

大数据环境下视频内容智能推荐算法优化研究

刘 涛

浙江通见科技有限公司 浙江 杭州 310000

摘 要: 大数据环境下视频内容智能推荐对提升用户体验和视频平台竞争力至关重要,本文围绕视频内容智能推荐算法优化展开研究,阐述传统推荐算法在大数据环境下的局限性,分析基于协同过滤、内容基及深度学习的推荐算法原理与应用。结合实例与数据,提出融合多种算法、引入多源数据及利用强化学习等优化策略,提升推荐准确性、多样性与实时性,研究结果为视频平台优化推荐算法提供有价值的参考,助力其在大数据时代满足用户需求,提升平台影响力。

关键词: 大数据; 视频内容; 智能推荐算法; 算法优化

引言

互联网技术飞速发展下视频数据呈爆炸式增长,数据显示,截至2024年全球互联网视频流量占总流量比例超80%,海量视频内容中用户难快速找到感兴趣的视频,视频平台也急需高效推荐算法提升用户体验与竞争力。智能推荐算法作为关键技术,能根据用户行为数据、兴趣偏好等信息为其推送符合需求的视频内容,大数据环境下如何优化视频内容智能推荐算法,提高推荐的准确性、多样性和实时性,成为当前研究的热点与重点。

1 传统视频推荐算法在大数据环境下的局限

1.1 协同过滤算法的不足

协同过滤算法是传统视频推荐领域核心方法,通过剖析用户行为数据定位与目标用户兴趣相符的相似群体,依群体偏好为目标用户推荐视频,大数据浪潮下,该算法局限性凸显,用户规模和视频资源指数级增长,数据稀疏性制约性能。某综合性视频平台千万级用户、百万量级视频库,用户实际交互数据不足总数据量3%,用户-项目矩阵稀疏,传统余弦相似度等度量方法难准确算用户间相似度,影响推荐准确性,算法处理大规模数据时,计算复杂度高且实时性差,ItemCF算法新视频上线,要重新计算与全量视频相似度,传统方式在百万量级视频库更新一次耗时数小时,无法满足实时推荐,新内容冷启动难,用户难及时获取最新优质视频资源。

1.2 基于内容的推荐算法瓶颈

基于内容的推荐算法通过分析视频标题、描述、标签等文本特征及音频、图像等多媒体特征,依用户历史内容偏好实现推荐,大数据环境下,其缺陷显现。视频内容特征提取遇技术瓶颈,非结构化数据处理难度大。处理影视剧中复杂场景时,传统计算机视觉算法难准确识别动态画面中人物关系与情感表达;嘈杂环境下语音识

别技术准确率不足85%,音频特征提取有偏差,算法对用户兴趣理解单一,仅靠内容特征难捕捉用户兴趣动态变化。社交媒体热点事件中,算法缺社交数据和实时热点感知能力,难及时调整推荐策略,推荐内容与用户即时需求脱节,推荐多样性评分低于行业均值15%-20%,难满足用户多元化内容消费需求。

2 大数据环境下视频内容智能推荐算法分析

2.1 基于协同过滤的推荐算法原理与应用

协同过滤算法分基于用户的协同过滤(UserCF)和基于物品的协同过滤(ItemCF),UserCF计算用户间相似度,找与目标用户兴趣相近的邻居用户,把邻居用户喜欢的视频推荐给目标用户,相似度计算常用余弦相似度、皮尔逊相关系数等方法,ItemCF计算视频间相似度,依目标用户已观看的视频,找相似视频推荐。实际应用中,ItemCF在电商等领域应用多,UserCF在社交网络相关推荐场景表现好,某知名视频平台用UserCF算法,根据用户观看历史、点赞、评论等行为数据构建用户兴趣模型,计算用户间余弦相似度,筛选出与目标用户相似度高的邻居用户^[1]。若目标用户A喜欢看科幻类视频,与A相似度高的邻居用户B、C等除喜欢科幻类视频外,还对悬疑类视频兴趣高,系统就会把悬疑类视频推荐给用户A。

许多视频平台运用协同过滤算法收获成效,Netflix借助大规模用户观影数据,通过该算法为用户推荐个性化影视剧,研究数据显示,采用协同过滤算法后,Netflix用户观看时长平均提升20%,用户留存率提高15%。但如前文所讲,协同过滤算法在大数据环境中存在数据稀疏性和冷启动问题。实际应用里,部分平台会结合其他算法,像基于内容的推荐算法,对协同过滤结果补充优化,遇到新用户或新视频时,先利用基于内容的推荐算法

为新用户推荐热门且契合平台主流内容的视频,同时为新视频找到可能感兴趣的初始用户群体,积累一定数据后,再切换至协同过滤算法进行更精准推荐。

2.2 基于内容基础的推荐算法原理与应用

基于内容的推荐算法着重对视频内容进行特征提取与分析,针对文本类视频描述信息,借助自然语言处理技术提取关键词、主题等特征;针对视频图像部分,运用计算机视觉技术提取颜色、纹理、形状等特征;针对音频部分,提取音频频率、节奏等特征。依据用户对视频内容特征的偏好构建用户兴趣模型,当有新视频时,计算其与用户兴趣模型的相似度,将相似度高的视频推荐给用户,以一部电影为例,系统先对电影剧情介绍、演员信息等文本内容进行分词、去停用词处理,提取“爱情”“冒险”“主演XXX”等关键主题和关键词。同时对电影海报、精彩片段等图像进行特征提取,得到颜色直方图、纹理特征向量等,综合这些特征形成电影的内容特征向量,对于用户,系统根据其以往观看电影的内容特征偏好,构建用户的兴趣特征向量,并通过计算电影与用户兴趣特征向量的余弦相似度等,判断用户对该电影的感兴趣程度。

部分聚焦特定领域的视频平台,像知识类视频平台,普遍运用基于内容的推荐算法,某知识类视频平台会深入剖析课程视频的标题、知识点描述等内容,借此为用户推荐相关知识视频^[2]。该平台统计数据显示,采用基于内容的推荐算法后,用户对推荐视频的点击率提升18%,在平台的平均学习时长增加15分钟,此算法存在一定局限,例如对视频内容的理解较为单一,难以捕捉用户复杂多变的兴趣,为解决这一问题,部分平台开始探索将基于内容的推荐算法与其他算法结合。以融合深度学习算法为例,可对视频内容展开更深入的语义理解,提升特征提取的准确性与全面性,通过构建深度神经网络模型,对视频的图像、音频、文本等多模态数据进行联合学习,挖掘数据间的潜在关联,进而更精准地把握用户兴趣,优化推荐效果。

3 大数据环境下视频内容智能推荐算法优化策略

3.1 融合多种推荐算法

发挥协同过滤算法与内容基算法的优势并规避各自短板,可采取融合策略,融合时,先借助内容基算法完成视频初步筛选,依据视频内容特征向用户推送一批契合其兴趣的视频。继而运用协同过滤算法,按照用户间相似性及用户对视频的行为数据,对初步推荐结果实施优化与排序。具体可采用加权融合方式,结合不同场景与数据特性,为协同过滤算法和内容基算法的推荐结果设

定差异化权重,当用户对视频内容存在明确偏好且平台对视频内容特征提取较为精准时,可适度提升内容基算法的权重;若用户间兴趣相似度较高且行为数据充足,可增大协同过滤算法的权重。此种方式有助于提升推荐的准确性与多样性,某视频平台应用协同过滤与内容基算法融合策略后,用户对推荐视频的满意度提升25%,用户流失率降低12%。

深度学习算法在处理复杂数据结构与挖掘潜在模式方面展现出显著效能,将其与传统视频推荐算法有机融合,可进一步强化推荐效果的优化空间,具体而言,深度学习算法能够针对视频包含的多模态数据(如视觉层面的图像信息、听觉层面的音频信息、语义层面的文本描述等)展开多层次特征提取与深度学习^[3]。通过构建卷积神经网络(CNN)对视频图像的空间特征进行精细化捕捉,利用循环神经网络(RNN)对视频内容的时间序列信息进行动态建模,结合自然语言处理技术对视频文本描述的语义特征进行解析,以此形成覆盖多维度信息的视频特征表示体系,在此基础上,将深度学习提取的高阶抽象特征与传统算法中所依赖的用户行为特征、视频内容特征进行有机融合,共同输入至推荐模型中完成训练与预测过程。以国内某头部短视频平台的实践为例,其通过整合深度学习算法与协同过滤框架,对用户的历史观看记录、点赞收藏行为、评论互动数据以及视频的图像视觉特征、音频情感特征、文本语义特征进行联合建模,成功构建起具备深度语义理解能力的推荐系统,实测数据表明,该融合算法在推荐准确性评价指标上较传统协同过滤算法提升幅度达30%,用户留存率指标实现18%的增长,有效推动了平台用户体验的升级与市场竞争力的增强。

3.2 引入多源数据提升推荐效果

除传统用户观看历史、点赞、评论等行为数据外,深入剖析用户在视频平台的搜索记录、观看时长分布、暂停与快进时间点等多元行为数据,可更全面洞察用户兴趣与需求,分析用户搜索记录能捕捉即时兴趣点与潜在需求,若用户高频搜索某主题视频,表明其对此主题兴趣浓厚,推荐系统可针对性推送该主题优质内容^[4]。研究用户观看时长分布可揭示对不同类型视频的兴趣程度,若用户在某类视频平均观看时长较长,说明兴趣浓厚,系统可提升此类视频推荐权重,解析用户暂停与快进时间点,能明确其对视频内容的关注焦点与偏好,进而为用户精准推荐契合兴趣的视频内容,国内某视频平台通过深度挖掘用户行为数据,将搜索意图与观看历史结合优化推荐算法,数据显示,优化后用户对推荐视频点击转化

率提升20%，平台平均停留时间延长10分钟。

社交数据在视频推荐场景中具备显著价值，将用户的社交关系网络（包括好友列表、关注与被关注关系链、兴趣社群等多维度数据）深度融入推荐算法，能够有效捕捉用户间的社交影响力权重与关系亲密度差异。当用户A的好友群体呈现出对某一类型视频的集中偏好时，基于社交传播的协同效应，用户A对该类型内容产生兴趣的可能性显著提升，推荐系统可依据此类社交关联特征动态调整推荐权重，增强推荐内容与用户潜在兴趣的匹配度。结合外部环境数据维度（如实时热点事件演变、季节气候特征、用户地理位置坐标等动态信息），能够进一步提升推荐服务的场景适配性：在突发热点事件周期内，系统可实时抓取相关视频内容并进行优先级推送，满足用户对时效性信息的获取需求；针对季节更替特点，精准推荐应季主题内容（如夏季聚焦游泳、户外旅游等清凉题材视频，冬季侧重滑雪运动、居家保暖等主题内容）；基于用户地理位置信息，智能推送本地特色文化视频或区域性活动相关内容，强化推荐服务的地域关联性^[5]。以某头部社交视频平台的实践为例，其通过构建社交关系图谱与外部环境特征的双重驱动模型，实现了个性化推荐服务的升级迭代。数据表明，在引入社交数据模块后，用户间互动率提升30%，视频分享次数增长25%；叠加外部环境数据融合策略后，用户对推荐视频的满意度提升22%，平台的社交属性与用户粘性均得到显著增强。

结语

大数据环境下优化视频内容智能推荐算法是提升平台竞争力与用户体验的关键。突破传统算法局限可通过融合多元算法及引入多源数据：联邦学习驱动的分布式协同过滤算法能在隐私保护下实现跨平台数据协同，解决数据孤岛问题；生成式对抗网络（GAN）与推荐模型结合可动态挖掘用户潜在兴趣，提升推荐多样性。未来技术迭代中，边缘计算将为推荐模型提供实时更新能力，知识图谱则通过构建用户-内容-场景的多维关联网络强化推荐的场景化与精准度。这些优化路径推动视频推荐向智能化、生态化演进，持续满足用户动态需求，助力视频产业在数据驱动下实现高质量发展。

参考文献

- [1]周婵.课程思政案例智能推荐工具构建及应用研究[D].西北师范大学,2024.
- [2]陈志玲.把关人视角下算法推荐的负面问题及应对研究[D].安徽财经大学,2023.
- [3]任颖姗.短视频用户抵抗智能推荐算法行为意愿影响因素研究[D].暨南大学,2023.
- [4]吕书影.基于大数据的顺网数字娱乐云平台个性化推荐策略研究[D].浙江理工大学,2022.
- [5]温凤鸣,解学芳.短视频推荐算法的运行逻辑与伦理隐忧——基于行动者网络理论视角[J].西南民族大学学报(人文社会科学版),2022,43(02):160-169.