

电子商务推荐系统研究综述

曹俊伟, 李军祥*, 李玉璐

上海理工大学管理学院, 上海

收稿日期: 2025年11月3日; 录用日期: 2025年11月19日; 发布日期: 2025年12月18日

摘要

随着互联网技术的飞速发展和电子商务规模的不断扩大, 信息过载问题日益严重。推荐系统作为解决信息过载问题的有效工具, 已成为电子商务平台提升用户体验、增加销售额和增强用户粘性的核心技术。本文旨在对电子商务推荐系统的研究现状进行系统性综述。首先介绍了推荐系统在电子商务中的重要性及其基本架构, 其次详细梳理和分析了基于协同过滤、基于内容、基于知识以及混合推荐等主流推荐算法的基本原理、优势与局限性, 然后介绍了用于训练模型的主流电子商务平台的数据集, 再探讨了推荐系统在实战中面临的关键挑战如冷启动、数据稀疏性等问题; 最后展望了电子商务推荐系统未来的发展趋势, 包括深度学习与大模型的深度融合、多目标与序列化推荐、可解释性与公平性、跨域与联邦学习推荐, 以及多模态信息融合等方向。

关键词

电子商务, 推荐系统, 协同过滤, 深度学习, 冷启动

A Review of Research on E-Commerce Recommender Systems

Junwei Cao, Junxiang Li*, Yulu Li

Business School, University of Shanghai for Science and Technology, Shanghai

Received: November 3, 2025; accepted: November 19, 2025; published: December 18, 2025

Abstract

With the rapid development of Internet technology and the continuous expansion of the e-commerce scale, the information overload problem has become increasingly severe. As an effective tool to alleviate the information overload problem, recommendation systems have become a core technology for e-commerce platforms to enhance user experience, increase sales, and strengthen user stickiness. This paper aims to provide a systematic review of the current research status of e-com-

*通讯作者。

merce recommendation systems. It first introduces the importance and fundamental architecture of recommendation systems in e-commerce. Subsequently, it provides a detailed organization and analysis of the basic principles, advantages, and limitations of mainstream recommendation algorithms, including collaborative filtering-based, content-based, knowledge-based, and hybrid recommendation approaches. Then, it introduces mainstream e-commerce platform datasets commonly used for training models. Furthermore, it discusses key challenges faced by recommendation systems in practical applications, such as the cold-start problem and data sparsity. Finally, the paper prospects the future development trends of e-commerce recommendation systems, including the deep integration of deep learning and large models, multi-objective and sequential recommendation, explainability and fairness, cross-domain and federated learning-based recommendation, and multi-modal information fusion.

Keywords

E-Commerce, Recommendation System, Collaborative Filtering, Deep Learning, Cold Start

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

电子商务的蓬勃发展使得在线商品和服务的数量呈指数级增长。用户在面对海量选择时，往往难以快速找到自己真正感兴趣的的商品，这种现象被称为“信息过载”^[1]。为了解决这一问题，推荐系统应运而生。它通过分析用户的历史行为(如浏览、购买、评分)、物品属性以及上下文信息，主动为用户预测并推荐其可能感兴趣的物品。

在电子商务领域，一个高效的推荐系统不仅能显著提升用户的购物体验和满意度，还能直接为平台带来巨大的商业价值，包括提高转化率、增加客单价、增强用户忠诚度等，其中跨境电商平台还有其特有的数据特性，也带来了其独特的数据价值和建模挑战^[2]。因此，对推荐系统的研究具有重要的理论意义和实际应用价值。本文旨在系统性地梳理电子商务推荐系统的核心技术、挑战与未来趋势，为相关研究者和实践者提供参考。

2. 推荐系统的基本架构

典型的电子商务推荐系统通常由三个核心功能模块构成，分别为数据采集与预处理层、推荐模型层以及结果生成与反馈层，各模块在推荐流程中承担不同的职能，共同构成一个完整的推荐闭环^[3]。

2.1. 数据采集与预处理层

该模块负责从多源异构数据中采集原始信息，并进行清洗、集成与转换，为上层推荐模型提供高质量的结构化输入。其数据源主要包括三类：(1) 用户数据，涵盖用户静态画像(如年龄、性别、地域等)及动态行为序列(如点击、浏览、收藏、加购、购买与评分等)；(2) 物品数据，包括商品本身的属性信息，例如标题、类别、品牌、价格及详细描述等；(3) 上下文数据，用于捕捉推荐场景的上下文信息，如时间、地理位置、终端设备及当前会话状态等^[4]。

2.2. 推荐模型层

作为推荐系统的核心，该模块基于预处理后的数据，运用各类推荐算法(如协同过滤、基于内容的推

荐、深度学习模型等)学习用户与物品的潜在特征表示，并预测用户对未交互物品的偏好程度。本模块的设计与实现直接决定了推荐系统的性能与效果，后续章节将系统性地分析与比较不同算法的原理及其适用场景[5]。

2.3. 结果生成与反馈层

该模块依据推荐模型输出的偏好评分或排序结果，生成最终面向用户的推荐列表，并将其呈现在相应终端界面。与此同时，系统会实时记录用户对推荐结果的反馈行为(如点击、转化、购买等)，并将其作为新一轮的训练数据反馈至模型层，从而形成一个持续优化的闭环学习机制，不断提升推荐的准确性与用户满意度[6]。

3. 主流电子商务平台的推荐算法综述

3.1. 基于协同过滤的推荐

基于协同过滤的推荐算法是推荐系统领域中最经典、应用最广泛的算法之一。其核心思想非常直观：“物以类聚，人以群分”。它假设过去有相似喜好的用户在未来也会有相似的喜好，并且相似的商品会被相似的用户喜欢。该算法主要分为两大类：

(1) 基于用户的协同过滤：首先找到与目标用户兴趣相似的一组“邻居用户”，然后将这些邻居用户喜欢而目标用户未曾接触过的物品推荐给目标用户。其关键在于如何计算用户之间的相似度(如余弦相似度、皮尔逊相关系数等)[7]。

(2) 基于物品的协同过滤：主要分析用户行为记录(如评分、购买)来计算物品之间的相似度。然后，根据目标用户历史喜欢的物品，找出与这些物品相似的物品，并将其推荐给用户。亚马逊的“购买了此商品的顾客也购买了……”就是该方法的经典应用[8]。

协同过滤的优势在于它不依赖于物品本身的属性信息，仅通过用户的历史行为数据就能进行有效推荐，能够发现复杂的、潜在的物品关联性。然而，它也面临着一些著名的挑战，如冷启动问题(新用户或新物品缺乏行为数据)、数据稀疏性问题(用户 - 物品评分矩阵非常稀疏，导致相似度计算不准确)以及可扩展性问题(用户和物品数量巨大时，计算相似度开销高昂)[9]。

协同过滤的奠基性研究可追溯到上世纪 90 年代。Goldberg 等人[10]在 Tapestry 系统中首次实践了“协同”的理念，依赖于用户明确标注的喜好。随后，Resnick 等人[1]提出的 GroupLens 系统奠定了现代协同过滤的基础，它通过收集用户对新闻文章的隐式评分，自动预测目标用户对未读文章的兴趣，开启了自动化、大规模的推荐时代。早期的研究主要集中在相似度计算方法的优化上，如 Sarwar 等人[11]系统地比较了基于用户和基于物品的方法，并指出在数据稀疏的场景下，基于物品的方法通常表现更稳定、可扩展性更好。

为了应对协同过滤固有的数据稀疏性和可扩展性挑战，学术界涌现出大量创新解决方案。其中，矩阵分解技术成为了一个里程碑式的突破。Koren 等人在著名的 Netflix Prize 竞赛中总结的论文指出，将隐语义模型(如 SVD、SVD++)与传统的邻域方法相结合，能显著提升预测精度[12]。矩阵分解通过将高维稀疏的用户 - 物品矩阵映射到低维潜在空间，有效地捕捉了用户和物品背后隐藏的抽象特征，从而缓解了数据稀疏性问题。与此同时，为了解决冷启动问题，研究者们提出了将协同过滤与基于内容的推荐相结合的方法，例如，Melville 等人[13]提出的“内容提升协同过滤”利用内容信息来填充稀疏的评分矩阵，为仅有少量行为的用户生成更可靠的推荐。这些研究共同推动了协同过滤从基础的邻域方法向更复杂、更精确的模型化方向发展。

随着深度学习技术的兴起，协同过滤的研究进入了新的阶段。神经协同过滤(NCF)框架[14]利用神经

网络取代传统的矩阵分解内积操作，以学习用户和物品之间更复杂的非线性交互关系。此外，图神经网络(GNN) [15]也被广泛应用于对用户 - 物品交互图进行建模，通过消息传播机制更有效地挖掘高阶协同信号。这些基于深度学习的模型进一步提升了推荐的性能，并成为当前研究的前沿热点。

3.2. 基于内容的推荐

基于内容的推荐算法是另一类基础的推荐技术，其核心思想与协同过滤截然不同。它不依赖于用户社群或物品群体的集体行为，而是专注于分析用户过去喜欢的物品本身的内容特征，以及待推荐物品的内容特征，通过计算两者之间的相似度来进行推荐。具体来说，该系统首先需要为每个物品构建一个内容档案，通常由一系列描述性特征(如关键词、类别、标签、演员、导演、文本内容等)组成。同时，系统会为用户构建一个用户画像，这个画像通常通过聚合该用户历史上喜欢过的物品的内容特征来形成(例如，一个用户经常看科幻电影，那么他的画像中“科幻”这个特征的权重就会很高)。当进行推荐时，算法会计算待推荐物品的内容档案与用户画像的相似度，并将最相似的物品推荐给用户。

基于内容的方法主要有以下优势：首先，它不存在冷启动中的新物品问题，因为只要一个新物品有内容描述，就可以立即被推荐。其次，它具有很好的可解释性，可以轻松地告诉用户：推荐此内容是因为您喜欢包含某些特征的项目。此外，它避免了协同过滤中常见的“同温层”效应，能够为用户发掘小众、独特的兴趣。然而，它也存在明显的局限性，主要是特征提取依赖领域知识、容易过度专业化(推荐内容过于单一，缺乏惊喜)以及对于新用户(用户冷启动)，由于缺乏历史行为，难以构建准确的用户画像[16]。

基于内容的推荐算法其思想根源可以追溯到信息检索领域。早期的系统大量借鉴了文本信息检索的技术，其中最具代表性的是 TF-IDF (词频 - 逆文档频率)方案。Mooney 和 Roy [17]在图书推荐领域的开创性工作，清晰地展示了如何利用书籍的文本内容(如目录、简介)来构建物品档案，并通过计算用户已喜欢书籍与候选书籍在 TF-IDF 向量空间中的余弦相似度来生成推荐。他们的研究证实了仅凭内容信息构建有效推荐系统的可行性。与此同时，Pazzani 和 Billsus [18]的工作将基于内容的方法推向了一个更智能的方向，他们提出使用朴素贝叶斯分类器等机器学习算法来学习用户喜好模型，将推荐问题形式化为一个二元分类任务(判断用户是否会喜欢某物品)，这为后续更复杂的机器学习模型的应用奠定了基础。

3.3. 基于知识的推荐

基于知识的推荐系统其理论基础深深植根于人工智能中的知识表示和推理领域。早期的代表性工作如 Ricci 和 Werthner [19]对基于案例的旅游推荐系统的研究，清晰地展示了如何通过定义领域本体(Ontology)来形式化旅游产品的复杂属性，并利用相似性度量来匹配用户约束与产品案例，这为处理复杂、多属性的商品推荐提供了范例。与此同时，Felfernig 和 Burke [20]对基于约束的推荐系统进行了系统性的梳理，他们阐述了如何将用户需求、产品目录和商业规则形式化为一个约束满足问题，并通过高效的约束求解器来找出符合条件的推荐结果。这些研究共同确立了基于知识的推荐作为一种解决数据稀疏和冷启动场景的有效范式，并强调了其在支持用户深度交互和透明决策方面的独特价值。

为了克服构建和维护大规模知识库的挑战，并增强系统的智能性与适应性，后续研究主要朝着两个方向发展。其一是与其它推荐技术的融合。研究者们探索将基于知识的方法作为混合推荐系统中的一个核心组件，用以补充数据驱动方法的不足。例如，在会话式推荐系统中，基于知识的推理被用于引导对话、解释推荐结果，并与协同过滤模型结合以平衡用户明确需求与潜在兴趣。其二是引入更先进的知识表示与学习技术。近年来，随着知识图谱的兴起，将其作为推荐系统的背景知识库成为了研究热点。知识图谱能够以图结构的形式自然地表达实体、属性及其间复杂的语义关系。通过利用图嵌入、图神经网络等技术，系统可以从知识图谱中学习到用户和物品的丰富语义表示，从而在保留可解释性的同时，实

现更灵活、更深层次的推理，极大地拓展了基于知识推荐的能力边界[21]。

3.4. 深度学习驱动的混合推荐模型

深度学习的兴起，为推荐系统突破了传统方法的瓶颈，开启了新的发展阶段。其核心贡献在于能够自动学习高维稀疏特征的稠密表示，并高效地捕捉复杂的非线性及高阶特征交互。这一领域的模型演进，清晰地呈现出一条从浅层模型与深度模型并行，到深度融合，再到针对用户动态兴趣与多任务学习等特定问题深化发展的技术路径。

3.4.1. 浅层模型与深度模型的融合与并行

早期的深度学习推荐模型致力于如何将传统的浅层模型与深度神经网络有机结合。Wide & Deep [22]是这一思想的典范，其“Wide”部分利用广义线性模型处理记忆性特征，保证模型的鲁棒性和可解释性；“Deep”部分则通过多层感知机学习特征的深层交互与泛化能力。该架构成功解决了记忆与泛化的平衡问题，成为工业界的标准基线。在 Wide & Deep 的启发下，研究者们提出了多种改进架构。DeepFM [23]是对其的一项重要发展，它用因子分解机取代了 Wide 部分，使得浅层模块无需精细的特征工程即可自动学习二阶特征交互，与 Deep 部分共享特征输入，形成了更紧密的联合训练架构。为进一步提升特征交叉的显式建模能力。DCN [24]引入了交叉网络，通过高效的逐层外积操作显式地构建有限阶的特征交叉，并与一个并行的深度网络相结合。而 xDeepFM [25]则进一步推进一步，设计了压缩交互网络，旨在同时实现向量级交互的显式性与高阶交叉的隐式学习，被认为是更“纯粹”的深度因子分解机。

3.4.2. 用户动态兴趣序列建模

电子商务场景中，用户的兴趣是动态演变的，而非静态不变的。为此，一系列基于序列建模的深度网络被提出，以捕捉用户兴趣的动态变化。

DIN [26]首次引入了注意力机制，针对候选商品自适应地激活并加权用户丰富的历史行为，模拟了人类在决策时对不同历史信息赋予不同权重的认知过程。在此基础上，DIEN [27]更进一步，通过引入双层循环神经网络与辅助损失，不仅捕捉用户的兴趣，更模拟了兴趣随时间“演化”的过程，使兴趣表示更具时序性与动态性。DSIN [28]则从用户行为会话的角度切入，利用自注意力机制与 Bi-LSTM 分别捕捉会话内兴趣和会话间兴趣的演化，对用户短期意图的建模更为精准。

3.4.3. 多任务学习与样本选择偏差问题

在复杂的电商业务中，优化单一目标(如点击率)往往不足，需要同时优化多个相关任务(如点击、转化、停留时长等)。多任务学习成为解决这一问题的关键技术。

ESMM [29]通过创新的“CTR-CTCVR”概率连乘结构，利用点击任务作为中间桥梁，直接在全部样本空间(包括曝光和未曝光样本)上建模转化率，有效解决了传统 CVR 模型面临的样本选择偏差与数据稀疏两大难题。

随着任务复杂度的增加，简单的共享底层参数结构可能导致任务间的“跷跷板”现象。MMoE [30]通过引入多个专家网络和任务特定的门控机制，允许不同任务以不同权重选择和组合共享的知识，增强了模型在多任务间的表达能力与灵活性。PLE [31]模型在 MMoE 的基础上，进一步设计了任务特定的专家网络和分层渐进的信息分离机制，显式地减少了任务间的负向干扰，在更复杂的任务关系(如任务相关性低或存在序列依赖)下表现优异。

3.4.4. 双塔结构与门控网络增强

在召回阶段或需要高效匹配的场景中，双塔模型因其结构清晰、线上服务性能高而广受欢迎。DSSM

[32]及其变体是此类模型的代表，通过将用户和物品分别编码为独立的向量表示，并通过计算向量相似度进行匹配，实现了海量候选集下的高效检索。

为了增强模型的特征选择与信息流动能力，门控机制被广泛应用。GateNet [33]通过设计门控单元来调控网络中各层间的信息流，提升了模型的表征能力。类似地，GemNN [34]则将门控机制深度融入多任务学习框架，通过门控网络来精细化地控制不同任务间信息的共享与隔离，进一步优化了多任务学习的性能。

4. 用于训练模型的电子商务平台数据集

在数据驱动的研究范式下，高质量、大规模的数据集构成了推荐系统、社会计算与电子商务分析等领域的实证基础。这些数据集不仅为算法模型的训练与评估提供了必要的土壤，其本身所蕴含的独特属性和挑战也常常定义并推动着特定研究方向的发展。下文将针对在学术界享有盛誉且应用广泛的数据集进行系统性的介绍。本节将超越基础的数据结构描述，深入探讨其数据生成机理、固有的数据偏差与研究挑战，以及它们在推动学科前沿中所扮演的角色。

4.1. Yelp

作为商业点评平台的典型代表，其核心学术价值在于它提供了一个多模态、多关系的社会经济系统快照。该数据集并非实验室环境下受控收集的结果，而是 Yelp 平台真实用户活动的自然沉淀，因而具备了极高的生态效度。从数据结构观之，它精巧地交织了结构化数据(如数值评分、商家属性类别)、非结构化文本(用户撰写的长篇评论)、明确的社交图谱(用户间的好友关系)以及时空序列数据(带有时间戳的签到记录)。这种丰富的异构性使其成为检验复杂混合推荐算法的理想试验场。例如，研究者可以探索如何将评论文本中的语义信息通过自然语言处理技术进行嵌入，并与表征用户信任关系的社交网络拓扑结构、以及制约用户选择的商家地理邻近性进行协同建模。然而，使用该数据集时必须审慎考虑其固有的选择偏差，即活跃用户与热门商家的过度代表性问题，以及用户生成内容中可能存在的极端意见偏差。这些偏差要求研究者在模型构建与结果泛化中采取必要的校正措施。

4.2. Gowalla 数据集

源于一个已停止运营的基于位置的社交网络服务，它在学术史上的地位由其独特的高分辨率时空轨迹与显性社交链接的结合所奠定。该数据集的核心记录单元是用户在某一个特定兴趣点的“签到”行为，每一个记录都包含了用户标识、位置标识与精确到秒的时间戳。这种数据形态使得对人类社会移动性的定量化、细粒度研究成为可能。在研究方法论上，Gowalla 数据集强有力地推动了推荐系统研究从静态的、仅依赖共现关系的协同过滤，向动态的、序列敏感的下一代推荐模型的演进。研究者利用此数据集，广泛开发并验证了基于马尔可夫链、循环神经网络以及注意力机制的顺序位置推荐模型，旨在预测用户下一个可能访问的地点。此外，该数据集的高度稀疏性(绝大多数用户仅访问过极少量的地点)和显著的长尾分布(少数热门地点集中了大部分访问)为研究带来了核心挑战，如何在这种挑战下有效捕捉用户的个性化移动模式并实现精准推荐，成为了该领域一个持久的研究议题。其社交链接数据则允许学者探究同质性效应——即好友之间是否在移动模式上表现出相似性——在位置预测中的作用。

4.3. MovieLens-10M (ML-10M)

该数据集由 GroupLens 研究实验室长期维护，在推荐系统文献中扮演着类似于“基准砝码”的角色。与许多来自商业平台、噪声显著且需要复杂预处理的数据集不同，ML-10M 以其高度的洁净性、稳定的结构和适中的规模而备受青睐。它主要包含用户对电影的 1 至 5 星评分，并辅以电影题材信息和用户自

由添加的标签。该数据集最大的优势在于其作为协同过滤算法基准测试床的不可撼动的地位。由于其用户 - 物品交互矩阵相较于其他真实世界数据集(如 Gowalla 或 Amazon)更为稠密, 它能够更稳健地评估模型在捕捉潜在用户偏好方面的核心能力。从早期的矩阵分解技术(如 SVD、SVD++)到现代的神经网络模型(如 NeuMF、LightGCN), 无数推荐算法的性能都是在 ML-10M 上进行了首次报告和横向比较。然而, 这一优势也伴随着其局限性: ML-10M 缺乏丰富的上下文信息(如用户的社会关系或详细的时序行为), 并且其评分数据存在严重的选择偏差(用户主动选择观看并评分电影), 这限制了对某些更复杂推荐场景(如序列推荐或捆绑推荐)的探索。

4.4. Amazon Reviews

该数据集作为全球最大电子商务平台的用户反馈档案, 其规模与范围在学术界是前所未有的。该数据集跨越了从图书到电子产品等数十个商品类别, 并随时间推移不断更新, 形成了一个动态演化的生态系统。每条数据记录不仅包含评分, 更重要的是包含了用户撰写的评论文本, 这为研究开辟了至关重要的文本驱动的研究方向。该数据集是评分预测和推荐系统研究, 特别是处理极端数据稀疏性和冷启动问题的经典选择。由于亚马逊的商品海量, 任何单一用户仅能接触其中极小一部分, 导致用户 - 物品矩阵极度稀疏。此外, 该数据集蕴含着丰富的社会经济现象研究素材。例如, 通过分析评论的有用性投票, 可以研究在线声誉系统的形成机制; 通过追踪价格、评分和销售排名的时序变化, 可以分析供需动态; 而跨类别购买行为则为研究跨领域推荐和用户兴趣迁移提供了宝贵资源。使用此数据集的主要挑战在于其庞大規模带来的计算复杂性以及对数据进行有效采样和清洗的必要性。

4.5. Tmall

该数据集由阿里巴巴集团发布, 代表了中国电子商务背景下的大规模用户行为数据, 其特点与西方的 Amazon 数据集形成了有趣的对比与补充。该数据集的核心并非围绕静态的评分或评论, 而是捕捉了用户在平台内进行商品浏览、收藏、加购与购买等一系列隐式反馈和动态行为序列。这种数据形态更贴近真实的商业决策场景, 因为购买行为远比评分行为更能直接反映用户偏好。因此, Tmall 数据集极大地促进了基于隐式反馈和序列行为的推荐模型的发展。在该数据集上的研究典型地侧重于建模用户的动态意图, 即如何从用户连续的行为日志(例如, 先浏览多个品牌的手机, 然后将其中一款加入购物车)中推断其当前的购物目标, 并预测其最终的购买决策。这一特性使其特别适合于评估下一代电子商务推荐技术, 如会话式推荐和实时重新排序。同时, 该数据集也常被用于研究诸如“双十一”等大型促销活动期间的用户行为突变、流量峰值预测以及营销策略的有效性等独特的商业智能问题。

5. 推荐系统面临的关键挑战

尽管推荐技术不断演进, 但在实际应用中仍面临若干根本性挑战。这些挑战并非孤立存在, 而是相互交织, 深刻影响着推荐系统的性能与可信度。本节将深入探讨其中三个核心挑战: 用户兴趣的动态性与可塑性、准确性与多样性的内在博弈, 以及推荐结果的公平性与可解释性。

5.1. 用户兴趣的动态性与可塑性挑战

传统推荐模型在很大程度上建立在“用户兴趣是静态”的假设之上, 通过历史行为的聚合来构建一个固定的用户表征。然而, 真实场景中的用户兴趣本质上是动态演变且高度情境依赖的。一个用户可能长期对数码产品抱有浓厚兴趣, 这构成了其稳定的长期画像; 但在某个特定的会话中, 其核心意图可能瞬间转变为家人选购节日礼物。这种动态性和可塑性构成了推荐系统面临的首要难题, 即兴趣漂移与深层动机挖掘的困境。兴趣漂移体现在用户兴趣会随时间自然变化, 一个典型的例子是, 用户在特定生命

周期阶段(如孕期)的消费模式会集中爆发，而后迅速衰减，若系统不能及时感知这种转变，仍固守过时的历史数据，就会导致刻板印象式的无效推荐，严重损害用户体验。更深层次的挑战在于，用户表面的点击、浏览行为只是其兴趣的显式反馈，背后所隐藏的复杂、细微的深层动机却难以被捕捉。用户连续点击多个不同品牌的运动鞋，其表层行为是“浏览运动鞋”，但深层动机可能是“为马拉松比赛寻找一款轻便的竞速鞋”。仅基于行为共现的传统模型无法理解这种语义层面的意图差别，导致推荐流于表面。正因如此，推荐系统研究的前沿方向之一，便是如何从静态建模转向动态感知，例如通过 DIN、DIEN 等引入注意力与序列建模的机制，来模拟用户兴趣的演化轨迹，从而更敏锐地响应其当前需求。

5.2. 准确性与多样性的内在博弈

推荐系统的核心目标之一是提供精准的推荐，但过度追求准确性在实践中会引致“信息茧房”与“同质化”的负面效应，这本质上是一个深刻的探索与利用的博弈难题。精准推荐意味着向用户呈现与其已知兴趣高度吻合的物品，这在短期内无疑能提升点击率与用户满意度。然而，其长期风险在于，它可能将用户禁锢在一个狭窄的兴趣领域内，使其无法接触到自身潜在兴趣之外的新鲜内容，最终导致用户体验的僵化与审美疲劳。从系统生态健康的角度看，若流量过度集中于少数头部热门商品，那些处于长尾的、新生的商品将永无曝光之日，这无疑会抑制平台的创新活力与生态多样性。因此，纯粹的准确性导向在长期看来反而可能损害用户忠诚与商业机会。与之相对，多样性旨在打破这种僵局，通过引入新颖、跨品类的物品来拓宽用户的视野，激发潜在需求。但多样性若不加约束，又会泛滥成无关内容的堆砌，直接冲击推荐结果的相关性。面对这一两难境地，现代推荐系统的设计哲学已从追求单一指标的极致，演进为对多目标的协同优化。正如 MMoE、PLE 等多任务学习模型所倡导的，系统需要在统一的框架内智慧地平衡点击率、转化率、多样性、新颖性等多个目标，其终极目的不仅是预测用户“想要什么”，更是要判断用户“可能喜欢什么但还不知道”，从而在满足与启发之间找到那个微妙的平衡点。

5.3. 推荐公平性与算法可解释性的挑战

随着推荐系统在社会经济活动中扮演的角色日益重要，其决策过程的公平性与透明性已从伦理考量演变为必须解决的实践难题。公平性挑战呈现出双重维度。其一关乎于用户，算法不应因用户的性别、年龄、地域等受保护属性而产生系统性歧视，例如在招聘或信贷场景中，有偏的历史数据可能导致模型对特定群体机会的不公平剥夺。其二关乎于物品或内容提供者，算法需要为新卖家、小众内容创造相对公平的曝光机会，避免流量垄断扼杀生态的竞争与创新。与公平性紧密交织的是可解释性挑战。一个性能卓越的“黑箱”模型，即便其推荐结果在统计上是公平的，若无法提供令人信服的理由，也难以建立用户信任。当推荐出现偏差或错误时，用户迫切需要理解“为何我会收到此推荐”，以便调整自身行为或判断系统的可靠性。电商平台中“因买 A 而推荐 B”的基于协同过滤的解释直观易懂，而面对 xDeepFM、DIN 等复杂深度学习模型内部错综复杂的非线性计算，生成清晰、简洁的事后解释变得异常困难。解决这一系列挑战呼唤技术与社会规范的双重创新。在技术层面，研发模型无关的事后解释方法或设计自身具备强可解释性的模型架构是重点方向。在社会与机制层面，则需建立有效的算法审计流程与伦理规范，推动推荐系统从纯粹的性能驱动，向负责任、可信赖的人工智能系统演进。

6. 未来发展趋势

展望未来，电子商务推荐系统的演进将不再局限于单一模型或指标的优化，而是呈现出一种系统性、融合性的发展态势。其核心驱动力来自于对用户理解更深层次、对业务目标更全面把握，以及对技术伦理更高度重视的内在需求。未来的研究与实践将围绕以下几个关键方向展开。

首先，语义理解与认知推理的深化将成为下一代推荐系统的核心竞争力。当前基于深度学习的模型虽能自动学习特征交互，但对商品内容与用户意图的语义理解仍停留在浅层。大型语言模型的崛起为解决这一问题提供了全新路径。通过注入 LLMs 的强大语义能力，推荐系统能够超越传统的关键词匹配和行为共现，真正理解商品描述的功能特性、情感色彩乃至文化隐喻，并能从用户冗长的评论和搜索词中精准提炼其深层需求与偏好。这意味着推荐系统将从一个“行为预测器”向“语义理解者”乃至“认知伙伴”演进，能够进行基于知识的推理，例如，理解“适合海边度假的连衣裙”不仅关乎品类，更涉及材质、风格与场景的复杂匹配。

其次，推荐系统的架构设计正面临从静态优化到动态交互的范式转移。未来的系统必须更好地适应信息的实时性和用户兴趣的流动性。这要求序列化推荐模型进一步进化，从仅仅捕捉历史行为序列，发展到实时响应当前会话内的微秒级意图变化。同时，推荐不再是一个单向的、一次性的信息推送过程，而将演变为一个支持多轮交互、允许用户反馈与纠正的对话式系统。在这种动态范式中，系统能够通过主动提问、澄清或试探性推荐来逐步明确用户意图，从而实现决策过程的共同建构，极大地提升了推荐的精准性与用户参与感。

最后，在技术能力飞速进步的同时，负责任的人工智能原则将深刻嵌入推荐系统的设计基因。这具体体现在三个层面：在价值权衡上，系统需在统一的多任务学习框架下，更加精巧地平衡用户体验、商业价值与社会效益，例如在追求点击率的同时，主动管理信息茧房效应与平台生态健康；在数据利用上，跨域推荐与联邦学习等技术将在保护用户隐私和数据安全的前提下，突破数据孤岛，实现知识的安全流动与协同增益；在系统透明度上，可解释性将不再是一个事后附加功能，而是模型内在的设计要求，通过生成自然语言解释和可视化决策路径，构建用户与算法之间的信任桥梁。最终，一个成功的推荐系统，将是技术卓越性与社会责任感的高度统一。

7. 结论

推荐系统是电子商务生态系统中不可或缺的智能组件。从早期的协同过滤到如今基于深度学习的复杂模型，推荐技术不断演进，以应对日益复杂的业务场景和用户需求。尽管在冷启动、数据稀疏性、可解释性等方面仍面临挑战，但随着深度学习、自然语言处理、隐私计算等技术的进步，未来的电子商务推荐系统将更加智能、精准、个性化且可信。它不仅是一个技术工具，更是连接用户与商品、提升商业效率与用户体验的核心桥梁。

基金项目

上海市大学生创新项目(SH2025080)。

参考文献

- [1] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J. (1994) GroupLens: An Open Architecture for Collaborative Filtering of Netnews. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work—CSCW'94*, Chapel Hill, 22-26 October 1994, 175-186. <https://doi.org/10.1145/192844.192905>
- [2] 张艳菊, 余冰冰. 考虑情境异质图神经网络的跨境电商推荐模型研究[J/OL]. 计算机工程与应用: 1-13. <https://link.cnki.net/urlid/11.2127.tp.20251028.1716.016>, 2025-11-03.
- [3] Ricci, F., Rokach, L. and Shapira, B. (2010) Introduction to Recommender Systems Handbook. In: Ricci, F., Rokach, L., Shapira, B. and Kantor, P.B., Eds., *Recommender Systems Handbook*, Springer, 1-35.
- [4] Adomavicius, G. and Tuzhilin, A. (2005) Toward the Next Generation of Recommender Systems: A Survey of the State-Of-The-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, **17**, 734-749. <https://doi.org/10.1109/tkde.2005.99>
- [5] Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., et al. (2012) Context-Aware Recom-

- mender Systems for Learning: A Survey and Future Challenges. *IEEE Transactions on Learning Technologies*, **5**, 318-335. <https://doi.org/10.1109/tlt.2012.11>
- [6] Jannach, D., Zanker, M., Felfernig, A. and Friedrich, G. (2010) Recommender Systems: An Introduction. Cambridge University Press. <https://doi.org/10.1017/cbo9780511763113>
- [7] Herlocker, J.L., Konstan, J.A., Borchers, A. and Riedl, J. (1999) An Algorithmic Framework for Performing Collaborative Filtering. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, 15-19 August 1999, 230-237. <https://doi.org/10.1145/312624.312682>
- [8] Linden, G., Smith, B. and York, J. (2003) Amazon.com Recommendations: Item-To-Item Collaborative Filtering. *IEEE Internet Computing*, **7**, 76-80. <https://doi.org/10.1109/mic.2003.1167344>
- [9] Schein, A.I., Popescul, A., Ungar, L.H. and Pennock, D.M. (2002) Methods and Metrics for Cold-Start Recommendations. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval—SIGIR'02*, Tampere, 11-15 August 2002, 253-260. <https://doi.org/10.1145/564418.564421>
- [10] Goldberg, D., Nichols, D., Oki, B.M. and Terry, D. (1992) Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, **35**, 61-70. <https://doi.org/10.1145/138859.138867>
- [11] Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. (2001) Item-Based Collaborative Filtering Recommendation Algorithms. *Proceedings of the 10th International Conference on World Wide Web*, Hong Kong SAR, 1-5 May 2001, 285-295. <https://doi.org/10.1145/371920.372071>
- [12] Koren, Y., Bell, R. and Volinsky, C. (2009) Matrix Factorization Techniques for Recommender Systems. *Computer*, **42**, 30-37. <https://doi.org/10.1109/mc.2009.263>
- [13] Melville, P., Mooney, R.J. and Nagarajan, R. (2002) Content-Boosted Collaborative Filtering for Improved Recommendations. *Proceedings of the Eighteenth National Conference on Artificial Intelligence*, Edmonton, July 2002, 187-192.
- [14] He, X., Liao, L., Zhang, H., Nie, L., Hu, X. and Chua, T. (2017) Neural Collaborative Filtering. *Proceedings of the 26th International Conference on World Wide Web*, Perth, 3-7 April 2017, 173-182. <https://doi.org/10.1145/3038912.3052569>
- [15] Wang, X., He, X., Wang, M., Feng, F. and Chua, T. (2019) Neural Graph Collaborative Filtering. *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Paris, 21-25 July 2019, 165-174. <https://doi.org/10.1145/3331184.3331267>
- [16] Pazzani, M.J. and Billsus, D. (2007) Content-Based Recommendation Systems. In: Brusilovsky, P., Kobsa, A. and Nejdl, W., Eds., *The Adaptive Web*, Springer, 325-341.
- [17] Mooney, R.J. and Roy, L. (2000) Content-Based Book Recommending Using Learning for Text Categorization. *Proceedings of the Fifth ACM Conference on Digital Libraries*, Denver, 7-11 June 2005, 195-204. <https://doi.org/10.1145/336597.336662>
- [18] Pazzani, M. and Billsus, D. (1997) Learning and Revising User Profiles: The Identification of Interesting Web Sites. *Machine Learning*, **27**, 313-331. <https://doi.org/10.1023/a:1007369909943>
- [19] Ricci, F. and Werthner, H. (2001) Case Base Querying for Travel Planning Recommendation. *Information Technology & Tourism*, **4**, 215-226.
- [20] Felfernig, A. and Burke, R. (2008) Constraint-Based Recommender Systems: Technologies and Research Issues. *Proceedings of the 10th International Conference on Electronic Commerce*, Innsbruck, 19-22 August 2008, 1-10. <https://doi.org/10.1145/1409540.1409544>
- [21] Zhang, F., Yuan, N.J., Lian, D., Xie, X. and Ma, W. (2016) Collaborative Knowledge Base Embedding for Recommender Systems. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 353-362. <https://doi.org/10.1145/2939672.2939673>
- [22] Cheng, H., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., et al. (2016) Wide & Deep Learning for Recommender Systems. *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, Boston, 15 September 2016, 7-10. <https://doi.org/10.1145/2988450.2988454>
- [23] Guo, H., Tang, R., Ye, Y., Li, Z. and He, X. (2017) DeepFM: A Factorization-Machine Based Neural Network for CTR Prediction. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, Melbourne, 19-25 August 2017, 1725-1731. <https://doi.org/10.24963/ijcai.2017/239>
- [24] Wang, R., Fu, B., Fu, G. and Wang, M. (2017) Deep & Cross Network for Ad Click Predictions. *Proceedings of the ADKDD'17*, Halifax, 14 August 2017, 1-7. <https://doi.org/10.1145/3124749.3124754>
- [25] Lian, J., Zhou, X., Zhang, F., Chen, Z., Xie, X. and Sun, G. (2018) xDeepFM: Combining Explicit and Implicit Feature Interactions for Recommender Systems. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 1754-1763. <https://doi.org/10.1145/3219819.3220023>
- [26] Zhou, G., Zhu, X., Song, C., Fan, Y., Zhu, H., Ma, X., et al. (2018) Deep Interest Network for Click-Through Rate Prediction. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*,

- London, 19-23 August 2018, 1059-1068. <https://doi.org/10.1145/3219819.3219823>
- [27] Zhou, G., Mou, N., Fan, Y., Pi, Q., Bian, W., Zhou, C., et al. (2019) Deep Interest Evolution Network for Click-Through Rate Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 5941-5948. <https://doi.org/10.1609/aaai.v33i01.33015941>
- [28] Feng, Y., Lv, F., Shen, W., Wang, M., Sun, F., Zhu, Y., et al. (2019) Deep Session Interest Network for Click-Through Rate Prediction. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, Macao SAR, 10-16 August 2019, 2301-2307. <https://doi.org/10.24963/ijcai.2019/319>
- [29] Ma, X., Zhao, L., Huang, G., Wang, Z., Hu, Z., Zhu, X., et al. (2018) Entire Space Multi-Task Model: An Effective Approach for Estimating Post-Click Conversion Rate. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, Ann Arbor, 8-12 July 2018, 1137-1140. <https://doi.org/10.1145/3209978.3210104>
- [30] Ma, J., Zhao, Z., Yi, X., Chen, J., Hong, L. and Chi, E.H. (2018) Modeling Task Relationships in Multi-Task Learning with Multi-Gate Mixture-Of-Experts. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, 19-23 August 2018, 1930-1939. <https://doi.org/10.1145/3219819.3220007>
- [31] Tang, H., Liu, J., Zhao, M. and Gong, X. (2020) Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations. *Fourteenth ACM Conference on Recommender Systems*, 22-26 September 2020, 269-278. <https://doi.org/10.1145/3383313.3412236>
- [32] Huang, P., He, X., Gao, J., Deng, L., Acero, A. and Heck, L. (2013) Learning Deep Structured Semantic Models for Web Search Using Clickthrough Data. *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, San Francisco, 27 October-1 November 2013, 2333-2338. <https://doi.org/10.1145/2505515.2505665>
- [33] Zhou, C., Bai, J., Song, J., Liu, X., Zhao, Z., Chen, X., et al. (2018) ATRank: An Attention-Based User Behavior Modeling Framework for Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **32**, 4564-4571. <https://doi.org/10.1609/aaai.v32i1.11618>
- [34] Ma, J., Zhao, Z., Chen, J., Li, A., Hong, L. and Chi, E.H. (2019) SNR: Sub-Network Routing for Flexible Parameter Sharing in Multi-Task Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **33**, 216-223. <https://doi.org/10.1609/aaai.v33i01.3301216>