

# 大数据技术在电商用户行为分析与精准推荐中的应用研究

杨国波 袁月婷 苏宇航

蒙古国民族大学, 蒙古国乌兰巴托, 999097;

**摘要:** 在数字经济时代, 电商行业竞争日趋激烈, 用户需求的个性化与多元化成为行业发展的核心驱动力。大数据技术凭借其海量数据处理、深度分析挖掘的能力, 为电商平台实现用户行为的精准解析与个性化推荐提供了关键支撑。本文以大数据技术在电商领域的应用为核心, 首先构建了电商用户行为分析与精准推荐的大数据技术架构, 明确数据采集、存储、处理及应用各环节的技术选型逻辑; 其次深入剖析基于大数据的用户行为分析体系, 从行为维度划分、特征提取及用户画像构建三个层面展开研究; 再次聚焦精准推荐的核心模型, 探讨协同过滤、内容推荐及深度学习推荐模型的应用场景与优化策略; 最后分析大数据应用的实际成效与面临的挑战, 并提出针对性发展建议。

**关键词:** 大数据技术; 电商平台; 用户行为分析; 精准推荐; 用户画像

**DOI:** 10.64216/3080-1516.25.12.091

## 引言

随着互联网技术的飞速发展, 电商行业已从早期的流量争夺阶段迈入“以用户为中心”的精细化运营阶段。传统电商模式下的“千人一面”推荐策略, 因无法匹配用户的个性化需求, 导致用户转化率偏低、留存率不足等问题日益凸显。在此背景下, 大数据技术的兴起为破解这一难题提供了有效路径。电商平台通过对用户在平台上产生的浏览, 点击, 购买, 评价等海量行为数据进行采集与深度分析, 精准把握用户的消费偏好与潜在需求, 进而实现精准推荐。这种数据驱动运营模式, 不仅能够优化用户的购物体验, 也帮助电商平台优化库存结构, 提高营销效率。本文基于大数据技术的核心特性, 深入探讨其在电商用户行为分析与精准推荐中的应用逻辑与实践路径, 力求为电商行业的高质量发展提供有益借鉴。

## 1 大数据技术架构的电商用户行为分析与精准推荐

### 1.1 数据采集层: 全面的多源数据汇聚

数据采集层是架构的基础, 主要对电商平台多源异构数据进行全面实时的采集。电商用户行为数据来源广泛, 包括用户在平台内的线上行为数据、线下消费关联数据以及外部第三方数据。线上行为数据是核心数据源, 主要包括用户的浏览日志, 点击记录, 搜索关键词, 购物车操作, 订单信息, 评价反馈等, 此类数据实时性强, 维度丰富, 能够直接反映用户的消费意图。线下消费关联数据主要通过线上线下融合渠道获取, 如用户线下门店消费记录, 扫码互动数据等, 可以弥补纯线上数据对用户消费场景刻画的不足。外部第三方数据, 比如行业趋势数据, 社交平台的用户偏好数据等, 都能够为用户

行为分析提供更广阔的视角。

### 1.2 数据存储层: 异构数据的高效存储

电商用户行为数据呈现出“4V”特性, 即海量 (Volume)、高速 (Velocity)、多样 (Variety)、低价值密度 (Value), 传统的关系型数据库难以满足其存储需求。数据存储层需根据数据的不同类型与特征, 采用多元化的存储方案, 实现异构数据的高效存储与快速访问。对于用户基本信息、订单详情等结构化数据, 由于其数据格式规范、查询需求频繁, 采用传统的关系型数据库如 MySQL 进行存储, 确保数据查询的高效性与一致性。对于用户浏览日志、点击流等非结构化或半结构化数据, 采用分布式文件系统如 HDFS 进行存储, 其具备高容错性、高扩展性的特点, 能够满足海量数据的存储需求。对于需要频繁读写的实时数据, 如用户实时浏览记录、购物车动态变化数据, 采用内存数据库如 Redis 进行存储, 通过将数据存储在内存中, 大幅提升数据的读写速度, 保障实时推荐的响应效率。此外, 为实现对多源数据的统一管理, 采用数据仓库技术如 Hive 构建统一的数据仓库, 将分散在不同存储系统中的数据进行整合, 为后续的用户行为分析提供统一的数据支撑。

### 1.3 数据处理层: 数据价值的深度挖掘

数据处理层是大数据技术架构的核心, 其任务是对存储层汇聚的数据进行清洗、转换、整合与深度分析, 挖掘数据中蕴含的用户行为规律与消费偏好。该层采用分布式计算框架与数据挖掘算法, 实现对海量数据的高效处理与价值提取。

在数据处理环节, 首先通过数据清洗工具剔除采集数据中的噪声数据、补全缺失值, 确保数据的准确性;

然后通过数据转换将非结构化数据转换为结构化数据,实现数据格式的统一;最后通过数据整合将不同来源的数据进行关联,形成完整的用户行为数据链路。在数据分析环节,采用分布式计算框架如 Spark、Flink 实现对海量数据的并行计算,提升分析效率。同时,运用数据挖掘算法如聚类分析、关联规则挖掘、时序分析等,对用户行为数据进行深度挖掘。例如,通过聚类分析将具有相似行为特征的用户划分为同一用户群体,通过关联规则挖掘发现用户购买行为之间的关联关系,通过时序分析预测用户的未来消费趋势,为精准推荐提供数据支撑。

#### 1.4 应用服务层:分析结果的落地转化

应用服务层是大数据技术架构的终端,其核心任务是将数据处理层的分析结果转化为具体的应用服务,实现对用户行为分析与精准推荐的落地。该层主要包括用户画像系统、精准推荐系统、营销决策系统等核心应用模块。用户画像系统将数据处理层挖掘的用户行为特征进行整合,构建多维度的用户画像,全面刻画用户的消费偏好、购买能力、消费习惯等信息。精准推荐系统基于用户画像数据与商品数据,通过推荐算法为用户生成个性化的商品推荐列表,并通过 API 接口嵌入到电商平台的首页、商品详情页、购物车页面等关键位置。营销决策系统将用户行为分析结果提供给电商平台的运营人员,为其制定营销策略、优化商品布局、调整库存结构提供数据支持。例如,根据用户画像数据为不同用户群体制定差异化的营销活动,根据商品关联分析结果优化商品的陈列布局,提升用户的购买概率。

### 2 基于大数据的电商用户行为分析体系构建

#### 2.1 用户行为维度的科学划分

电商用户行为具有多样性与复杂性,为实现对用户行为的全面解析,需从不同维度对用户行为进行科学划分。根据用户在电商平台的行为路径与行为目的,可将用户行为划分为浏览行为、交互行为、购买行为、售后行为四个核心维度,各维度相互关联,共同构成完整的用户行为链路。浏览行为是用户与电商平台交互的起点,主要包括用户的访问时长、访问频率、浏览页面类型、页面跳转路径等特征。通过对浏览行为的分析,可初步判断用户的潜在消费兴趣。例如,用户长时间浏览某类商品的详情页,说明其对该类商品具有较高的兴趣;用户频繁访问平台的促销活动页面,说明其对价格较为敏感。交互行为是用户在浏览过程中与平台产生的深度互动行为,主要包括搜索关键词、收藏商品、加入购物车、评价商品、分享商品等特征。交互行为更能反映用户的真实消费意图,例如,用户将商品加入购物车但未立即购买,说明其具有购买意愿但可能存在价格顾虑或等待

促销活动;用户对已购买商品进行高评分并详细评价,说明其对商品满意度较高,可能成为该商品的复购用户。

#### 2.2 用户行为特征的精准提取

用户行为特征提取是将原始的用户行为数据转化为可用于分析的结构化特征的过程,是用户行为分析体系的核心环节。基于大数据技术的特征提取,采用“原始数据—衍生特征—核心特征”的三级提取模式,实现对用户行为特征的精准提取。在原始数据层面,通过数据采集层获取用户的浏览、交互、购买、售后等原始行为数据,确保数据的完整性与实时性。在衍生特征层面,对原始数据进行加工处理,生成具有一定业务意义的衍生特征。例如,基于用户的购买频率与购买金额,生成“用户活跃度”“用户消费等级”等衍生特征;基于用户的浏览时长与点击次数,生成“商品兴趣度”“页面停留深度”等衍生特征;基于用户的退换货次数与评价分数,生成“用户满意度”“商品信任度”等衍生特征。在核心特征层面,采用特征选择算法如互信息法、决策树算法等,从衍生特征中筛选出与用户消费行为关联性强的核心特征,降低后续分析的计算复杂度。例如,通过特征选择发现“商品兴趣度”“用户消费等级”“购买频率”等特征与用户的复购行为关联性最强,将其作为核心特征用于用户画像构建。

#### 2.3 多维度用户画像的构建与更新

用户画像是基于用户行为特征构建的用户虚拟模型,是精准推荐的核心依据。基于大数据技术的用户画像构建,采用“静态特征+动态特征”的多维度构建模式,并建立实时更新机制,确保用户画像的精准性与时效性。静态特征是指用户相对稳定的基本信息,主要包括用户的性别、年龄、地域、职业、收入水平等,这类特征通过用户注册信息与第三方数据整合获取,具有较强的稳定性。动态特征是指用户随时间变化的行为特征,主要包括用户的近期兴趣偏好、消费习惯、活跃度等,这类特征通过对用户实时行为数据的分析提取获得,具有较强的时效性。将静态特征与动态特征相结合,构建多维度的用户画像标签体系,标签类型包括人口属性标签(如“25-35 岁女性”“一线城市白领”)、兴趣偏好标签(如“喜欢户外运动”“关注智能家居”)、消费行为标签(如“高频复购用户”“价格敏感型用户”)、满意度标签(如“高满意度用户”“售后投诉用户”)等。

### 3 大数据在电商精准推荐中的核心应用与优化

#### 3.1 协同过滤推荐:基于相似性的匹配推荐

协同过滤推荐是电商精准推荐中应用最早、最成熟的推荐模型之一,其核心思想是“物以类聚,人以群

分”，通过分析用户与商品之间的交互关系，挖掘具有相似行为特征的用户或相似属性的商品，实现推荐。根据推荐逻辑的不同，协同过滤推荐可分为基于用户的协同过滤与基于物品的协同过滤两种类型。基于用户的协同过滤通过计算用户之间的行为相似度，为目标用户推荐与其相似用户喜欢的商品。其核心步骤包括：计算用户之间的相似度，采用皮尔逊相关系数、余弦相似度等算法，根据用户的购买历史、浏览记录等行为数据，计算目标用户与其他用户的相似度；筛选相似用户群体，选取与目标用户相似度较高的Top-N用户作为相似用户群体；生成推荐列表，将相似用户群体喜欢的商品中，目标用户未浏览或未购买的商品推荐给目标用户。该模型的优势在于能够发现用户的潜在兴趣，推荐结果具有较强的多样性，但当用户数量庞大时，计算用户相似度的复杂度较高，且对新用户存在“冷启动”问题。

### 3.2 内容-based 推荐：基于特征匹配的精准推荐

内容-based 推荐是基于商品与用户兴趣的特征匹配实现推荐，其核心思想是为用户推荐与其兴趣特征最匹配的商品。该模型的关键在于对商品特征与用户兴趣特征的精准提取与匹配，能够有效解决协同过滤推荐的“冷启动”问题，且推荐结果具有较强的个性化。商品特征提取是内容-based 推荐的基础，通过文本挖掘、图像识别等技术，从商品的标题、描述、图片、分类等信息中提取商品的核心特征。例如，对于服装类商品，提取其款式（如“休闲风”“商务风”）、颜色（如“黑色”“红色”）、尺码（如“S码”“M码”）、材质（如“棉麻”“聚酯纤维”）等特征；对于图书类商品，提取其主题（如“科幻小说”“历史传记”）、作者、出版社、评分等特征。用户兴趣特征提取基于用户的历史行为数据，通过分析用户喜欢的商品特征，构建用户的兴趣特征模型。例如，用户多次购买“休闲风棉麻服装”，则其兴趣特征可定义为“偏好休闲风、棉麻材质的服装”。

### 3.3 深度学习推荐：基于深层特征的智能推荐

随着大数据技术与人工智能技术的融合发展，深度学习推荐模型逐渐成为电商精准推荐的核心趋势。深度学习模型通过构建多层神经网络，能够自动学习用户与商品的深层特征，解决传统推荐模型无法捕捉复杂非线性关系的问题，大幅提升推荐的精准性与多样性。在电商精准推荐中，常用的深度学习模型包括卷积神经网络（CNN）、循环神经网络（RNN）、深度学习协同过滤（D

eeepCF）等。CNN 模型主要用于商品图像特征的提取，通过对商品图片进行卷积运算，提取商品的颜色、纹理、款式等深层图像特征，解决传统模型仅依赖文本特征的局限性。例如，对于服装商品，CNN 模型能够精准提取其设计细节、图案风格等深层特征，提升与用户兴趣特征的匹配精度。RNN 模型主要用于处理用户的时序行为数据，通过对用户的历史行为序列进行建模，捕捉用户行为的时间依赖性与序列关联性，预测用户的下一步行为。例如，RNN 模型能够分析用户在不同时间节点的浏览、购买行为，发现用户的季节性消费规律，为用户推荐符合其季节需求的商品。

## 4 结论

大数据技术为电商用户行为分析与精准推荐提供了强大的技术支撑，通过构建“采集-存储-处理-应用”的全链路大数据技术架构，实现了对用户行为数据的深度挖掘与价值转化。基于大数据的用户行为分析体系，通过多维度行为划分、精准特征提取与动态用户画像构建，为精准推荐提供了核心依据。协同过滤、内容-based、深度学习等推荐模型的应用与融合，大幅提升了推荐的精准性与个性化。大数据技术的应用不仅提升了用户购物体验与平台运营效率，更推动了电商行业的转型发展。

## 参考文献

- [1] 邓梅霜. 基于大数据技术的电商用户行为分析与精准营销策略研究[J]. 营销界, 2025, (11): 46-48.
- [2] 李旭方. 大数据技术驱动下的大模型在电商求购交易模式中的应用[J]. 大数据时代, 2025, (05): 70-74.
- [3] 王宁宁. 大数据与人工智能在电商运营模式中的应用及影响[J]. 商业经济研究, 2025, (02): 38-41.
- [4] 王晓莹. 收益管理视角下大数据技术在电商平台动态定价中的应用[J]. 中国管理信息化, 2024, 27(18): 95-97.
- [5] 陈梅芳. 大数据技术在电商企业中的应用分析[J]. 现代商业研究, 2024, (16): 179-181.

作者简介：杨国波（1994.01.07-），男，汉，广东揭阳人，大学本科，研究方向为大数据社交网络分析。  
袁月婷（1997.10.21-），女，汉，四川广安人，大学本科，研究方向为机器学习及数据挖掘。  
苏宇航（2001.03.22-），男，汉，山东济南人，大学本科，研究方向为大数据可视化分析。