

基于评分规则的电商平台评论系统仿真研究*

• 杨 磊¹ 张伟杰¹ 郝彩霞²

(1 华南理工大学电子商务系 广州 510006; 2 常州大学商学院 常州 213159)

【摘要】随着网购普及,消费者日益依赖在线评论来获取信息,特别是产品评分,其已成为衡量卖家水平的重要指标。因此,深入理解电商平台的评分规则,即平台是如何统计评分的,显得尤为重要。现有评分规则主要有两种:传统评分规则和动态评分规则。尽管评分规则的影响深远,但相关研究仍显匮乏。本研究采用仿真实验方法,对比两种规则对评论系统设计的影响。研究发现,传统评分规则在提升评论系统效率性方面通常更佳,例如加快评论数量的积累和减少尾部卖家的客户流失。动态评分规则在促进竞争公平性方面总是更有效,特别是当计分周期极短时,动态评分规则反而能推动搜索型产品更快地积累评论数量。因此,平台应根据自身的经营目标,选择适合的评分规则。

【关键词】评分规则 产品评分 评分近期性 评论系统设计 仿真实验

中图分类号: F224 文献标识码: A

1. 引言

随着电商的蓬勃发展,网购已成为消费者青睐的购物方式。其中,在线评论作为消费者获取产品信息的重要渠道,发挥着关键作用。根据 BrightLocal 的调查,只有 3% 的消费者“从未”读过评论^①,这深刻揭示了评论在消费者心中的核心地位。诸多研究亦表明,评论帮助消费者了解产品详情 (Chen & Xie, 2008),减少购买中的不确定性 (Dellarocas, 2003; Feng et al., 2019; Wang et al., 2018),并引导他们做出更理性的决策 (Kwark et al., 2014; Wu & Wu, 2016; Zimmermann et

* 基金项目:国家自然科学基金 (No. 72471096); 广东省基础与应用基础研究基金联合基金 (No. 2025A1515012184); 中央高校基本科研业务费专项资金 (No. CXTD202406); 国家自然科学基金青年基金 (No. 72301038); 江苏省社会科学基金 (No. 23GLC017)。

通讯作者: 郝彩霞, E-mail: hcx94@cczu.edu.cn.

^① BrightLocal. Local consumer review survey 2024 [EB/OL]. <https://www.brightlocal.com/research/local-consumer-review-survey/>.

al., 2018)。

产品评分是在线评论中的核心量化指标,对买卖双方都至关重要。对于消费者来说,评分是快速判断产品质量的重要依据。BigCommerce 的调查显示,对于评分优秀的卖家,消费者愿意额外支付 31% 的价格^①。这表明更高的评分往往与更多的销售额相关 (Chevalier & Mayzlin, 2006; Ren et al., 2018; Scholz et al., 2017)。对于卖家而言,评分是电商平台促销活动的参与门槛之一,决定着产品的竞争力^②。更重要的是,它还直接影响产品在搜索结果中的排序位置 (Jin et al., 2023; Zhang et al., 2016), 高分产品会被优先展示,而低分产品则被减少甚至失去曝光机会 (Derakhshan et al., 2022)。

考虑到产品评分的重要性,卖家需要深入了解电商平台的评分规则,即平台是如何统计评分的。按评分统计的时间划分,目前存在两种常见的评分规则。一种被本研究命名为“传统评分规则”,该规则累计计算所有已发表评分的算术平均值。在进行评分建模时,现有研究都遵循这一符合直觉的传统办法,如 Park 等 (2021)、Besbes 和 Scarsini (2018),以及 Huang 等 (2021)。在现实中,中国的苏宁易购^③、美国的百思买和亚马逊^④也在应用这种规则。另一种被商界称为“动态评分规则”,该规则强调评分的近期性,在每次统计评分时滑动计算最近一个计分周期内的评论数据。目前,这种规则更受电商平台的青睐。在国内,天猫的店铺评分^⑤、京东的产品评分^⑥都只统计最近 180 天内的评分,而拼多多的店铺评分只计算最近 90 天内的评分^⑦。在国外,eBay 会分别展示最近一个月、最近六个月和最近一年内各个评论效价的数量,以及最近一年内的多维评分。亚马逊自 2015 年使用新算法以来,尽管产品评分依然从第一条起统计,但是算法显然为近期评分赋予了更多权重^⑧。实践者认为,在动态评分规则下,消费者能获取更新鲜的评论信息,卖家受历史低分的影响更低^⑨。

然而,评分规则的潜在影响并未明确,学术界对此的探讨也尚显不足。评分规则的不同可能导致评分数值产生差异,进而左右消费者的购买决策与卖家的销量分配,最终影响电商平台的业绩。因此,本研究旨在深入探索平台的评分规则,特别是动态评分规则,以揭示它们各自对评论系统绩

① BigCommerce. The inside scoop on e-commerce reviews: Why they matter and how to make the most of them [EB/OL]. https://www.bigcommerce.com/blog/online-reviews/#h2_who_is_reading_online_reviews.

② 幕思城. 抖音小店运营规范是什么 (抖音小店运营规则介绍) [EB/OL]. <https://www.musiccheng.com/news/i449964.html>.

③ 苏宁易购帮助中心. 本地商户常见问题 [EB/OL]. <http://image.suning.cn/OtherArea/bzzx/bdshejw.html>.

④ Feedbackwhiz. How are Amazon product ratings calculated? [EB/OL]. <https://www.feedbackwhiz.com/blog/how-does-amazon-calculate-product-ratings>.

⑤ 商家服务大厅-淘宝网. 店铺动态评分 (DSR) 计算逻辑是什么? [EB/OL]. https://helpcenter.taobao.com/servicehall/knowledge_detail?knowledgeId=1119291.

⑥ 商家帮助中心-京东. 【商智使用指南】经营分析功能模块 (下) [EB/OL]. <https://helpcenter.jd.com/vendor/issue/887-4352.html>.

⑦ 幕思城. 拼多多 dsr 评分多久出来 (店铺出动态评分规则) [EB/OL]. <https://www.musiccheng.com/news/i431000.html>.

⑧ WIRED. What do Amazon's star ratings really mean? [EB/OL]. <https://www.wired.com/story/amazon-stars-ratings-calculated/>.

⑨ 动态评分_百度百科 [EB/OL]. https://baike.baidu.com/item/%E5%8A%A8%E6%80%81%E8%AF%84%E5%88%86?fromModule=lemma_search-box.

效的具体影响, 为平台在评论系统设计方面提出管理建议。

2. 文献综述

在线评论和产品评分在电商研究中的重要作用已受到广泛认可。本研究基于两个主要假设展开研究: 一是在线评论的失验效应, 二是产品评分影响排名。首先, 根据期望失验理论, 失验效应 (disconfirmation effect) 即消费者对同一产品的预期估值与实际体验之间存在的差异 (Ho et al., 2017), 会影响他们的评论发表意愿 (Guan et al., 2023; Li et al., 2020)。当消费者感知到强烈的失验效应时, 出于利他主义, 他们会更有动机留下评论, 为其他潜在消费者提供更多信息。因此, 在为消费者的网购行为建模时, 本研究特别考虑了这一因素。此外, Ho 等 (2017) 的研究还揭示了多种因素对失验效应感知的调节作用, 如购买与到货之间的时间距离、同行消费者的评价分歧等。本研究则进一步发现, 作为一种量化展示工具, 评分规则也会对这种效应的感知产生间接影响, 尤其是动态评分规则, 其往往起到缓和作用, 这一发现对 Ho 等 (2017) 的研究进行了有益补充。其次, 产品评分在产品排名机制中扮演着核心角色。尽管学者们常利用评论信息优化产品排序和推荐算法 (张瑾等, 2019; 余本功等, 2022; Ren et al., 2021; Poppink et al., 2023), 但实际的排名机制远比这复杂。平台在构建排名机制时, 除了评分外, 还综合考虑了销量、线下信誉等多种因素 (Jin et al., 2023)。此外, 基于产品多重属性的中心性排名算法 (Scholz et al., 2017) 等算法思路也进一步丰富了排名机制。因此, 在为排名建模时, 本研究将综合考虑这些多元因素, 旨在提高模型的普适性和实用性。

本研究致力于剖析评分近期性的潜在影响, 即近期评分如何影响总体评分的数值呈现。以往研究大多在评论的时间距离 (李启庚等, 2017)、近期评论的价值 (Tandon et al., 2021; Fernandes et al., 2022; Lee et al., 2017) 等方面展开探讨, 这些方面多侧重于消费者对评论时效性的感知。本研究则进一步将焦点细化至评分近期性, 旨在揭示评分规则如何直接影响总体评分的显示, 进而影响产品排名的变动与卖家销量的分配, 并最终作用于评论系统的绩效表现。在评分近期性方向上, 与本研究最为相关的是 Carnehl 等 (2023) 的研究, 他们使用内生的价格和计分周期, 发现提高最近评分的权重可能总是有利于卖家, 但会给消费者带来负面影响, 如更高的价格和更慢的社会学习。然而, 本研究的仿真结果认为, 在外生价格与计分周期的条件下, 这种做法通常有助于缓解失验效应感知, 这可能对消费者有益。同时, 本研究还特别关注了动态评分规则, 这一规则强调近期评分的权重。现有文献在计算总体评分时, 往往遵循学术上的建模传统, 即采用传统评分规则, 对所有评论一视同仁, 不区分时间因素。例如, Besbes 和 Scarsini (2018) 在探讨在线评论的信息传递作用时, 使用了这一规则来建模, 未考虑评论的新旧差异。同样, Huang 等 (2021) 在研究评论对供应链动态的影响时, 也采用了类似方法。在这些研究中, Park 等 (2021) 的研究与本研究尤为相关, 他们基于传统评分规则, 发现了早期评论对产品口碑的演变具有决定性作用。然而, 通过仿真分析剔除早期评论的干扰后, 本研究认为, 他们的结论可能并非完全归因于早期评论的特殊地位, 或许是由传统评分规则本身的性质所致。早期评论的作用在传统评分规则下确实较为突出, 但在动态评

分规则下, 由于近期评分被赋予更多权重, 这种先发优势可能会受到相当程度的削弱。

本研究隶属于评论系统设计领域。当前研究已深入探讨了多维评分系统 (Schneider et al., 2021; Liu et al., 2023; Kong et al., 2020; Chen et al., 2018) 的影响, 并从评分尺度 (Jiang & Guo, 2014)、评分显示格式 (Wang et al., 2019; Jia et al., 2023) 等视觉角度提出了对评论系统的改进策略。然而, 关于评分近期性这一时间因素的研究尚显不足, 采用强调评分近期性的动态评分规则, 是否真的能持续提升电商平台的整体效益, 仍值得深入探讨。因此, 本研究将关注点放在评分规则上, 即近期评分对总体评分数值的影响, 而非近期文本评论, 探讨如何利用这一因素来优化评论系统设计, 以期为平台运营提供新思路。本研究建议平台应根据其特定的运营目标来选用适合的评分规则, 以实现评论系统绩效的间接调控。

3. 研究方法

3.1 模型设置

本节首先对电商平台的评论系统进行建模, 该系统由两个卖家、一个平台和多期消费者共同参与。基于产业组织理论和消费者心理学等理论, 这一模型描述了消费者的网购行为与评论系统的运作机制。图 1 呈现了这一机制, 其中实线代表消费者的网购流程, 而虚线则表示评论系统的影响路径。

具体而言, 整个机制可分为以下七个关键步骤: (1) 市场进入: 在每一期期初, 一定数量的新消费者进入市场。(2) 产品估值: 进入市场后, 消费者根据产品 A 和 B 的现有在线评论, 分别对二者的质量和匹配程度进行估值。他们的估值受上一期的评分方差与累计评论数量影响, 而评分的近期性则由评论数量随时间推移的积累来刻画。(3) 购买决策: 消费者根据预期效用做出购买决策。产品排名将决定他们的购买优先级, 排名靠前的产品会被优先考虑。此步骤将产生两种结果, 即购买某一产品或者都不买。对于那些选择不购买的消费者, 他们将直接进入第五步。(4) 退货决策与评分决策: 那些购买了产品的消费者需要根据实现效用选择是否退货和是否进行评分, 以及评多少分 (好评、中评或差评)。其中, 退货行为会撤回已成交的销量, 并影响产品排名, 而评论发表意愿则会受失验效应的影响。(5) 市场退出: 无论消费者是否购买产品, 他们最终都会退出市场。(6) 评论数据更新: 在消费者退出后, 平台根据设定的评分规则, 更新所有产品的评论数据, 包括计算显示评分、评分方差和累计评论数量, 并将这些最新信息展示给下一期的消费者。其中, 显示评分反映了消费者对已买产品的满意度, 评分方差代表了消费者口碑的分化程度, 而累计评论数量则体现了消费者对产品信息的掌握程度。请注意, 评分规则将直接影响评分及其方差的显示, 并间接影响下一期的产品排名与消费者估值。(7) 市场刷新与下一期进入: 随着评论数据的更新, 两个产品将进入下一期, 同时消费者市场刷新。这个过程循环往复, 促成了消费者网购的不断进行。

本节将对各个步骤的具体假设进行详细介绍。在第一步, 对消费者市场进行建模。假设在每一期期初都有固定数量 M 的潜在消费者进入平台来查看产品 A 与 B 。遵循霍特林模型, 每个消费者 i 的

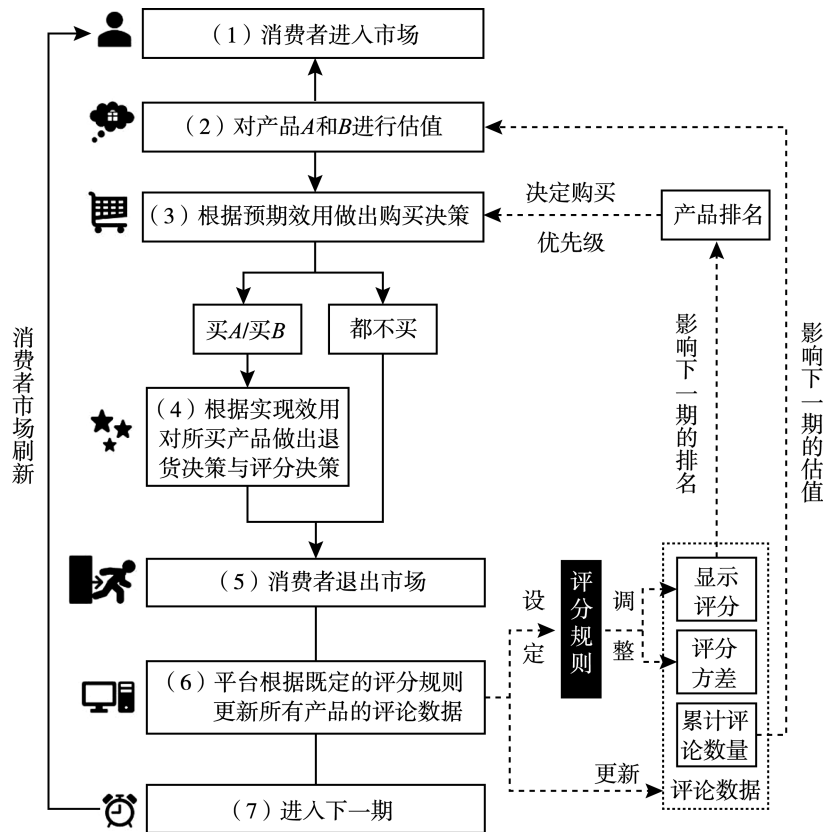


图1 消费者参与网购与评论系统运作的流程

真实偏好位置 x_i 都相互独立, 并均匀分布于长度为 1 的线性市场内, 即 $x_i \sim U[0, 1]$ 。同时, 两个竞争性卖家在同一平台上分别销售产品 A 和 B, 二者的位置 z 分别固定在市场的两端, 即 $z_A = 0, z_B = 1$ 。假设这两个产品的质量 q 和价格 p 均为相等的外生常数, 这意味着卖家不会调整这些变量。为便于后续分析, 质量与价格合并为价值 $v (v = q - p)$ 。

在第二步, 假设消费者在购买前对产品的价值和匹配程度存在不确定性, 因此他们需要依赖评论信息来对产品价值和自身偏好位置进行估值。对于产品的价值估值, 假设每个消费者 i 在购买前都会参考产品 A 和 B 的口碑分化信息来分别形成自己对两个产品真实价值 v_A, v_B 的估计 $\hat{v}_{A,i}, \hat{v}_{B,i}$ 。这种价值估值 \hat{v} 遵循正态分布 (Archak et al., 2011), 其均值等于产品的真实价值 v , 方差 $\sigma^2(\hat{v})$ 则受评分方差 $\sigma^2(R)$ 的影响。考虑到价值估值标准差 $\sigma(\hat{v})$ 与评分标准差 $\sigma(R)$ 之间存在 $\sigma(\hat{v}) = c \cdot \sigma(R)$ 的对应关系 (Chen et al., 2021), 以产品 A 为例 (产品 B 同理), 在时期 t , 某个消费者 i 对该产品的价值估值为 $\hat{v}_{A,i,t} \sim N(v, c^2 \cdot \sigma^2(R_{A,t-1}))$ 。其中, c 表示消费者的单位不匹配成本。其数值越大, 意味着消费者在购买决策中为匹配程度给予越大权重。

对于产品的匹配程度估值, 假设消费者在查看评论前完全不清楚自己的真实偏好位置 \hat{x} 。他们会先认为自己的初始偏好位置为 \hat{y} , \hat{y} 均匀分布于 $[0, 1]$ 上, 并与 \hat{x} 独立。随后, 通过阅读产品 A 和 B 的文本评论, 他们能够逐渐修正自己的偏好位置。为了简化模型, 假设每个消费者都会查看所有

评论, 且每一条评论都是有用的。以产品 A 为例 (产品 B 同理), 在时期 t , 某个消费者 i 在查看产品 A 的评论后对该产品的匹配程度估值为 (Zhao et al., 2022):

$$\hat{x}_{A, i, t} = (1 - w_{A, t}) \cdot \hat{y}_{i, t} + w_{A, t} \cdot x_{i, t} \quad (1)$$

其中, w 表示文本评论的帮助程度, 值域为 $[0, 1]$ 。该变量刻画了评论的近期性, 其数值 $w_{A, t}$ (或 $w_{B, t}$) 随着累计评论数量 $N_{A, t}$ (或 $N_{B, t}$) 的增加而增大, 表明评论的丰富性有助于消费者更准确地评估产品的匹配程度 (Park et al., 2021; Liu et al., 2017), 使估值 \hat{x}_i 更接近真实值 x_i 。

此外, 本研究引入产品体验程度 (degree of experience) 的概念, 用 $\frac{c}{v}$ 来表示 (Chen et al., 2021)。当体验程度较高时, 产品类型表现为体验型产品。这意味着消费者会更加关注产品与个人需求的匹配程度。在这种情况下, 文本评论的重要性会相应提升, 而表面上的评分数据 (如显示评分和评分方差) 对购买决策的影响力则会减弱。相对而言, 当体验程度较低时, 产品类型表现为搜索型产品, 此时消费者会更加关注产品质量的表面数据, 评分的影响力将增强。需要注意的是, 当体验程度适中时, 产品将处于上述两种类型的中间状态, 本研究将这种产品类型称为“半体验型产品”。引入体验程度的设定有助于深入探讨在线评论对不同类型产品的影响。

在第三步, 阐述了消费者购买决策的机制。总的来说, 消费者会先根据产品排名来选择优先查看哪个产品, 再根据预期效用来决定是否购买当前产品。首先, 消费者在平台上搜索产品, 搜索结果会按照产品排名依次展示 A 和 B 两个产品, 而产品排名会根据上一期期末的排名得分 K 确定。排名靠前的产品会显示在搜索结果的第一页, 而排名靠后的则显示在第二页。这种排名机制会导致排名优势的产品抢占排名劣势产品的销量。然后, 消费者会优先查看排名靠前的产品, 并根据产品详情页中的评论信息对该产品进行估值。预期效用 EU 即消费者在购买前对产品的主观感知, 成为他们选择是否购买当前产品的判断标准。在时期 t , 某个消费者 i 对产品 A 和 B 的预期效用 $EU_{A, i, t}$ 与 $EU_{B, i, t}$ 的计算公式分别为:

$$\begin{cases} EU_{A, i, t} = \hat{v}_{A, i, t} - c \cdot \hat{x}_{A, i, t} \\ EU_{B, i, t} = \hat{v}_{B, i, t} - c \cdot (1 - \hat{x}_{B, i, t}) \end{cases} \quad (2)$$

当 EU 的值大于 0 时, 消费者会直接购买该产品; 反之, 则会放弃购买并转向下一页的产品。如果下一页产品的预期效用为正, 则选择购买; 否则, 放弃购买所有产品。此外, 当两个产品的排名得分相等时, 消费者将选择预期效用更高且数值为正的产品。图 2 (a) 描述了这一购买决策流程。

在第四步, 构建消费者退货决策与评分决策的模型。简而言之, 消费者会根据实现效用, 先做出退货决策, 再进行评分决策, 即选择是否发表评分, 并决定评多少分。实现效用 RU 是消费者基于自己的客观需要对产品的主观满意度, 反映了他们在购买后的真实感受。假设消费者在购买后能够了解到产品的真实质量 v 和自己的真实偏好位置 x , 不再存在产品不确定性。在时期 t , 某个消费者 i 对其所买产品 A 或 B 的实现效用 $RU_{A, i, t}$ 与 $RU_{B, i, t}$ 的计算公式为:

$$\begin{cases} RU_{A, i, t} = v - c \cdot x_{i, t} \\ RU_{B, i, t} = v - c \cdot (1 - x_{i, t}) \end{cases} \quad (3)$$

退货决策与评分决策的具体流程如下。首先, 消费者需进行退货决策。当 RU 值低于好评效用临

界值 u 时, 消费者会倾向于考虑退货。换言之, 只有意图给予中评或差评的消费者才可能产生退货意愿^①。其中, 差评消费者的退货行为通常源于对产品价值与匹配度的不满意, 而中评消费者的退货行为则可能归因于平台退货流程的便捷性导致的无理由退货现象。需注意的是, 本研究预设给予好评的消费者不会考虑退货, 尽管现实中存在退货后给予好评的个别情况, 但这往往受卖家“退货需先给好评”等不当要求的影响。本研究专注于未受人为因素干扰的自然交易情境。消费者退货的可能性由概率 $\varphi_{i,t}$ 来量化表示。与直觉相符, 消费者的实现效用 RU 越低, 即满意度越低, 其退货的可能性越大。退货行为将直接导致已成交的销量记录被撤销。然而, 值得注意的是, 即便在退货后, 消费者仍保留发表评论的权利。以淘宝平台为例, 消费者虽然在退货后无法直接进行评论, 但是可以采取先确认收货、发表评论后再退货的操作, 以此实现既退货又发表评论的目的^②。

然后, 消费者会根据失验效应, 即他们预期估值与实际效用之间的差距, 来决定是否发表评论。使用 EU 与 RU 之间的差值绝对值 ΔU 来衡量失验效应强度。在时期 t , 对于某个已买消费者 i , 失验效应强度为:

$$\Delta U_{i,t} = |RU_{i,t} - EU_{i,t}| \quad (4)$$

ΔU 的值越大, 表明消费者越有强烈的意愿留下评论。同时, 鉴于消费者的评论发表倾向与失验效应强度呈非线性 U 形的关系 (Ho et al., 2017), 采用抛物线来近似刻画消费者发表评论的概率。对于某个在时期 t 的已买消费者 i , 其感知到的失验效应强度 $\Delta U_{i,t}$ 与评论发表概率 $\delta_{i,t}$ 之间的函数关系为:

$$\delta_{i,t} = \delta(\Delta U_{i,t}) = \begin{cases} \left(\frac{\Delta U_{i,t}}{d}\right)^2, & 0 \leq \Delta U_{i,t} < d \\ 1, & d \leq \Delta U_{i,t} \end{cases} \quad (5)$$

其中, d 代表 $\Delta U_{i,t}$ 的最大临界值, 以确保 $\delta_{i,t}$ 的值始终落在区间 $[0, 1]$ 内。各个消费者之间的发表行为相互独立。需要注意的是, 只有已买消费者才能进行评分。如果消费者决定发表评论, 则根据 RU 的不同区间给出相应的评分: RU 达到临界值 u 则好评 (1 分), 介于 0 到 u 之间则中评 (0.5 分), 低于 0 则差评 (0 分)。因此, 显示评分 R 的值域为 $[0, 1]$ 。图 2 (b) 总结了已买消费者的退货决策与评分决策流程。

在第五步, 无论是否购买、是否退货与是否评分, 所有消费者在每一期期末都将退出市场。

在第六步, 平台整理评论数据, 并根据设定的评分规则更新产品的评论信息和排名。首先, 统计评论数据与系统数据。评论数据是指产品 A 和 B 各自的当期好评数量 $h_{A,t}$ 与 $h_{B,t}$ 、当期中评数量 $m_{A,t}$ 与 $m_{B,t}$ 、当期差评数量 $n_{A,t} - h_{A,t} - m_{A,t}$ 与 $n_{B,t} - h_{B,t} - m_{B,t}$, 以及当期评论数量 $n_{A,t}$ 与 $n_{B,t}$, 这些数据将用于输出下一期的评分均值及方差。同时, 评论数据还包括累计评论数量 $N_{A,t} = \sum_{i=1}^t n_{A,t}$ 与 $N_{B,t} = \sum_{i=1}^t n_{B,t}$, 这些数据将在下一期帮助消费者进一步降低购买前的不确定性。此外, 系统数据是指产品 A 和 B 各自的当期退货数 $b_{A,t}$ 与 $b_{B,t}$ 、减去退货数后的当期实际销量 $s_{A,t}$ 与 $s_{B,t}$ 、

^① 本研究还曾考虑只有差评消费者才会考虑退货的情况, 受篇幅所限未列示, 读者如需要请联系通讯作者。

^② 商家服务大厅-淘宝网. 已经退款了, 平台能处理评价吗? [EB/OL]. https://helpcenter.taobao.com/servicehall/knowledge_detail?knowledgeId=13067407.

当期浏览量 $f_{A, t}$ 与 $f_{B, t}$ ，这些数据将用于计算下一期的排名得分。

其次，平台根据既定的评分规则来更新两个产品的显示评分 $R_{A, t}$ 与 $R_{B, t}$ 及其评分方差 $\sigma^2(R_{A, t})$ 与 $\sigma^2(R_{B, t})$ 。最新的显示评分将参与排名得分计算，而评分方差则将为下一期的新到消费者提供估值参考。这两种变量的计算都采用了加权平均的方式。例如，显示评分即为各个评论效价（即好评、中评和差评三种效价）与其数量所占比例之积的总和，见公式 (6-7)，而评分方差则是对每个评论效价的方差求加权平均值，见公式(8)-(9)。

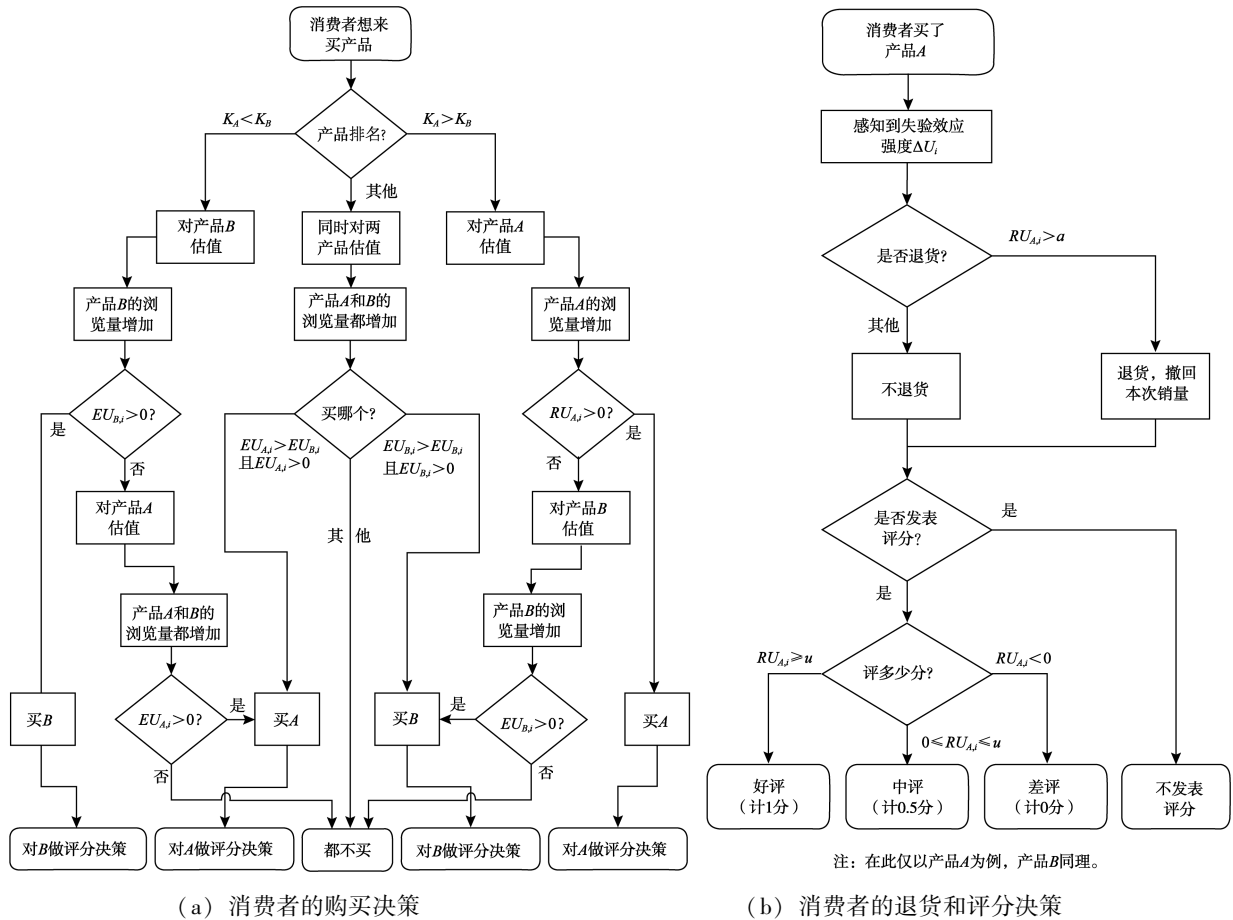


图 2 消费者进行购买决策与退货和评分决策的过程

不同评分规则在计算公式和理念上存在差异。传统评分规则强调所有已买消费者的话语权平等，无论他们是早期还是近期的购买者。在这种规则下，显示评分 R_t^{trad} 及其方差 $\sigma^2(R_t^{\text{trad}})$ 使用公式 (6) 与公式 (8) 来计算。这种规则认为，评论数量的累计会使产品口碑更加稳定和准确。相比之下，动态评分规则更关注近期消费者的评价。它设定了一个统计周期 T ，仅在该周期内的评论数据才会参与显示评分及其评分方差的计算。在这种规则下，显示评分 R_t^{dyn} 及其方差 $\sigma^2(R_t^{\text{dyn}})$ 使用公式 (7) 与公式 (9) 来计算。这种规则的逻辑在于，随着评论数量的积累，近期消费者在购买时能够参考更多的评论信息，因此他们的评价比早期购买者更加准确和有参考价值。

$$R_t^{\text{trad}} = \frac{\sum_{j=1}^t h_j}{\sum_{j=1}^t n_j} + 0.5 \cdot \frac{\sum_{j=1}^t m_j}{\sum_{j=1}^t n_j} \quad (6)$$

$$R_t^{\text{dyn}} = \frac{\sum_{j=t-T+1}^t h_j}{\sum_{j=t-T+1}^t n_j} + 0.5 \cdot \frac{\sum_{j=t-T+1}^t m_j}{\sum_{j=t-T+1}^t n_j} \quad (7)$$

$$\begin{aligned} \sigma^2(R_t^{\text{trad}}) &= (1 - R_t^{\text{trad}})^2 \cdot \frac{\sum_{j=1}^t h_j}{\sum_{j=1}^t n_j} + (0.5 - R_t^{\text{trad}})^2 \cdot \frac{\sum_{j=1}^t m_j}{\sum_{j=1}^t n_j} \\ &+ (0 - R_t^{\text{trad}})^2 \cdot \frac{\left(\sum_{j=1}^t n_j - \sum_{j=1}^t h_j - \sum_{j=1}^t m_j \right)}{\sum_{j=1}^t n_j} \end{aligned} \quad (8)$$

$$\begin{aligned} \sigma^2(R_t^{\text{dyn}}) &= (1 - R_t^{\text{dyn}})^2 \cdot \frac{\sum_{j=t-T+1}^t h_j}{\sum_{j=t-T+1}^t n_j} + (0.5 - R_t^{\text{dyn}})^2 \cdot \frac{\sum_{j=t-T+1}^t m_j}{\sum_{j=t-T+1}^t n_j} \\ &+ (0 - R_t^{\text{dyn}})^2 \cdot \frac{\left(\sum_{j=t-T+1}^t n_j - \sum_{j=t-T+1}^t h_j - \sum_{j=t-T+1}^t m_j \right)}{\sum_{j=t-T+1}^t n_j} \end{aligned} \quad (9)$$

再次, 平台通过综合多方面因素来计算产品的排名得分, 以输出下一期排名。排名得分 K 综合考虑显示评分 R 、退货率 β 、实际销量 s 、转化率 F 和其他因素 ε 这五个维度, 各个维度均占 20% 的权重。这一设定旨在体现产品排名是多个因素综合作用的结果, 而非仅由显示评分决定。尽管现实中各平台的排名算法可能更为复杂且精细, 但本研究通过降低评分的权重并引入多元环境因素, 旨在探讨评分规则的抗干扰能力, 这种模型简化在一定程度上仍能有效反映现实情况。本研究亦尝试了不同的权重调配方案, 例如显示评分 R (40%)、退货率 β (10%)、实际销量 s (15%)、转化率 F (15%)、其他因素 ε (20%) 的方案, 但发现对现有主要结论并未产生显著影响。^① 该产品排名权重模型如图 3 所示。以产品 A 为例 (产品 B 同理), 在时期 t 期末, 该产品的排名得分 $K_{A,t}$ 为各个维度的加权平均值, 由以下公式计算得出:

$$K_{A,t} = 20\% \times R_{A,t} + 20\% \times (1 - \beta_{A,t}) + 20\% \times \frac{s_{A,t}}{M} + 20\% \times F_{A,t} + 20\% \times \varepsilon \quad (10)$$

排名得分更高的产品将在下一期获得更高排名。除了显示评分外, 其他四个维度的相关介绍如下:

^① 受篇幅所限, 具体方案未列示, 读者如需要请联系通讯作者。

(1) 实际销量 (s) : 指扣除退货数后的实际销售数量。在计算排名得分时, 将销量 s 标准化为 $\frac{s}{M}$, 以确保各输入因素的数值上限统一为 1。在现实中, 拼多多的综合排名算法会根据一段时间内的产品评分和销量等因素来计算产品搜索的排名结果^①。

(2) 产品退货率 (β) : 指退货量 b 与成交销量 ($b + s$) 的比值^②。以产品 A 为例 (产品 B 同理), 在时期 t 期末, 该产品的退货率为 $\beta_{A, t}$ 由以下公式算出:

$$\beta_{A, t} = \frac{b_{A, t}}{b_{A, t} + s_{A, t}} \quad (11)$$

退货率是电商实践中的重要因素之一。根据 Invesp 的调查, 线上销售的退货率超过 30%, 远高于线下实体店的 8.9%。^③ 鉴于此, 将退货率纳入考量, 使模型设定更为完善且贴近现实。退货率对排名得分具有负向影响, 高退货率会导致产品排名权重下降, 进而抑制销售^④。以现实中的淘宝平台为例, 退款纠纷率是影响其产品综合排序的关键因素之一^⑤。

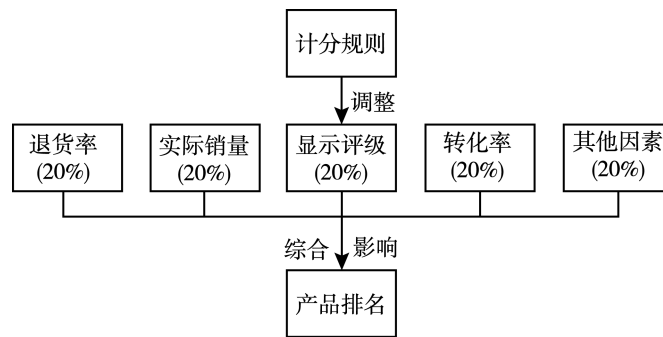


图3 产品排名权重模型

(3) 产品转化率 (F) : 指浏览某一产品的用户中实际购买该产品的用户所占比例^⑥。转化率被定义为未扣除退货数的成交销量 ($b + s$) 与浏览量 f 之间的比值。以产品 A 为例 (产品 B 同理), 在时期 t 期末, 该产品的转化率 $F_{A, t}$ 由以下公式算出:

$$F_{A, t} = \frac{b_{A, t} + s_{A, t}}{f_{A, t}} \quad (12)$$

① 卖家网. 拼多多搜索排名规则, 怎么查? [EB/OL]. <https://www.maijia.com/article/494228>.
 ② Make Influence. E-commerce return rate; What is it, why is it important and how to decrease it [EB/OL]. <https://blog.makeinfluence.com/en/blog/product-return-rate-what-is-it-and-how-to-decrease-it>.
 ③ Invesp. E-commerce product return rate-Statistics and trends [EB/OL]. <https://www.invespro.com/blog/e-commerce-product-return-rate-statistics/>.
 ④ 店查查. 淘宝退货率高有什么影响 [EB/OL]. <https://www.dianchacha.com/news/info/id/1806>.
 ⑤ 商家服务大厅-淘宝网. 影响综合搜索排序的因素有哪些呢? [EB/OL]. https://helpcenter.taobao.com/servicehall/knowledge_detail?kwdContentId=10235731701138432.
 ⑥ 数字营销网. 电商转化率是什么意思? 电商转化率如何提高? [EB/OL]. <https://www.shuziyingxiao.net/dzsw/yddzsw/102298.html>.

在现实中, 平台会将转化率作为产品排序的考虑因素之一。例如, 亚马逊的产品排名算法侧重考虑相关性、转化率和客户满意度这三个指标^①。需要注意的是, 上述的实际销量、转化率和退货率在统计时都只考虑最近一个时期内的当期数据。该设定与现实中淘宝退货率只计算 30 天内的数据相似^②。

(4) 其他因素 (ε): 指除了前四种因素以外的其他因素, 例如相关性、价格、服务质量、商品人气等。为了控制变量, 假设竞争性产品中的其他因素都相同, 即 ε 恒等于 1。同时, 还假设消费者不会发表虚假评论, 卖家不存在虚假订单行为, 以及产品排名遵从自然排序, 即不考虑广告位。

在最后一步, 待评论信息更新后, 两个产品随即进入市场的新一期。届时, 又将有 M 个全新的消费者涌入市场。所谓“全新”, 是指这些消费者对于先前的任何信息均一无所知, 包括历史价格、历史销量以及先前消费者的估值等, 他们的估值将仅依赖于最新的评论信息。

参数 v 、 c 、 u 的取值范围界定如下: 第一, 当消费者与其已买产品的位置之间的距离最近时, 他一定会给好评, 即 $v > u$, 以确保满意的消费者不会只给中评。第二, 当二者的距离最远时, 他一定会给差评, 即 $v - c < 0$, 以确保负面评价的存在。第三, 为了增强排名竞争的重要性, 当消费者处于线性市场的中点时, 他对两个产品都会给好评, 即 $v - \frac{1}{2}c > u$ 。

3.2 符号说明

表 1 总结了产品排名权重模型中的所有符号及其含义。

表 1

参数	定 义
M	每个时期市场上的潜在消费者数量 ($M > 0$)
v	产品的真实价值 ($v > 0$)
c	消费者偏好的单位不匹配成本 ($c > 0$)
u	消费者给予好评与不考虑退货的效用临界值 ($u > 0$)
T	动态评分规则的一个统计周期
d	失验效应强度的最大临界值 ($d > 0$)
ε	除了产品评分、转化率、实际销量和退货率以外的用于计算排名得分的其他因素 ($\varepsilon \equiv 1$)
变量	定 义
$x_{i,t}$	在第 t 期消费者 i 的真实偏好位置 ($x_{i,t} \sim U[0, 1]$)
$\hat{v}_{A,i,t}, \hat{v}_{B,i,t}$	在第 t 期消费者 i 对产品 A 、 B 的价值估值 $\hat{v}_{A,i,t} \sim N(v, c^2 \cdot \sigma^2(R_{A,t-1}))$, $\hat{v}_{B,i,t} \sim N(v, c^2 \cdot \sigma^2(R_{B,t-1}))$

^① Sellerapp. Amazon sales rank explained: What it is? & 12 steps to improve BSR [EB/OL]. <https://www.sellerapp.com/blog/amazon-sales-rank/>.

^② 店查查. 淘宝退款率的计算公式是什么 [EB/OL]. <https://ext.dianchacha.com/news/info/id/2866>.

续表

变量	定义
$\hat{y}_{i,t}$	在第 t 期消费者 i 的初始偏好位置估值 ($\hat{y}_{i,t} \sim U[0, 1]$)
$\hat{x}_{A,i,t}, \hat{x}_{B,i,t}$	在第 t 期消费者 i 在阅读产品 A 、 B 的评论后对自身偏好位置的估值 ($0 \leq \hat{x}_{A,i,t}, \hat{x}_{B,i,t} \leq 1$)
$\Delta U_{i,t}$	在第 t 期消费者 i 所遇到的失验效应强度, 定义为 EU 与 RU 差值的绝对值
$\delta_{i,t}$	在第 t 期消费者 i 愿意发表评论的概率, 由失验效应强度决定
$\varphi_{i,t}$	在第 t 期消费者 i 决定退货的概率, 由实现效用决定
$EU_{A,i,t}, EU_{B,i,t}$	时期 t 消费者 i 在购买前对产品 A 、 B 的预期效用
$RU_{A,i,t}, RU_{B,i,t}$	时期 t 消费者 i 在购买后对产品 A 、 B 的实现效用
$s_{A,t}, s_{B,t}$	产品 A 、 B 在第 t 期减去退货数后的当期实际销量
$n_{A,t}, n_{B,t}$	产品 A 、 B 在第 t 期的当期评论数量
$N_{A,t}, N_{B,t}$	产品 A 、 B 在第 t 期的累计评论数量
$h_{A,t}, h_{B,t}$	产品 A 、 B 在第 t 期的当期好评数量
$m_{A,t}, m_{B,t}$	产品 A 、 B 在第 t 期的当期中评数量
$b_{A,t}, b_{B,t}$	产品 A 、 B 在第 t 期的当期退货数
$\beta_{A,t}, \beta_{B,t}$	产品 A 、 B 在第 t 期的退货率
$f_{A,t}, f_{B,t}$	产品 A 、 B 在第 t 期的浏览量
$F_{A,t}, F_{B,t}$	产品 A 、 B 在第 t 期的转化率
$K_{A,t}, K_{B,t}$	产品 A 、 B 在第 t 期的排名得分
$R_{A,t}, R_{B,t}$	产品 A 、 B 在第 t 期期末的显示评分 ($0 \leq R \leq 1$)
$\sigma(R_{A,t}), \sigma(R_{B,t})$	产品 A 、 B 在第 t 期期末的显示评分标准差 ($0 \leq \sigma(R) < 1$)
上标 trad, dyn	分别表示传统评分规则与动态评分规则

3.3 绩效指标体系

本节建立绩效指标体系以比较评论系统在不同评分规则下的绩效。该体系先从宏观层面设置一级指标, 进而为每个一级指标设置对应的结果指标与过程指标。结果指标旨在量化特定维度的具体绩效, 而过程指标则旨在揭示结果指标背后的形成原因。具体而言, 该体系设定了两个一级指标: 公平性指标与效率性指标。对于公平性指标, 采用产品间累计销量差 (ΔS_{AB}) 作为结果指标, 以衡量不同评分规则下的竞争公平性。同时, 通过排名变动概率 ($P(RC)$) 这一过程指标, 来解释公平性差异背后的原因。对于效率性指标, 设置两个结果指标: 累计评论数量总量 (ΣN_{AB}) 和平均流失客户数 (\overline{LC})。前者用于衡量评分规则在促进评论数量方面的效率, 而后者则用于测量在改善产品

推荐准确性方面的绩效。此外, 引入平均失验效应强度 ($\overline{\Delta U}$) 作为过程指标, 以揭示评分规则对评论积累速度的影响路径。表 2 总结了这一绩效指标体系。

一级指标	过程指标	结果指标
公平性	排名变动概率 ($P(\text{RC})$)	产品间累计销量差 (ΔS_{AB})
效率性	平均失验效应强度 ($\overline{\Delta U}$)	累计评论数量总量 (ΣN_{AB})
	—	平均流失客户数 ($\overline{\text{LC}}$)

在评论系统中, 潜在的偏差可能倾向于特定卖家, 进而引发机会不平等现象。例如, 已积累一定成功度的卖家往往更易获得推荐机会 (Horan & Bullen-Smith, 2023)。鉴于此, 本研究致力于设立明确的公平性指标, 以量化评估这种偏差可能带来的不平等性。同时, 评论系统还需确保效率性, 以满足平台在关键绩效指标上的要求, 诸如通过增加评论总量来丰富信息内容^①, 以及使用客户流失数来衡量产品推荐的准确性^②。下文将对各个指标进行详细介绍。

3.3.1 公平性结果指标——产品间累计销量差

电商平台肩负着维护市场公平的重任, 这要求平台制定公正的竞争规则, 设计合理的运作系统, 以确保市场竞争的公平性^③。对于两个同质同价的竞争性产品而言, 如果二者的销量差距扩大仅仅是因为平台使用了不同评分规则, 那么这显然是违反公平竞争原则的。为了衡量不同评分规则下的竞争公平性, 设置结果指标“产品间累计销量差” (ΔS_{AB}), 该指标通过计算两个产品累计销量的差值绝对值得出, 其计算公式为:

$$\Delta S_{AB} = |S_A - S_B| \quad (13)$$

其中, 两个产品的累计销量 S_A 和 S_B 的计算公式分别为 $S_A = \sum_{i=1}^l s_A$ 和 $S_B = \sum_{i=1}^l s_B$ 。 ΔS_{AB} 的数值越小, 意味着两个产品间的销量分布越均衡, 市场竞争也就越公平。

3.3.2 公平性过程指标——排名变动概率

与 Eber (2011) 使用获胜概率来衡量竞争公平性类似, 定义过程指标“排名变动概率” ($P(\text{RC})$)。该指标表示当期显示评分更低的产品在下一期能反超其对手的概率。排名变动概率

^① MARA. Understanding the ROI of online review ratings: Metrics, KPIs, and more [EB/OL]. <https://www.mara-solutions.com/post/understanding-the-roi-of-online-review-ratings-metrics-kpis-and-more>, 2024-04-29.

^② Paddle. Customer churn analysis: One of SaaS' most important processes [EB/OL]. <https://www.paddle.com/resources/customer-churn-analysis>.

^③ 例如, 《中华人民共和国电子商务法》第三十二条规定, 电子商务平台经营者应当遵循公开、公平、公正的原则, 制定平台服务协议和交易规则。中国人大网. 中华人民共和国电子商务法 [EB/OL]. http://www.npc.gov.cn/zgrdw/npc/lfzt/rlyw/2018-08/31/content_2060827.htm.

的提升,反映了竞争双方排名的更迭更频繁。对于两个同质同价的竞争性产品来说,频繁的排名变动意味着销量的分配更为均衡,从而增强了竞争公平性。本文基于仿真实验中的样本数据来估算排名变动概率。具体而言,本研究定义了一个虚拟变量 RC_t ,用以标识在第 t 期是否发生了产品排名的变动。若该事件成功发生,则 RC_t 记为 1,否则记为 0。该变量的判定条件如以下公式所示:

$$RC_t = \begin{cases} 1, & \text{若 } (K_{A,t} \geq K_{B,t} \text{ 且 } K_{A,t-1} < K_{B,t-1}) \text{ 或 } (K_{A,t} \leq K_{B,t} \text{ 且 } K_{A,t-1} > K_{B,t-1}) \\ 0, & \text{其他} \end{cases} \quad (14)$$

最后,本研究计算 RC_t 在所有重复实验中的平均值,该均值即为排名变动的样本概率 $P(RC)$ 。

3.3.3 效率性结果指标——累计评论数量总量

电商平台期望评论系统能激发消费者积极发表更多评论。评论数量不仅关乎产品人气和浏览量,更体现用户活跃度和产品信息丰富度^①。同时,评论数量的增加反映了平台受消费者信赖,提升了信息透明度和意见公正性。最终,这些评论构成宝贵的商业数据,助力平台搜索引擎优化,提升浏览量及销量^②。因此,评论数量是评估平台运营成效的关键指标。引入结果指标“累计评论数量总量”($\sum N_{AB}$),旨在评估评论系统在促进评论数量增长方面的效率性。该指标代表平台上所有产品的累计评论数量总和,其计算公式如以下公式所示:

$$\sum N_{AB} = N_A + N_B \quad (15)$$

该指标的数值较高意味着当前评分规则有助于提升平台评论信息丰富度。

3.3.4 效率性过程指标——平均失验效应强度

在本节模型中,消费者发表评论的意愿受其失验效应感知影响。当评论系统所呈现的信息与消费者实际体验存在显著偏差时,这种感知会变得尤为强烈,进而激发消费者更强的发表评论的冲动,最终促使产品收获更多评论。然而,这一过程的代价是牺牲了评论信息的准确性。鉴于此,本研究将“平均失验效应强度”作为衡量“累计评论数量总量”的过程指标。平均失验效应强度指标反映了评论信息的准确性水平,其数值越高,意味着评论信息越不准确,但同时也预示着评论数量的增加。为了测量这一指标,本研究基于仿真实验中的样本数据,先算出每一期所有 $\Delta U_{i,t}$ 的均值 ΔU_t ,再求出 ΔU_t 在所有重复实验中的均值。这一最终的均值即为平均失验效应强度 ($\overline{\Delta U}$)。

3.3.5 效率性结果指标——平均流失客户数

为了评估评分规则在提升评论系统推荐产品准确性方面的效率,提出另一个结果指标“平均流失客户数”(LC)。该指标用于量化排名靠后的产品因其排名劣势而流失的客户数量。具体来说,当

① 财论道场_百家号. 随着互联网经济的发展,在线评论对网络购物会产生哪些影响 [EB/OL]. <https://baijiahao.baidu.com/s?id=1733784837679232569>.

② CSP Global. The importance of online reviews for e-commerce businesses [EB/OL]. <https://online.csp.edu/resources/article/ecommerce-online-reviews/>.

消费者在某一时期原本更适合购买排名靠后的产品, 但因搜索排序的影响而选择购买排名靠前的产品时, 他就被视为排名靠后产品的流失客户。这种情况的发生, 往往意味着消费者未能获得最大的实现效用, 从而影响了评论系统为消费者推荐合适产品的准确性。在后续的仿真实验中, 本研究运用以下公式来识别流失客户:

$$\begin{aligned} \text{或} \quad & (K_{A,t} > K_{B,t} \text{ 且 } EU_{A,i,t} > 0) \text{ 且 } (RU_{B,i,t} > RU_{A,i,t} \text{ 且 } RU_{B,i,t} > 0) \\ & (K_{B,t} > K_{A,t} \text{ 且 } EU_{B,i,t} > 0) \text{ 且 } (RU_{A,i,t} > RU_{B,i,t} \text{ 且 } RU_{A,i,t} > 0) \end{aligned} \quad (16)$$

当该公式的计算结果为真时, 即意味着某消费者成为排名靠后产品的流失客户。随后, 本研究将这些消费者计入当期流失客户数 LC_t 。为了得到更具代表性的结果, 本研究进一步对每一期的所有重复实验数据进行平均处理, 得出平均流失客户数 \overline{LC} 。该指标的数值越大, 说明在既定评分规则下评论系统通过产品排名为消费者推荐合适产品的准确度越低。

3.4 仿真实验设计

本研究利用 Python 编程语言对所设计的模型进行仿真模拟。为确保结果的可靠性, 针对每一种评分规则重复了 20000 次实验, 每次实验的时间轴长度固定为 50 个时期。考虑到产品类型的影响, 根据产品的体验程度将其分为三类: 搜索型产品、半体验型产品和体验型产品, 它们分别对应低、中、高三种不同的体验程度, 即 $\frac{c}{v} \in \{1.1, 1.3, 1.5\}$ 。这一设定是基于 Chen 等 (2021) 的研究, 他们发现所测试产品的体验程度均落在 $[1.1, 1.5]$ 区间内。接着, 针对每一种产品类型, 测试了五种评分规则, 包括传统评分规则 Trad-R 与四种计分周期不同的动态评分规则 Dyn-R ($T \in \{1, 2, 3, 4\}$)。综上所述, 共计进行了 $20000 \times 3 \times 5$ 次实验, 以确保研究结果的全面性和准确性。

表 3 概括了这一仿真实验设计, 包括参数的设定、自变量的设置, 以及因变量的收集。仿真实验对每一自变量组合进行测试, 然后收集因变量数据, 这些数据将用于导出指标。此外, 通过以下措施防止仿真结果出现异常: (1) 参数和变量初始化: 将变量 w 、 R 和 $\sigma(R)$ 的初始值分别设为 0、0 和 1。将评分方差的初始值设为 1 是因为该变量的上界为 1^①。同时, 针对搜索型产品, 将参数 a 调整为 -0.08。这是为了防止没有消费者选择退货的情况发生。(2) 清零设置: 在动态评分规则下, 当

^① 这里要证明的是, 显示评分标准差 $\sigma(R)$ 的上界为 1。不论平台采用什么评分规则, 在评分的统计范围内, 本研究将采用以下的简化符号: 显示评分标准差 $\sigma(R)$ 、显示评分 $R(\in [0, 1])$ 、好评数 $h(\geq 0)$ 、中评数 $m(\geq 0)$, 则差评数 $(n - m - h) \geq 0$, 那么就有 $h + m \leq n$ 。首先, 我们知道:

$$\sigma(R) = \sqrt{\frac{h}{n}(1-R)^2 + \frac{m}{n}(0.5-R)^2 + \left(1 - \frac{h}{n} - \frac{m}{n}\right)(0-R)^2} = \sqrt{R^2 - \left(\frac{2h+m}{n}\right)R + \frac{4h+m}{4n}}$$

将 $R = \frac{h+0.5m}{n}$ 代入该式并化简, 得 $\sigma(R) = \frac{1}{2n} \sqrt{4h(n-m-h) + m(n-m)}$, 则有 $\sigma(R) < \frac{1}{2n} \sqrt{(4h+m)n}$ 。

又因为 $R \leq 1$, 就有 $4h+m \leq 4n-m$, 则: $\sigma(R) < \frac{1}{2n} \sqrt{(4n-m)n} < 1$

证毕。

产品在计分周期内没有收到评论时,其显示评分将重置为 0,评分方差回归初始值。这一设定旨在模拟现实中“暂无评分”时消费者估值最不准确的情况。此外,对转化率和退货率也进行清零设置,以防止零除异常。(3)热身期内不设置排名:为了控制初始排名对实验结果的影响,设置初始 5 个时期为热身期。在热身期内不进行产品排名,且统一使用传统评分规则。消费者在此期间的购买决策过程将遵循图 2 (a) 中“同时为两产品估值”的路径进行。

表 3 仿真实验设计概况表

参数	取值	自变量	取值	因变量	符号
M	30	评分规则	{ Trad-R,	显示评分	$R_{A,t}, R_{B,t}$
v	1		Dyn-R ($T=1$),	当期评论数量	$n_{A,t}, n_{B,t}$
d	1		Dyn-R ($T=2$),	当期实际销量	$s_{A,t}, s_{B,t}$
u	0.2		Dyn-R ($T=3$), Dyn-R ($T=4$) }	失验效应强度	ΔU_t
		产品类型	$\frac{c}{v} \in \{1.1, 1.3, 1.5\}$	当期排名得分	$K_{A,t}, K_{B,t}$
				当期流失客户数	LC_t

4. 结果与分析

本节从公平性和效率性两个指标维度来分别呈现绩效差异的时间序列结果。对于公平性维度,图 4 和图 5 分别展示了结果指标产品间累计销量差和过程指标排名变动概率的结果。对于效率性维度,图 6 和图 8 分别展示了两个结果指标累计评论数量总量与平均流失客户数的结果,图 7 则对累计评论数量总量的过程指标平均失验效应强度进行了可视化。

4.1 公平性指标结果分析

图 4 表明,评分规则会影响产品间累计销量差,进而影响竞争的公平性。首先,无论产品类型如何,所有评分规则下的产品间累计销量差都会随着时间的推移而增大。然而,值得注意的是,传统评分规则下的销量差距最为显著,说明这一规则会加剧竞争不公平性。相比之下,动态评分规则在缩小产品间销量差距方面表现更为出色。这一优势主要源于动态评分规则的滚动计算特性,它能够在一定程度上重置历史评分,从而避免小优势累积成大差距的情况。其次,一味地缩短动态评分规则的计分周期并不总是能缩小产品间销量差距。特别是对于半体验型产品与体验型产品,计分周期过短的动态评分规则如 Dyn-R ($T=1$) 可能导致其优势不再明显。这可能是由于短周期下样本容量过小,使得产品评分容易受其他因素的影响,从而改变了评分规则对销量分配的效果。总的来说,动态评分规则的公平性表现总是优于传统评分规则。

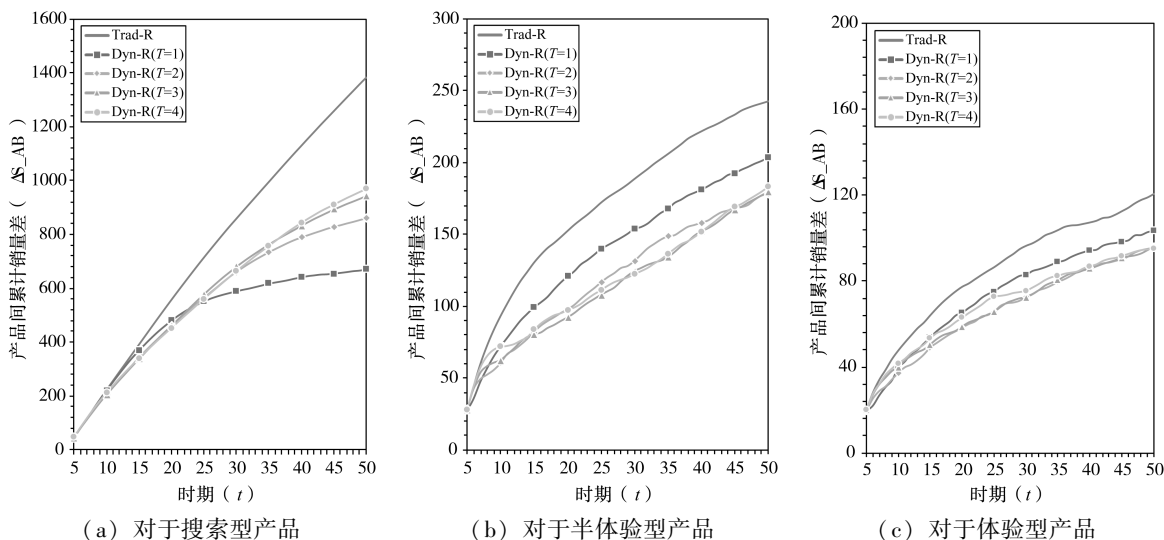


图4 产品间累计销量差指标在不同评分规则下的时序图

接着, 通过排名变动概率指标, 揭示评分规则对产品间累计销量差异的影响机制。图5表明, 较低的排名变动概率与较大的产品间累计销量差存在密切关系。具体而言, 传统评分规则下的排名变动概率始终较低, 这意味着在该规则下, 产品排名相对固定。这种固化的排名格局导致平台销量持续集中在少数排名靠前的产品上, 从而加剧了竞争产品之间的销量差距。随着时间的推移, 这种差异逐渐扩大, 进一步削弱了市场竞争的公平性。相比之下, 动态评分规则通过提高排名变动频率, 有效促进了销量在竞争产品之间的均匀分配。在这种规则下, 即使初期排名靠后的产品也有机会通过未来的良好口碑获得更高的排名。此外, 虽然该图中排名变动概率的发展趋势存在波动, 但这些波动并未改变不同评分规则对排名变动概率的基本差异。无论产品类型如何, 传统评分规则下的排名变动概率始终最低, 而动态评分规则则表现出较高的变动性。这一结果进一步强调了评分规则在影响产品排名变动频率方面的关键作用。

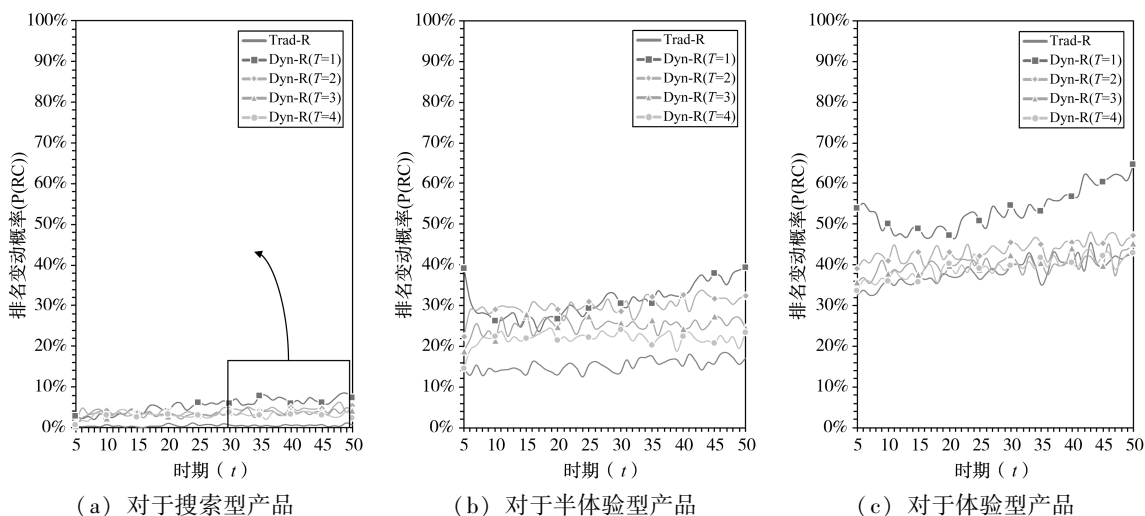


图5 排名变动概率指标在不同评分规则下的时序图

4.2 效率性指标结果分析

图 6 表明, 各个评分规则对评论数量产生速度的影响存在差异。从时间序列的角度来看, 传统评分规则往往能比动态评分规则更有效地触发消费者的评论意愿。以半体验型产品与体验型产品为例, 如图 6 (b) (c) 所示, 传统评分规则下的累计评论数量总量始终最高。然而, 当产品类型转向搜索型时, 见图 6 (a), 情况可能发生转变。与传统评分规则相比, 周期为 1 的动态评分规则似乎更能激发消费者发表评价的意愿。这一发现与直觉相悖, 通常认为动态评分规则的计分周期越短, 其与传统评分规则之间的差距就越显著。毕竟, 在某种程度上, 传统评分规则可以视作计分周期无限延长的动态评分规则。综上所述, 该结果表明, 在促进评论数量方面, 尽管传统评分规则在多数情况下优于动态评分规则, 但对于搜索型产品, 计分周期极短的动态评分规则有可能展现出其独特的优势。

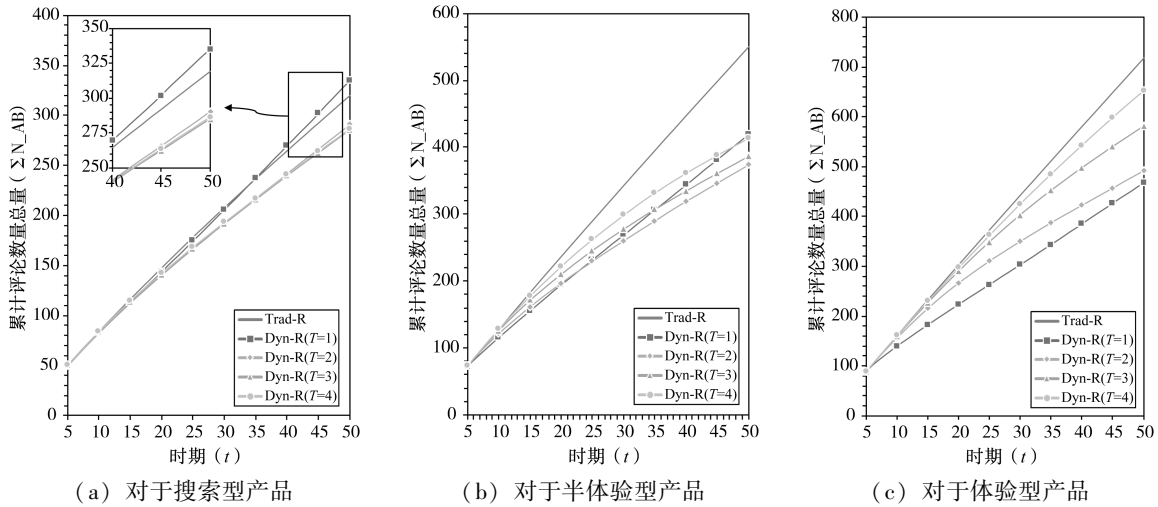


图 6 累计评论数量总量指标在不同评分规则下的时序图

为了探究各个评分规则在累计评论数量总量上产生差异的机制, 进一步分析它们在平均失验效应强度上的变化趋势。根据模型设定, 当失验效应强度增加时, 虽然评论信息的准确性会降低, 但是愿意发表评论的消费者会增多。因此, 更高的平均失验效应强度总是与更快的评论积累速度相关。图 7 揭示了平均失验效应强度在不同评分规则下的时间序列变化。首先, 一般情况下, 与传统评分规则相比, 动态评分规则下的平均失验效应强度普遍更低, 并且这一强度会随着时间的推移逐渐减小。这表明, 在动态评分规则下, 评论信息的准确性得到了逐步提升, 尽管消费者的评论产生速度有所减缓。然而, 一个有趣的现象是, 当产品类型搜索型时, 见图 7 (a), 如果动态评分规则的计分周期过短, 它可能诱发比传统评分规则更大的平均失验效应强度。这说明, 过分强调评分的近期性有时可能适得其反。另外, 当动态评分规则的计分周期设置适中时, 例如当产品体验程度高时, 图 7 (c), 将周期设置为 2, 评论系统虽然在促进评论数量方面表现较差, 见图 6 (c), 但在减轻失验效应方面效率较高。

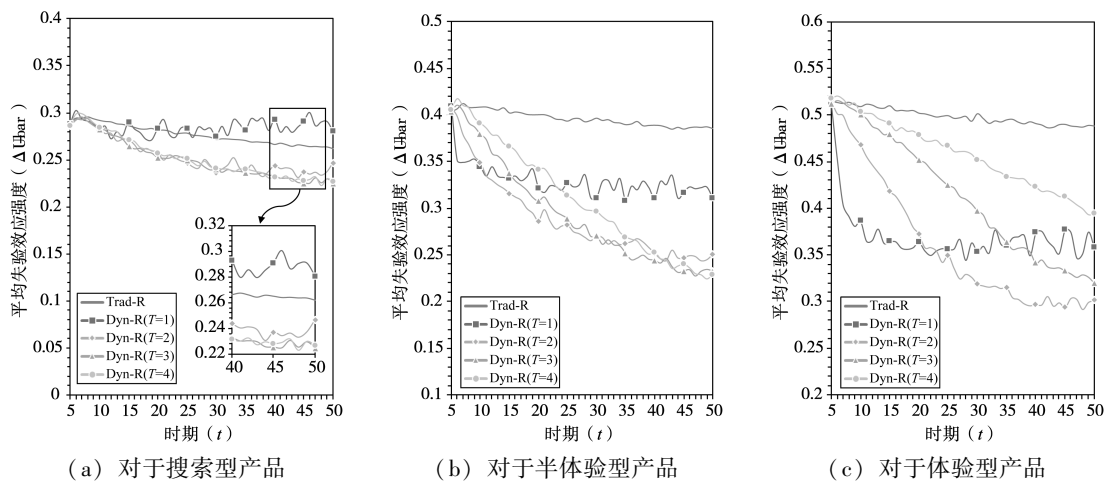


图7 平均失验效应强度指标在不同评分规则下的时序图

总的来说, 图7的结果揭示了各个评分规则在评论数量产生的效率方面产生差异的原因。累计评论数量更多是由消费者感知到更强的失验效应所致, 更强的失验效应意味着更不准确的评论信息。这一发现解释了当产品类型搜索型时计分周期很短的动态评分规则比传统评分规则引发更多评论的原因, 即在这种情况下, 这种动态评分规则引发了消费者更强的失验效应感知。

图8揭示了不同评分规则在提升产品推荐准确性方面的动态变化趋势。从该图可见, 通常情况下, 动态评分规则下的平均流失客户数往往高于传统评分规则, 这意味着在改善产品推荐准确性方面, 动态评分规则相较于传统评分规则表现较为逊色。其原因可能是, 传统评分规则由于涵盖了所有消费者的历史评分, 其样本容量相对较大。这使得传统评分规则在进行产品推荐时能够参考更为全面和稳定的数据, 因此其流失客户数的发展走势相对平稳。相比之下, 动态评分规则主要关注近期数据, 其样本容量相对较小。这种局限性导致产品评分在推荐效率方面的影响力降低, 因为基于较少数据的推荐可能不够准确和全面。因此, 在动态评分规则下, 平均流失客户数更容易受到其他因素的影响, 从而导致该指标数值的明显增加。值得注意的是, 对于体验型产品, 见图8(c), 尽管在传统评分规则下的平均流失客户数可能略高于动态评分规则, 但这一差异并不显著。综上, 动态评分规则在改善产品推荐准确性方面的表现往往不如传统评分规则。

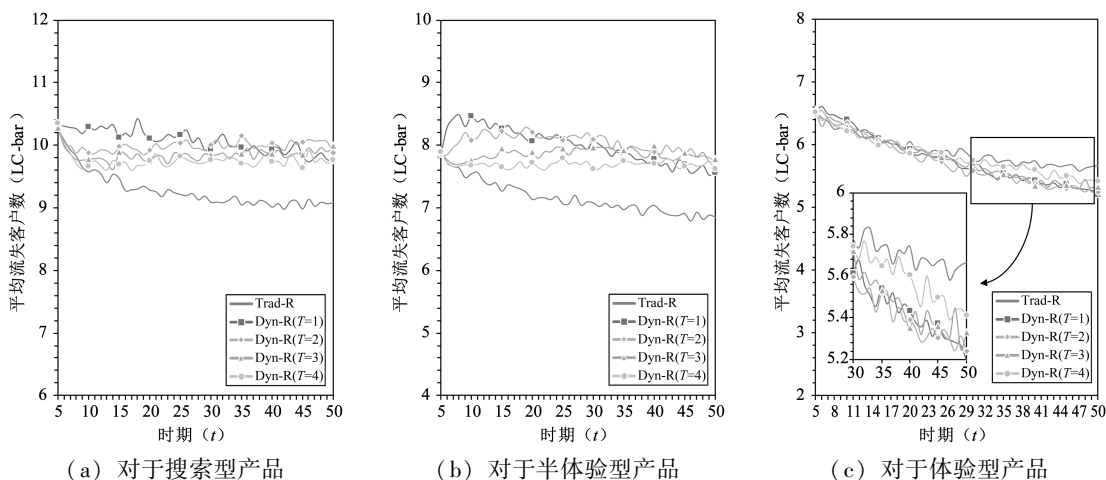


图8 平均流失客户数指标在不同评分规则下的时序图

5. 结论与讨论

5.1 结论

本研究探讨了传统评分规则与动态评分规则如何通过调控评分显示来影响评论系统表现。经过仿真实验结果分析,本研究对比了不同评分规则的作用效果,剖析了背后的作用机制,并探讨了如何选择最适宜的评分规则。本研究得出两大核心结论。

首先,传统评分规则与动态评分规则在竞争环境公平性和评论系统效率性方面展现出不同影响。对于公平性维度,传统评分规则往往加剧市场马太效应,即早期优势产品销量持续增长,与竞争对手差距进一步拉大,这在搜索型产品中尤为明显。相比之下,动态评分规则有助于促进销量公平分配,抑制销量过度集中于头部卖家的现象,从而缓和马太效应。对于效率性维度,传统评分规则通常更为高效,无论是在评论数量增长还是在提升产品推荐准确性方面。然而,当计分周期极短时,仅适用于搜索型产品,动态评分规则才有可能在评论数量增长方面表现更佳。

本研究深入剖析了这些影响背后的作用机制,为平台通过评分规则调控评论系统提供了有力支撑。对于公平性维度,评分规则通过调整产品评分显示,影响排名变动频率,进而调控销量分配,最终影响市场竞争公平性。动态评分规则通过定期清除历史评分数据,增加排名变动,防止头部卖家稳定占据销量优势,从而增强竞争公平性。对于效率性维度,在评论数量增长方面,评分规则通过调整评分数据计算周期,影响评论信息准确性,从而作用于消费者失验效应感知,进而影响其发表评价意愿,最终影响评论的积累速度。评论数量的增加往往意味着消费者的失验效应感知增强。在提升产品推荐准确性方面,动态评分规则由于仅考虑最近的数据,样本容量较小,导致产品评分易受波动,进而使得评分规则对产品推荐准确性的影响力不如其他因素。因此,一味强调评分近期性的动态评分规则在这方面的表现反而不如全面考虑历史评论的传统评分规则。

5.2 讨论

基于对先前研究的回顾总结,本研究认识到评分规则在评分计算建模中的重要性,这一因素对研究结论的可靠性至关重要。因此,对于那些基于传统评分规则得出的研究结论,其鲁棒性可能需要更多的案例和数据来进一步验证和探讨。以 Park 等 (2021) 的研究为例,他们强调了早期评论对产品口碑的显著影响,但是这一结论可能未充分考虑评分规则的差异。本研究注意到,在传统评分规则下,早期评论的作用可能确实较为突出,然而在动态评分规则下,这种影响可能相对减弱。

本研究将评分规则纳入评论系统设计的考量,并从平台角度分析了其作用机制。研究发现,动态评分规则虽然在提升竞争公平性方面发挥积极作用,但是对于评论系统效率性的提升,例如加快评论积累速度或改善产品推荐准确性,其效果可能并不显著。因此,在设定评分规则时,平

台需要权衡公平性与效率性之间的微妙平衡。此外,本研究还对 Carnehl 等 (2023) 的观点进行了深入思考。他们主张,提高近期评论的权重对卖家总是有利的,但这一做法可能导致更高的价格和更慢的评论学习速度,从而对消费者产生负面影响。然而本研究的仿真实验结果显示,这一举措或许能降低消费者的失验效应感知,并减缓卖家的评论积累速度。因此,这一策略并不总是对消费者不利,反而可能对卖家造成一定挑战。本研究期待未来研究能够进一步探讨这一复杂而有趣的现象。

本研究旨在深化平台运营者对动态评分规则的理解。先前,平台运营者普遍认为动态评分规则能够帮助卖家摆脱历史低分的影响,同时也有利于消费者及时获取最新评价信息^①。然而,本研究对此持有不同看法。尽管动态评分规则确实有助于增加产品排名的波动性,进而提升市场竞争的公平性,但这并不意味着最新的评论信息总是最可靠的。尤其是在计分周期设置得极短的情况下,对于搜索型产品而言,动态评分规则可能加剧评论信息的失真。这是因为过短的计分周期会使统计随机性增加,从而导致评论信息的准确性下降。此外,动态评分规则在改善产品推荐准确性方面的效果也值得商榷。实际上,由于近期评分的随机性较大,其他影响排名的因素可能因此产生更大的干扰作用,这反而可能导致排名靠后的卖家面临更大的客户流失风险。

基于上述讨论,本研究为电商平台的评论运营提供了重要的管理启示。首先,本研究鼓励平台积极调整评论系统的评分规则,以实现对市场生态和评论系统效率的间接调控。因此,平台运营者应深刻认识到评分规则在塑造平台生态中的关键作用。其次,本研究为平台运营者提供了选择评分规则时的指导策略。传统评分规则与动态评分规则各有利弊,并无绝对优劣之分。因此,平台应基于自身的发展战略,为评论系统量身打造合适的评分规则。具体来说,若平台的目标是打造公平的市场竞争环境,或为消费者提供更精准的评论信息,那么动态评分规则将是一个理想的选择。反之,若平台更侧重于提升评论系统的效率,如增加评论数量或优化产品推荐算法,那么传统评分规则可能更为适用。此外,对于搜索型产品,将动态评分规则的计分周期设置得较短,可能有助于在保障评论积累效率的同时,维护市场竞争的公平性。这样的策略既能够鼓励消费者积极发表评价,又能够避免销售过度集中,促进市场的健康发展。综上所述,本研究强调,在设计评论系统时,平台运营者应充分利用调整评分规则这一有力手段,以实现多元化的管理目标。

此外,本研究虽力求完善,但仍不可避免地存在一些局限性。首先,本研究模型建立在消费者评分自然发生的基础之上,未考虑虚假评论对消费者购买决策的潜在误导。其次,为了简化模型,本研究也未考虑信息过载对消费者决策的影响。在实践中,评论数量过多可能导致消费者处理信息的负担加重,从而干扰其决策过程 (Jacoby et al., 1974)。再次,评论的文本长度也可能影响评论的有用性 (Liu et al., 2017),对文本评论的建模还可以做进一步的分类讨论。最后,尽管评分体系已间接反映了差评的影响,但本研究在排名权重模型中尚未直接纳入差评率这一直观反映消费者不满程度的因素。差评率不仅可能单独影响产品排名,还可能与其他因素产生协同效应,产生独特影响。

因此,未来研究可以从以下几个方向进行拓展。第一,为弥补本研究的局限性,后续研究可在

^① 卖家网. 淘宝店铺刷动态评分很重要吗? 有哪些作用? [EB/OL]. <https://www.maijia.com/article/489818>.

仿真实验中纳入更多因素,例如考虑更加复杂的产品排名机制,以更准确地模拟电商实践中真实的市场环境。第二,可以优化研究方法,或考虑其他产品类型,例如日常必需品与高档耐用品等分类,从多个角度更全面地揭示评分规则的影响机制。第三,探索评分规则的其他潜在影响。虽然现有研究探讨了在线评论对社会学习 (Besbes & Scarsini, 2018) 以及供应链的影响 (Huang et al., 2021),但本研究尚未采用动态评分规则对这些研究展开深入讨论。最后,进一步探讨评论系统设计的创新策略,例如根据投票数为评论赋予不同权重,从而为平台运营提供更为丰富的策略支持。

◎ 参考文献

- [1] 李启庚,赵晓虹,余明阳.服务型产品在线评论信息特征对评论感知有用性与购买意愿的影响 [J]. 工业工程与管理, 2017, 22 (6).
- [2] 余本功,张子薇,王惠灵.一种融合多层次情感和主题信息的 TS-AC-EWM 在线商品排序方法 [J]. 计算机科学, 2022, 49 (S1).
- [3] 张瑾,尤天慧,樊治平.基于多属性在线评价信息的商品购买推荐排序方法 [J]. 东北大学学报 (自然科学版), 2019, 40 (1).
- [4] Archak, N., Ghose, A., Ipeiritos, P. G. Deriving the pricing power of product features by mining consumer reviews [J]. *Management Science*, 2011, 57 (8).
- [5] Besbes, O., Scarsini, M. On information distortions in online ratings [J]. *Operations Research*, 2018, 66 (3).
- [6] Carnehl, C., Stenzel, A., Schmidt, P. Pricing for the stars: Dynamic pricing in the presence of rating systems [J]. *Management Science*, 2023, 70 (3).
- [7] Chen, P., Hitt, L. M., Hong, Y., et al. Measuring product type and purchase uncertainty with online product ratings: A theoretical model and empirical application [J]. *Information Systems Research*, 2021, 32 (4).
- [8] Chen, P. Y., Hong, Y., Liu, Y. The value of multidimensional rating systems: Evidence from a natural experiment and randomized experiments [J]. *Management Science*, 2018, 64 (10).
- [9] Chen, Y., Xie, J. Online consumer review: Word-of-mouth as a new element of marketing communication mix [J]. *Management Science*, 2008, 54 (3).
- [10] Chen, Y. C., Shang, R. A., Kao, C. Y. The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment [J]. *Electronic Commerce Research and Applications*, 2009, 8 (1).
- [11] Chevalier, J. A., Mayzlin, D. The effect of word of mouth on sales: Online book reviews [J]. *Journal of Marketing Research*, 2006, 43 (3).
- [12] Dellarocas, C. The digitization of word of mouth: Promise and challenges of online feedback mechanisms [J]. *Management Science*, 2003, 49 (10).
- [13] Derakhshan, M., Golrezaei, N., Manshadi, V., et al. Product ranking on online platforms [J].

- Management Science, 2022, 68 (6).
- [14] Eber, N. Fair play in contests [J]. Journal of Economics, 2011 (103).
- [15] Feng, J. , Li, X. , Zhang, X. M. Online product reviews-triggered dynamic pricing: Theory and evidence [J]. Information Systems Research, 2019, 30 (4).
- [16] Fernandes, S. , Panda, R. , Venkatesh, V. G. , et al. Measuring the impact of online reviews on consumer purchase decisions—A scale development study [J]. Journal of Retailing and Consumer Services, 2022 (68).
- [17] Guan, Y. , Tan, Y. , Wei, Q. , et al. When images backfire: The effect of customer-generated images on product rating dynamics [J]. Information Systems Research, 2023, 34 (4).
- [18] Ho, Y. C. , Wu, J. , Tan, Y. Disconfirmation effect on online rating behavior: A structural model [J]. Information Systems Research, 2017, 28 (3).
- [19] Horan, T. , Bullen-Smith, I. Online ranking system effects on perceived fairness: Gender, income and education [J]. The Journal of Social Media in Society, 2023, 12 (2).
- [20] Huang, S. , Potter, A. , Eyers, D. Using simulation to explore the influence of online reviews on supply chain dynamics [J]. Computers & Industrial Engineering, 2021 (151).
- [21] Huang, S. , Potter, A. , Eyers, D. , et al. The influence of online review adoption on the profitability of capacitated supply chains [J]. Omega, 2021 (105).
- [22] Jacoby, J. , Speller, D. E. , Kohn, C. A. Brand choice behavior as a function of information load [J]. Journal of Marketing Research, 1974, 11 (1).
- [23] Jia, H. , Wan, E. W. , Zheng, W. Stars versus bars: How the aesthetics of product ratings “shape” product preference [J]. Journal of Consumer Research, 2023, 50 (1).
- [24] Jiang, Y. , Guo, H. Design of consumer review systems and product pricing [J]. Information Systems Research, 2015, 26 (4).
- [25] Jin, C. , Yang, L. , Hosanagar, K. To brush or not to brush: Product rankings, consumer search, and fake orders [J]. Information Systems Research, 2023, 34 (2).
- [26] Kong, D. , Yang, J. , Duan, H. , et al. Helpfulness and economic impact of multidimensional rating systems: Perspective of functional and hedonic characteristics [J]. Journal of Consumer Behaviour, 2020, 19 (1).
- [27] Kwark, Y. , Chen, J. , Raghunathan, S. Online product reviews: Implications for retailers and competing manufacturers [J]. Information Systems Research, 2014, 25 (1).
- [28] Lee, S. K. , Lindsey, N. J. , Kim, K. S. The effects of news consumption via social media and news information overload on perceptions of journalistic norms and practices [J]. Computers in Human Behavior, 2017 (75).
- [29] Li, H. , Meng, F. , Pan, B. How does review disconfirmation influence customer online review behavior? A mixed-method investigation [J]. International Journal of Contemporary Hospitality Management, 2020, 32 (11).

-
- [30] Liu, X. , Li, C. , Nicolau, J. L. , et al. The value of rating diversity within multidimensional rating system: Evidence from hotel booking platform [J]. *International Journal of Hospitality Management*, 2023 (110).
- [31] Liu, Y. , Feng, J. , Liao, X. When online reviews meet sales volume information: Is more or accurate information always better? [J]. *Information Systems Research*, 2017, 28 (4).
- [32] Park, S. , Shin, W. , Xie, J. The fateful first consumer review [J]. *Marketing Science*, 2021, 40 (3).
- [33] Poppink, B. , Frasinca, F. , Robal, T. An experimental study on re-ranking web shop search results using semantic segmentation of user profiles [J]. *Electronic Commerce Research and Applications*, 2023 (62).
- [34] Ren, J. , Yeoh, W. , Shan, Ee M. , et al. Online consumer reviews and sales: Examining the chicken-egg relationships [J]. *Journal of the Association for Information Science and Technology*, 2018, 69 (3).
- [35] Ren, L. , Zhu, B. , Xu, Z. Continuous exp strategy for consumer preference analysis based on online ratings [J]. *IEEE Transactions on Fuzzy Systems*, 2021, 30 (7).
- [36] Schneider, C. , Weinmann, M. , Mohr, P. N. C. , et al. When the stars shine too bright: The influence of multidimensional ratings on online consumer ratings [J]. *Management Science*, 2021, 67 (6).
- [37] Scholz, M. , Pfeiffer, J. , Rothlauf, F. Using pagerank for non-personalized default rankings in dynamic markets [J]. *European Journal of Operational Research*, 2017, 260 (1).
- [38] Tandon, A. , Aakash, A. , Aggarwal, A. G. , et al. Analyzing the impact of review recency on helpfulness through econometric modeling [J]. *International Journal of System Assurance Engineering and Management*, 2021 (12).
- [39] Wang, C. , Zhang, X. , Hann, I. H. Socially nudged: A quasi-experimental study of friends' social influence in online product ratings [J]. *Information Systems Research*, 2018, 29 (3).
- [40] Wang, W. , Li, F. , Yi, Z. Scores vs. stars: A regression discontinuity study of online consumer reviews [J]. *Information & Management*, 2019, 56 (3).
- [41] Wu, Y. , Wu, J. The impact of user review volume on consumers' willingness-to-pay: A consumer uncertainty perspective [J]. *Journal of Interactive Marketing*, 2016, 33 (1).
- [42] Zhang, M. , Guo, X. , Chen, G. Prediction uncertainty in collaborative filtering: Enhancing personalized online product ranking [J]. *Decision Support Systems*, 2016 (83).
- [43] Zhao, C. , Wang, X. , Xiao, Y. , et al. Effects of online reviews and competition on quality and pricing strategies [J]. *Production and Operations Management*, 2022, 31 (10).
- [44] Zimmermann, S. , Herrmann, P. , Kundisch, D. , et al. Decomposing the variance of consumer ratings and the impact on price and demand [J]. *Information Systems Research*, 2018, 29 (4).

**The Impact of Rating Rules on E-Commerce Platforms'
Review Systems Under a Simulation Experiment**

Yang Lei¹ Zhang Weijie¹ Hao Caixia²

(1 Department of Electronic Business, South China University of Technology,
Guangzhou, 510006; 2 School of Business, Changzhou University, Changzhou, 213159)

Abstract: With the prevailing of online shopping, consumers increasingly depend on online reviews for product information, particularly product ratings, which have become a key metric for assessing the performance of online sellers. Therefore, it is crucial to have a thorough understanding of the rating rules established by e-commerce platforms, specifically how ratings are calculated. Presently, there are two main rating rules: the traditional rating rule (Trad-R) and the dynamic rating rule (Dyn-R). Despite the far-reaching influence of rating rules, there is a lack of comprehensive research in this stream. This study employs a simulation experiment approach to compare the effects of the two rating rules on the design of review systems. The findings indicate that Trad-R is usually superior in enhancing the efficiency of review systems, such as accelerating the accumulation of reviews and reducing customer attrition for lower-ranked sellers. Dyn-R is always more effective in promoting competitive fairness, but especially, when its rating cycle is set to extremely short, Dyn-R can drive a faster accumulation of reviews for search products. Therefore, e-commerce platforms should select the appropriate rating rules based on their business objectives.

Key words: Rating rules; Product ratings; Recency of reviews; Review system design; Simulation experiment

专业主编: 许明辉