

节,智能传感器结合 AI 算法可实时监测土壤中重金属铅、镉与有机磷农药残留,数据异常时及时预警,预警准确率 96.3%;通过电子鼻结合机器学习,可快速识别食品中微生物污染产生的挥发性有机物,实现食源性致病菌沙门氏菌的早期检测,准确率 97.8%。

营养成分分析:利用机器学习模型可快速分析食品中蛋白质、碳水化合物、维生素等营养成分含量。一般在葡萄酒品质分析中,结合感官与化学分析数据构建机器学习模型,可实现对葡萄酒香气浓郁度的客观化评价,关联其有机酸、多酚类成分与感官指标,决定系数 $R^2 \geq 0.921$ 。食品掺假检测:通过光谱技术结合 AI 算法,可实现对食品掺假的精准识别。利用拉曼光谱结合 1D-CNN 模型实现芝麻油掺假检测准确率达 100%;在蜂蜜检测中,通过机器学习算法分析其化学成分,可有效识别蔗糖掺假行为,实现地理溯源准确率达 95.28%。内蒙古地区乳制品企业采用“AI+近红外光谱”技术,结合本地奶源成分大数据,实现蛋白质、脂肪含量的实时检测,检测周期从 2 小时缩短至 5 分钟。

3.3 多模态数据协同检测

通过外观与内在品质的同步评估,融合图像数据与成分检测数据,构建多模态 AI 模型,实现食品外观缺陷与内在成分的同时检测。在食品的肉类检测中,通过可见光图像识别表面淤血、异味等外观缺陷,结合近红外光谱数据分析蛋白质、脂肪含量及兽药残留,通过大数据关联两者特征,构建综合评价模型,全面评估肉类品质与安全性,检测效率较传统方法提升 3 倍以上,综合准确率达 97.3%。

4 当前面临的挑战

数据方面:隐私与法律顾虑导致食品企业间形成“数据壁垒”,数据共享不足;多模态检测数据缺乏统一标准格式与标注规范,不同设备数据难以融合,影响模型跨场景适配;部分地区数据采集意识薄弱,且高温、高温等极端环境下数据采集难度大,训练数据不足(中小企业样本量常 <500 组)直接限制模型准确性。

技术方面:图像识别易受光照、食品形态差异影响,对微小缺陷(<0.1mm)的识别准确率 <85%;成分分析设备成本高昂(单台售价 ≥ 50 万元),且局限于实验室,难以实现现场快速检测;AI 模型存在“黑箱问题”,可解释性差,不利于结果溯源验证,且复杂模型对硬件资源要求高 Transformer 模型需 ≥ 16 GB 显存,边缘部署适配性差,难以在中小食品企业简易设备上运行。

应用方面:基层检测机构缺乏 AI 与食品安全检测复合型人才,人才短缺制约技术普及;部分 AI 检测设备操作复杂、对环境要求高,落地难度大;行业缺乏统一的 AI 检测技术标准与资质认证体系,检测结果认可度有待提升。

5 未来发展趋势

5.1 构建“AI+物联网+区块链”一体化监管体系

未来将进一步推动 AI 与物联网、区块链技术的协同整

合,通过物联网实时采集食品供应链环境数据,AI 模型进行风险预警与质量预测,区块链保障数据不可篡改,实现从农田到餐桌的全链条透明化监管。构建全国性食品安全大数据平台,整合各地检测数据、生产数据、流通数据,通过 AI 模型实现跨区域、跨环节的风险预警。

5.2 提升现场检测能力,发展设备微型化与模型轻量化

针对成分分析设备成本高、便携性差的问题,未来将重点发展微型化检测传感器,集成微处理器实现数据实时采集与分析,提升现场检测能力。同时,优化轻量化 MobileNet、YOLOv8-tiny 等 CNN 模型的边缘部署效率,降低硬件资源依赖,开发无需专用学习终端、支持离线运行的 AI 检测系统,满足中小食品企业的实际需求。

5.3 增强检测结果的可信度,提升模型优化与加强标准化

加强多模态数据融合算法研究,提升模型对复杂食品基质的适应性;建立 AI 检测技术的 ISO/TC 34 相关标准标准化操作规范,构建统一数据格式与标注规范,提高检测结果的重现性与可比性。同时探索注意力机制、梯度可视化模型可解释性方法,平衡模型复杂度与可解释性,增强检测结果的行业认可度。

5.4 联邦学习的协同训练,隐私保护与数据共享,

不同地区的食品安全检测机构可通过联邦学习共享模型训练成果,无需传输原始数据,既保障数据安全,又解决训练数据不足的问题。采用联邦学习技术,在保护企业数据隐私的前提下,实现多机构数据协同训练,整合多方数据提升模型性能。

6 结语

本文系统梳理大数据与人工智能在食品安全检测中的融合路径,通过图像识别与成分分析技术的协同应用,实现从“被动检测”到“主动预警”的转型。当前技术已在食品缺陷检测、掺假识别、品质分级等场景实现规模化应用,显著提升了检测效率与精度,但仍面临数据标准化缺失、模型泛化能力不足、边缘部署适配性差等挑战。

未来需重点突破数据标准化、边缘部署、隐私保护三大核心问题,通过构建“AI+物联网+区块链”一体化体系、优化轻量化模型、推广联邦学习技术,推动技术在中小食品企业的规模化应用。对于相关研究者而言,需聚焦实际应用中的痛点问题,加强多模态数据融合、模型可解释性等方向的研究,推动技术创新与成果转化,为食品安全全链条监管提供技术支撑。

参考文献

- [1] 朱孝山,刘伟伟.人工智能技术在食品加工中的应用研究[J].肉类研究, 2021.
- [2] 王琳琳,王昌伟,张素霞.人工智能技术在食品质量检测中的应用现状及策略[J].现代食品, 2025(7).
- [3] 林君萍.人工智能技术在食品安全监管中的应用——评《我国食品安全监管机制与效率研究》[J].食品工业, 2020, 41(12):2.

A multimodal data-based tourist destination recommendation algorithm and its application

Yuqing Hao

School of Economics, Hangzhou Dianzi University, Hangzhou, Zhejiang, 310018, China

Abstract

With the rapid expansion of user bases on online travel platforms, accurately identifying user preferences and enhancing personalized recommendation effectiveness have become critical challenges. This study addresses the cold start problem and feature utilization limitations in travel recommendation systems by proposing a novel recommendation method based on semantic retrieval and multimodal gating fusion (MAG). The approach involves collecting and preprocessing review texts and accompanying images, extracting multimodal features using MiniLM and OpenCLIP, aligning them through two-layer MLP networks, and achieving unified semantic space alignment for user-to-attraction matching via adaptive gating weight fusion. Experimental results demonstrate that this method significantly outperforms fixed fusion baselines in multimodal retrieval tasks, delivering higher recommendation accuracy and enhanced interpretability.

Keywords

travel recommendation algorithm; multimodal fusion; cross-modal alignment; adaptive gating (MAG); semantic retrieval

基于多模态数据的旅游目的地推荐算法及其应用

郝雨青

杭州电子科技大学经济学院, 中国·杭州浙江 310018

摘要

随着在线旅游平台用户规模的扩大,如何精准挖掘用户偏好、提升个性化推荐效果成为重要课题。本文针对旅游推荐中的冷启动与特征利用不足问题,提出一种基于语义检索与跨模态门控融合(MAG)的推荐方法。通过采集并预处理评论文本与配图,利用MiniLM与OpenCLIP提取多模态特征,经两层MLP对齐并以门控权重自适应融合,构建统一语义空间实现用户—景点匹配。实验结果表明,该方法在跨模态检索中显著优于固定融合基线,具备更高的推荐精度与可解释性。

关键词

旅游推荐算法;多模态融合;跨模态对齐;自适应门控(MAG);语义检索

1 引言

近年来,随着数字经济的快速发展,全球旅游业正经历由线下向线上转型的关键阶段。根据世界旅游城市联合会(WTCF)2023年报告,全球在线旅游市场规模已达8950亿美元,年增长率保持在12%-15%^[1]。从技术演进角度看,旅游推荐算法大致经历了三个发展阶段:基于内容的推荐(Content-based)、协同过滤推荐(Collaborative Filtering)以及近年来逐渐兴起的基于深度建模与多模态融合的智能推荐^[2]。早期工作以基于内容的方法与邻域式或项目式协同过滤为主,强调相似度与共现关系的刻画,但在数据稀疏与冷启动情境下易受限^[3]。随后,矩阵分解与排序学习等方法成为主流,通过潜在因子建模与隐式反馈优化有效缓解稀

疏性问题^[4]。进入深度学习阶段,工业界与学术界广泛采用Wide&Deep与NCF等端到端框架,会话/序列推荐中基于自注意力的SASRec也成为稳健基线^[5-7]。面向旅游POI的时空属性,研究进一步将地理距离与时间因素纳入建模(如GeoMF),强调空间邻近性与时序性对用户偏好的影响^[8]。上述技术脉络及其在旅游场景的落地,国内文献从基于内容过滤、协同过滤、混合推荐、深度学习与上下文感知等维度也做了系统梳理^[9-12]。

多模态方向方面,典型方法如VBPR将视觉线索纳入排序优化,以缓解冷启动与图文错配;而跨模态预训练CLIP显著提升文本—图像对齐能力,使“以文找图/项目”的语义检索成为可行基础^[13-14]。同时,列表多样性重排常用MMR在“相关性—冗余”之间折中,缓解同质化扎堆问题^[15]。总体看,现有研究仍存在三方面不足:其一,现有研究对旅游场景中的时空和地域差异考虑不足,影响了推荐结果的准确性和泛化性能;其二,现有研究对多模态数据多

【作者简介】郝雨青(1995-),男,中国辽宁丹东人,硕士,研究方向:数据挖掘。

利用不足,未能充分发挥不同模态信息之间的互补性;其三,现有研究对工程落地性的讨论仍然有限,尤其在轻量化召回与多样性可控方面缺乏系统验证。

基于此,本文以评论文本与景区图片为两大模态,采用跨模态对齐与门控融合获得统一景点向量,并以用户偏好描述形成查询向量,完成 Top-K 检索与推荐和可解释排序;相较仅依赖单模态或固定权重融合的做法,本文的多模态融合策略在保证实现复杂度可控的同时,提升了跨模态检索一致性与推荐排序质量。相关设计既呼应了国内外对“上下文—多源—可解释”方向的共识,也契合了在线旅游平台在提升转化与用户满意度方面的现实诉求。

2 数据采集与数据预处理

2.1 数据获取

为构建高质量的推荐算法训练数据集,本文主要采集携程旅游平台的用户一目的地交互行为数据与用户生成内容(User Generated Content, UGC)。选择携程作为数据源的理由是:(1)结构化程度高,景点页与评论页字段稳定,包含评分、评论时间、配图、标签等,利于解析与对齐;(2)评论密度与覆盖度高,热门景点评论量充足,便于构建中等规模多模态数据样本;(3)可复现性好,按城市→景点→评论分页的 URL 组织清晰;(4)多模态可用,同一评论通常可获得文本和配图,满足多模态融合需求。

文中数据通过 Python 编写的网络爬虫程序自动化抓取,选取了华东地区具有代表性的 34 个热门景点作为样本集合。每个景点采集约 20-50 条用户评论,总计获得评论文本 8160 条内容。结构化数据字段包含景点名称、评论人、评论时间、评论内容、IP 属地、评分等信息。同时,为丰富模态信息并辅助图像特征建模,采集了 3761 张用户上传的评论附图,用于后续的多模态分析与建模。用户评论结构化数据的部分字段如表 1 所示。

表 1 用户评论数据部分字段

字段名	字段说明	样例
景点 ID	旅游景点编号	80261
景点名称	旅游景点名称	拙政园
评论人	评论用户的昵称	苍宇行者
评论时间	用户评论时间	2025-07-20
评论内容	用户评论内容	非常好这次体验
IP 属地	用户 IP 归属地	浙江
评分	用户对景点的打分	3

图 1 展示了用户评论页面的内容信息,包含了用户昵称、评分、评论内容及配图。

2.2 数据预处理

为保证后续模型训练的稳定性与推荐结果的有效性,本文从文本、图像与跨模态稳健性三方面进行标准化预处理。

2.2.1 文本侧预处理

为保证数据质量,先对缺失评分或评论内容的记录直接剔除,IP 属地/用户昵称等辅助字段缺失可保留;同时识别并去除内容相同或时间戳近似的重复评论,避免样本偏置。之后采用标准化清洗流程:去除 HTML/特殊符号与冗余标点,编码统一为 UTF-8,繁体统一转为简体,随后用 jieba 分词并结合停用词表过滤无效词。

截断与汇总:针对每个景点的用户评论,按时间或相关性排序后,最多保留 60 条并拼接,以抑制长尾评论对语义的噪声放大,同时保留主体偏好。

向量化与归一化:采用 Sentence-BERT 对拼接语料进行编码,并在编码前后执行 L2 归一化,确保不同样本、不同长度文本的表示处于可比尺度,利于语义检索与后续融合的数值稳定。



图 1 用户评论区页面示例

2.2.2 图像侧预处理

首先,在有效性检查与解码阶段,对采集到的图片 URL 再次做规范化与去重,剔除重复链接与明显失效的外链;下载成功的图片统一解码为 RGB 三通道,并记录失败样本以便聚合时剔除,避免异常数据进入候选库。之后使用 OpenCLIP 的官方流程进行 resize 中心裁剪归一化,获得单图向量表示。其中,尺寸调整与中心裁剪将图片对齐到模型默认输入分辨率(如 224 × 224),通道归一化采用 OpenCLIP 的默认均值与标准差(典型设置为 mean=(0.481, 0.458, 0.408), std=(0.269, 0.261, 0.276)),以消除不同来源图片在亮度与色彩分布上的系统性差异。

针对同一景点对应的多张图片,本文在编码完成后对其特征进行均值池化以得到单一的景点级图像表示,便于与文本侧一一对齐并进入统一的融合流程。若某景点缺失图片,则在数据侧以零向量占位,避免样本被丢弃并便于统一处理。

2.2.3 预处理后数据规模

经数据清洗与规范化处理后,形成与实验一致的候选

集合。共保留景点数 34 个，评论文本合计 8008 条，文本 / 图像 / 融合向量维度分别为 384/512/512。

3 特征建模与算法构建

本文采用两阶段流程。首先从用户的偏好描述或历史文本中提取语义表示，经同一映射得到用户画像向量，作为检索查询。随后对每个候选景点，将其文本特征用轻量两层 MLP 映射到与图像一致的空间，依据文本与图像的一致性产生门控权重（主设定 $\tau=2.0$ ），将两模态自适应融合为单一景点向量，无图样本自动退化为仅文本。最终以用户画像与融合后景点向量的余弦相似度进行排序，结合城市或关键词等规则与可选的最大边际相关（MMR）去冗余，输出 Top-K 推荐结果。

3.1 用户画像属性构建

用户给出一段偏好描述 q （无显式偏好时，可将其历史评论拼接为 q ）。采用 paraphrase-multilingual-MiniLM-L12-v2 作为文本编码器，记为 Enc_T 。对用户偏好描述进行编码，并在输出后执行 L2 归一化，经同一投影映射映入到共享空间，得到用户画像向量：

$$z_{user}=g\left(L2\left(Enc_T(q)\right)\right) \quad (1)$$

其中与景点文本分支共享参数，确保用户向量与景点向量处于同一语义空间。

3.2 旅游景点特征提取

本文对每个景点同时建模文本与图像，并在同一潜在空间做自适应融合。首先，使用 paraphrase-multilingual-MiniLM-L12-v2 作为文本编码器，将文本编码为 e_{text} ，图像部分采用 OpenCLIP ViT-B/32 作为基础编码器，每张图像在经过模型的视觉编码器后，会得到一个全局特征向量 e_{image} 。为消除不同模态的分布差异，分别用两层 MLP 将全局向量投影到共享维度，记

$$z_{text}=\text{Linear}_2\left(\text{LayerNorm}\left(\text{GELU}\left(\text{Linear}_1\left(e_{text}\right)\right)\right)\right) \quad (2)$$

$$z_{image}=\text{Linear}_2\left(\text{LayerNorm}\left(\text{GELU}\left(\text{Linear}_1\left(e_{image}\right)\right)\right)\right) \quad (3)$$

其中， z_{text} 、 z_{image} 为投影后的向量，为后续的跨模态融合创造条件。

跨模态对齐阶段训练时使用联合对齐损失，使两模态在该空间内幅度与方向一致：

$$\mathcal{L}_{align}=0.5 \cdot \text{MSE}\left(z_{text}, z_{image}\right)+0.5 \cdot\left(1-\cos\left(z_{text}, z_{image}\right)\right) \quad (4)$$

在得到对齐表征后，依据样本级模态一致性做门控融合（Multimodal Adaptive Gating, MAG）门控：令

$$\alpha=\sigma\left(\tau \cdot \cos\left(z_{text}, z_{image}\right)\right) \quad (5)$$

$$z=\alpha z_{text}+(1-\alpha) z_{image} \quad (6)$$

其中 τ 为温度参数， $\sigma(\cdot)$ 为 Sigmoid， α 为按样本动态计算的门控权重。为对照，本文设置固定融合基线：

$$z_{base}=\lambda z_{text}+(1-\lambda) z_{image}, \lambda=0.5 \quad (7)$$

固定权重在样本间模态信息量不均衡时易产生“平均化”损失、解释性不足，实验上也显著弱于动态门控。

3.3 推荐算法流程

旅游目的地推荐算法的大致流程为，以 3.1 得到的用户画像向量 z_{user} 为查询，对之前得到并缓存的全部景点融合向量集合 z_{item} 进行排序；如需限定范围，先按城市 / 关键词做布尔预过滤得到候选集合 $C \in z_{item}$ ，随后以余弦相似度 $s(u, i)=\cos\left(z_{user}, z_{item}\right)$ 计分并取 Top-K，该有序列表即为推荐结果的基础输出。

为避免“相似景点扎堆”的同质化，本文在已按余弦相似度排好的列表上采用最大边际相关（Maximal Marginal Relevance, MMR）做二次重排，以在相关性与多样性之间取得折中。其核心思想是，在保证与用户偏好相似度较高的前提下，惩罚与已选结果过于相似的新候选，从而拉开风格与地域等维度的差异。

MMR 的计算方法为：

$$\text{MMR}(i)=\lambda s(u, i)-(1-\lambda) \max _{j \in S} \cos\left(z_i, z_j\right) \quad (8)$$

其中 S 为当前已选 Top 列表，第一项保证与用户相关，第二项抑制候选间冗余。本文默认 $\lambda=0.7$ （可在 0.5–0.9 内调节），仅改变 Top-K 的内部顺序，不修改主体评分函数 $s(u, i)$ 与候选集合 C 。

推荐算法的执行路径如图 2 所示。左侧为用户偏好建模流程，右侧为景点多模态特征提取与融合流程，二者在相似度计算阶段汇合后，依次经过过滤与 MMR 去冗余，最终输出 Top-K 推荐结果。

4 推荐算法实现与评估

4.1 实验环境与评价指标

在完成推荐算法模型构建与特征设计后，本文进一步在本地实验环境中实现了推荐算法的部署与评估，旨在验证基于语义检索和跨模态对齐—门控融合的推荐方法在推荐效果与数值稳健性方面的实际表现。开发环境基于 Python 编程语言，采用 scikit-learn、Surprise、jieba 等主流工具构建各功能模块。数据处理、建模与评估过程均在本地多核环境（Windows 11, Python 3.10, 32GB RAM）中完成，确保流程可复现且结果稳定。

评价指标采用 HR@K 与 NDCG@K。

命中率（Hit Ratio, HR@K）

$$\text{HR@K}=\frac{\text{Number of Hits@K}}{\text{Targets in Testing Sets}} \quad (9)$$

其中, Number of Hits@K 表示在测试集中, 系统为每个用户的推荐结果命中了其真实感兴趣的项目的总次数, Targets in Testing Sets 表示测试集中参与评估的用户总数。

归一化折损累计增益 (Normalized Discounted Cumulative Gain, NDCG@K)

$$DCG@K = \sum_{i=1}^K \frac{2^{rel_i-1}}{\log_2(i+1)}, NDCG@K = \frac{DCG@K}{IDCG@K} \quad (10)$$

其中, DCG 表示实际推荐结果的累计增益, IDCG 表示理想排序下的累计增益。NDCG 值越接近 1, 说明模型的推荐排序越优。

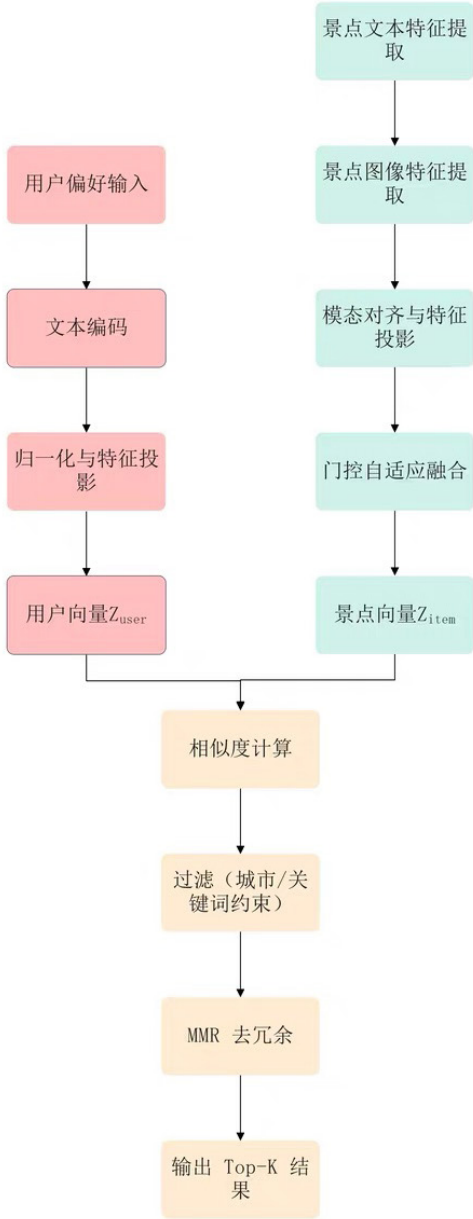


图 2 推荐算法流程图

4.2 跨模态检索与推荐结果

4.2.1 跨模态检索结果

跨模态检索的目标是以某一模态的数据作为查询, 能否在候选集中准确检索到与之对应的另一模态数据。在本文的实验中, 以景点评论文本作为输入查询, 要求模型能够在所有候选景点图像 / 融合表示中准确返回其对应的景点。若检索准确率和排序效果良好, 则表明文本与图像模态在共享空间中实现了有效对齐, 从而为后续的推荐算法提供了可靠的物品向量表示。

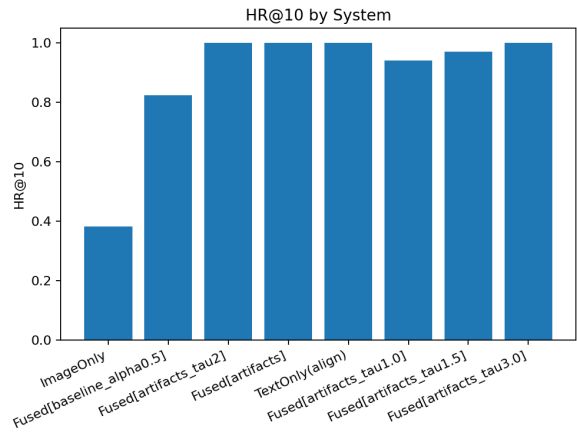


图 3 HR@10 柱状图

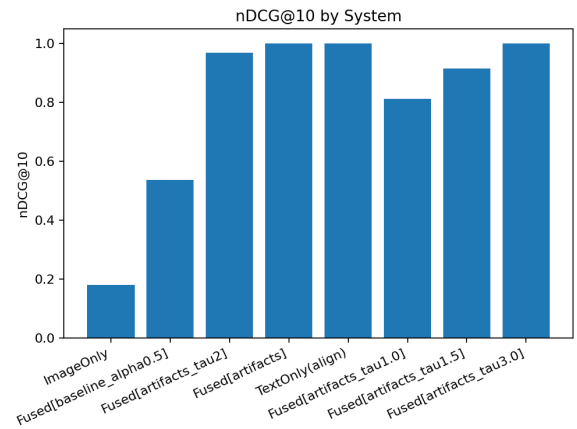


图 4 NDCG@10 柱状图

从图 3 和图 4 的整体结果看, Text-only 明显优于 Image-only, 说明在旅游评论场景中, 文本承载的可检索语义更强。固定加权 (Fixed- $\alpha=0.5$) 相较单模态有提升, 但在样本间模态信息量不均衡时受限。采用 MAG 后 ($\tau=2.0$), 在 HR@10 指标上达到 1.000, 显著优于 Fixed- α ; 在 NDCG@10 上约为 0.97, 与最优水平接近, 明显高于 Fixed- α 。这些结果表明, 动态门控融合能依据样本本级模态一致性自适应分配权重, 避免固定融合的“平均化”损失, 从而提升命中率与排序质量, 验证了自适应机制在跨模态检索中的有效性。

4.2.2 推荐算法结果

以用户偏好描述生成 Z_{user} ，在全部景点融合向量集合上按余弦相似度排序输出 Top-K。为便于解释结果，结果同时给出门控权重 α 以指示文本/图像在该样本下的相对贡献。其中， α 越接近 1 表示推荐结果更依赖文本模态信息，越接近 0 表示更依赖图像模态，下列推荐结果示例同时给出 score 与 α 以便直观对照。

用户偏好描述 1：“江南水乡、夜景、安静、人少”——案例展示。Top-10 中排序前两项为：拙政园 (score=0.981, $\alpha=0.969$)、周庄 (score=0.979, $\alpha=0.972$)。两者 α 均接近 1，表明该查询下文本与图像语义高度一致，融合更偏向文本信号；推荐结果与 Text-only 上限接近。

用户偏好描述 2：“江南水乡、古镇、安静悠闲”——有/无 MMR 对照。无 MMR 的 Top-5 结果为：周庄、拙政园、西湖、乌镇、普陀山 ($\alpha \approx 0.96-0.98$)。结果呈“风格相近聚集”的特征。MMR ($\lambda=0.7$, Top-5) 的 Top-5 结果为：周庄、西湖、灵隐寺、拙政园、瘦西湖。在保持相关性的同时，列表在城市/类型上更分散，推荐结果的多样性提升。

结果显示，在用户偏好描述驱动的推荐场景中，门控融合在模态一致性较高的查询下与文本上限表现接近；结合 MMR 可在不显著影响相关性的前提下提升列表多样性。

4.3 参数敏感性与消融实验

4.3.1 参数敏感性分析

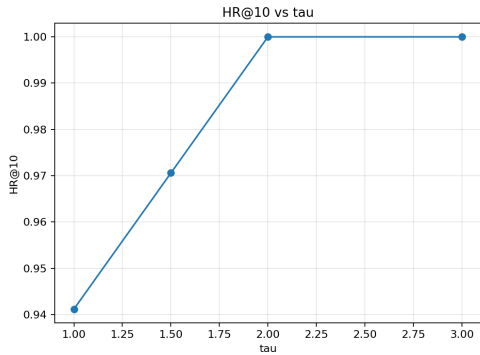


图 5 HR@10 随温度参数 τ 的变化曲线

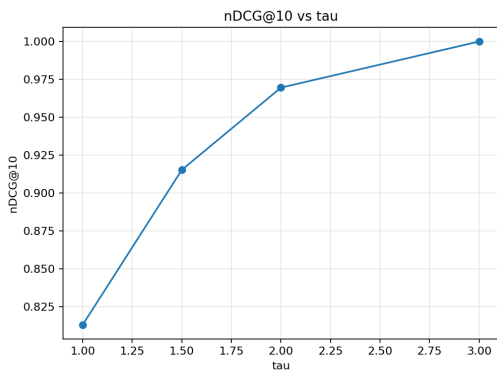


图 6 NDCG@10 随温度参数 τ 的变化曲线

在保持模型与候选集合不变的前提下，仅在推理阶段对 $\tau \in 1.0, 1.5, 2.0, 3.0, 5.0$ 做网格扫描并计算 α 与融合向量 z 。主指标曲线如图 5 与图 6 所示，性能随温度参数 τ 从 1.0 \rightarrow 2.0 持续提升；当 $\tau \geq 3.0$ 后曲线趋于平缓，部分指标开始逼近 Text-only 的上限。该趋势表明，适宜的门控敏感度有助于按样本一致性自适应分配权重，而过大的敏感度会使门控过于极端，削弱多模态融合带来的收益。据此，本文默认选择 $\tau=2.0$ 。

4.3.2 消融实验

本节在 K=10 截断下给出完整模型 (Full, MAG $\tau=2.0$) 与关键对照固定融合 (Fixed- $\alpha=0.5$) 的主结果及相对提升。

表 2 MAG 与固定加权融合的性能对比

方法	HR@10	Δ HR@10	NDCG@10	Δ NDCG@10
MAG	1.000	-	0.970	-
Fixed- α	0.823	+0.177	0.537	+0.433

由表 2 可见，与固定权重融合相比，MAG ($\tau=2.0$) 在 HR@10 上提升 0.177，在 NDCG@10 上提升 0.433，说明样本级自适应门控能显著改善检索命中率与前列排序质量，避免固定融合对强模态信号的稀释。

4.4 定性分析

为补充数值结果的解释性，本节从门控权重分布与代表性样本两方面给出直观分析。

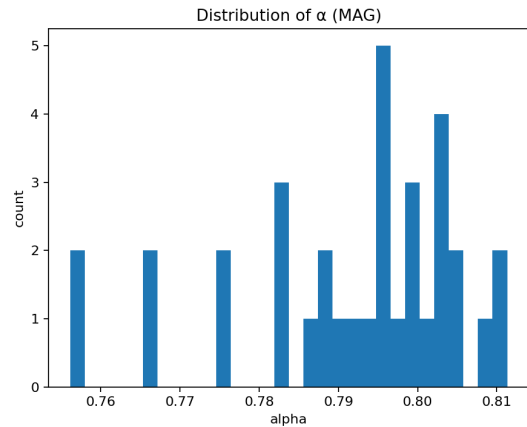


图 7 门控权重 α 的分布

如图 7 所示，在主设定 $\tau=2.0$ 下，门控权重 α 的直方图覆盖低—中—高区间，未出现集中于 0.5 的“平均化”峰值，说明模型能够依据样本的文本—图像一致性进行样本级自适应分配。由于样本量有限且分箱较粗，局部区间出现计数略集中的现象，属统计离散造成的可视化差异，不影响整体分布趋势与结论。当景点配图呈现明显的视觉线索（如夜景、地标外立面等）时，门控权重 α 倾向较低，融合表示更依赖图像模态，相较固定加权可获得更高的前列排序；当评论语义清晰、主题词密集（如历史园林、博物馆类）