

电子商务用户行为可视化分析

吴淑梅, 周 宓

(泉州师范学院数学与计算机科学学院, 福建 泉州 362000)

摘要: 用户行为数据是电商企业决策的重要依据。阐释数据分析领域中的AARRR(获取用户、提高活跃度、提高留存率、获取收入、自传播)模型、漏斗分析模型以及用户行为分析方法。针对用户行为的基础指标和偏好,通过多个维度进行全面分析,并将得出的结论以图形化方式展示。最终,通过漏斗图的形式展现用户购物流程中的转化率。电子商务企业可以利用用户行为研究,针对个体提供定制化服务,进而实现精细化的市场决策。

关键词: 用户行为; AARRR(获取用户、提高活跃度、提高留存率、获取收入、自传播)模型; 漏斗图; 转化率

中图分类号: TP18; TP274 **文献标志码:** A **文章编号:** 1671-1807(2025)03-0335-07

在互联网时代,用户行为数据已成为企业决策的关键参考因素。用户行为指的是消费者在电商平台上产生的行为,如登录、浏览商品、加入购物车、收藏、点赞、分享、观看视频、跳过视频等。用户行为研究是一种探究用户在使用产品或接受服务过程中的行为模式、习惯以及偏好倾向的手段。通过对用户行为数据的深入剖析,能够深刻洞察用户需求,掌握用户偏好,进而精准推出符合用户期望的产品与服务。用户行为受到诸多因素影响,如产品特性、服务品质、用户需求、市场竞争等。通过对用户行为的变化进行深入研究,能够洞察这些因素如何影响用户行为,从而对产品和服务进行优化,提升用户体验。此外,用户行为的变化也能反映市场的动态,进而揭示市场趋势及企业战略调整,为企业决策制定提供关键支撑。

用户行为可视化分析是一种利用图形和图表等手段,以直观方式呈现用户行为分析结果的方法。此方法有助于深入洞察用户行为的演变,挖掘用户行为变化的规律,识别用户兴趣和需求。热力图、词云图、时间序列图等是目前广泛应用的可视化分析技术。

国内在电子商务用户行为可视化分析的研究主要集中在点击流分析、购买路径分析、用户评价分析、异常行为检测等方面。李强等^[1]提出,通过热图分析用户点击热点,可以直观地看到用户在页面上的关注区域,从而优化网页设计。王明等^[2]通过

漏斗图分析发现,用户在加入购物车和支付环节的流失率最高,建议改进结算流程,如简化支付步骤和增加多种支付方式,以减少用户流失。张华等^[3]提出,利用可视化技术可以直观展示异常用户的行为模式,如频繁的订单取消或大量的低价值订单,开发了一种基于图表的监控系统,有效识别了电商平台上的异常行为。

国外在电子商务用户行为可视化分析方面的研究相对更加系统和多样,主要集中在综合用户行为分析、实时分析与预测、用户行为模式识别等方面。国外研究通常结合多种数据源,如社交媒体、浏览器日志等,进行综合分析。Huang等^[4]提出一种多源数据融合的方法,通过结合社交媒体数据和点击流数据,全面分析用户行为。Gao和Wu^[5]利用深度学习算法,分析了用户在电商平台上的行为模式,发现了用户偏好的潜在规律。

尽管国内外在这一领域已有大量研究,但仍有诸多可改进之处和新的研究方向。比如传统的用户行为分析往往忽略了用户的情境信息,如时间、地点和设备类型;如何在保护用户隐私的前提下进行有效的行为分析成为一个重要课题;不同文化背景的用户在电子商务平台上的行为模式可能存在显著差异,但相关研究较少。

综合以上分析,电子商务用户行为分析还需关注以下几方面:①加强数据收集与分析,通过多渠道获取用户行为数据,并利用大数据分析技术深入

收稿日期: 2024-06-25

基金项目: 福建省中青年骨干教师教育科研项目(科技类)(JAT220828, JAT210312)

作者简介: 吴淑梅(1981—),女,福建泉州人,硕士,讲师,研究方向为计算机程序设计及人工智能;通信作者周宓(1981—),女,福建泉州人,硕士,讲师,研究方向为计算机程序设计及人工智能。

挖掘用户需求和市场趋势。②提升个性化推荐算法的准确性,优化用户体验,提高转化率。③强化隐私保护措施,确保用户信息的安全,建立完善的监管机制。④推动跨平台合作与数据共享,打破数据孤岛,促进整个行业的创新和发展。⑤培养专业人才和团队,加大对相关领域的投入和支持力度。这些建议旨在促进电子商务行业的健康发展,提高用户满意度和忠诚度,为行业创造更多价值。

本文结合这些问题,在整理分析国内外现有文献的基础上,结合用户的情境信息,从用户行为的基础指标和偏好等多维度分析电子商务的用户行为,并将结果可视化,以精确地捕捉到用户行为的变化,探究用户行为变化的成因及其所产生的影响。

1 基本理论

1.1 AARRR 模型

AARRR 模型是一种针对用户生命周期的框架,它旨在通过收集和获取用户数据来优化产品或服务。该模型涵盖了以下 5 个方面:用户获取(acquisition)、用户激活(activation)、用户留存(retention)、用户付费(revenue)以及用户推荐(referral)^[6]。这些环节共同构成了一个完整的用户生命周期管理策略,有助于提升用户体验和业务成长。

用户获取,亦称为客户获得,其核心目标在于吸引新用户进入产品或服务领域,即通常所述的市场推广活动。关键指标涵盖了新增用户数量、获客成本(CAC)以及用户转化率等方面。

在用户激活阶段,目标是指导用户完成产品的核心功能体验,从而将其转化为产品的活跃用户。普遍应用的策略涵盖了改善产品界面与交互体验、提供新手导航教程、设置有吸引力的首次使用激励以及个性化推荐内容。关键指标涵盖了活跃用户数量、激活率以及用户完成核心功能的比例。

用户留存的主要目标是确保用户持续参与并防止用户流失。常用策略涵盖定期发布有价值的资讯、优化产品功能及性能、提供卓越的客户服务、举办用户互动活动等。核心指标有留存率、用户流失率、用户满意度调查结果以及用户反馈情况。

在用户付费阶段,主要目标是通过多种手段将用户转化为收益,从而提升产品的盈利能力,并实现商业价值的最大化。在数字化时代,变现途径呈现多样化,主要包括广告营收、订阅费用、虚拟产品交易以及实体商品销售等。

用户推荐阶段的目标是借助现有用户的社交影响力实现新用户的引进,用户推荐营销被认为是

一种具有持久性和效益性的增长策略。关键性能指标涵盖了推荐用户的数量以及新用户的活跃程度。

1.2 漏斗分析模型

漏斗分析模型是一种普遍应用于用户体验分析的策略,该方法通过将用户行为流程细分为若干环节^[7],并对每个环节的转化率进行深入研究,以此来衡量用户体验的质量。漏斗分析模型在日常数据运营领域,如流量监控和产品目标转化等方面得到广泛应用。该模型能够帮助企业掌握用户在各个阶段的使用情况,从而优化产品或服务。

漏斗分析方法一般包含以下几个环节。

(1)定义漏斗。首先需要确定用户行为的主要过程,并明确每个步骤的目标和转化目标。

(2)数据收集。需要收集和分析用户行为数据,包括用户进入网站或应用程序的时间、浏览页面、完成特定任务等。

(3)计算转化率。根据漏斗中的各个步骤,计算每个步骤的转化率,即成功完成特定步骤的用户比例。

(4)分析流失点。通过漏斗分析,可以看到用户在哪一个环节流失了,并分析其原因。

(5)优化产品或服务。根据分析结果,对产品或服务进行优化,提高用户满意度和忠诚度。

漏斗分析模型是一种高效的数据分析技术,能够协助产品团队深入洞察用户行为模式,进而实施更加精准的策略决策。通过持续改善产品或服务,优化用户体验,能够帮助用户更高效地实现目标任务,进而促成更大的商业收益。

2 如何进行用户行为分析

2.1 事件分析

行为事件分析是数据分析领域中的关键研究方向,旨在探讨行为事件对产品的影响及其程度。用户注册、激活、浏览产品详情页以及购买等行为的发生频率、顺序和持续时间,都是需要关注的。行为分析在网络平台或应用程序的运营中扮演着至关重要的角色,旨在了解用户如何使用产品及服务,以及哪些因素会对用户行为产生影响。

行为事件分析法展现出卓越的筛选、分组和聚合功能,逻辑结构清晰且操作简便^[8],因此在金融领域、网络安全、电子商务等多个行业中得到了广泛应用。行为事件分析法通常包括事件定义与选择、下钻分析、解释与结论等阶段。

事件指某一用户在特定时刻、特定地点,通过

特定方式实现了具体行为。这个概念涵盖了5个核心要素,分别是“who”“when”“where”“what”和“how”。

而参与事件的主体是 who,对于未登录用户,则可能为 Cookie、设备 ID 等匿名 ID;在网络环境中,用户身份可通过实际的用户标识符进行验证。对于电子商务平台而言,“who”一般代表登录的用户名。当谈论事件发生的具体时间时,通常会以年、月、日、秒、时、分的形式进行记录。事件发生地点,即 where,通过分析用户登录时的 IP 地址,可以推断出用户所处的省份和城市。在分析用户参与特定事件的方式时,需考虑多种因素,如所使用的设备类型、浏览器版本、应用程序版本以及来源渠道等。“what”描述用户进行的活动的具体内容。例如,在电子商务平台中,“购买”行为涵盖了但不限于消费者选购的商品种类、商品名称、商品分类、购买数量、支付金额、支付方式等。

下钻分析和精细化条件筛选是高效的行为事件分析。钻取在数据分析过程中是一个不可或缺的环节。该技术通过调整数据维度的层次和改变分析粒度,协助企业从数据中挖掘更加全面、隐蔽的信息。钻取操作可分为向上钻取(roll up)和向下钻取(drill down)。向下钻取是在分析过程中增加维度,深入探索数据层面,发掘更丰富的数据价值,从而及时做出更为精准的决策。

通过事件分析,企业能够深入洞察用户行为,从而精准调整产品或服务,进一步提升用户满意度和忠诚度。除此之外,事件分析还可以用于监测安全事故、发现异常行为等。

2.2 页面点击分析

页面点击分析技术被广泛应用于展示页面区域内各元素的点击密度,并通过图形化表现。此技术能够精确评估用户与产品交互的深层次联系,实现产品跳转路径分析,完成产品页面之间的深层次需求挖掘。通过结合页面点击分析和其他数据模型,可以从全方位的角度挖掘数据价值。这一过程中,将直观地对比并分析用户在页面上的关注度、浏览频次、人数以及页面内各个可点击元素所占的百分比。页面点击分析常被应用于主页、活动页面、产品详情页等包含复杂交互逻辑的网页。大部分企业通过固定埋点的方式来收集用户页面点击信息。所涉及的数据维度主要包括页面访问量(PV)、浏览总人数(UV)以及点击人数比例等。

3 数据预处理

本文以淘宝用户行为数据为例,采用 Windows 10 操作系统,选用 Python 3.11 作为编程语言,以 Anaconda3 作为数据分析工具。在此基础上,运用 numpy 和 pandas 库对数据进行处理,通过 matplotlib 库和 pyecharts 库完成数据可视化实现。

3.1 数据描述

本文使用的数据集是来自阿里云天池大赛提供的公开数据集。该数据集涵盖了2017年11月25日至2017年12月3日期间,约一百万个具有行为的随机用户的所有行为。数据集内有987 994名不同用户,9 439种商品类别,以及4 162 024件商品^[9]。用户行为的总数达到了100 150 807次。用户的行为包含了点击、购买、将商品加入购物车以及收藏等。鉴于数据集规模庞大,实验过程的运行效率较低,因此本文仅选取原始数据集中前50万条用户行为记录作为研究对象。表1展示了淘宝用户行为数据集所记录的一部分信息。

表1显示,数据集共包含5个字段,分别为用户ID(userid)、商品ID(itemid)、商品类目ID(categoryid)、用户行为类型(type)以及行为发生时间戳(timestamp)。用户行为种类被归纳为点击(pv)、购买(buy)、加入购物车(cart)以及收藏(fav)^[10]。用户行为数据的字段信息见表2。

由表2可知,该数据集中的各个字段均未出现缺失值,其中userid、itemid、categoryid以及timestamp 4个字段均采用int64数据类型。

表1 天猫用户数据记录(部分)

用户ID	商品ID	商品类目ID	用户行为类型	行为发生时间戳
1	2 268 318	2 520 377	pv	1511544070
1	2 333 346	2 520 771	pv	1511561733
1	2 576 651	149 192	pv	1511572885
1	3 830 808	4 181 361	pv	1511593493
1	4 365 585	2 520 377	pv	1511596146

表2 用户行为数据的字段信息

序号	列名	记录数	类型
0	userid	500000 non-null	int64
1	itemid	500000 non-null	int64
2	categoryid	500000 non-null	int64
3	type	500000 non-null	object
4	timestamp	500000 non-null	int64

3.2 数据预处理

在进行数据分析之前,需要对原始数据进行一些预处理操作。鉴于时间戳类型为 int64,不利于从中提取时间信息进行数据分析,因此有必要将其转化为日期型年月日格式,以便于后续根据各个时间维度进行分析。同时,需要排除掉不在研究时间范围内的数据,并对索引进行重置。经过处理的数据。处理后的数据集见表 3。

表 3 显示,时间戳类型已经转换为 datetime64,数据值包含了年、月、日、时、分、秒等信息,并被存储在名为 buytime 的新列中。数据集规模为 499 778,与处理前相比,仅减少了 222 条,因此选择直接删除。

随后,对于购买时间列 buytime,提取其月份、日期、时间、小时和星期,分别以 month、date、time、hour 和 weekday 作为新的列名。同时,剔除不必要的时间戳列 timestamp。结果见表 4。

3.3 用户行为的整体分析

3.3.1 用户行为统计

首先对用户行为数据进行统计,包括点击(pv)、购买(buy)、添加到购物车(cart)和收藏(fav) 4 种行为的数量及其在总体中的占比,并以饼图形式展示,如表 5 和图 1 所示。

从图 1 可以发现,有点击行为的人数占了大多数,占比高达 89.87%,而最后有购买行为的人数非常少,仅仅占了 2.04%。

3.3.2 跳失率

跳失率是评估网站或店铺访问质量的关键指标,其定义为用户在进入后仅浏览一个页面即离开

所产生的访问次数,与该页面总访问次数之比^[11]。网站的吸引力、用户体验和营销成效可以通过跳失率的高低来反映。较高的跳失率,反映出网站或店铺的页面缺乏吸引力或营销策略效果欠佳,导致用户体验不佳;相反,则意味着这些方面表现较为优异。

流量指标 UV 指独立访客的数量,而 pv 则表示点击次数。一位独立访客可能会产生多个点击行为。平均访问量,即 pv/UV,是常用的一个度量指标。简而言之,平均访问量是指每位用户在浏览网页时平均点击的次数。平均访问量是衡量网站吸引力的一个重要指标。黏度越高,点击率亦随之提升。通过对 UV 与 pv 的数据分析,电商企业能够洞察用户行为模式和流量状况,进而实施有针对性的策略,以提升用户忠诚度。UV 和 pv 的统计数据见表 6。该数据

表 5 用户行为的数量

用户行为类型	数量
pv	449 135
cart	27 509
fav	12 950
buy	10 184

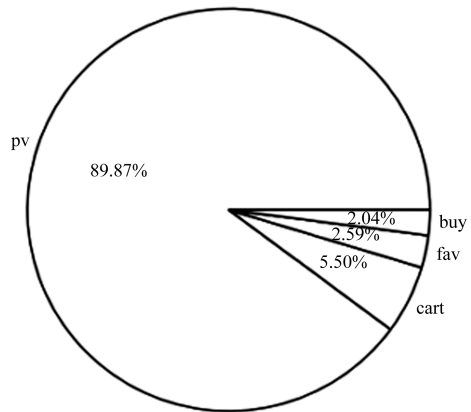


图 1 各类人员行为的比例

表 6 流量指标数据

流量指标	数据
独立访客数	4 856
点击量	449 135
平均访问量	92.490 733

表 3 处理后的数据集 1

序号	列名	记录数	类型
0	userid	499778non-null	int64
1	itemid	499778non-null	int64
2	categoryid	499778non-null	int64
3	type	499778non-null	object
4	timestamp	499778non-null	int64
5	buytime	499778non-null	datetime64[ns]

表 4 处理后的数据集 2

用户 ID	商品 ID	商品类目 ID	用户行为类型	行为发生时间戳	月份	日期	时间
1	2268318	2520377	pv	2017-11-25 01:21:10	11	2017-11-25	01:21:10
1	2333346	2520771	pv	2017-11-25 06:15:33	11	2017-11-25	0:15:33
1	2576651	149192	pv	2017-11-25 09:21:25	11	2017-11-25	09:21:25
1	3830808	4181361	pv	2017-11-25 15:04:53	11	2017-11-25	15:04:53
1	4365585	2520377	pv	2017-11-25 15:49:06	11	2017-11-25	15:49:06

集的平均访问量大约为 92.49 次。

基于前述分析,计算出点击行为用户占总用户数的比例,即跳失率,结果为 5.91%。据此,可以推断,在浏览商店的用户群体中,约有 5.91%的人在进入网页后立即退出,未展开任何其他互动或购买行为。商家可通过提升页面响应速度、降低弹窗频次、精细化商品展示以及优化购物体验等策略,降低用户流失率,增强网站黏性,从而实现转化率的提升。

3.4 用户偏好分析

用户偏好体现为用户在选购商品或接受服务过程中所展现出的倾向性,这种倾向性通常与个人需求、价值观念、经历等因素密切相关。用户偏好可以体现为对特定商品类别、品牌或价格区间的倾向,亦可表现为浏览登录习惯。本节通过分析用户行为发生的时间来分析用户偏好。

3.4.1 按日期进行 pv、UV 统计

分别计算网站单日点击数及独立访客总数,并绘制折线图,如图 2 所示。

通过图形对比分析,可以观察到 pv 和 UV 的变化趋势大致一致,这与常规现象相吻合。通过网页

点击次数、独立访客人数随时间的变化趋势可以看出,从 2017 年 11 月 25 日到 2017 年 12 月 1 日点击次数和独立访客人数变化较为稳定,只有少量波动;2017 年 12 月 2 日日点击量和独立访客数均出现大幅上升,2 日、3 日两天保持高点击量和高访客数。对于该现象的原因分析,可能与 2017 年 12 月 2 日及 12 月 3 日为周末休息日,或者 2 日、3 日商家举办了特定的促销活动有关。

3.4.2 按小时进行 pv、UV 统计

点击量 pv 与独立访客数 UV 在 24 h 内的变化趋势如图 3 所示。

在对点击量和独立访客数进行按小时统计的变化趋势图中,观察到每日 00:00—06:00 的点击量与独立访客数相对较低,其最低点均出现在 04:00 左右。用户活跃度最高的时段是 21:00—22:00。在 10:00,网络流量呈现轻微上升趋势,而点击数和独立访客数则在 20:00—22:00 时达到顶峰。该现象揭示了网站使用者的行为模式,其中用户活跃的时间段主要集中在 20:00—22:00,而 04:00 则是用户相对较为不活跃的时间。

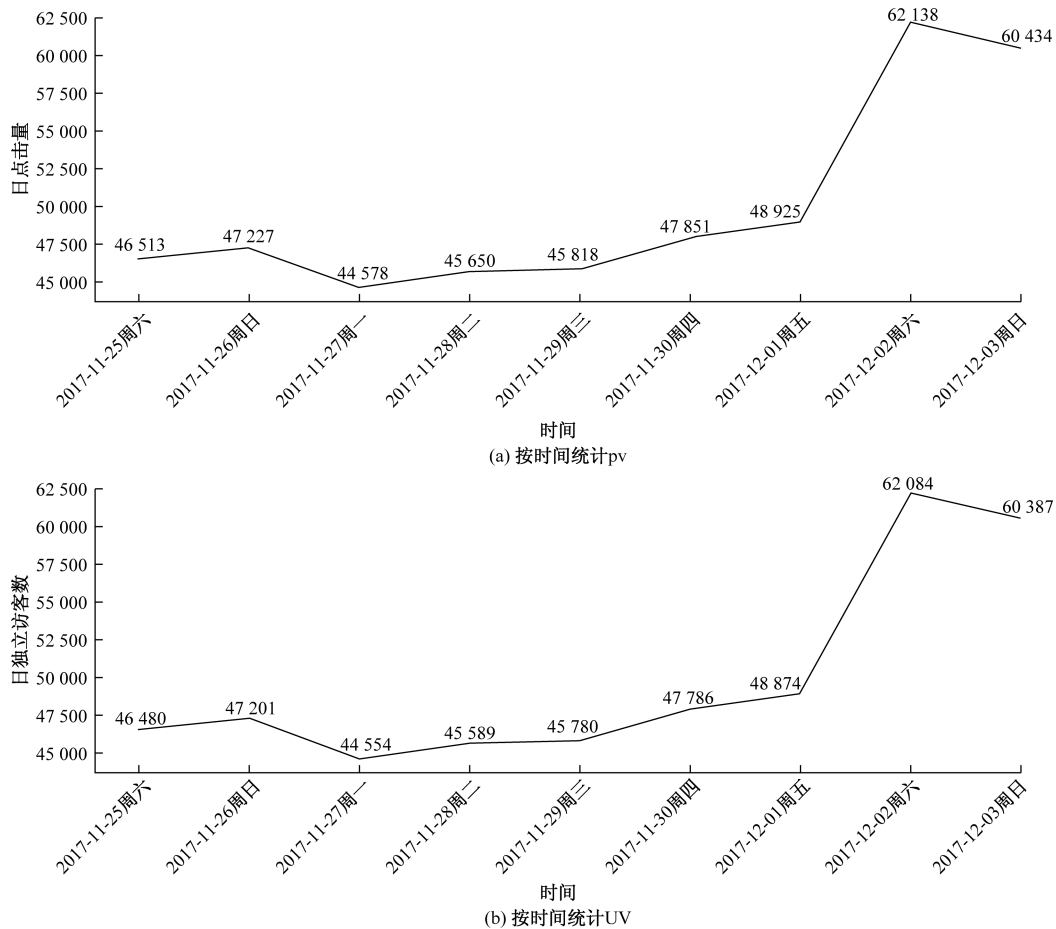


图 2 按时间计算 pv、UV 数据统计

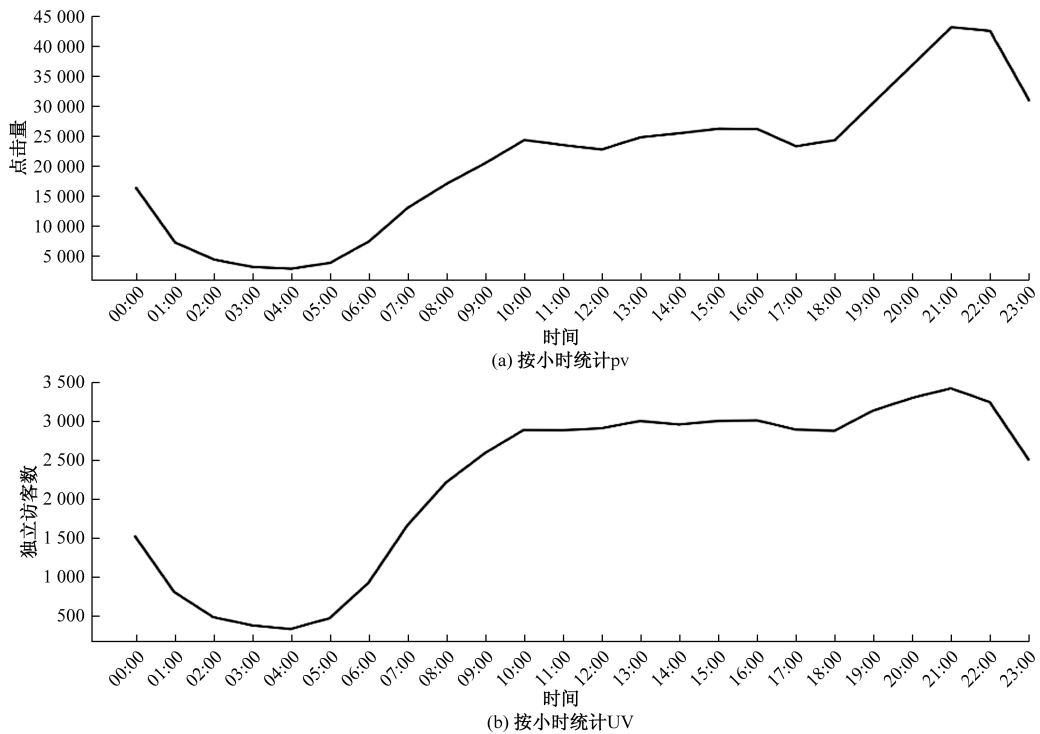


图 3 按小时进行 pv、UV 统计

3.4.3 留存率

留存率是衡量网站、互联网应用或网络游戏运营状况的关键指标,具体定义为在统计周期(周/月)内,每日活跃用户数在第 N 日仍然使用该应用的用户占比的平均值。 N 通常选取 2、4、8、15、31,这些数值分别对应于次日留存率、3 日留存率、周留存率、半月留存率以及月留存率。用户留存率作为衡量用户黏性的指标,其表现在当 N 值较大、留存率较高时,反映出用户黏性增强。

在本文中,对 3 日留存率进行计算,结果见表 7。

表 7 显示,用户留存率超过 65%,这意味着大约有 2/3 的根据上述图表分析,用户留存率超过 65%,这意味着大约有 2/3 的用户在首次购买后的 3 天内会再次进行购物。留存率在 2017 年 11 月 29 日及 11 月 30 日达到高峰,而随后的 3 天恰逢周末。此前的研究结果与本次观察相吻合,周末期间消费者购买行为的数量呈上升趋势,推测这可能与商家

表 7 日留存率

日期	留存率
2017-11-25	0.772 170
2017-11-26	0.673 858
2017-11-27	0.656 667
2017-11-28	0.673 913
2017-11-29	0.965 909
2017-11-30	1.808 800

在周末所开展的促销活动有关。这一现象揭示了部分消费者对淘宝平台的忠诚程度,然而从整体来看,消费者复购行为仍有较大的提升空间。商家可依据消费者购买偏好,优化购物流程,提升客户服务水平,定期开展推广活动,为消费者提供更精确的推荐服务,增强用户黏性,从而提高复购率。

3.5 绘制用户行为转化漏斗图

漏斗图通过梯形面积来反映某个环节业务量与前一个环节之间的差距。在漏斗图中,初始值始终为 100%,并且在各个阶段依次递减。在实际应用中,漏斗图对于监测用户转化率和留存率具有较高的适用性,同时能够揭示线性流程中的关键瓶颈。通过分析漏斗各阶段的用户数量变化,可以直观地识别出问题所在的环节,并据此做出相应决策。各阶段用户行为统计数据见表 8。

采用漏斗图的方式来展示浏览、加入购物车、下单以及完成购买的数量,如图 4 所示。通过图表,可以直观地观察到在各个环节中用户数量的变化趋势。

表 8 用户行为统计数据

序号	行为	数量
0	浏览	449 135
1	加入购物车	27 509
2	下单	12 950
3	购买	10 184

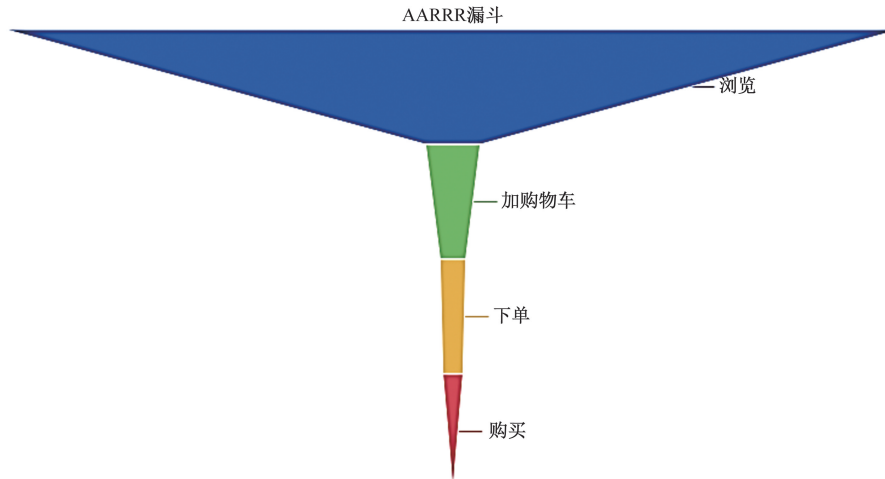


图4 用户行为转化漏斗

4 结论

本文对真实淘宝用户行为数据进行了深入剖析,详尽阐述了数据预处理与分析流程,并将结果以可视化的形式展现。

用户行为可视化分析是探究用户需求和行为变化的关键方法。借助于用户行为的可视化剖析,得以深入洞察用户需求,从而对产品与服务进行优化调整,以提升用户体验。在未来,伴随数据技术的持续进步,用户行为可视化分析的精确度和全面性将得到显著提升,从而为企业决策制定提供更为坚实的依据。伴随大数据与人工智能技术的进步,对用户行为的分析应用将更加广泛和深入。

参考文献

- [1] 李强,王明,刘红. 基于热图分析的电商用户行为研究[J]. 电子商务研究, 2020, 36(4): 487-501.
- [2] 王明,张华,李强. 电商平台用户购买路径分析及优化[J]. 管理科学学报, 2021, 42(1): 74-89.
- [3] 张华,李强,王明. 电商平台异常用户行为分析[J]. 计算机科学与应用, 2018, 40(2): 123-135.
- [4] HUANG J, WANG X, LIU Y. Heatmap and path analysis in user behavior visualization[J]. Information Visualization, 2021, 20(1): 56-72.
- [5] GAO H, WU Y. User behavior analytics in e-commerce: challenges and future directions[J]. Electronic Commerce Research, 2021, 22(2): 345-367.
- [6] 吉祥. 探索“增长黑客”理论在信用卡数字化经营中的应用[J]. 中国信用卡, 2021(8): 40-46.
- [7] 用户行为分析模型实践(二): 漏斗分析模型[EB/OL]. (2024-01-22) [2024-05-20]. <https://developer.baidu.com/article/detail.html?id=2851250>.
- [8] 解析常见的数据分析模型: 行为事件分析[EB/OL]. (2024-05-20) [2024-06-25]. <https://developer.aliyun.com/article/230259?spm=5176.26934562.main.6.11d07749enNdd5>.
- [9] 杨非非. 基于客户旅程分析的淘宝用户购买行为预测研究[D]. 大连: 东北财经大学, 2022.
- [10] 王治博. 基于 Hadoop 的电商平台用户数据挖掘研究[D]. 北京: 华北电力大学, 2021.
- [11] 付森,吴慧佳,李子木. 数据驱动下网络消费者购买行为分析研究[J]. 中国管理信息化, 2022, 25(13): 60-62.

Visualization Analysis of E-commerce User Behavior

WU Shumei, ZHOU Mi

(Mathematics and Computer Science, Quanzhou Normal University, Quanzhou 362000, Fujian, China)

Abstract: User behavior data is an important basis for decision-making in e-commerce enterprises. The AARRR (acquisition, activation, retention, revenue, referral) model, funnel analysis model, and user behavior analysis methods in the field of data analysis were explained. Based on the basic indicators and preferences of user behavior, a comprehensive analysis was conducted through multiple dimensions, and the conclusions drawn were presented in a graphical manner. Finally, the conversion rate in the user shopping process was presented in the form of a funnel chart. E-commerce enterprises can utilize user behavior research to provide customized services for individuals, thereby achieving refined market decision-making.

Keywords: user behavior; AARRR(acquisition, activation, retention, revenue, referral) model; funnel plot; conversion rate