



ARTICLE

基于深度学习的视频推荐研究综述

王东培^{1,*}¹北京工商大学, 计算机与人工智能学院, 北京 100048

学术编辑: 马慧璿; 收稿日期: 2025-02-26; 录用日期: 2025-03-24; 发布日期: 2025-03-29

*通讯作者: ✉王东培, 2320742208@qq.com

摘要

随着视频平台的迅速发展, 如何精准地推荐用户感兴趣的内容成为研究热点。基于深度学习的推荐算法因其强大的特征表示能力, 逐渐成为视频推荐系统中的核心技术。本文综述了基于深度学习的视频推荐研究进展, 阐述了深度学习的基本概念及其在视频推荐领域应用的背景, 对不同视频推荐中的深度学习模型进行分析与比较。在此基础上, 展望了深度学习在视频推荐中的发展趋势, 尤其是在提升推荐精度和实时性方面的应用潜力。

关键词: 视频推荐, 深度学习, 神经网络

A review of video recommendation research based on deep learning

Dongpei Wang^{1,*}¹School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China

Academic Editor: Huijun Ma; Submitted: 2025-02-26; Accepted: 2025-03-24; Published: 2025-03-29

*Correspondence Author: ✉Dongpei Wang 2320742208@qq.com

Abstract

With the rapid development of video platforms, how to accurately recommend the content that users are interested in has become a research hotspot. Deep learning-based recommendation algorithms have gradually become the core technology in video recommendation systems due to their powerful feature representation capability. This paper reviews the research progress of video recommendation based on deep learning, describes the basic concept of deep learning and its background of application in the field of video recommendation, and analyzes and compares the deep learning models in different video recommendations. On this basis, it looks forward to the development trend of deep learning in video recommendation, especially the application potential in improving recommendation accuracy and real-time performance.

Keywords: video recommendations, deep learning, neural networks

文章引用

王东培. 基于深度学习的视频推荐研究综述. 自然科学学报, 2025, 2(1): 19–31.

Citation

Dongpei Wang (2025). A review of video recommendation research based on deep learning. Chinese Journal of Natural Science, 2(1), 19–31.

© 2025 ASP (Advancing Science Press Limited). Personal use is permitted, but republication/redistribution requires ASP permission.

1 引言

随着信息技术的迅猛发展，不同类型的视频如短视频、电影、电视剧等迅速发展，成为全球用户日常娱乐、社交和学习的重要组成部分。视频内容的爆发性增长导致平台用户兴趣呈现高度多样化，如何在海量数据中为用户精准推荐相关内容，已成为视频平台面临的核心挑战。尽管传统的推荐算法，如协同过滤和基于内容的推荐方法，在一定程度上能够提供个性化推荐，但它们在处理大规模数据、捕捉用户兴趣动态变化以及提供多样化推荐等方面存在显著局限性 [1, 2]。因此，基于深度学习的推荐技术逐渐成为解决这些问题的重要途径。

深度学习是一种模拟人脑神经网络结构的机器学习方法，通过多层神经网络自动从大规模数据中学习特征并进行有效表示 [3]。卷积神经网络（CNN）、循环神经网络（RNN）和图神经网络等，能够通过端到端的训练，从复杂的非结构化数据中提取出高级特征 [4]。这种强大的特征学习能力，使深度学习在图像、语音、文本等领域取得了显著成果，并展现出在推荐系统中的巨大潜力 [5]。

在视频推荐领域，深度学习的引入使得推荐系统能够从多模态数据中提取更深层次的特征，结合视频内容、用户行为和社交网络等信息，提供更加精准和个性化的推荐 [6]。与传统推荐算法依赖浅层特征或显式行为数据不同，深度学习能够通过深层网络结构，综合考虑用户的历史行为、视频内容特征以及用户与平台的交互模式，从而显著提升推荐的准确性和实时性 [7]。此外，深度学习在处理多样化推荐需求、解决冷启动问题以及增强推荐系统鲁棒性等方面也展现了独特优势。

本文旨在综述基于深度学习的视频推荐研究，回顾近年来该领域的主要研究进展。重点分析当前主流的深度学习模型及其在视频推荐中的具体应用，探讨各类模型在不同推荐场景下的优缺点。本文将对对比不同方法的效果，并评估其在实际应用中的可行性与挑战。最后，本文将展望深度学习在视频推荐领域的未来发展趋势，分析其潜在的研究方向和应用前景 [91, 92]。

2 传统推荐模型

传统推荐模型可以分为基于协同过滤的推荐（Collaborative Filtering-based Recommendation）、基于内容的推荐（Content-based Recommendation）以及混合推荐（Hybrid Recommendation） [8]。协同过滤推荐通过分析用户行为的相似性来生成个性化的推荐结果；基于内容的推荐侧重于视频内容特征与用户兴趣的匹配；混合推荐则通过结合不同推荐策略，旨在提高推荐的精准度与多样性。采用这些传统推荐方法，不仅有效改善了视频推荐的质量，也为后续基于深度学习的推荐技术发展提供理论依据与参考。

协同过滤推荐模型是视频推荐中最为常见的方法，最早于上世纪 90 年代 PaloAlto 研究中心开发的系统——Tapestry [9] 中首次引入。然而，协同过滤方法在短视频推荐中仍然面临数据稀疏和冷启动问题，尤其是当用户或视频数据稀缺时，推荐效果可能受到显著影响。Kim 等 [10] 提出了一种基于协同标签的协同过滤方法，

针对稀疏数据和冷启动用户问题。该方法通过用户生成的标签捕捉用户偏好,利用协同标签系统中的标签频率和用户相似度构建候选标签集(CTS),通过朴素贝叶斯分类器为用户推荐项目。Fidel等[11]比较多种协同过滤算法,评估其在不同数据稀疏度条件下的推荐效果,将协同过滤作为推荐系统的核心方法,依赖用户与视频的交互数据,通过识别相似用户或相似视频实现个性化推荐。Bobadilla等[12]提出一种改进方法,将协同过滤扩展至用户群体,并将推荐范围限制在特定的视频集合内。该方法在满足用户群体观看偏好方面具有更高的精确度,能够更有效地应对大规模数据的推荐需求。Choi等[13]提出了一种基于用户面部表情变化的在线视频推荐方法,通过实时捕捉用户的情感反应来动态识别其偏好,从而解决了传统推荐系统中的新用户问题。这一方法不依赖历史评分或购买记录,为推荐提供了新的思路。Yang等[14]针对群体推荐和参考项目推荐中的不足,通过结合群体用户的联合推荐和与参考项目相似的推荐,设计了新的模型架构来优化推荐结果。该方法在提升推荐质量的同时,成功解决了传统方法的局限性。Najafabadi等[15]针对协同过滤(CF)在数据稀疏情况下的不足,提出了一种改进方案,借助用户与物品的隐式交互数据有效处理大规模数据集。通过关联规则挖掘与聚类技术,优化了用户兴趣模式的发现与数据维度压缩,从而提高了推荐系统的性能。该方法在精度和召回率上表现出显著优势。协同过滤推荐无需深入理解视频内容,仅依赖用户行为数据即可提供个性化推荐,能够发现潜在的用户兴趣,推荐结果多样性较好。但协同过滤推荐依赖大量用户数据,容易受到数据稀疏和冷启动问题的影响,且计算复杂度较高,难以在大规模应用中保持高效性。

基于内容的推荐模型通过分析短视频的内容特征与用户的历史偏好进行匹配,这种方法不依赖其他用户的数据,因此能有效避免冷启动问题。Yashar等[16]强调了基于内容的视频推荐方法在提升推荐精度和应对冷启动问题中的重要作用。该方法通过分析视频的多模态特征(如视觉、音频、文本),使系统能够在用户交互数据不足的情况下,依然提供个性化推荐。Du等[17]通过引入协同嵌入回归(CER)模型,将用户交互与内容特征结合,提高了推荐的精准度和鲁棒性。研究提出的优先级晚期融合(PRI)方法,能够优化多模态特征的整合,进一步改善推荐效果。基于内容的推荐能够解决新视频推荐的问题,但它也面临内容单一化的挑战,可能导致推荐结果过于集中在用户已观看的内容类型,从而缺乏多样性。此外,该方法对视频的元数据和特征信息依赖较大,而视频平台上的内容复杂且多样,如何从中提取有效特征仍然是一个难题。

为了弥补单一推荐方法的不足,混合推荐方法逐渐被引入到短视频推荐中,能够结合协同过滤和基于内容的推荐各自的优势,从而提升推荐系统的性能。Pazzani等[18]提出的方法主要采用协同过滤算法聚类用户。但保留了用户的文本描述信息,用来补充计算那些共同观看数很低的用户间相关性。Ioannis等[19]提出了一种商业级的混合推荐系统 AMORE,采用多线程架构,利用混合推荐方法通过结合用户观看历史和视频内容特征,能够有效克服冷启动问题,并在数据稀疏的情况下仍能提供准确推荐。Walek等[20]提出了一种结合协同过滤、基于内容的推荐和模糊专家系统的混合方法。该混合推荐方法可以利用协同过滤识别相似用户和内容推荐分析视频特征,同时结合模糊专家系统处理用户偏好和评分的不确定性,提高推荐的精准度和可靠性。Wei等[21]提出了一种基于标签和评分的混合电影推荐方法,结合社交电影网络和偏好主题模型,整合用户的历史评分与社交标签信息,有效提升了推荐系统的精度与鲁棒性。Wen等[22]提出了一种混合推荐方法,通过计算网页术语权重对新闻内容进行分类,并基于用户的浏览历史构建兴趣与偏好模型。结合协同过滤技术,系统依据网页内容和用户偏好,向用户推荐与其兴趣相关的新闻,该方法有效提升了推荐系统的精度与效率。Sahu等[23]提出了一种结合情感分析与混合推荐系统的框架,用于对尚未发布的电影进行个性化推荐。该方法通过分析预告片评论情感并预测电影评分,进而生成用户偏好的推荐列表,显著提升了未来电影推荐的精度。Kaššák等[24]提出了一种混合推荐方法,该方法将内容推荐与协同过滤策略相结合,通过为每个用户生成内容和协同推荐,并应用聚合策略优化群体偏好,最终对协同推荐进行重排序。有效解决了群体推荐中的偏好冲突问题,特别是在 Top-N 推荐项上。Amolochitis等[25]提出了 AMORE 混合推荐系统,融合了基于用户、物品和内容的推荐算法。通过多线程实现的 k-NN 用户与物品算法结合定制化内容推荐方法,有效支持

视频点播服务的日常推荐请求，推动了租赁收入增长及客户保持率提升。混合推荐方法能够有效缓解数据稀疏、冷启动和过度集中推荐的问题，且能够增加推荐内容的多样性，但其计算复杂度较高，在视频平台中如何高效融合多种推荐策略仍然是一个研究挑战。

尽管以协同过滤推荐、基于内容的推荐和混合推荐为代表的传统推荐算法在短视频推荐系统中取得了显著的应用成效。但这些方法仍面临着一些固有的挑战。首先，数据稀疏问题在短视频平台中尤为突出，传统算法往往缺乏足够的用户行为数据，导致推荐的准确性和相关性较低。其次，冷启动问题依然是传统方法的主要瓶颈，往往缺乏足够的历史数据进行有效的预测和推荐。最后，计算复杂度问题也限制了传统算法的可扩展性和实时性，尤其是在面对大规模数据时，诸如协同过滤等算法的效率难以满足短视频平台对实时推荐的高要求，具体如表 1所示。

表 1. 传统推荐方法优缺点及应用场景比较

传统推荐算法	优点	局限性	应用场景
协同过滤推荐	无需内容分析，具备多样性	冷启动问题	用户数据丰富的平台
基于内容的推荐	缓解冷启动，多模态分析	特征依赖大	新视频
混合推荐	兼具协同过滤和基于内容推荐	复杂度高	大规模、多样化需求

随着深度学习技术的快速发展，未来视频推荐系统有望通过神经网络等先进模型，自动从海量数据中提取特征，减少对人工特征工程的依赖，进而有效缓解数据稀疏和冷启动问题。此类基于深度学习的推荐方法不仅有望提升推荐的精准度与效率，还能更好地适应大规模数据的处理需求，为视频平台的个性化推荐服务提供更加高效且智能的解决方案。

3 深度学习推荐模型

近年来，深度学习推荐算法发展迅速，得益于神经网络架构的创新、大规模数据的可用性以及计算能力的提升。深度学习方法，如卷积神经网络（CNN）、循环神经网络与长短期记忆网络（RNN 与 LSTM）、图神经网络（GNN）、自编码器等，已广泛应用于个性化推荐系统，显著提升了推荐准确性和用户满意度。在图 1中展示了历年来的出版物数量。

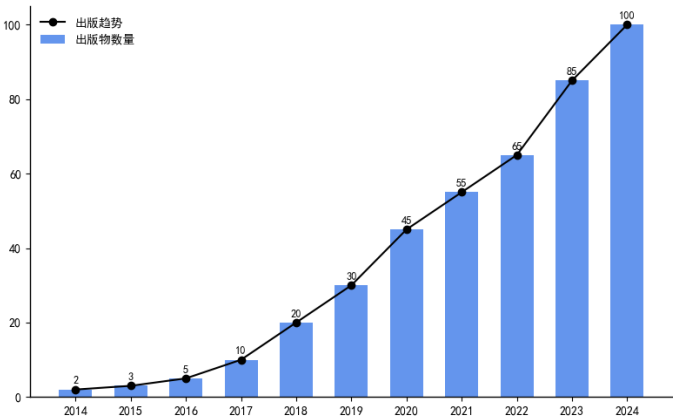


图 1. 近十年出版物数量

3.1 RNN 与 LSTM 循环神经网络

循环神经网络 (Recurrent Neural Networks, RNN) 及长短期记忆网络 (Long Short-Term Memory, LSTM) 因其在处理时序数据方面的卓越能力, 已成为视频推荐系统的核心技术。近年来, 研究人员通过优化网络结构和引入注意力机制, 提升了 RNN 和 LSTM 在视频推荐中的适用性, 使其在个性化推荐、用户行为建模及长时序依赖处理中展现出显著的优势 [26]。RNN 模型结构如图 2 所示。

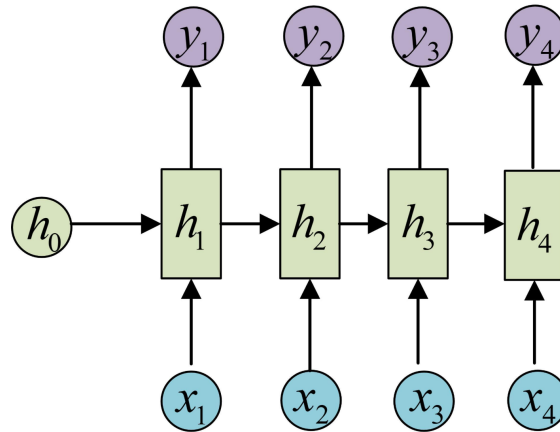


图 2. RNN 模型结构图

Ibrahim 等 [27] 通过整合来自多个外部资源 (如评分、投票和评论) 的多维度数据, 并运用基于 RNN/LSTM 的情感分析模型, 提出了一种优化的电影推荐方法。该方法结合用户电影兴趣注意力机制, 以更精准地捕捉用户偏好, 提升推荐系统的效果。Choe 等 [28] 提出了一种新型推荐系统, 结合了用户历史行为中的时间序列特性, 设计了具有分层结构的递归神经网络 (RNN), 分别处理长短期时间间隔与序列信息。该方法通过考虑时间依赖性, 显著提升了推荐精度。Ali 等 [29] 提出了一种改进的 LSTM 模型 (LSTM-IIMA), 结合了内外部路径分析以增强电影推荐系统的效果。内路径分析聚焦于单一路径内的用户互动, 而外路径分析则捕捉多路径间的复杂依赖关系, 从而提升了模型对用户、电影及其他实体间深层次联系的建模能力。Tripathi 等 [30] 提出情感感知视频推荐引擎 (EmoWare), 融合强化学习与深度双向递归神经网络 (DBRNN), 以建模用户的情感偏好。通过分析用户的非言语情绪反馈, 并结合上下文感知协同过滤方法, 优化推荐策略, 实现视频库的动态更新与个性化推荐。Cao 等 [31] 提出多模态哈希标签推荐模型 (LOGO), 采用并行长短时记忆网络 (LSTM) 建模多模态序列结构, 并引入注意力机制分配单元权重, 将多模态特征映射到公共空间并施加模态相似性约束, 结合神经协同过滤框架提升标签推荐的准确性和泛化能力。Zhou 等 [26] 构建基于聚类与长短时记忆网络 (LSTM) 的个性化学习路径推荐模型, 通过特征相似性度量对学习者的学习行为进行聚类, 并利用 LSTM 预测学习路径与学习表现, 以生成个性化的学习路径推荐方案。

3.2 基于 Transformer 模型

Transformer 模型是一种基于注意力机制的深度学习架构, 主要用于序列到序列的任务。其创新在于引入了自注意力机制 (Self-Attention), 能够在处理输入序列时, 通过计算各位置之间的相似性动态调整信息传递的权重, 从而有效捕捉输入元素之间的长程依赖关系。Transformer 架构无需顺序处理数据, 从而实现了高度的并行化计算, 同时其架构的设计也能够更好地处理长距离依赖问题, 因此自然语言处理任务中表现出色, 成为 BERT 等许多先进模型的基础。Transformer 模型结构如图 3 所示。

Hu 等 [32] 提出了一种基于多模态多视角注意力网络 (CRMMAN) 的协同推荐模型, 该模型对用户历史交互进行正负交互划分, 以分别建模用户的偏好与厌恶特征。此外, 模型融合场景中的语义与结构信息, 以增

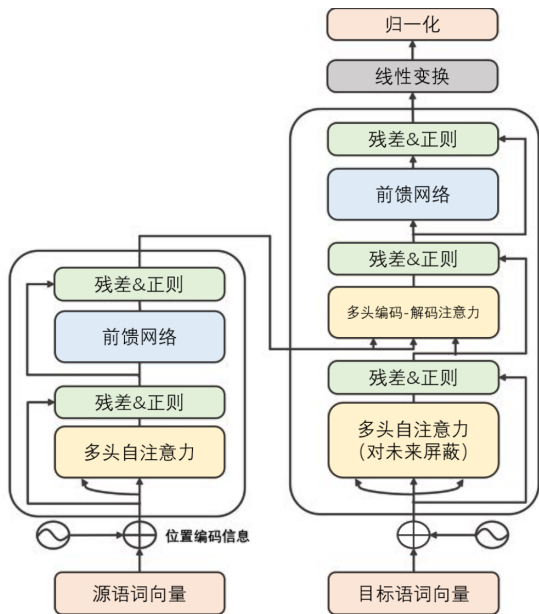


图 3. Transformer 模型结构图

强物品表示能力，从而优化视频推荐性能。Zhang 等 [33] 提出了一种时间感知 Transformer 模型 (TAT4SRec)，融合自注意力机制以提升序列推荐效果。该模型采用编码器-解码器架构，对时间戳与交互物品进行独立建模，并构建差异化嵌入模块以处理连续与离散数据，从而更精确地刻画用户偏好动态变化。Zhu 等 [34] 提出特征交互双重自注意力网络 (FIDS)，利用多头注意力机制建模物品特征间的高阶关系，并通过双自注意力网络分别学习物品序列与特征序列的转换模式。该方法有效融合特征交互信息与序列依赖性，从而提升视频推荐的精准度。Wang 等 [35] 提出一种融合自注意力机制与协同度量学习 (SACML) 的群体推荐模型，自注意力机制用于建模群体成员与服务之间的相似性关系，并动态分配成员权重。结合协同度量学习，该模型构建群体与服务的度量空间，以更精准地聚合个体偏好，优化群体推荐效果。Zhao 等 [36] 提出层次化多模态注意力网络 (HMAN)，利用时间同步评论增强视频推荐。模型采用多级文本压缩模块提取评论语义，结合范围卷积块解析视频内容，再通过层次化多模态融合机制构建完整视频表示，从而提升推荐效果。Huang 等 [37] 提出多尺度建模时间层次注意力方法 (MMTHA)，针对用户历史行为序列进行层次化建模，以优化微视频推荐效果。该方法通过时间窗口机制提取短期动态偏好，结合类别级与物品级注意力机制刻画用户的粗粒度与细粒度兴趣，并利用多头自注意力捕捉长周期偏好关联，以增强用户兴趣建模的精确性。Bai 等 [38] 提出四通道序列推荐模型 (SR-MAG)，融合多级自注意力网络与门控脉冲神经网络 (GSPN)，分别建模用户的长期与短期行为序列。通过自注意力机制捕获序列中的上下文信息，增强用户兴趣表征能力，从而提升视频推荐的精准度。Ma 等 [39] 构建社交关系驱动的动态推荐模型 (SRDPR)，通过挖掘用户公共偏好进行兴趣群体划分，以缓解交互数据稀疏问题。结合时间序列信息与自注意力机制，建模用户长期与短期兴趣演化，从而提升微视频推荐的个性化精准度。Siet 等 [40] 提出了一种结合深度学习的电影推荐系统，旨在解决数据稀疏性和冷启动问题。该方法通过整合用户信息与电影序列嵌入，并采用 Transformer 结构和多头注意力机制，提升推荐精度。此外，利用 KMeans 聚类对电影类型进行分组，进一步增强个性化推荐的多样性。

3.3 GNN 图神经网络

图神经网络 (GNN) 是一种处理图结构数据的神经网络模型，通过在图中传播节点信息来学习节点和图的表示 [41]。其核心思想是通过聚合邻居节点的特征更新每个节点的表示，使得节点能够有效捕捉图的结构信息。GNN 广泛应用于节点分类、图分类等任务，并在处理复杂图数据方面展现了强大的能力。GNN 神经网络

结构如图 4 所示。

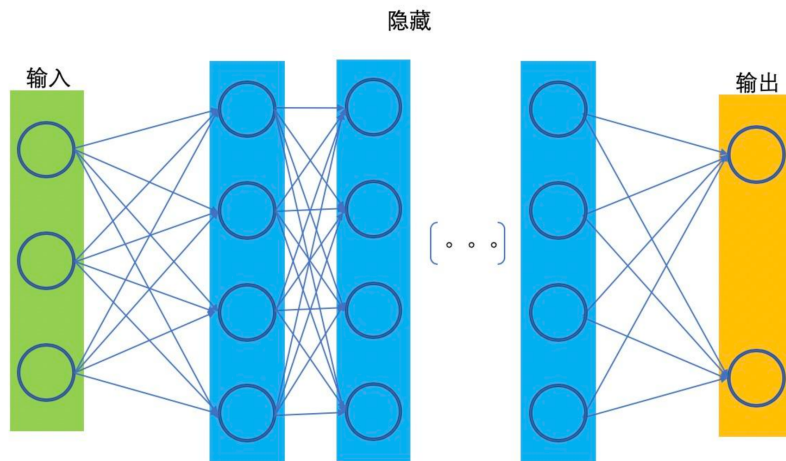


图 4. GNN 神经网络结构图

Wang 等 [42] 提出了一种基于双图神经网络 (DualGNN) 的微视频推荐框架, 通过构建用户-视频二分图和用户共现图, 挖掘用户间的关联性并协同学习每个用户的特定模态融合模式。框架采用单模态与多模态表示学习模块, 分别建模用户在各模态上的偏好, 从而提升推荐效果。Wei 等 [43] 提出层次化用户意图图网络 (HUI-GN) 框架, 通过用户与物品的共交互模式挖掘多层次用户意图。该框架通过层内和层间聚合操作, 从细粒度到粗粒度逐步构建用户意图表示, 从而更精确地刻画用户偏好和物品特征。Cai 等 [44] 提出了一种异构层次特征聚合网络 (HHFAN) 用于个性化微视频推荐, 构建异构信息图以深入挖掘用户、微视频及多模态信息之间的复杂关系。通过层次化特征聚合网络, 框架能够有效提取异构图中的结构特征与语义信息, 进而优化推荐效果。Tao 等 [45] 提出了一种多模态图注意力网络 (MGAT), 旨在通过多模态交互图上应用图神经网络 (GNN), 细致地捕捉用户在不同模态上的偏好差异。该方法通过引入门控注意力机制, 动态调整各模态信息的权重, 从而更精确地刻画用户的兴趣并优化推荐性能。He 等 [46] 提出了一种基于元路径的图对比学习网络 (MPGCL), 该方法通过构建同质图来有效捕捉异质图中的多样性, 并采用对比学习机制增强节点表示的鲁棒性, 以应对噪声干扰。通过最大化不同视角下图表示的一致性, 显著提升了微视频推荐的效果。Li 等 [47] 提出了一种创新的多模态推荐框架 (GNNMR), 该框架结合了图神经网络和深度互学习技术, 旨在解决不同模态间固有偏差的问题。通过在单模态用户-物品图上独立训练 GNN, 并采用互知识蒸馏机制, 模型能够同步不同模态的潜在语义关系, 从而提升推荐效果。Wang 等 [48] 提出超边图神经网络 (HGNN), 通过超图结构建模学习者之间的关联, 并引入超边注意力机制以优化学习者表征。该方法构建跨学习者的课程序列图, 融合长短期序列关系, 并采用 GRU 编码短期模式, 从而增强视频推荐的精准性。Yan 等 [49] 提出特征交互图神经网络 (FIKGRec), 构建基于知识图谱的物品图, 以显式建模异构结构中的特征交互关系, 从而优化信息传播与聚合过程。引入偏好感知注意力机制获取用户精细化偏好, 并利用深度神经网络挖掘用户与物品间的复杂关联, 以提升推荐性能。Darban 等 [50] 提出基于图模型的混合推荐方法 (GHRs), 融合用户评分相似性、人口统计特征及地理位置信息, 以构建更精细化的推荐框架。通过自编码器提取综合特征, 生成高维用户表示, 并结合聚类方法优化用户分组, 从而提升推荐精度及冷启动问题的解决能力。

3.4 其他深度学习推荐算法

在前文中, 本文系统地阐述了四类具有代表性的深度学习推荐算法。除此之外, 仍有众多其他深度学习模型被广泛应用于视频推荐系统的研究领域。这些方法涵盖了多种神经网络结构及其变体, 进一步丰富和拓展了

视频推荐系统的研究范式。鉴于篇幅所限，本文将在本节对现有的其他深度学习推荐算法进行归纳性综述，具体算法及其相关参考文献详见表 2。

表 2. 其他深度学习模型文献汇总表

分类	出版物
深度神经网络	[51],[52],[53],[54],[55],[56],[57],[58],[59],[60]
卷积神经网络	[61],[62],[63],[64],[65],[66],[67],[68],[69],[70]
自编码器	[71],[72],[73],[74],[75],[76],[77],[78],[79],[80]
受限玻尔兹曼机	[81],[82],[83],[84],[85],[86],[87],[88],[89],[90]

4 对比分析

递归神经网络（RNN）、长短时记忆网络（LSTM）、变换器（Transformer）和图神经网络（GNN）是深度学习中常用的模型，它们在时序数据、多模态数据等任务上的表现各具优势。总体来看，RNN 和 LSTM 适用于时序数据建模，Transformer 具有更强的长距离依赖建模能力，GNN 则在结构化数据的时间建模中表现突出。在计算效率、多模态数据处理能力和长序列建模等方面，这些模型各有优缺点，适用于不同的应用场景。

具体而言，在时序数据处理中，RNN 由于其循环结构，能够捕捉短期依赖关系，但梯度消失问题限制了其对长时依赖的建模能力。LSTM 通过引入门控机制，缓解了这一问题，使其在时间序列预测、语音识别等任务中表现更佳。相比之下，Transformer 采用自注意力机制，能有效建模长距离依赖，且计算并行度更高，但需要额外的位置编码来保留时间信息。而 GNN 主要处理图结构数据，在时序任务中通常需要结合时间卷积或递归单元。对于多模态数据，Transformer 由于其注意力机制，能更好地融合不同模态信息，如文本-图像对齐等，而 RNN 和 LSTM 计算顺序性较强，难以高效整合多模态数据。GNN 在结构化多模态任务（如知识图谱）中表现较好，但对非结构化数据的适应性较弱。在计算效率方面，Transformer 并行性强，适合大规模数据，而 RNN 和 LSTM 受限于串行计算，GNN 在大规模图数据上则面临较高计算成本。因此，Transformer 在多模态数据上优势明显，LSTM 仍是强大的时序建模工具，而 GNN 适用于结构化数据的时序任务。

5 未来展望

基于深度学习的视频推荐系统在未来具有广泛的应用前景，尤其随着技术的不断进步和数据量的快速增长。一方面，深度学习能够有效挖掘视频内容的复杂特征，如视觉、语音和文本信息，帮助系统更加精准地理解用户需求。近年来，循环神经网络（RNN）、Transformer 模型以及图神经网络（GNN）等深度学习技术的应用，使得视频推荐系统能够处理更为复杂的输入数据，提升推荐的个性化和准确性。另一方面，基于深度学习的视频推荐系统有望进一步提升用户体验。例如，通过多模态学习，系统能够结合用户的观看历史、评论、点赞等行为数据，更加全面地了解用户偏好，并提供更具针对性的推荐内容。随着实时数据流的增加，深度学习技术也为动态推荐策略提供了新的可能，能够在用户行为发生变化时，快速调整推荐结果，增强系统的自适应能力。

未来，随着大数据和计算能力的不断提升，深度学习模型有望实现更加精准的实时视频推荐，进一步拓宽其在个性化广告、电商平台、内容创作和娱乐领域的应用范围。此外，视频推荐系统的透明度和可解释性问题也将成为研究的重点，如何提升模型的可理解性，进而优化用户信任，是未来的重要挑战之一。

参考文献

- [1] Mu, Y., & Wu, Y. (2023). Multimodal movie recommendation system using deep learning. *Mathematics*, 11(4), 895. [CrossRef]
- [2] Xia, H., Li, J. J., & Liu, Y. (2020). Collaborative filtering recommendation algorithm based on attention GRU and adversarial learning. *IEEE Access*, 8, 208149–208157. [CrossRef]
- [3] Hassan, A., & Mahmood, A. (2018). Convolutional recurrent deep learning model for sentence classification. *IEEE Access*, 6, 13949–13957. [CrossRef]
- [4] Ye, W., Askarisichani, O., Jones, A., & Singh, A. (2020). Learning deep graph representations via convolutional neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 34(5), 2268–2279. [CrossRef]
- [5] Oprea, S., Martinez-Gonzalez, P., Garcia-Garcia, A., Castro-Vargas, J. A., Orts-Escolano, S., Garcia-Rodriguez, J., & Argyros, A. (2020). A review on deep learning techniques for video prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6), 2806–2826. [CrossRef]
- [6] Lei, F., Cao, Z., Yang, Y., Ding, Y., & Zhang, C. (2023). Learning the user's deeper preferences for multi-modal recommendation systems. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(3s), 1–18. [CrossRef]
- [7] Steck, H., Baltrunas, L., Elahi, E., Liang, D., Raimond, Y., & Basilico, J. (2021). Deep learning for recommender systems: A Netflix case study. *AI magazine*, 42(3), 7–18. [CrossRef]
- [8] Li, Y., Lu, L., & Xuefeng, L. (2005). A hybrid collaborative filtering method for multiple-interests and multiple-content recommendation in E-Commerce. *Expert systems with applications*, 28(1), 67–77. [CrossRef]
- [9] David, G. (1992). Using Collaborative Filtering to Weave an Information Tapestry. *Communications of the ACM*, 35(12), 61–70. [CrossRef]
- [10] Kim, H. N., Ji, A. T., Ha, I., & Jo, G. S. (2010). Collaborative filtering based on collaborative tagging for enhancing the quality of recommendation. *Electronic Commerce Research and Applications*, 9(1), 73–83. [CrossRef]
- [11] Cacheda, F., Carneiro, V., Fernández, D., & Formoso, V. (2011). Comparison of collaborative filtering algorithms: Limitations of current techniques and proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web (TWEB)*, 5(1), 1–33. [CrossRef]
- [12] Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J. (2012). Generalization of recommender systems: Collaborative filtering extended to groups of users and restricted to groups of items. *Expert systems with Applications*, 39(1), 172–186. [CrossRef]
- [13] Choi, I. Y., Oh, M. G., Kim, J. K., & Ryu, Y. U. (2016). Collaborative filtering with facial expressions for online video recommendation. *International Journal of Information Management*, 36(3), 397–402. [CrossRef]
- [14] Yang, Z., Wu, B., Zheng, K., Wang, X., & Lei, L. (2016). A survey of collaborative filtering-based recommender systems for mobile internet applications. *IEEE Access*, 4, 3273–3287. [CrossRef]
- [15] Najafabadi, M. K., Mahrin, M. N. R., Chuprat, S., & Sarkan, H. M. (2017). Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data. *Computers in Human Behavior*, 67, 113–128. [CrossRef]
- [16] Deldjoo, Y., Schedl, M., Cremonesi, P., & Pasi, G. (2020). Recommender systems leveraging multimedia content. *ACM Computing Surveys (CSUR)*, 53(5), 1–38. [CrossRef]
- [17] Du, X., Yin, H., Chen, L., Wang, Y., Yang, Y., & Zhou, X. (2018). Personalized video recommendation using rich contents from videos. *IEEE Transactions on Knowledge and Data Engineering*, 32(3), 492–505. [CrossRef]
- [18] Pazzani, M. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, 13, 393–408. [CrossRef]

- [19] Barragáns-Martínez, A. B., Costa-Montenegro, E., Burguillo, J. C., Rey-López, M., Mikic-Fonte, F. A., & Peleteiro, A. (2010). A hybrid content-based and item-based collaborative filtering approach to recommend TV programs enhanced with singular value decomposition. *Information Sciences*, 180(22), 4290-4311. [[CrossRef](#)]
- [20] Walek, B., & Fojtik, V. (2020). A hybrid recommender system for recommending relevant movies using an expert system. *Expert Systems with Applications*, 158, 113452. [[CrossRef](#)]
- [21] Wei, S., Zheng, X., Chen, D., & Chen, C. (2016). A hybrid approach for movie recommendation via tags and ratings. *Electronic Commerce Research and Applications*, 18, 83-94. [[CrossRef](#)]
- [22] Wen, H., Fang, L., & Guan, L. (2012). A hybrid approach for personalized recommendation of news on the Web. *Expert Systems with Applications*, 39(5), 5806-5814. [[CrossRef](#)]
- [23] Sahu, S., Kumar, R., MohdShafi, P., Shafi, J., Kim, S., & Ijaz, M. F. (2022). A hybrid recommendation system of upcoming movies using sentiment analysis of youtube trailer reviews. *Mathematics*, 10(9), 1568. [[CrossRef](#)]
- [24] Kaššák, O., Kompan, M., & Bielíková, M. (2016). Personalized hybrid recommendation for group of users: Top-N multimedia recommender. *Information Processing & Management*, 52(3), 459-477. [[CrossRef](#)]
- [25] Amolochitis, E., Christou, I. T., & Tan, Z. H. (2014). Implementing a commercial-strength parallel hybrid movie recommendation engine. *IEEE Intelligent Systems*, 29(2), 92-96. [[CrossRef](#)]
- [26] Zhou, Y., Huang, C., Hu, Q., Zhu, J., & Tang, Y. (2018). Personalized learning full-path recommendation model based on LSTM neural networks. *Information sciences*, 444, 135-152. [[CrossRef](#)]
- [27] Ibrahim, M., Bajwa, I. S., Ul-Amin, R., & Kasi, B. (2019). A Neural Network-Inspired Approach for Improved and True Movie Recommendations. *Computational intelligence and neuroscience*, 2019(1), 4589060. [[CrossRef](#)]
- [28] Choe, B., Kang, T., & Jung, K. (2021). Recommendation system with hierarchical recurrent neural network for long-term time series. *IEEE Access*, 9, 72033-72039. [[CrossRef](#)]
- [29] Ali, Z., Muhammad, A., Al-Shamayleh, A. S., Qureshi, K. N., Alrawagfeh, W., & Akhunzada, A. (2023). Enhancing performance of movie recommendations using LSTM with meta path analysis. *IEEE Access*, 11, 119017-119032. [[CrossRef](#)]
- [30] Tripathi, A., Ashwin, T. S., & Guddeti, R. M. R. (2019). EmoWare: A context-aware framework for personalized video recommendation using affective video sequences. *IEEE Access*, 7, 51185-51200. [[CrossRef](#)]
- [31] Cao, D., Miao, L., Rong, H., Qin, Z., & Nie, L. (2020). Hashtag our stories: Hashtag recommendation for micro-videos via harnessing multiple modalities. *Knowledge-Based Systems*, 203, 106114. [[CrossRef](#)]
- [32] Cao, D., Miao, L., Rong, H., Qin, Z., & Nie, L. (2020). Hashtag our stories: Hashtag recommendation for micro-videos via harnessing multiple modalities. *Knowledge-Based Systems*, 203, 106114. [[CrossRef](#)]
- [33] Zhang, Y., Yang, B., Liu, H., & Li, D. (2023). A time-aware self-attention based neural network model for sequential recommendation. *Applied Soft Computing*, 133, 109894. [[CrossRef](#)]
- [34] Zhu, Y., Yao, S., & Sun, X. (2024). Feature Interaction Dual Self-attention network for sequential recommendation. *Frontiers in Neurorobotics*, 18, 1456192. [[CrossRef](#)]
- [35] Wang, H., Li, Y., & Frimpong, F. (2019). Group recommendation via self-attention and collaborative metric learning model. *IEEE Access*, 7, 164844-164855. [[CrossRef](#)]
- [36] Zhao, W., Wu, H., He, W., Bi, H., Wang, H., Zhu, C., ... & Chen, E. (2023). Hierarchical multi-modal attention network for time-sync comment video recommendation. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(4), 2694-2705. [[CrossRef](#)]
- [37] Huang, N., Hu, R., Wang, X., & Ding, H. (2023). Multi-scale modeling temporal hierarchical attention for sequential recommendation. *Information Sciences*, 641, 119126. [[CrossRef](#)]
- [38] Bai, X., Huang, Y., Peng, H., Wang, J., Yang, Q., Orellana-Martín, D., ... & Pérez-Jiménez, M. J. (2024). Sequence

- recommendation using multi-level self-attention network with gated spiking neural P systems. *Information Sciences*, 656, 119916. [[CrossRef](#)]
- [39] Ma, J., Bian, K., Wen, J., Xu, Y., Zhong, M., & Zhu, L. (2023). SRDPR: Social Relation-driven Dynamic network for Personalized micro-video Recommendation. *Expert Systems with Applications*, 226, 120157. [[CrossRef](#)]
- [40] Siet, S., Peng, S., Ilkhomjon, S., Kang, M., & Park, D. S. (2024). Enhancing sequence movie recommendation system using deep learning and kmeans. *Applied sciences*, 14(6), 2505. [[CrossRef](#)]
- [41] Kumar, S., Mallik, A., Khetarpal, A., & Panda, B. S. (2022). Influence maximization in social networks using graph embedding and graph neural network. *Information Sciences*, 607, 1617-1636. [[CrossRef](#)]
- [42] Wang, Q., Wei, Y., Yin, J., Wu, J., Song, X., & Nie, L. (2021). Dualgnn: Dual graph neural network for multimedia recommendation. *IEEE Transactions on Multimedia*, 25, 1074-1084. [[CrossRef](#)]
- [43] Wei, Y., Wang, X., He, X., Nie, L., Rui, Y., & Chua, T. S. (2021). Hierarchical user intent graph network for multimedia recommendation. *IEEE Transactions on Multimedia*, 24, 2701-2712. [[CrossRef](#)]
- [44] Cai, D., Qian, S., Fang, Q., & Xu, C. (2021). Heterogeneous hierarchical feature aggregation network for personalized micro-video recommendation. *IEEE Transactions on Multimedia*, 24, 805-818. [[CrossRef](#)]
- [45] Tao, Z., Wei, Y., Wang, X., He, X., Huang, X., & Chua, T. S. (2020). Mgat: Multimodal graph attention network for recommendation. *Information Processing & Management*, 57(5), 102277. [[CrossRef](#)]
- [46] He, Y., Wu, G., Cai, D., & Hu, X. (2023). Meta-path based graph contrastive learning for micro-video recommendation. *Expert Systems with Applications*, 222, 119713. [[CrossRef](#)]
- [47] Li, J., Yang, C., Ye, G., & Nguyen, Q. V. H. (2024). Graph neural networks with deep mutual learning for designing multi-modal recommendation systems. *Information Sciences*, 654, 119815. [[CrossRef](#)]
- [48] Wang, X., Ma, W., Guo, L., Jiang, H., Liu, F., & Xu, C. (2022). HGNN: Hyperedge-based graph neural network for MOOC course recommendation. *Information Processing & Management*, 59(3), 102938. [[CrossRef](#)]
- [49] Yan, S., Li, C., Wang, H., Lin, B., & Yuan, Y. (2024). Feature interactive graph neural network for KG-based recommendation. *Expert Systems with Applications*, 237, 121411. [[CrossRef](#)]
- [50] Darban, Z. Z., & Valipour, M. H. (2022). GHRS: Graph-based hybrid recommendation system with application to movie recommendation. *Expert Systems with Applications*, 200, 116850. [[CrossRef](#)]
- [51] Jain, G., Mahara, T., Sharma, S. C., Agarwal, S., & Kim, H. (2022). TD-DNN: A time decay-based deep neural network for recommendation system. *Applied Sciences*, 12(13), 6398. [[CrossRef](#)]
- [52] Vukotić, V., Raymond, C., & Gravier, G. (2018). A crossmodal approach to multimodal fusion in video hyperlinking. *IEEE MultiMedia*, 25(2), 11-23. [[CrossRef](#)]
- [53] Tung, T. Y., & Gündüz, D. (2022). DeepWiVe: Deep-learning-aided wireless video transmission. *IEEE Journal on Selected Areas in Communications*, 40(9), 2570-2583. [[CrossRef](#)]
- [54] Gong, J., Zhang, X., Li, Q., Wang, C., Song, Y., Zhao, Z., & Wang, S. (2021). A top-N movie recommendation framework based on deep neural network with heterogeneous modeling. *Applied Sciences*, 11(16), 7418. [[CrossRef](#)]
- [55] Lee, H. C., Kim, Y. S., & Kim, S. W. (2024). Real-time movie recommendation: Integrating persona-based user modeling with NMF and deep neural networks. *Applied Sciences*, 14(3), 1014. [[CrossRef](#)]
- [56] Zhang, W., Du, Y., Yoshida, T., & Yang, Y. (2019). DeepRec: A deep neural network approach to recommendation with item embedding and weighted loss function. *Information sciences*, 470, 121-140. [[CrossRef](#)]
- [57] Feng, X., Liu, Z., Wu, W., & Zuo, W. (2022). Social recommendation via deep neural network-based multi-task learning. *Expert Systems with Applications*, 206, 117755. [[CrossRef](#)]
- [58] Zuo, Y., Zeng, J., Gong, M., & Jiao, L. (2016). Tag-aware recommender systems based on deep neural networks.

- Neurocomputing*, 204, 51-60. [CrossRef]
- [59] Wu, C., Wang, J., Liu, J., & Liu, W. (2016). Recurrent neural network based recommendation for time heterogeneous feedback. *Knowledge-Based Systems*, 109, 90-103. [CrossRef]
- [60] Wang, X., Tan, Q., & Zhang, L. (2020). A deep neural network of multi-form alliances for personalized recommendations. *Information Sciences*, 531, 68-86. [CrossRef]
- [61] Suresha, M., Kuppa, S., & Raghukumar, D. S. (2020). A study on deep learning spatiotemporal models and feature extraction techniques for video understanding. *International Journal of Multimedia Information Retrieval*, 9(2), 81-101. [CrossRef]
- [62] Dudekula, K. V., Syed, H., Basha, M. I. M., Swamykan, S. I., Kasaraneni, P. P., Kumar, Y. V. P., ... & Azar, A. T. (2023). Convolutional neural network-based personalized program recommendation system for smart television users. *Sustainability*, 15(3), 2206. [CrossRef]
- [63] Wang, Z., Chen, H., Li, Z., Lin, K., Jiang, N., & Xia, F. (2021). VRConvMF: Visual recurrent convolutional matrix factorization for movie recommendation. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(3), 519-529. [CrossRef]
- [64] Yu, S., Guo, M., Chen, X., Qiu, J., & Sun, J. (2023). Personalized movie recommendations based on a multi-feature attention mechanism with neural networks. *Mathematics*, 11(6), 1355. [CrossRef]
- [65] Chen, X., Zhao, P., Liu, Y., Zhao, L., Fang, J., Sheng, V. S., & Cui, Z. (2019). Exploiting aesthetic features in visual contents for movie recommendation. *IEEE Access*, 7, 49813-49821. [CrossRef]
- [66] Liang, Z., Yang, Z., & Cheng, J. (2023). Weight normalization optimization movie recommendation algorithm based on three-way neural interaction networks. *Complex & Intelligent Systems*, 9(6), 6241-6254. [CrossRef]
- [67] Chen, X., Zhao, P., Xu, J., Li, Z., Zhao, L., Liu, Y., ... & Cui, Z. (2018). Exploiting visual contents in posters and still frames for movie recommendation. *IEEE Access*, 6, 68874-68881. [CrossRef]
- [68] Zhang, Q., Wang, L., Xu, K., Lu, W., Ma, X., & Huang, Y. (2022). Learning coupled latent features via review texts for IOT service recommendation. *Computers and Electrical Engineering*, 101, 108084. [CrossRef]
- [69] Zhang, Q., Wang, L., Meng, X., Xu, K., & Hu, J. (2019). A generic framework for learning explicit and implicit user-item couplings in recommendation. *IEEE Access*, 7, 123944-123958. [CrossRef]
- [70] Yang, D., Shen, C., Liu, B., Xue, L., & Xiao, Y. (2020). Generating knowledge-based attentive user representations for sparse interaction recommendation. *IEEE Transactions on Knowledge and Data Engineering*, 34(9), 4270-4284. [CrossRef]
- [71] Yan, W., Wang, D., Cao, M., & Liu, J. (2019). Deep auto encoder model with convolutional text networks for video recommendation. *IEEE Access*, 7, 40333-40346. [CrossRef]
- [72] Ferreira, D., Silva, S., Abelha, A., & Machado, J. (2020). Recommendation system using autoencoders. *Applied Sciences*, 10(16), 5510. [CrossRef]
- [73] Geng, Y., Zhu, Y., Li, Y., Sun, X., & Li, B. (2022). Multi-feature extension via semi-autoencoder for personalized recommendation. *Applied Sciences*, 12(23), 12408. [CrossRef]
- [74] Zhu, Y., Wu, X., Qiang, J., Yuan, Y., & Li, Y. (2021). Representation learning with collaborative autoencoder for personalized recommendation. *Expert Systems with Applications*, 186, 115825. [CrossRef]
- [75] Zhong, S. T., Huang, L., Wang, C. D., Lai, J. H., & Yu, P. S. (2020). An autoencoder framework with attention mechanism for cross-domain recommendation. *IEEE Transactions on Cybernetics*, 52(6), 5229-5241. [CrossRef]
- [76] Da' u, A., & Salim, N. (2020). Recommendation system based on deep learning methods: a systematic review and new directions. *Artificial Intelligence Review*, 53(4), 2709-2748. [CrossRef]
- [77] Li, L., Xiahou, J., Lin, F., & Su, S. (2023). Distvae: distributed variational autoencoder for sequential recommenda-

- tion. *Knowledge-Based Systems*, 264, 110313. [CrossRef]
- [78] Yan, W., Wang, D., Liu, J., Ma, L., & Li, Z. (2019). Multi-Channel and Fusion Encoding Strategy Based Auto Encoder Model for Video Recommendation. *IEEE Access*, 7, 86004-86017. [CrossRef]
- [79] Yi, J., & Chen, Z. (2021). Multi-modal variational graph auto-encoder for recommendation systems. *IEEE Transactions on Multimedia*, 24, 1067-1079. [CrossRef]
- [80] Conceição, F. L., Pádua, F. L., Lacerda, A., Machado, A. C., & Dalip, D. H. (2019). Multimodal data fusion framework based on autoencoders for top-N recommender systems. *Applied Intelligence*, 49, 3267-3282. [CrossRef]
- [81] Bao, L., Sun, X., Gong, D., & Zhang, Y. (2021). Multisource heterogeneous user-generated contents-driven interactive estimation of distribution algorithms for personalized search. *IEEE Transactions on Evolutionary Computation*, 26(5), 844-858. [CrossRef]
- [82] Harshvardhan, G. M., Gourisaria, M. K., Rautaray, S. S., & Pandey, M. (2022). UBMTR: Unsupervised Boltzmann machine-based time-aware recommendation system. *Journal of King Saud University-Computer and Information Sciences*, 34(8), 6400-6413. [CrossRef]
- [83] Chaturvedi, A., Tiwari, A., & Chaturvedi, S. (2020). Sysevorecomd: Network reconstruction by graph evolution and change learning. *IEEE Systems Journal*, 14(3), 4007-4014. [CrossRef]
- [84] Pujahari, A., & Sisodia, D. S. (2019). Modeling side information in preference relation based restricted boltzmann machine for recommender systems. *Information Sciences*, 490, 126-145. [CrossRef]
- [85] Chen, Z., Ma, W., Dai, W., Pan, W., & Ming, Z. (2020). Conditional restricted Boltzmann machine for item recommendation. *Neurocomputing*, 385, 269-277. [CrossRef]
- [86] Wu, X., Yuan, X., Duan, C., & Wu, J. (2019). A novel collaborative filtering algorithm of machine learning by integrating restricted Boltzmann machine and trust information. *Neural Computing and Applications*, 31(9), 4685-4692. [CrossRef]
- [87] Li, L., Huang, H., Li, Q., & Man, J. (2023). Personalized movie recommendations based on deep representation learning. *PeerJ Computer Science*, 9, e1448. [CrossRef]
- [88] Ning, L., Pittman, R., & Shen, X. (2018). LCD: A fast contrastive divergence based algorithm for restricted Boltzmann machine. *Neural Networks*, 108, 399-410. [CrossRef]
- [89] Hazrati, N., Shams, B., & Haratizadeh, S. (2019). Entity representation for pairwise collaborative ranking using restricted Boltzmann machine. *Expert Systems with Applications*, 116, 161-171. [CrossRef]
- [90] Mu, R. (2018). A survey of recommender systems based on deep learning. *IEEE Access*, 6, 69009-69022. [CrossRef]
- [91] Cheng, L., Du, L., Liu, C., Hu, Y., Fang, F., & Ward, T. (2024). Multi-modal fusion for business process prediction in call center scenarios. *Information Fusion*, 108, 102362.
- [92] Zhang, L., Wang, H., Cheng, L., & Fang, F. (2025). Tighter regulation is needed for AI companions. *Nature*, 642(8068), 572-572.



王东培, 男, 河北省石家庄市人, 北京工商大学计算机与人工智能学院电子商务专业本科生, 主要研究方向为计算机技术及其应用。曾获得“挑战杯”首都大学生创业计划竞赛三等奖、全国英语翻译大赛全国二等奖等荣誉。

Dongpei Wang, a native of Shijiazhuang, Hebei Province, is currently an undergradu-

ate student majoring in E-commerce at the School of Computer and Artificial Intelligence, Beijing Technology and Business University. His research interests focus on computer technology and its applications. He has received several honors, including the Third Prize in the “Challenge Cup” Capital College Student Entrepreneurship Competition and the Second Prize in the National English Translation Contest.