大模型驱动的多模态信息生成与信息推荐

吴晔[1,2],陆俊霖[1]

(1.北京师范大学 新闻传播学院,北京 100875;2.北京师范大学 计算传播学研究中心,广东 珠海 519000)

摘 要:随着人工智能技术的快速发展,大语言模型在多模态信息生成和推荐系统中扮演核心角色.介绍了大模型如何通过跨模态学习,实现文本、图像、音频和视频数据的融合,推动信息生成的自动化和多样化,显著提升内容生成质量.在推荐系统中,大模型通过嵌入匹配、token 表示和直接作为推荐引擎,提升了个性化推荐的精准度和多样性.未来的研究可以聚焦于提升多模态模型的推理能力和生成质量,同时加强数据安全和透明性,进一步拓展大语言模型在信息生成与推荐中的应用潜力.

关键词:大语言模型;多模态信息;个性化推荐;智能传播

中图分类号:G202 **文献标志码**:A **文章编号**:1000-2367(2025)05-0145-07

人工智能(artificial intelligence,AI)技术的迅猛发展正在深刻重构传播学的理论体系与实践范式,其中大语言模型(large language model,以下简称“大模型”)在自然语言理解与生成领域的突破性进展,已成为驱动信息传播变革的核心技术.大模型的技术突破主要依托于多项关键技术创新,其中 transformer 架构作为其核心技术框架[1],相较于传统的循环神经网络(recurrent neural network)具有显著优势.该架构不仅实现了序列数据的并行化处理,还通过自注意力机制(self-attention mechanism)有效捕捉文本中的长距离依赖关系,从而显著提升模型在复杂语言模式理解与生成任务中的表现,特别是在长文本处理和多层次语义解析方面,其连贯性与准确性均得到显著提升[2].此外,自监督学习(self-supervised learning)范式的引入进一步增强了模型的泛化能力,使其能够从海量未标注数据中自主学习语言特征,从而更好地适应传播领域中日益增长的多样化与个性化需求[3].同时,大语言模型通过预训练(pre-training)与微调(fine-tuning)相结合的训练策略,不仅能够从大规模文本语料中学习通用语言表征,还能快速适配特定任务场景下的信息处理需求,展现出强大的领域适应性[4].

2023 年全球范围内共发布了 149 个通用大模型,其中美国以 61 个模型的发布量位居全球首位,中国则以 15 个模型的发布量紧随其后[5].从技术发展路径来看,美国依托其开放的技术生态系统、强大的算力基础设施以及丰富的数据资源,在全球大模型领域保持着显著的技术优势与广泛影响力;而中国则通过国家战略层面的政策引导与资源投入,重点推进大模型在垂直领域的深度应用与产业化落地,加速了该技术在信息传播领域的深度融合与创新实践.尽管两国在大模型技术的发展路径上呈现出不同的战略取向与实践模式,但均充分彰显了大模型技术在信息生成、传播与交互方面的巨大潜力与应用价值.

1 大模型驱动的多模态信息生成

多模态大语言模型(multimodal large language model)是一种深度学习模型,能够处理并生成多种数据

收稿日期:2024-12-11;**修回日期**:2025-02-27.

基金项目:国家自然科学基金(11875005);北京社会科学基金重点项目(21DTR040).

作者简介(通信作者):吴晔(1982—),男,福建莆田人,北京师范大学教授,博士,博士生导师,研究方向为计算传播学,E-mail:wuye@bnu.edu.cn.

引用本文:吴晔,陆俊霖.大模型驱动的多模态信息生成与信息推荐[J].河南师范大学学报(自然科学版),2025,53(5):145-151.(Wu Ye,Lu Junlin.Multimodal information generation and recommendation system driven by large language models[J].Journal of Henan Normal University(Natural Science Edition),2025,53(5):145-151.DOI:10.16366/j.cnki.1000-2367.2024.12.11.0002.)

模态,如文本、图像、音频和视频[6].通过跨模态学习和大规模数据训练,这类模型实现了多模态数据的联合建模与交互[7].多模态内容生成是利用人工智能模型,将不同模态的数据进行融合,通过跨模态的理解与生成,创造出更丰富的内容[6].这是大模型在自动化内容生成方面的一项巨大突破.

1.1 多模态大模型的典型架构

多模态大语言模型的典型架构通常由编码器(encoder)、连接器(connector)、大语言模型以及生成器(generator)4 部分构成,能够实现多模态内容的生成与输出(图 1)[8].其中,模态编码器是多模态处理的基础模块,负责将原始模态数据(如图像、音频、视频等)编码为高维特征表示.为实现跨模态语义对齐,模态编码器通常采用预训练模型,例如 CLIP 模型中的视觉编码器通过大规模图像-文本对数据的预训练,实现了视觉特征与文本语义的高效对齐[9-10].连接器作为多模态特征转换的关键模块,主要负责将模态编码器生成的特征映射为大语言模型可理解的表示形式.根据特征融合方式的不同,连接器可分为 3 类:1)基于投影的连接器(projection-based connector),其通过线性或非线性投影将多模态特征转换为与文本嵌入空间相兼容的表示[11].2)基于查询的连接器(query-based connector),通过引入可学习的查询向量(如 Q-Former)从多模态特征中提取关键信息[12].这 2 类连接器主要在 token 级别进行特征融合,将多模态特征处理为与文本 token 类似的表示形式,以便与大语言模型的文本输入兼容[11-12].3)是基于融合的连接器(fusion-based connector),通过在大语言模型的 transformer 层中插入跨模态注意力机制来实现多模态特征与文本特征的深度融合[13].大语言模型作为多模态系统的核心组件,承担着逻辑推理与语义理解的关键任务,负责处理经过连接器转换的多模态信息[14].在大语言模型的架构中,多头注意力机制作为 transformer 模型的核心组件,能够通过接受来自不同模态的查询(query)、键(key)和值(value)向量,实现多模态信息的深度整合与交互[15-16].生成器作为输出模块,能够将大语言模型处理后的多模态信息转化为文本、图像、音频和视频等形式的输出内容.目前,序列生成模型(sequence generation model)和扩散模型(diffusion model)是 2 类最常用的生成方法[17-18].在序列生成模型中,OpenAI 发布的 DALL-E 模型是一个典型代表.该模型基于 4 亿对图文数据训练集,采用 VQ-VAE 图像离散自编码器与 GPT 相结合的架构,在文本生成图像任务上实现了高质量的生成效果与强大的泛化能力,因此被誉为"图像版 GPT"[19].在扩散模型领域,stable diffusion 作为开源模型的代表,将潜在扩散模型成功拓展至开放领域的文本到图像生成任务[20].此外,在闭源扩散模型中,OpenAI 的 DALL-E 2 和谷歌的 Imagen 也展现了卓越的性能[21].

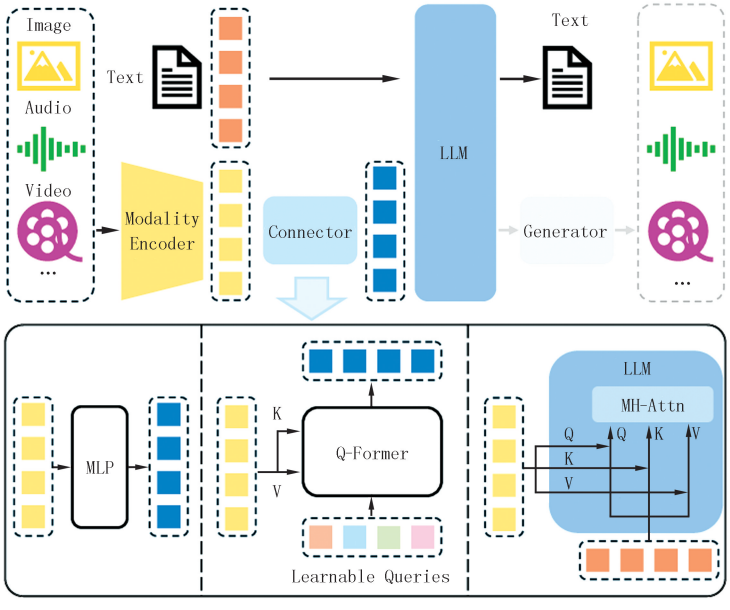

图1 多模态大模型的典型架构[8]

Fig. 1 Typical MLLM architecture[8]

1.2 多模态大模型的训练过程

在多模态大模型的训练过程中,作为核心组件的大语言模型通常采用参数冻结策略,其参数更新率通常控制在 0.1%以下[22],以显著降低训练成本并提高训练效率.这种参数高效微调(parameter-efficient fine-tuning)方法使得多模态大模型能够在保持较低计算资源消耗的同时,实现对多模态任务的强大支持.多模态大模型的训练流程主要分为 2 个关键阶段:多模态预训练和多模态指令微调[23].

在多模态预训练阶段,模型需要利用大规模的图像-文本对数据进行训练,以实现跨模态语义对齐,并将视觉信号转化为大语言模型可理解的特征表示或 token 序列[24].常用的训练数据集包括 X-text 数据集[25],其中涵盖多种模态组合,如图像-文本、视频-文本、语音-文本等.图像-文本数据通常呈现 2 种形式:单一图像-文本对(<img1><txt1>)和交替图像-文本序列(<img1><txt1><txt2><img3><txt3>).对于多模态理解模型,训练目标主要集中于文本生成损失函数的优化[26],可表示为:$\mathcal{L}_{\text{text}}=-\sum_{t=1}^{T}\ln P(w_t \mid w_{<t}, v)$. 对于多模态生成模型,则需要同时优化文本生成损失、模态生成损失和输出对齐损失[27],其总损失函数可表示为:$\mathcal{L}_{\text{tatal}}=\mathcal{L}_{\text{text}}+\lambda_1\mathcal{L}_{\text{modal}}+\lambda_2\mathcal{L}_{\text{align}}$,其中 $\mathcal{L}_{\text{align}}$ 采用 CLIP-style 对比损失[28].在多模态指令微调阶段,模型通过指令格式化的数据集对预训练的多模态大模型进行进一步优化,以提升其遵循复杂指令的能力.指令微调主要包括监督微调(supervised fine-tuning)和基于人类反馈的强化学习(reinforcement learning from human feedback)[29].监督微调使用格式化为<instruction,input,output>的 3 元组数据集,通过最大似然估计优化响应生成质量[30]:$\mathcal{L}_{\text{SFT}}=-E_{(x,y)\sim D_{\text{SFT}}}\sum_{t=1}^{T}\ln P(y_t \mid y_{<t}, X)$. 人类反馈学习采用近端策略优化(PPO)算法[31],通过奖励模型 $R_{\varnothing}(v,t)$ 引导生成策略:$\pi_\theta:\max_\theta E_{(v,t)\sim D,y\sim\pi_\theta}[R_{\varnothing}(v,t,y)-\beta KL(\pi_\theta \| \pi_{\text{ref}})]$. 随着多模态数据越来越多地融入大模型中,视觉语言指令调优(visual language instruction tuning)受到了越来越多的关注,与纯文本指令调优相比,它呈现出更复杂的特征.在训练过程中,使用多模态数据进行联合训练,常见的损失函数包括分类损失、回归损失和对比学习损失等[32].

1.3　多模态大模型的发展方向

尽管当前的多模态大语言模型在视觉信息推理任务中已取得显著进展,但在处理复杂多模态应用场景时,其性能仍存在明显不足[33].为提升多模态模型对复杂问题的推理能力,研究者提出了构建更广泛且更具挑战性的视觉指令集的策略,通过增加任务的多样性和复杂性来增强模型的视觉推理性能[29].然而,这一领域更核心的挑战在于多模态大模型的构建方法与学习机制的优化[12].此外,随着伪造内容形式的日益多样化,多模态融合检测技术逐渐成为研究热点.其中,跨模态一致性分析是一种典型的研究方向,该方法通过检测不同模态之间的逻辑冲突和特征不一致性来识别伪造内容.例如,在视频伪造检测场景中,研究者通过分析图像帧与音频信号的时间同步性和语义一致性来识别潜在的篡改痕迹[34].然而,视频、音频和文本的伪造特征往往相互交织,如何有效解构和分析这种复杂的多模态数据成为当前研究的重点和难点.

2　大模型驱动的信息推荐

推荐系统的核心任务在于精准捕捉并深入理解用户的潜在偏好,从而为其推送个性化的信息资源[35].当前,大多数推荐系统的研究主要依赖于用户的显式或隐式交互行为日志(如商品点击、购买记录、评分数据及评论内容)来训练推荐模型,其中深度学习方法因其强大的特征提取和非线性建模能力而成为主流技术范式[36].然而,推荐系统在实际应用中仍面临诸多挑战.首先,冷启动问题(cold start)是一个长期存在的难题.新用户或新物品缺乏足够的交互数据,系统难以准确推断其偏好,从而导致推荐效果显著下降[37].其次,用户偏好往往具有跨领域特性,例如用户在电商平台上的购物偏好可能与其在视频平台上的观看偏好存在潜在关联.这种跨领域推荐不仅增加了模型的复杂性,也对数据的整合与迁移提出了更高要求[38].此外,推荐的动态性和时效性也是关键挑战,用户偏好可能随时间、情境或外部因素的变化而发生显著改变,这要求推荐系统具备实时学习和快速适应的能力[39].

2.1　3 个代表性建模范式和 2 种分类

基于大语言模型的推荐系统为解决上述挑战提供了新的技术路径.这类系统通常由以下 3 种方式构建(图 2)[40].第 1 种方法是使用嵌入进行推荐(LLM Embeddings+RS).在这种模式下,用户特征(如用户 ID、人口统计学信息、历史行为偏好等)和物品特征(如类别标签、评分统计、文本描述等)通过大语言模型编码为低维稠密向量(embedding),随后输入到传统推荐系统模型中进行匹配[41].推荐系统通过向量匹配计算用户嵌入与物品嵌入之间的相似度,从而评估推荐的相关性[42].常用的相似度度量方法包括余弦相似度(cosine similarity)和内积(dot product)等[43].这种方法的优势在于其高效性,因为嵌入向量是固定长度的,能够充分利用传统推荐系统的高效计算框架.然而,其局限性在于可能丢失部分细粒度的语义信息.第 2 种方法是

使用 token 表示进行推荐(LLM Tokens + RS).大语言模型将用户描述和物品描述转换为 token 序列,而非直接生成嵌入向量.推荐系统通过分析这些 token 序列来实现个性化匹配,从而保留更丰富的语义信息[44].这种方法允许推荐系统直接对文本进行自然语言处理或其他文本分析操作,特别适用于需要深度理解文本语义的场景.然而,由于 token 序列可能具有不定长度,其计算开销通常较大,且对模型的序列处理能力提出了更高要求[33].在这种范式中,大语言模型直接作为推荐系统的核心组件,根据任务指令、用户信息和物品描述进行推理,并生成完整的推荐响应[35].根据模型参数是否需要更新,可分为基于特定提示的方法(prompt-based method)和基于指令微调的方法(instruction fine-tuning method)[45].基于特定提示的方法通过设计一系列自然语言提示来引导大语言模型完成推荐任务[46].例如,将推荐任务转化为"根据用户 X 的历史行为,推荐可能感兴趣的商品"的形式.基于指令微调的方法则通过微调大语言模型使其适配推荐任务.其核心在于构建适合推荐任务的指令数据集,这些指令通常基于用户与物品的交互数据以及定制化的提示模板构建,为模型提供明确的任务指导[47].例如,指令数据可能包括"用户 A 在过去一周内购买了商品 B 和 C,请推荐相关商品"等形式.

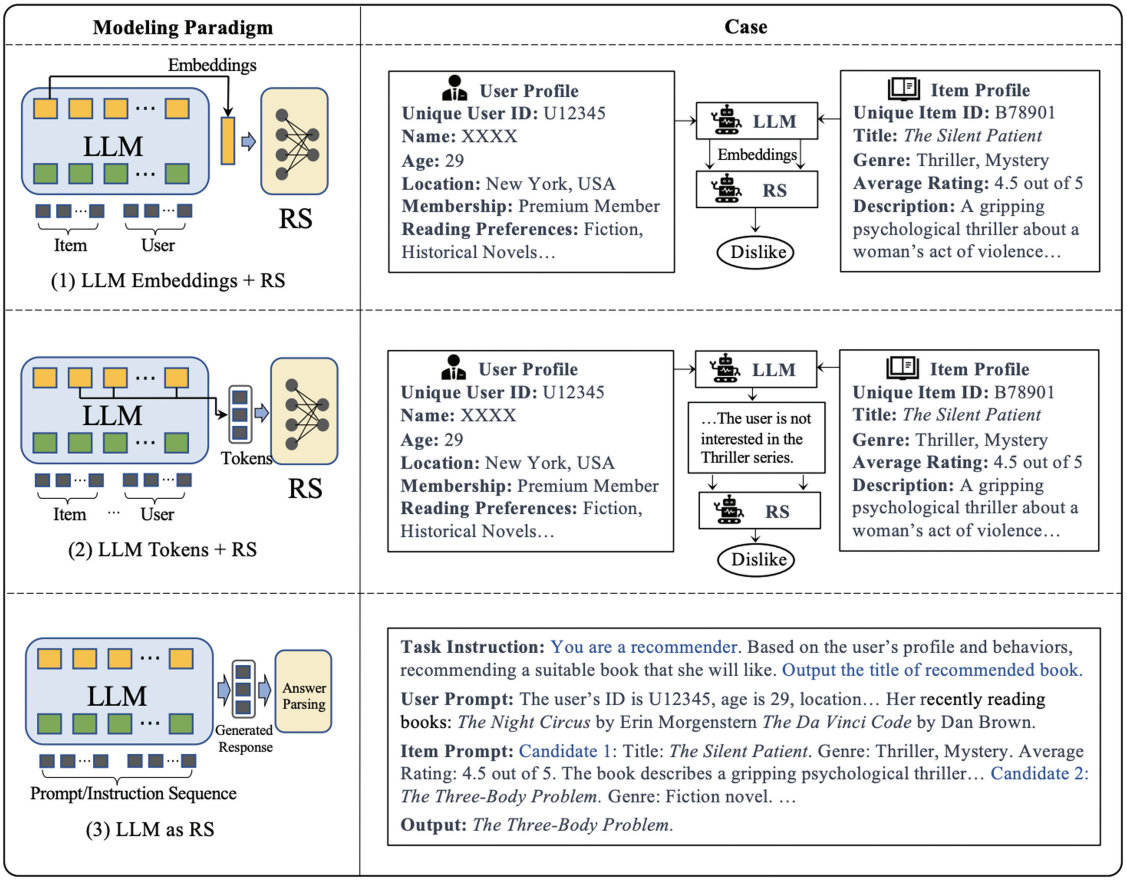

图2 推荐系统中大型语言模型研究的3个代表性建模范式[41]

Fig. 2 Three representative modeling paradigms of the research for large language models on recommendation systems[41]

推荐系统可以分为生成式推荐(generative recommendation)和判别式推荐(discriminative recommendation)[48].生成式推荐通常采用生成模型,如变分自编码器或生成对抗网络,其核心目标是生成符合用户潜在兴趣的推荐内容[49].在训练过程中,生成式推荐模型通过最大化生成数据的对数似然来优化模型参数[50]:$\mathscr{L}=\sum_{u\in U}\ln(\sum_{i\in I}P(i\mid U))$.判别式推荐使用判别模型(如逻辑回归、支持向量机等)[51],通过模型对用户 u- 项目 i 对进行评分,通常用一个评分函数 $f(x_{u,i})$ 来预测用户对项目的兴趣.判别式推荐直接学习每个用户对项目的兴趣匹配度.训练过程中,目标是最小化损失函数(如交叉熵损失或均方误差损失)[52]:$\mathscr{L}=\sum_{(u,i)\in D}\mathscr{L}(r(u,i),y_{u,i})$.

2.2　大模型驱动推荐系统的发展方向

随着人工智能技术的不断演进，智能推荐系统从传统的文本交互模式提升至多模态数据交互的新维度.以教育场景为例，对话式推荐机制能够为学生精准匹配最优学习资源与课程体系，同时借助生成式技术实现学习内容的个性化定制[53].在医疗健康领域，这类技术不仅能够为患者提供精准的健康资讯与专业建议，更能生成定制化的健康评估报告与诊疗方案[54].随着技术迭代升级，智能推荐系统将在更广泛的领域展现其应用价值.

从技术演进的历史维度来看，人类始终将开发具有自主决策能力的智能体(AI agent)作为重要研究方向，以期通过智能化手段解决各类复杂任务.智能推荐系统依托深度学习技术，多维度分析用户行为特征，整合历史行为数据、实时状态信息及环境变量，构建精准的用户意图识别模型.在商业消费领域，系统不仅基于用户历史消费记录进行推荐，更能综合市场动态、品牌策略及供应链信息，为用户提供最优购物决策建议.在数字娱乐领域，通过情感计算与社交数据分析，智能体能够精准识别用户情绪状态，深度理解用户兴趣偏好演变，从而提供更具情感共鸣的音乐、影视及阅读推荐[55].

3　总结和展望

大语言模型正在成为信息传播和推荐系统领域的核心力量，凭借其在文本理解、跨模态数据处理和复杂推理能力上的优势，为信息生成与分发注入了新的活力.尽管目前的多模态大语言模型已初步具备生成和推理能力，但在复杂应用场景中仍存在诸多挑战.未来的研究应着重探索更广泛且复杂的视觉指令集和高效的多模态融合方法，以进一步提升模型的推理深度与生成质量.此外，大模型的推荐系统在数据隐私保护和可解释性方面仍需加强.随着大语言模型技术的不断进步和多模态数据处理能力的提升，其在信息传播和推荐系统中的应用前景将更加广阔.

参　考　文　献

[1] CHANG Y P, WANG X, WANG J D, et al. A survey on evaluation of large language models[J]. ACM Transactions on Intelligent Systems and Technology, 2024, 15(3): 1-45.

[2] ISLAM S, ELMEKKI H, ELSEBAI A, et al. A comprehensive survey on applications of transformers for deep learning tasks[J]. Expert Systems with Applications, 2024, 241: 122666.

[3] MOHAMED A, LEE H Y, BORGHOLT L, et al. Self-supervised speech representation learning: a review[J]. IEEE Journal of Selected Topics in Signal Processing, 2022, 16(6): 1179-1210.

[4] DING N, QIN Y J, YANG G, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models[J]. Nature Machine Intelligence, 2023, 5(3): 220-235.

[5] STANFORD UNIVERSITY. 2024 AI Index report[EB/OL]. [2024-11-20]. https://aiindex.stanford.edu.

[6] WU J Y, GAN W S, CHEN Z F, et al. Multimodal large language models: a survey[C]//2023 IEEE International Conference on Big Data (BigData). December 15-18, 2023. Sorrento: IEEE, 2023: 2247-2256.

[7] WANG Y. Survey on deep multi-modal data analytics: collaboration, rivalry, and fusion[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2021, 17(1s): 1-25.

[8] YIN S K, FU C Y, ZHAO S R, et al. A survey on multimodal large language models[EB/OL]. [2024-11-20]. https://arxiv.org/abs/2306.13549v4.

[9] ZHANG C, YANG Z C, HE X D, et al. Multimodal intelligence: representation learning, information fusion, and applications[J]. IEEE Journal of Selected Topics in Signal Processing, 2020, 14(3): 478-493.

[10] XUE H W, SUN Y C, LIU B, et al. CLIP-ViP: adapting pre-trained image-text model to video-language representation alignment[EB/OL]. [2024-11-20]. https://arxiv.org/abs/2209.06430v4.

[11] ANDO A, GIDARIS S, BURSUC A, et al. RangeViT: towards vision transformers for 3D semantic segmentation in autonomous driving[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). June 17-24, 2023. [S.l.]: IEEE, 2023: 5240-5250.

[12] LIANG P P, ZADEH A, MORENCY L P. Foundations & trends in multimodal machine learning: principles, challenges, and open questions[J]. ACM Computing Surveys, 2024, 56(10): 1-42.

[13] XU P, ZHU X T, CLIFTON D A. Multimodal learning with transformers: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(10): 12113-12132.

[14] HOLZINGER A,MALLE B,SARANTI A,et al.Towards multi-modal causability with Graph Neural Networks enabling information fusion for explainable AI[J].Information Fusion,2021,71:28-37.

[15] WU T,PENG J J,ZHANG W Q,et al.Video sentiment analysis with bimodal information-augmented multi-head attention[J].Knowledge-Based Systems,2022,235:107676.

[16] CHEN X,ZHANG N Y,LI L,et al.Hybrid transformer with multi-level fusion for multimodal knowledge graph completion[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval.Madrid Spain:ACM,2022:904-915.

[17] GUIMARAES G L,SANCHEZ-LENGELING B,OUTEIRAL C,et al.Objective-reinforced generative adversarial networks(ORGAN)for sequence generation models[EB/OL].[2024-11-20].https://arxiv.org/abs/1705.10843v3.

[18] YANG L,ZHANG Z L,SONG Y,et al.Diffusion models:a comprehensive survey of methods and applications[J].ACM Computing Surveys,2024,56(4):1-39.

[19] KONDRATYUK D,YU L J,GU X Y,et al.VideoPoet:a large language model for zero-shot video generation[EB/OL].[2024-11-20].https://arxiv.org/abs/2312.14125v4.

[20] HUANG Y,HUANG J C,LIU Y F,et al.Diffusion model-based image editing:a survey[EB/OL].[2024-11-20].https://arxiv.org/abs/2402.17525v4.

[21] RAHMAN A,AL-MAMOON F,SAQIB M N,et al.Implementation of diffusion model in realistic face generation[D].Dhaka:Brac University,2024.

[22] BEN ZAKEN E,RAVFOGEL S,GOLDBERG Y.BitFit:simple parameter-efficient fine-tuning for transformer-based masked language-models[EB/OL].[2024-11-20].https://arxiv.org/abs/2106.10199v5.

[23] HUANG J X,ZHANG J Y,JIANG K,et al.Visual instruction tuning towards general-purpose multimodal model:a survey[EB/OL].[2024-11-20].https://arxiv.org/abs/2312.16602v1.

[24] HAN X,WANG Y T,FENG J L,et al.A survey of transformer-based multimodal pre-trained modals[J].Neurocomputing,2023,515:89-106.

[25] RASHNO E,ESKANDARI A,ANAND A,et al.Survey:transformer-based models in data modality conversion[EB/OL].[2024-11-20].https://arxiv.org/abs/2408.04723v1.

[26] LUO H S,JI L,SHI B T,et al.UniVL:a unified video and language pre-training model for multimodal understanding and generation[EB/OL].[2024-11-20].https://arxiv.org/abs/2002.06353v3.

[27] LI J Y,TANG T Y,ZHAO W X,et al.Pre-trained language models for text generation:a survey[J].ACM Computing Surveys,2024,56(9):1-39.

[28] RADFORD A,KIM J W,HALLACY C,et al.Learning transferable visual models from natural language supervision[EB/OL].[2024-11-20].https://arxiv.org/abs/2103.00020.

[29] OUYANG L,WU J,JIANG X,et al.Training language models to follow instructions with human feedback[J].Advances in Neural Information Processing Systems,2022,35:27730-27744.

[30] JIN Y Z,LI J,LIU Y X,et al.Efficient multimodal large language models:a survey[EB/OL].[2024-11-20].https://arxiv.org/abs/2405.10739v2.

[31] SCHULMAN J,WOLSKI F,DHARIWAL P,et al.Proximal policy optimization algorithms[EB/OL].[2024-11-20].https://arxiv.org/abs/1707.06347v2.

[32] RAHATE A,WALAMBE R,RAMANNA S,et al.Multimodal co-learning:Challenges,applications with datasets,recent advances and future directions[J].Information Fusion,2022,81:203-239.

[33] UPPAL S,BHAGAT S,HAZARIKA D,et al.Multimodal research in vision and language:a review of current and emerging trends[J].Information Fusion,2022,77:149-171.

[34] FENG C,CHEN Z Y,OWENS A.Self-supervised video forensics by audio-visual anomaly detection[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).[S.l.]:IEEE,2023:10491-10503.

[35] PU P,CHEN L,HU R.A user-centric evaluation framework for recommender systems[C]//Proceedings of the Fifth ACM Conference on Recommender Systems.Chicago Illinois USA:ACM,2011:157-164.

[36] ZHANG S,YAO L N,SUN A X,et al.Deep learning based recommender system[J].ACM Computing Surveys,2020,52(1):1-38.

[37] GOPE J,JAIN S K.A survey on solving cold start problem in recommender systems[C]//2017 International Conference on Computing,Communication and Automation(ICCCA).[S.l.]:IEEE,2017:133-138.

[38] ZANG T Z,ZHU Y M,LIU H B,et al.A survey on cross-domain recommendation:taxonomies,methods,and future directions[J].ACM Transactions on Information Systems,2023,41(2):1-39.

[39] RAZA S,DING C.News recommender system:a review of recent progress,challenges,and opportunities[J].Artificial Intelligence Review,2022,55(1):749-800.

[40] WU L K,ZHENG Z,QIU Z P,et al.A survey on large language models for recommendation[J].World Wide Web,2024,27(5):60.

[41] ELKAHKY A M，SONG Y，HE X D. A multi-view deep learning approach for cross domain user modeling in recommendation systems [C]//Proceedings of the 24th International Conference on World Wide Web. Florence Italy. International World Wide Web Conferences Steering Committee，2015：278-288.

[42] ISINKAYE F O，FOLAJIMI Y O，OJOKOH B A. Recommendation systems：principles，methods and evaluation[J]. Egyptian Informatics Journal，2015，16(3)：261-273.

[43] SONG Y Q，ROTH D. Unsupervised sparse vector densification for short text similarity[C]//Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics：Human Language Technologies. Denver，Colorado. Stroudsburg，PA，USA：ACL，2015：1275-1280.

[44] ZHAO Z H，FAN W Q，LI J T，et al. Recommender systems in the era of large language models(LLMs)[EB/OL].[2024-11-20]. https://arxiv.org/abs/2307.02046v6.

[45] LIU P F，YUAN W Z，FU J L，et al. Pre-train，prompt，and predict：a systematic survey of prompting methods in natural language processing[J]. ACM Computing Surveys，2023，55(9)：1-35.

[46] LIU P，ZHANG L M，GULLA J A. Pre-train，prompt，and recommendation：a comprehensive survey of language modeling paradigm adaptations in recommender systems[J]. Transactions of the Association for Computational Linguistics，2023，11：1553-1571.

[47] ZHANG J J，XIE R B，HOU Y P，et al. Recommendation as instruction following：a large language model empowered recommendation approach[EB/OL].[2024-11-20]. https://arxiv.org/abs/2305.07001v1.

[48] DELDJOO Y，DI NOIA T，MERRA F A. A survey on adversarial recommender systems[J]. ACM Computing Surveys，2022，54(2)：1-38.

[49] DELDJOO Y，HE Z K，MCAULEY J，et al. Recommendation with generative models[EB/OL].[2024-11-20]. https://arxiv.org/abs/2409.15173v1.

[50] THEIS L，VAN DEN OORD A，BETHGE M. A note on the evaluation of generative models[EB/OL].[2024-11-20]. https://arxiv.org/abs/1511.01844v3.

[51] MAROCO J，SILVA D，RODRIGUES A，et al. Data mining methods in the prediction of dementia：a real-data comparison of the accuracy，sensitivity and specificity of linear discriminant analysis，logistic regression，neural networks，support vector machines，classification trees and random forests[J]. BMC Research Notes，2011，4：299.

[52] CHENG G，YANG C Y，YAO X W，et al. When deep learning meets metric learning：remote sensing image scene classification via learning discriminative CNNs[J]. IEEE Transactions on Geoscience and Remote Sensing，2018，56(5)：2811-2821.

[53] WU S Y，CAO Y，CUI J J，et al. A comprehensive exploration of personalized learning in smart education：from student modeling to personalized recommendations[EB/OL].[2024-11-20]. https://arxiv.org/abs/2402.01666v1.

[54] LI Y H，LI Y L，WEI M Y，et al. Innovation and challenges of artificial intelligence technology in personalized healthcare[J]. Scientific Reports，2024，14(1)：18994.

[55] PORIA S，CAMBRIA E，BAJPAI R，et al. A review of affective computing：from unimodal analysis to multimodal fusion[J]. Information Fusion，2017，37：98-125.

Multimodal information generation and recommendation system driven by large language models

Wu Ye[1,2]，Lu Junlin[1]

(1. School of Journalism and Communication，Beijing Normal University，Beijing 100875，China；
2. Computational Communication Research Center，Beijing Normal University，Zhuhai 519000，China)

Abstract：The rapid development of artificial intelligence technology has enabled large language models(LLMs) to play a significant role in multimodal information generation and recommendation systems. This paper introduces how LLMs achieve cross-modal learning，integrating text，image，audio，and video data to drive automation and diversification in information generation，greatly enhancing content quality. In recommendation systems，LLMs improve the accuracy and diversity of personalized recommendations through embedding matching，token representation，and functioning directly as recommendation engines. Future research should focus on enhancing the reasoning ability and generating quality of multimodal models，strengthening data security and transparency，and expanding the application potential of LLMs in information generation and recommendation.

Keywords：large language models；multimodal information；personalized recommendation；intelligent communication

[责任编校　杨浦　刘洋]