AIGC 语言模型分析及其高校图书馆应用场景研究 [J]. 农业图书情报学报 , 2023, 35(7): 27-38.
[15] 缪翌 , 张卫锋 , 徐领 . 基于 CLIP 的视频时刻检索预训练模型 [J]. 计算机应用研究 , 2024, 41(12): 3866-3872.
[16] 张思敏 , 刘新妹 , 殷俊龄 , 等 . 基于 YOLOv7 改进的 PCB 缺陷检测方法 [J]. 计算机与现代化 , 2024(12): 45-52.
[17] 李华明 . 机遇与挑战：ChatGPT 赋能图书馆知识服务中的应用研究 [J]. 农业图书情报学报 , 2024, 36(8): 96-105.
[18] 刘继明 , 张培翔 , 刘颖 , 等 . 多模态的情感分析技术综述 [J]. 计算机科学与探索 , 2021, 15(7): 1165-1182.
[19] 张继 . 基于任务自适应的小样本学习研究 [D]. 成都：电子科技大学 , 2024.
[20] 喻金龙 , 张磊 , 许宁 , 等 . ChatGPT 在风湿科中医电子病历症状信息预处理中的应用 [J]. 中华中医药学刊 , 2025, 43(3): 24-29, 301.
[21] 付永华 , 司俊勇 , 周九常 . 智慧图书馆多模态阅读情感研究 [J]. 图书馆工作与研究 , 2024(2): 25-33.
[22] 刘思得 , 李东升 . 面向高校图书馆智慧服务的大语言模型探索性研究 —— 以命名实体识别任务为例 [J]. 现代情报 , 2024, 44(12): 102-121.
[23] SHAZEER N, MIRHOSEINI A, MAZIARZ K, et al. Outrageously large neural networks: the sparsely-gated mixture-of-experts layer[J]. arXiv preprint arXiv:1701.06538v1, 2017.
[24] 何浩 , 张辉 . 基于改进主动学习的入侵检测方法 [J/OL]. 计算机科学 , 1-13 (2025-03-08) [2025-06-16]. http://kns.cnki.net/kcms/detail/50.1075.TP.20250317.1847.034.html.