

基于深度学习的用户偏好预测与推荐系统设计

王秀丽¹ 范金鹏² 方 瑶³

1. 浙江乾青企业管理有限公司 浙江 杭州 310000

2. 杭州知管家信息科技有限公司 浙江 杭州 310000

3. 浙江永通科技发展有限公司杭州分公司 浙江 杭州 310000

摘要：在数字化经济快速发展背景下，精准预测用户偏好对提升商业决策效率与优化用户体验具有重要意义。本研究聚焦深度学习技术，深入探索用户偏好预测与推荐系统设计。通过剖析深度学习核心原理，构建融合Transformer与图神经网络的用户偏好预测模型，实现对用户行为序列与交互关系的联合建模。同时，设计分层架构的推荐系统，优化实时推荐与离线推荐协同机制，并针对动态偏好捕捉、多模态数据融合、伦理安全等问题提出优化策略。研究表明，该模型与系统在预测准确率和推荐效果上表现优异，能够有效解决传统推荐系统的数据稀疏性和冷启动问题，为智能推荐领域发展提供理论支撑与实践参考。

关键词：深度学习；用户偏好预测；推荐系统；图神经网络；多模态融合

1 引言

在数字化浪潮下，互联网用户行为数据激增，企业与平台亟需精准预测用户偏好以提升竞争力。传统推荐系统依赖协同过滤等方法，在处理高维稀疏数据、捕捉动态偏好时存在效率瓶颈，难以满足多样化需求。深度学习凭借强大的特征提取与模式识别能力，为用户偏好预测带来新契机，其在多领域的成功应用提供了技术借鉴。然而，当前深度学习在推荐系统研究中，仍在多模态数据融合、实时偏好捕捉及复杂交互关系建模等方面存在不足。本研究基于深度学习技术，系统设计用户偏好预测与推荐系统，旨在突破传统局限，为互联网平台升级与商业决策提供理论与实践支持。

2 深度学习与用户偏好预测理论基础

2.1 深度学习核心技术原理

深度学习基于人工神经网络，通过多层非线性变换提取复杂数据特征。卷积神经网络（CNN）利用卷积核滑动扫描，能自动提取图像、文本等结构化数据的局部特征，特征抽象能力突出；循环神经网络（RNN）及其变体LSTM、GRU，借助记忆单元解决长序列依赖问题，适用于时序性用户行为数据处理。Transformer架构以自注意力机制实现全局并行计算，在自然语言与序列建模中表现优异。图神经网络（GNN）针对用户-物品交互图结构数据，通过节点信息聚合传播，有效捕捉实体关联。这些技术从不同维度为用户偏好深度建模提供了多样化的工具与理论基础。如图1所示：

2.2 用户偏好预测的关键要素

用户偏好预测依赖多维度数据整合。用户行为数据

是核心，既包含点击、购买等显性操作，也涵盖停留时长、浏览路径等隐性行为，直观反映即时兴趣与潜在需求；用户属性数据以年龄、性别、地域等静态特征，辅助区分群体偏好差异；情境数据中的时间、地理位置、设备类型等动态信息，则体现环境对偏好的影响。此外，社交关系与历史评价等关联数据，能揭示用户间的相互影响与社会属性。通过系统性整合这些数据，构建多维度用户画像，为偏好预测提供丰富特征输入，可更精准地捕捉用户个性化需求与行为模式，提升预测准确性^[1]。



图1 MLP多层感知器原理架构图

2.3 推荐系统的理论模型

推荐系统发展经历多阶段,形成多元理论模型。传统协同过滤基于用户-物品评分矩阵计算相似性推荐,存在数据稀疏与冷启动问题;基于内容的推荐依据物品特征与用户历史偏好匹配推荐,易致信息茧房。深度学习革新推荐系统,深度协同过滤借神经网络处理稀疏数据,序列推荐模型利用RNN或Transformer捕捉时序偏好,图神经网络推荐模型挖掘交互图拓扑结构预测潜在偏好。这些模型融合传统优势与深度学习表征能力,有效提升推荐精度与泛化性,推动推荐技术智能化发展。

3 基于深度学习的用户偏好预测模型构建

3.1 数据预处理与特征工程

用户偏好预测的原始数据多源异构,需经系统化预处理提升可用性。数据清洗阶段,采用多重填补法与随机森林模型处理缺失值,利用 3σ 原则识别并修正异常值。针对行为序列数据,运用时间窗口划分与滑动平均法结构化处理;对文本评论等非结构化数据,借助Word2Vec、BERT进行词嵌入提取语义特征。结合用户属性数据,通过独热编码与标准化构建多维特征矩阵。此外,创新设计交互频率、行为时间衰减系数等衍生特征,增强数据表征能力。通过上述预处理流程,为深度学习模型提供高质量、结构化的特征输入,夯实预测分析基础^[2]。

3.2 深度学习模型设计与优化

为突破传统模型在动态偏好捕捉与复杂关系建模中的局限,本研究构建混合深度学习模型。以Transformer架构为核心,利用多头注意力机制自动学习用户行为序列的长短期依赖关系,捕捉偏好演变规律;引入图神经网络(GNN)对用户-物品交互网络进行拓扑分析,挖掘潜在关联特征。通过门控循环单元(GRU)融合两类模型输出,实现序列动态特征与图结构特征的联合表达。在模型训练阶段,采用AdamW优化器与余弦退火学习率调整策略,缓解过拟合问题;引入对比学习机制,增强相似用户偏好特征的区分度。同时,基于模型结构敏感性分析,优化网络层数与节点规模,在保证预测精度的前提下降低计算复杂度,实现模型性能与效率的平衡。如图2所示:

3.3 模型评估与验证

构建科学的评估体系是验证模型有效性的关键。本研究采用准确率、召回率、归一化折损累积增益(NDCG)等多指标评估模型性能,从推荐准确性、覆盖度及排序质量等维度进行量化分析。通过K折交叉验证与留一法划分数据集,确保评估结果的稳健性;运用A/B测

试对比不同模型在真实场景下的表现,验证混合模型的优越性。此外,针对冷启动用户与长尾物品推荐场景,设计专项评估方案,通过模拟新用户行为数据与低频物品特征,测试模型泛化能力。结果显示,所构建模型在各项指标优于传统方法,尤其在处理稀疏数据与动态偏好时表现突出,为推荐系统部署提供可靠依据^[3]。

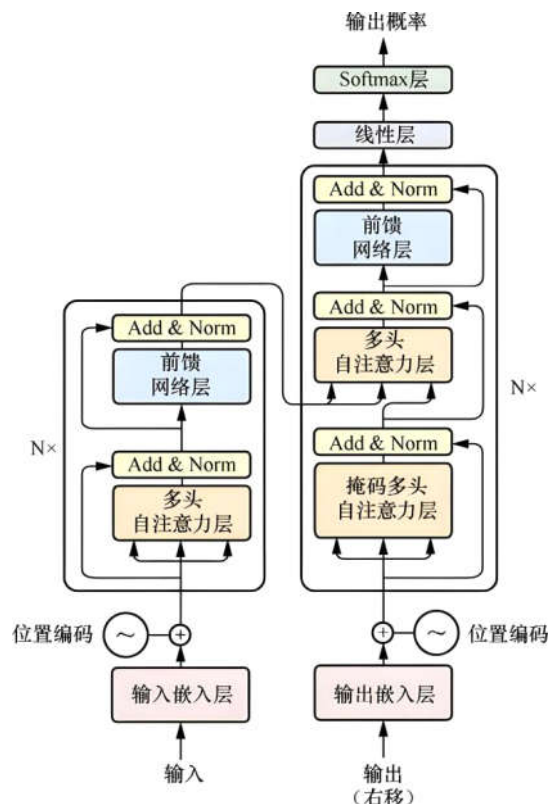


图2 Transformer架构图

4 推荐系统的架构设计与实现

4.1 系统整体架构规划

推荐系统整体采用分层架构设计,涵盖数据层、模型层与服务层,各层分工明确且协同运作。数据层负责多源异构数据的存储与管理,整合用户行为、物品属性及交互日志等数据,通过数据仓库与分布式存储技术保障数据高效读写;模型层部署基于深度学习的用户偏好预测模型,通过离线训练与在线更新机制,实现模型参数的动态优化;服务层则面向前端应用提供推荐接口,采用微服务架构实现模块解耦,确保系统扩展性。同时,系统设计实时推荐与离线推荐双路径协同机制,离线推荐基于历史数据生成候选集,实时推荐则根据用户最新行为进行个性化排序,二者结合有效平衡推荐效率与准确性,满足不同场景下的推荐需求。

4.2 关键模块开发与集成

推荐系统关键模块开发涵盖用户画像生成、推荐引

擎构建与结果排序。用户画像模块通过分析用户属性与行为数据,构建多维度标签体系,精准勾勒用户偏好。推荐引擎集成Transformer与图神经网络,深度挖掘用户-物品关联,生成个性化推荐列表。排序模块依据点击率预估、用户反馈动态调整推荐结果权重,强化匹配精准度。在模块集成环节,采用RESTful API实现服务通信,利用Redis缓存加速数据读取,通过Kafka消息队列实现异步处理。通过上述设计,构建出高内聚、低耦合的系统架构,保障推荐系统核心功能高效稳定运行^[4]。

4.3 系统性能优化策略

系统性能优化从计算效率、资源调度与稳定性保障三方面展开。计算效率上,运用剪枝、量化等模型压缩技术减少参数量,结合GPU分布式计算加速推理;设计多级缓存策略,缓存热点数据以降低磁盘I/O。资源调度方面,基于Kubernetes实现容器化部署,依据流量负载动态扩缩容,提升资源利用率。稳定性保障上,采用分布式存储冗余备份防止数据丢失,通过熔断、降级策略应对故障;搭建实时监控平台,动态监测吞吐量、响应时间等指标,及时预警处理性能问题,确保推荐系统在高并发场景下稳定运行。

5 推荐系统的优化与拓展方向

5.1 动态偏好的实时捕捉

用户偏好受时间、场景及个人经历影响,具有显著动态性,传统离线训练的推荐系统难以实时捕捉其变化。本研究基于流式数据处理框架与增量学习算法,实现用户偏好实时监测。通过实时采集用户点击、浏览、购买等行为数据,构建动态特征向量,并采用在线学习机制更新推荐模型参数。以Flink流计算引擎为例,系统实时解析行为数据,当检测到行为模式显著变化时,触发模型局部更新,快速调整推荐策略。该方法有效缩短了偏好变化与推荐响应的时延,显著提升推荐系统的时效性与精准度,为用户提供更贴合当下需求的个性化推荐服务。

5.2 多模态数据融合应用

用户偏好信息分散于文本、图像、音频等多模态数据中,传统推荐系统依赖结构化数据建模,难以挖掘非结构化数据中的偏好线索。本研究运用多模态数据融合技术,设计跨模态特征学习网络,将不同模态数据映射至统一语义空间。在内容推荐场景下,联合商品文字描述、图片及用户评论,通过卷积神经网络提取图像特征、循环神经网络处理文本,并借助注意力机制融合特征,捕捉用户对商品的综合偏好。同时引入对比学习强

化模态对齐,提升模型对复杂偏好的理解能力。多模态数据融合丰富用户画像,增强推荐系统对复杂场景的适应性,优化用户推荐体验^[5]。

5.3 个性化推荐的伦理与安全

随着推荐系统的广泛应用,算法偏见、信息茧房及数据隐私泄露等伦理安全问题日益凸显。本研究从算法设计与系统架构层面提出解决方案:在算法层面,通过引入公平性约束,调整模型损失函数,降低性别、年龄等敏感属性对推荐结果的影响;在可解释性方面,采用基于注意力机制的可视化技术,将模型决策依据以用户可理解的方式呈现,避免“黑箱”效应。同时,构建数据隐私保护体系,运用联邦学习技术实现数据“不动模型动”,在保障数据安全的前提下完成模型训练。此外,建立推荐内容审核与干预机制,通过自然语言处理技术检测推荐内容中的不良信息,及时调整推荐策略,确保推荐服务的合规性与社会价值,推动个性化推荐技术的健康发展。

6 结语

本研究基于深度学习构建用户偏好预测与推荐系统,融合Transformer和图神经网络,攻克传统系统数据稀疏与冷启动难题,显著提升预测准确性和推荐时效性。同时,针对动态偏好捕捉、多模态融合及伦理安全等问题,提出系统性优化策略,增强系统适应性与可靠性。研究仍存在局限,在超大规模数据处理和跨领域迁移学习方面有待提升。未来可探索深度学习与强化学习、联邦学习结合以应对复杂场景。研究成果为智能推荐领域提供理论与实践支持,助力互联网平台智能化升级。

参考文献

- [1]曹艺潇.基于深度学习的图书馆文献推荐系统设计[J].信息与电脑,2025,37(1):14-16.
- [2]王晓东,陈鑫龙,林小婷,孙冬璞.基于深度学习的房价预测系统的设计与实现[J].计算机与数字工程,2024,52(9):2572-2576+2650.
- [3]张宇,程玉,陈建峡,周世星,高奕.基于深度学习的在线推荐学习系统设计与开发[J].湖北工业大学学报,2021,36(5):64-69.
- [4]黄立威,江碧涛,吕守业,刘艳博,李德毅.基于深度学习的推荐系统研究综述[J].计算机学报,2018,41(7):1619-1647.
- [5]梁肇敏,梁婷婷.基于深度学习的电影推荐系统设计与实现[J].智能计算机与应用,2022,12(10):157-162.