传统模式与大语言模型下推荐系统的比较研究

黄 文[1,2],李振江[3]

(1. 中国科学院 文献情报中心;2. 中国科学院大学 经济与管理学院信息资源管理系,北京 100190;
3. 东北林业大学 计算机与控制工程学院,黑龙江 哈尔滨 150040)

摘 要: 面对信息爆炸与复杂用户的需求,推荐系统的优化至关重要,大语言模型提供了新机遇。为此,通过比较传统模式与大语言模型的推荐系统,明确各自优势与局限性,为未来研发策略提供有力的参考。首先,梳理了传统推荐系统的发展脉络,包括基于协同过滤、基于内容和基于知识图谱的方法。其次,重点探讨了以BERT与GPT模型为代表的大语言模型在推荐系统中的实际应用,尤其是在不同调整方式下的性能表现。实验表明,两种模式的推荐系统在性能、用户体验、系统复杂性及资源消耗等方面存在明显差异,传统模式推荐方法更适用于规则性较强、数据相对稀疏且稳定的场景,大语言模型推荐方法更适合需要理解复杂语义、提供创新性解决方案和动态内容生成的场景。

关键词: 推荐系统;传统模型;大语言模型;对比分析

DOI: 10. 11907/rjdk. 232280 **开放科学(资源服务)标识码(OSID):**

中图分类号: TP18 **文文献标识码:** A **文章编号:** 1672-7800(2025)002-0204-07

Recommendation Systems: A Comparative Study of Traditional Models and Pre-Trained Model Approaches

HUANG Wen[1,2], LI Zhenjiang[3]

(1. *National Science Library*, *Chinese Academy of Sciences*; 2. *Department of Information Resources Management*, *School of Economics and Management*, *University of Chinese Academy of Sciences*, *Beijing* 100190, *China*;
3. *School of Computer and Control Engineering*, *Northeast Forestry University*, *Harbin* 150040, *China*)

Abstract: In the face of information explosion and intricate user demands, the optimization of recommendation systems is crucial. Large language models present a new opportunity in this regard. This paper contrasts traditional methods with recommendation systems under large language models, aiming to clarify their advantages and limitations, and offer robust references for future R&D strategies. This study initially reviews the evolution of traditional recommendation systems, including methods based on collaborative filtering, content, and knowledge graphs. Subsequently, it delves into the practical applications of large language models, such as BERT and the GPT series, especially their performance under different adjustments. Through comparative analysis, we observed notable differences between the two system paradigms in terms of performance, user experience, system complexity, and resource consumption. The recommendation method in the traditional mode is more suitable for scenarios with strong rules and relatively sparse and stable data, while the recommendation method under the large language model is more suitable for scenarios that need to understand complex semantics, provide innovative solutions, and generate dynamic content.

Key Words: recommendation system; traditional model; large language models; comparative analysis

0 引言

大数据时代下,各种类型、模态、来源的数据急速增长,人们难以有效准确地寻找到所需信息,因此快速、准确地获取目标信息的系统需求愈发强烈。在这种背景下,信息检索与推荐系统作用日渐凸显,前者致力于从众多数据中挑选出与需求最匹配的内容,后者旨在为用户提供个性化建议,帮助其找到心仪的产品或服务[1]。目前,无论是在商品推介、社交媒体内容筛选还是娱乐行业的精准推荐

收稿日期: 2023-12-19
扫描二维码阅读全文:

作者简介: 黄文(2000-),男,中国科学院大学经济与管理学院信息资源管理系硕士研究生,研究方向为大数据情报分析方法与技术;李振江(1999-),男,东北林业大学计算机与控制工程学院硕士研究生,研究方向为RNA甲基化位点预测。本文通讯作者:黄文。

方面,推荐系统已成为了一个关键组件,可通过深入分析用户行为、喜好及历史记录为用户展示出最具相关性的选择,从而提升信息匹配精确度和用户体验[2,3]。

尽管现代推荐系统已使用了协同过滤、基于内容的方法和知识图谱等技术,但依然面临不少挑战。近年来,BERT、GPT 等系列预训练语言模型或生成式大语言模型(大语言模型)在自然语言处理领域的卓越表现也为推荐系统领域的发展带来了新的启示[4-9]。本文根据模式依赖的核心技术模型,将推荐系统分为传统模式和大语言模式,如图 1 所示。

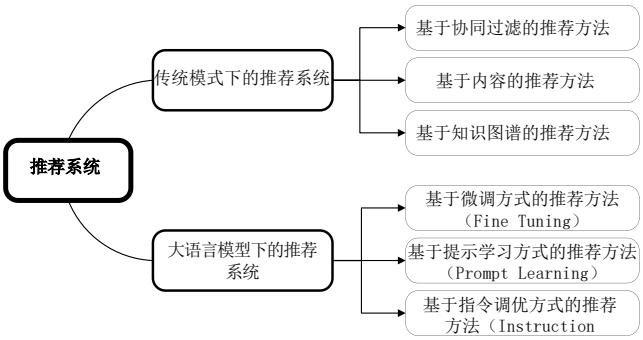

Fig. 1　Classification system of the recommender system

图 1　推荐系统分类体系

1　传统模式下推荐系统的发展

在协助用户探索与其个性化兴趣紧密相连的项目或内容上,推荐系统起到了至关重要的作用,既满足了用户对信息和体验的独特需求,又创造了愉悦且深入的互动体验。传统模式下推荐系统主要包含基于协同过滤的推荐方法、基于内容的推荐方法、基于知识图谱的推荐方法等,专注于利用用户行为和物品特征提供个性化推荐。

1.1　基于协同过滤的推荐方法

基于协同过滤的推荐方法是在传统模式下推荐系统的核心方法[10-12]。以商品推荐场景为例,模型根据用户行为与物品间的相互关系提供个性化推荐,优势在于提供了高度定制化的建议,在充分考虑相关联用户间的行为和偏好后帮助用户探索可能遗漏的物品[13]。例如,通过商品评分和用户购买记录构建关联模型来计算相似性[10]。目前,该方法已适用于 Netflix 和亚马逊等多家知名公司,可实时更新以满足于用户需求,既刺激了销售额,又极大提升了用户满意度[14]。然而,该方法的局限性包含:①在新用户或新物品场景中存在“冷启动”问题[15];②数据稀疏性会导致推荐准确性降低;③过度依赖历史行为导致推荐单一。尽管存在这一系列挑战,但该方法在处理大数据时仍然表现出色,为推荐系统的发展奠定了基础。

1.2　基于内容的推荐方法

基于内容的推荐方法是通过分析用户与物品的内容描述来确定相似性,从内容中挖掘关键特征,并使用余弦相似度等方法寻找用户与物品之间的关联,因此在电影、音乐和新闻等领域得到了广泛应用[16-18]。该方法的优势在于不受历史交互数据限制,能为用户提供清晰的推荐解释,例如 YouTube、Spotify 利用该方法分析视频元数据与音乐特征,成功提供了与用户兴趣相符的内容,但该方法也存在一些挑战,例如内容获取的成本较高、信息过载及内容描述的准确性等问题。

1.3　基于知识图谱的推荐方法

知识图谱(Knowledge Graph, KG)最早由谷歌在 2012 年提出,用于增强其搜索引擎效果,实现搜索的智能化与语义化,因其可利用丰富的语义信息实现个性化推荐而被应用于其他领域,特别是推荐系统[2,19-22]。基于知识图谱的推荐方法主要由知识图谱构建、用户建模和推荐生成 3 个核心流程组成,首先整合多种数据形成一个涵盖各实体及其关系的图谱;其次分析用户行为、资料,并与图谱进行匹配以精确识别用户需求;最后结合多种技术生成与用户兴趣相关的定制推荐方案[2,23,24]。例如,Facebook 使用知识图谱推荐朋友和兴趣群体,Google 的 Knowledge Graph 增强了搜索的语义相关性,其局限性主要在构建和维护知识图谱时需要大量时间和资源,涉及多个步骤,相关技术仍在发展和完善过程中。

2　大语言模型下推荐系统的发展

尽管在传统模式下推荐系统在众多领域都推动了科学研究的前进,但依然面临着一系列挑战,例如交互性差、可解释性不足等,不仅影响了用户体验,还关系到系统透明性和信任度等核心问题,从而制约了这些系统在实际应用中的广泛部署[25]。2022 年 ChatGPT 的推出标志着大语言模型的崛起,展示了其在多种应用任务中的强大潜力,尤其在推荐系统领域产生了深远影响[8,26-29]。

2.1　大语言模型的兴起

大语言模型的兴起在自然语言处理(Natural Language Processing, NLP)领域是一次划时代的突破,展现了神经网络技术在 NLP 任务中的突出进展。在此之前,传统 NLP 方法主要依赖手工设计的任务特征和规则模型,例如词袋模型、TF-IDF 模型及基于正则表达式的模型[30,31]。由于人工设计的任务特征与任务效果具有较强的关系,当面临更多样性的语言和负责任务时,传统 NLP 方法表现效果良莠不齐。为此,以 RNN、LSTM 技术为代表的深度学习模型凭借其捕获长文本的序列信息优势,为 NLP 任务带来了突破性进展,在语言建模、机器翻译和情感分析中都取得了出色表现,但该技术在面对长文本的上下文理解任务时依然存在一些挑战[32-35]。

Transformer 架构的提出和 BERT、GPT-2 等预训练语言模型的问世则解决了上述问题[7,8,36-38]。此类模型融合了自注意力机制和深度神经网络,能处理更长的文本序列和深入理解语言上下文,展现了出卓越的泛化能力[39]。其

中，泛化能力的体现通常需要两个步骤：①在大规模文本数据上经过自我监督的预训练；②经过特定任务的微调实现各种NLP任务。

近年来，ChatGPT等生成式大语言模型的推出进一步解决了预训练语言模型通用性不足的问题，可直接在多种NLP任务上使用，无需额外微调模型。各大公司纷纷发布类似竞品，例如LLaMa、百度文心一言和清华ChatGLM等在多个领域都取得突出成绩[40]。

2.2 大语言模型在推荐系统中的应用

在NLP模型不断更迭的背景下，推荐系统领域不断变革，相关NLP模型凭借其卓越的语言理解和生成能力，逐渐成为推荐系统的核心[41]。大型预训练语言模型（LLMs）适用于推荐系统的重要原因是其在理解复杂用户意图和偏好方面的出色能力，可分析用户搜索历史、购买记录、评价内容等方面捕捉用户的细微兴趣变化，并据此生成个性化推荐。例如，LLMs能从用户评价中提取传统推荐系统难以精确捕捉的情感倾向、特定的喜好和需求，甚至是隐含期望[42]。同时，LLMs还能在推荐过程中考虑语言的多样性和文化差异，为不同地区和语言背景的用户提供更精准的推荐，极大提升了推荐系统的相关性和用户满意度，为用户带来更丰富和贴合个人需求的体验。

本文主要阐述了大语言模型下推荐系统的实现过程，如图2所示。其中，大语言模型能精准解析用户自然语言输入，生成高效的文本表示，并深挖用户与待推荐项间的关系，推荐准确率得到了显著提升。尤其是，此类模型还经过了微调，通过提示学习等策略来优化特定推荐任务。

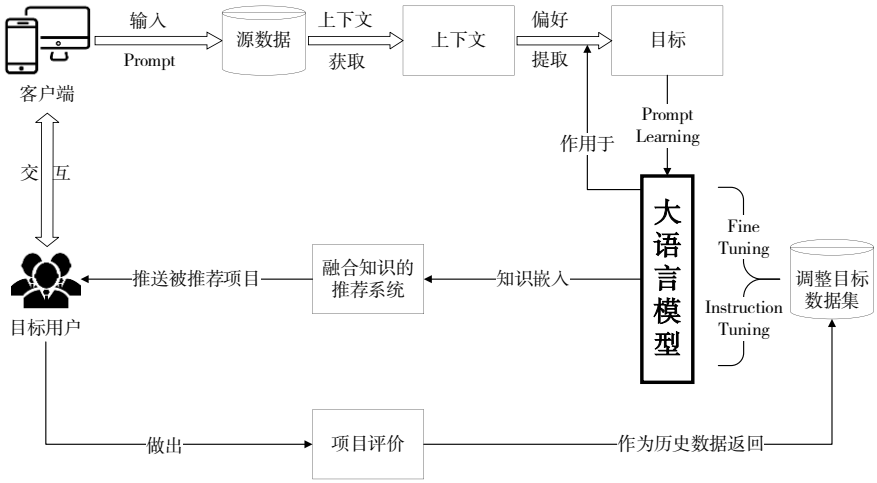

Fig. 2 Overall framework of large language model recommendation system

图2 大语言模型推荐系统整体框架

在大语言模型中，推荐系统主要从用户文本输入出发，利用先进的预处理流程优化输入数据，例如该系统能识别文本中的关键词、情感倾向，甚至是隐含的用户需求。系统基于优化后的数据生成一系列个性化的推荐候选项，并融合用户历史互动数据提升推荐内容的相关性，这一过程涉及复杂的算法，例如语义匹配和用户行为模式分析。候选项经过精心排序后形成最终的推荐清单，用户反馈将应用于循环优化模型性能。此外，系统还包括了指令调优、微调机制、细化推荐算法，以确保推荐结果持续精准。

最后，本文将深入探讨大语言模型在推荐系统中的先进应用，尤其是其多样化的调整策略，主要包括微调方式（Fine Tuning）[43-45]、提示学习方式（Prompt Learning）[46-49]和提示调优方式（Instruction Tuning）[50-52]，如图3所示。

2.2.1 基于微调方式的推荐方法

基于微调预训练的语言模型广泛应用于自然语言处理，应用领域遍及推荐系统等多个方面，核心思想是基于一个已在大规模文本数据上训练好的模型，针对特定任务或领域数据进行针对性的参数训练以满足任务需求[43]。具体到推荐任务中，可借助用户历史互动数据和项目描述数据等推荐数据集微调模型参数，以进一步捕捉用户与项目之间的关系，从而实现更个性化、精确的推荐。

Sun等[44]设计了一种名为BERT4Rec的顺序推荐模型，首先通过微调BERT模型并采用深度双向自注意力机制模拟用户行为序列，其此将完形填空任务引入序列推荐中，通过对序列中的随机掩盖项的左右上下文进行条件反射，从而预测序列中的随机掩盖项以解决信息泄漏问题并有效训练这一双向模型进行推荐[7]。在电子商务领域，Wu等[45]设计了一种新颖的端到端多任务学习框架，利用特定领域微调后的BERT模型进一步帮助模型理解特定领域的语义术语，其次采用多任务学习同时优化多个目标从而达到更好的推荐效果。

然而，上述研究的局限性为使用的传统Top-K推荐方法不够灵活。生成式语言模型能很好地解决了这一问题，例如Aleksandr等[53]提出GPTRec模型，基于GPT-2架构和奇异值分解（Singular Value Decomposition，SVD）标记化算法将物品ID分解为子标识符，以有效应对项目ID词汇量相较于语言模型词汇量大几倍的问题。同时，引入一种新的Next-K推荐策略充分考虑已推荐的项目，逐一生成推荐，使模型生成复杂的相互依赖的推荐列表。

总之，将预训练语言模型进行微调并整合到推荐系统

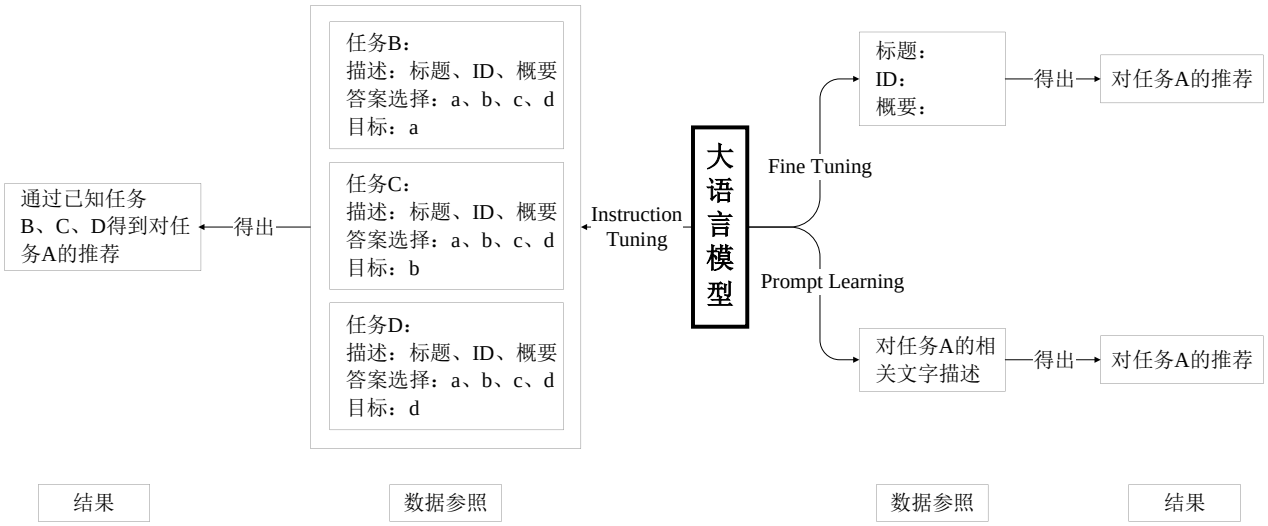

Fig. 3　Three prompt strategy concepts in the large language model

图3　大语言模型中3种提示策略概念

中不仅充分利用了强大的外部知识和个性化用户偏好,提升了推荐准确性,还具备了对有限历史数据新项目进行冷启动处理的能力,丰富了推荐系统的功能和效能,为用户提供了丰富多样的推荐体验,为推荐领域带来更大的前景和机遇[54]。

2.2.2　基于提示学习的推荐方法

大型语言模型在零样本或少样本推荐任务中也展现出了卓越能力,一些研究者尝试将提示学习模板及其学习模型的泛化能力(利用模型先验知识,在少量或无需额外数据的情况下提升模型在特定任务上的推荐能力)引入推荐系统,进而优化模型推荐效果[47]。其中,提示学习是一种 NLP 策略,已被认为是一种高效的推荐系统个性化方法,该策略允许在不大幅调整预训练语言模型结构、参数的前提下加入额外提示信息,能将下游任务转变为文本生成任务,以显著优化模型在特定任务上的表现,确保其输出更贴近任务目标[48-49]。

此外,提示信息还可能是任务描述、示例短语等,可将其整合至模型输入中引导模型生成类似示例。例如,Zhang 等[55]开发了一个 Prompt4NR 的实时新闻推荐框架,将原本用户是否会点击候选新闻的任务转化为一种完形填空式的掩码预测任务,精心设计了一系列提示模板,包括离散模板、连续模板及混合模板,构建了相应的答案空间,并整合多个提示模板的预测结果来提升模型性能。Gao 等[56]将用户配置文件与历史交互转换为提示信息,提出一种新范式 Chat-Rec,创新性的通过大型语言模型(Large scale language model, LLM)构建会话推荐系统,在 Chat-Rec 框架内通过聊天记录有效学习用户偏好,并基于上下文构建用户与产品之间的联系,从而让推荐过程更具互动性和可解释性。此外,Chat-Rec 还允许用户偏好的跨领域转移,用于不同领域的个性化推荐,并基于提示信息处理新项目的冷启动场景。综上,基于提示学习的 LLM 推荐方法为推荐系统的改进提供了全新思路,为推动人工智能生成内容(Artificial Intelligence Generates Content, AIGC)在推荐系统研究中的应用提供了新的实践场景[57]。

2.2.3　基于指令调优方式的推荐方法

指令调优(Instruction Tuning)是一种灵活、通用的微调方法,被看作是一种有监督微调(Supervised Fine Tuning),核心优势在于可调整任务描述,不依赖具体标注数据,因此在有监督或无监督情况下均能适应各种自然语言处理任务[50]。该方法强化了模型对人类意图的捕获和执行能力,尤其在零样本场景时表现卓越,即结合用户任务指令与大型语言模型可进一步优化模型的推荐个性化和效率。当模型接收用户自然语言描述或任务提示时,能准确解析用户需求,根据指令调整推荐内容,使其满足用户期望。

Geng 等[51]提出一种创新性的通用推荐框架 P5(预训练、个性化提示和预测范式推荐框架),将各种与推荐相关的任务集成到一个共享的条件语言生成框架中以提升推荐多样化和准确性。该框架的核心思想是将所有与推荐相关的数据,包括用户—项目交互历史、用户个人描述、项目元数据及用户评论等转化为通用的自然语言序列表示,有助于深入理解用户与项目间的语义关系,从而实现更加个性化推荐。在 P5 框架中,本文利用预训练阶段学习到的通用语言表示为各种不同推荐任务提供一个通用的基础模型,使任务间的知识共享和迁移更容易。此外,P5 还具备零样本学习和少样本学习能力,可根据不同用户需求生成自适应的个性化提示,面对新任务或冷启动问题时进行准确推荐,从而减少模型微调次数。

Zhang 等[52]采用以用户为中心的观点设计了一种创新的方法 InstructRec,在大语言模型中改进推荐系统,相较于其他方法的不同之处在于未依赖大语言模型的公共 API,而是对开源的大语言模型 3B Flan-T5-XL 进行个性化指令

调优，通过指令文本自由表达特定需求以更好的完成相关推荐任务。指令调优不仅增强了模型推荐精准度，还有效解决了冷启动与数据稀疏等传统推荐系统问题，从浅层模型推进到深度模型和大模型，既为用户带来了更精炼、更具个性化的推荐体验，又为推荐系统的深入研究创造了新的探索空间[51]。

3 传统模式与大语言模型比较

本文从性能、用户体验、系统复杂性与资源开销3个维度详细比较了传统模式与大语言模型的优缺点，如表1所示。

Table 1 Comparison of key dimensions between traditional models and large language model recommendation systems

表1 传统模式与大语言模型推荐系统的关键维度比较

优缺点	传统模式下的推荐系统	大语言模型下的推荐系统
优点	适合数据稀缺场景	深度理解指令、精度高
	离线计算“用户—项目”矩阵	拥有处理大模型数据的能力
	算法简单、维护直接	推荐内容丰富
缺点	冷启动问题	数据需求大
	受数据质量、特征工程影响	存在较高的隐私和安全问题
	推荐内容单调	训练、部署成本较高

3.1 性能比较

3.1.1 精度方面

传统推荐系统精度受数据质量与特征工程的直接影响，高效的特征工程通常需要深厚的领域专业知识，因此数据准确性与完整性对生成合适的推荐至关重要[15]。相较而言，基于大语言模型的推荐系统凭借其对自然语言指令的深入理解，在精度上更胜一筹，能更准确地捕获用户需求[41-42]。

3.1.2 应对大规模数据方面

传统推荐系统需要依赖强大的工程和资源，例如分布式计算和数据库优化，数据量过大将导致其性能降低，增加响应时间，而基于大语言模型的系统擅长处理海量文本和用户生成内容，能深入分析和理解用户行为，从大量文本中为用户提供更精准的建议，不受数据规模的约束[16,48,52]。

3.1.3 冷启动问题

传统推荐系统因大多依赖已有数据而存在较为明显的冷启动问题，即面对新用户或新项目时难以作出有效推荐。基于大语言模型的推荐系统通过解析用户自然语言指令或采取隐式特征学习，能在一定程度上缓解这一问题，甚至在少样本数据或缺乏历史数据的情境中也能提供针对性建议[49,51]。

3.2 用户体验比较

3.2.1 推荐多样性方面

传统推荐系统在多样性上通常存在局限性，基于协同过滤的方法可产生“过滤气泡”，导致推荐内容偏向用户历史偏好[10,11,15]。基于内容的方法主要依赖物品特征，可能导致推荐内容较为单调。相较而言，基于大语言模型的方法综合分析了用户历史行为与关联项目信息，并结合自然语言处理能力为用户呈现更广泛的内容，提升了模型推荐多样性。

3.2.2 用户行为分析方面

传统推荐系统通常依赖用户历史行为数据，例如点击、购买和评分历史以揭示用户兴趣，而基于大语言模型的推荐系统利用其出色的上下文解析能力，在分析用户历史行为的基础上进一步解读用户自然语言指令，提供了更深入和丰富的推荐[48,49,51,52]。

3.2.3 隐私与安全问题方面

传统推荐系统通常处理的用户数据规模较小，往往在本地存储或进行匿名化处理，因此在隐私安全方面风险较低。然而，基于大语言模型的系统因需要大量用户数据，可能存在较大的隐私安全风险[54]。

3.3 系统复杂性与资源开销

3.3.1 系统复杂性方面

传统推荐系统通常基于简单算法，实施与维护较为直接，协同过滤可能涉及计算用户—项目交互矩阵，优势是该操作可离线完成且无需过多计算资源[13]。基于内容的策略涉及特征工程，特征大多结构化、便于管理[16]。基于知识图谱方法虽要求构建和管理图谱数据，但复杂度尚可控[19]。基于大语言模型的推荐系统则十分复杂，需要大量文本数据和部分非结构化数据，对存储与计算资源的需求大，因此在部署此类模型时需要考虑硬件条件，可能会依赖云端或专用硬件。

3.3.2 数据处理方面

传统推荐系统大多依赖结构化数据，例如用户行为、物品特征等，处理与管理相对简单，而基于大语言模型的系统除了要处理结构化数据，还需应对非结构化的自然语言文本，涉及更多数据处理环节，例如文本清洗、分词、向量化，因此对数据收集、处理与存储的要求更高[41,42]。

3.3.3 模型训练与部署方面

传统推荐系统训练简洁，可在常规硬件上完成，例如协同过滤的用户—项目交互矩阵计算可通过常规数据库与服务器进行部署[13]。基于大语言模型的系统训练需求严格，通常需要多个GPU或TPU，为满足推理速度可能需要云计算或专业硬件，因此成本更高、处理时间时间更长[6]。

3.4 总结

传统推荐系统在数据质量和资源开销上往往表现优越，即便在数据条件不理想的情况下也能有效运作，对计算和存储资源需求较低，特别适合新兴市场或设施受限的场景，在推荐多样性并非核心关注点的应用中性能稳健；基于大语言模型的推荐系统在精确度、推荐多样性上更胜一筹，深度解析自然语言能提炼出更丰富的语义信息，进而更准确地捕获用户偏好，在社交媒体、新闻推荐或虚拟

助手等需要个性化与多样性推荐的场景中优势明显。

然而,基于大语言模型的推荐系统对数据和硬件资源需求较大,且隐私安全仍存在一定风险亟需解决,因此在选择推荐系统时需根据业务需求和资源进行考虑。此外,结合传统方法与大语言模型也是一种选择,能实现资源与性能的最佳平衡,从而在最大程度上优化用户体验。

4　发展趋势

4.1　强化提示策略的鲁棒性

鉴于用户交互的多种多样,鲁棒提示优化的重要性日益凸显[54]。该优化策略致力于提升系统对多样化用户指令的识别和响应能力,确保在任何情境中都能为用户提供准确、及时的建议,从而加强用户的信任和依赖。

4.2　构建面向用户的个性化推荐范式

随着用户行为与需求的多样性不断增加,个性化提示正成为推荐系统的关键发展方向[44,45]。通过对用户需求、行为和上下文进行深入分析和理解,既能为用户提供更精准、具有针对性和创造性的个性化推荐,又能提升用户满意度和系统应对能力[58]。

4.3　警惕大模型推荐的偏见问题

将大模型应用于推荐系统时必须高度警惕潜在的推荐偏见问题,原因为直接使用大模型可能会继承和放大现有社会偏见,导致不公平对待某些群体[59]。为此,在实际应用中特别需要考虑保护弱势群体,确保所有用户都能得到公平、无歧视的推荐服务。这一系列的创新和努力将推动大语言模型在推荐系统中的广泛应用,为用户提供更智能、贴近需求的个性化推荐体验。

5　结语

本文首先分析了传统推荐系统与基于大语言模型的推荐系统的优缺点与性能,其次探讨了他们的实际应用效果,最后从用户体验角度进行了比较。

实践表明,传统推荐方法更适合规则性较强、数据相对稀疏且稳定的场景,大语言模型模式下的推荐方法则更适于需要理解复杂语义、提供创新性解决方案和动态内容生成的场景。此外,本文还预测了未来推荐系统的发展趋势,旨在为推荐系统的发展提供参考与启示。

参考文献:

[1] WU Z Y, TANG Y, LIU H. Survey of personalized learning recommendation[J]. Journal of Frontiers of Computer Science and Technology, 2022, 16(1): 21-40.
吴正洋,汤庸,刘海. 个性化学习推荐研究综述[J]. 计算机科学与探索,2022,16(1):21-40.

[2] SCHAFER J B, KONSTAN J, RIEDL J. Recommender systems in e-commerce[C]// Proceedings of the 1st ACM Conference on Electronic Commerce, 1999: 158-166.

[3] WANG H, ZHANG F, XIE X, et al. DKN: deep knowledge-aware network for news recommendation[C]// Proceedings of the 2018 World Wide Web Conference, 2018: 1835-1844.

[4] TOUVRON H, LAVRIL T, IZACARD G, et al. LLaMa: open and efficient foundation language models [DB/OL]. https://arxiv. org/abs/2302. 13971.

[5] ZHANG S, ROLLER S, GOYAL N, et al. Opt: open pre-trained transformer language models[DB/OL]. https://arxiv. org/abs/2205. 01068.

[6] ZHAO W X, ZHOU K, LI J, et al. A survey of large language models [DB/OL]. https://arxiv. org/abs/2303. 18223.

[7] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[DB/OL]. https://arxiv. org/abs/1810. 04805.

[8] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 159: 1877-1901.

[9] ASAI A, SCHICK T, LEWIS P, et al. Task-aware retrieval with instructions[DB/OL]. https://arxiv. org/abs/2211. 09260.

[10] LINDEN G, SMITH B, YORK J. Amazon. com recommendations: item-to-item collaborative filtering[J]. IEEE Internet Computing, 2003, 7(1): 76-80.

[11] SARWAR B, KARYPIS G, KONSTAN J, et al. Item-based collaborative filtering recommendation algorithms[C]// Proceedings of the 10th International Conference on World Wide Web, 2001: 285-295.

[12] SHI Y, LARSON M, HANJALIC A. Collaborative filtering beyond the user-item matrix: a survey of the state of the art and future challenges [J]. ACM Computing Surveys, 2014, 47(1): 1-45.

[13] KOREN Y, BELL R, VOLINSKY C. Matrix factorization techniques for recommender systems[J]. Computer, 2009, 42(8): 30-37.

[14] LI B, WANG X L, DONG J T, et al. Contrastive learning and deep fusion recommendation model based on id features [J]. IEEE Access, 2024,12:163001-163015.

[15] XU S, GE Y, LI Y, et al. Causal collaborative filtering[C]// Proceedings of the 2023 ACM SIGIR International Conference on Theory of Information Retrieval, 2023: 235-245.

[16] PAZZANI M J, BILLSUS D. Content-based recommendation systems [M]. Heidelberg: Springer, 2007.

[17] SRIDHAR S, DHANASEKARAN D, LATHA G. Content-based movie recommendation system using MBO with DBN[J]. Intelligent Automation & Soft Computing, 2023, 35(3):3241-3257.

[18] DOMINGOS P. A few useful things to know about machine learning[J]. Communications of the ACM, 2012, 55(10): 78-87.

[19] SINGHAL A. Introducing the knowledge graph: things, not strings[J]. Official Google Blog, 2012, 5(16): 3.

[20] ZHANG F, YUAN N J, LIAN D, et al. Collaborative knowledge base embedding for recommender systems[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,2016: 353-362.

[21] GAO C, ZHENG Y, LI N, et al. A survey of graph neural networks for recommender systems: challenges, methods, and directions[J]. ACM Transactions on Recommender Systems, 2023, 1(1): 1-51.

[22] QIN C, ZHU H S, ZHUANG F Z, et al. A survey on knowledge graph-based recommender systems[J]. Scientia Sinica Informationis, 2020, 50(7):937-956.
秦川,祝恒书,庄福振,等. 基于知识图谱的推荐系统研究综述[J]. 中国科学:信息科学,2020,50(7):937-956.

[23] GOYAL P, CHHETRI S R, CANEDO A. Dyngraph2vec: capturing network dynamics using dynamic graph representation learning[J]. Knowledge-Based Systems, 2020, 187: 104816.

[24] GAO J, XIN X, LIU J, et al. Fine-grained deep knowledge-aware network for news recommendation with self-attention[C]// 2018 IEEE/WIC/ACM International Conference on Web Intelligence, 2018: 81-88.

[25] SONG W, DUAN Z, YANG Z, et al. Explainable knowledge graph-based recommendation via deep reinforcement learning[DB/OL]. https://arxiv.org/abs/1906.09506v1.

[26] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.

[27] GEHMAN S, GURURANGAN S, SAP M, et al. Realtoxicityprompts: evaluating neural toxic degeneration in language models [DB/OL]. https://arxiv.org/abs/2009.11462.

[28] STIENNON N, OUYANG L, WU J, et al. Learning to summarize with human feedback [J]. Advances in Neural Information Processing Systems, 2020, 33: 3008-3021.

[29] YANG J, JIN H, TANG R, et al. Harnessing the power of llms in practice: a survey on chatgpt and beyond [DB/OL]. https://arxiv.org/abs/2304.13712.

[30] ASTRAKHANTSEV N. Automatic term acquisition from domain-specific text collection by using Wikipedia[C]// Proceedings of the Institute for System Programming of the RAS, 2014: 7-20.

[31] LOSSIO-VENTURA J A, JONQUET C, ROCHE M, et al. Yet another ranking function for automatic multiword term extraction[C]// International Conference on Natural Language Processing, 2014: 52-64.

[32] HIDASI B, KARATZOGLOU A, BALTRUNAS L, et al. Session-based recommendations with recurrent neural networks[DB/OL]. https://arxiv.org/abs/1511.06939.

[33] LI J, REN P, CHEN Z, et al. Neural attentive session-based recommendation[C]// Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, 2017: 1419-1428.

[34] TANG J, WANG K. Personalized top-n sequential recommendation via convolutional sequence embedding [C]// Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 2018: 565-573.

[35] YUAN F, KARATZOGLOU A, ARAPAKIS I, et al. A simple convolutional generative network for next item recommendation[C]// Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, 2019: 582-590.

[36] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [DB/OL]. https://arxiv.org/abs/1409.0473.

[37] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[DB/OL]. https://arxiv.org/abs/1706.03762.

[38] XIONG R, YANG Y, HE D, et al. On layer normalization in the transformer architecture[C]// International Conference on Machine Learning, 2020: 10524-10533.

[39] LIN Z, FENG M, SANTOS C N, et al. A structured self-attentive sentence embedding[DB/OL]. https://arxiv.org/abs/1703.03130.

[40] HADI M U, QURESHI R, SHAH A, et al. A survey on large language models: Applications, challenges, limitations, and practical usage[DB/OL]. https://arxiv.org/abs/2303.18223.

[41] ZHANG Y, DING H, SHUI Z, et al. Language models as recommender systems: evaluations and limitations [DB/OL]. https://arxiv.org/abs/2411.10697.

[42] LI J, WANG M, LI J, et al. Text is all you need: learning language representations for sequential recommendation[DB/OL]. https://arxiv.org/abs/2305.13731.

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

[44] SUN F, LIU J, WU J, et al. BERT4Rec: sequential recommendation with bidirectional encoder representations from transformer[C]// Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019: 1441-1450.

[45] WU X, MAGNANI A, CHAIDAROON S, et al. A multi-task learning framework for product ranking with BERT[C]// Proceedings of the ACM Web Conference, 2022: 493-501.

[46] HOULSBY N, GIURGIU A, JASTRZEBSKI S, et al. Parameter-efficient transfer learning for NLP[C]// International Conference on Machine Learning, 2019: 2790-2799.

[47] PETRONI F, ROCKTÄSCHEL T, LEWIS P, et al. Language models as knowledge bases?[DB/OL]. https://arxiv.org/abs/1909.01066.

[48] BEN-DAVID E, OVED N, REICHART R. Pada: a prompt-based autoregressive approach for adaptation to unseen domains[DB/OL]. https://arxiv.org/abs/2102.12206v1.

[49] CHEN X, ZHANG N, XIE X, et al. Knowprompt: knowledge-aware prompt-tuning with synergistic optimization for relation extraction [C]// Proceedings of the ACM Web Conference, 2022: 2778-2788.

[50] ZHANG S, DONG L, LI X, et al. Instruction tuning for large language models: a survey[DB/OL]. https://arxiv.org/abs/2308.10792.

[51] GENG S, LIU S, FU Z, et al. Recommendation as language processing (RLP): a unified pretrain, personalized prompt & predict paradigm (p5) [C]// Proceedings of the 16th ACM Conference on Recommender Systems, 2022: 299-315.

[52] ZHANG J, XIE R, HOU Y, et al. Recommendation as instruction following: a large language model empowered recommendation approach [DB/OL]. https://arxiv.org/abs/2305.07001.

[53] PETROV A V, MACDONALD C. Generative sequential recommendation with GPTRec[DB/OL]. https://arxiv.org/abs/2306.11114.

[54] WU L, ZHENG Z, QIU Z, et al. A survey on large language models for recommendation[DB/OL]. https://arxiv.org/abs/2305.19860.

[55] ZHANG Z, WANG B. Prompt learning for news recommendation [DB/OL]. https://arxiv.org/abs/2304.05263.

[56] GAO Y, SHENG T, XIANG Y, et al. Chat-REC: towards interactive and explainable llms-augmented recommender system[DB/OL]. https://arxiv.org/abs/2303.14524.

[57] XIANG W, WANG Z, DAI L, et al. ConnPrompt: connective-cloze prompt learning for implicit discourse relation recognition[C]// Proceedings of the 29th International Conference on Computational Linguistics, 2022: 902-911.

[58] WANG W, LIN X, FENG F, et al. Generative recommendation: towards next-generation recommender paradigm [DB/OL]. https://arxiv.org/abs/2304.03516.

[59] ZHANG J, BAO K, ZHANG Y, et al. Is ChatGPT fair for recommendation? evaluating fairness in large language model recommendation[DB/OL]. https://arxiv.org/abs/2305.07609.

(责任编辑:刘嘉文)