
线上商家“标题党”行为与内卷式竞争

张 昊 严玉珊 王心怡 王巍霖*

内容提要 线上市场中存在以包含特定关键词的商品名称吸引消费者注意,进而影响购买选择的“标题党”现象。本文基于多主体囚徒困境进行理论分析,说明其实质是商品名称信息的模糊化,会使得竞争同质化并带来逆向选择,最终造成市场总体价格水平下降但销量无法增加的内卷式竞争后果。我们将信息模糊分为矛盾与重复两个维度,综合运用大语言模型“深度求索”智能评分与人工构建关键词词典匹配两种方法,对2023年4月到11月共计26周内某大型电商购物平台服装品类中涉及约20万个自营或第三方销售单品的40余万个名称进行分析,结果表明其中有接近10%存在用词矛盾,有大约40%存在至少1组关键词重复。本文在此基础上构建计量模型进行经验分析,结果验证了商品名称信息模糊化的理论后果。由此,深入认识线上市场不正当竞争的新形式,全面完善经营规则、优化商家竞争行为,应成为治理消费品领域内卷式竞争的必要举措。

关键词 线上市场 商品名称 信息模糊 多主体囚徒困境 内卷式竞争

DOI:10.19985/j.cnki.cassjwe.2026.05.008

一 引言

2026年《政府工作报告》明确指出,“加强反垄断、反不正当竞争,强化公平竞争审

* 张昊:中国社会科学院财经战略研究院 中国社会科学院大学应用经济学院;严玉珊(通讯作者):浙江师范大学经济与管理学院;王心怡、王巍霖:中国社会科学院大学应用经济学院。电子邮箱:georgezhangh@gmail.com(张昊),yanyushan4334@gmail.com(严玉珊),wangxinyi_0916@163.com(王心怡),3099470642@qq.com(王巍霖)。

作者感谢国家社会科学基金一般项目(25BJY042)的资助,感谢匿名审稿专家提出的宝贵修改意见,感谢张杭对经验研究工作的建议。当然,文责自负。

世界经济 * 2026年第5期 · 108 ·

查刚性约束,综合运用产能调控、标准引领、价格执法、质量监管等手段,深入整治‘内卷式’竞争”^①。“内卷”(involution)一词在经济学中主要指长时间无法进入更高层级的发展阶段,而只能转向内部复杂化的无效消耗,在原有水平不断重复的状态(Goldenweiser, 1936; Geertz, 1969)。对于“内卷式竞争”,现有研究多采取描述性界定,大致可以归纳出两个特点:一是竞争过程存在无意义的投入或消耗,常表现为低价、低质竞争,并且手段越来越复杂细微(王海燕和张占斌, 2025);二是竞争结果并不能带来真正的发展或创新,只会妨碍供需结构升级(刘志彪和王兵, 2024; 欧阳日辉和刘璇, 2025)。本文将这种具有内卷特征的竞争行为称为内卷式竞争。

一般而言,竞争能够实现优胜劣汰,有利于优化资源配置,也是创新的动力来源。但内卷式竞争意味着市场机制失效,必须找到其根源才能进行治理。已有研究表明,产业政策扭曲与“潮涌现象”、地区间市场分割与产业同构、国内外市场形势变化等因素都会造成价格竞争激化,而技术创新与知识产权保护等方面的短板则进一步将竞争固化在低水平层次(刘志彪和王兵, 2024; 李晓华, 2025; 张杰和任元明, 2025)。这些文献主要以钢铁、电解铝、锂电池、光伏、新能源汽车等生产环节规模经济特征较为明显的产品为分析对象,一方面,其资本投入本身是长期性的,形成的固定资产不能迅速调整;另一方面,保有一定的产能规模有助于适应经济周期波动(钟春平和翟乃森, 2022)。因此,在需求下降时企业有动机降低价格保持市场规模。但必须看到的是,包括食品(刘彤, 2025)、服装(魏银科, 2025)等在内的消费品领域也出现了内卷式竞争,而这些行业的资产规模要求与进出壁垒较低,甚至部分产品更新迭代也较为迅速。这意味着,在从宏观与产业层面分析内卷式竞争问题的同时,还应深入微观层面的竞争行为及市场主体所处的竞争环境中寻找市场机制扭曲的原因。

已有对消费品领域竞争内卷化的微观研究主要涉及数字平台间的竞争行为。近年来,线上市场对居民消费、产业运行及宏观价格表现的影响逐渐显现,加之其中比价便利、竞争激烈,以此作为研究消费品领域竞争内卷化的突破口具有重要意义。数字平台通过流量原则重塑市场规则形成低价竞争闭环,不仅会在制造业市场引发“柠檬市场”问题,还会导致生产资源配置效率下降和创新资源投入减少(欧阳日辉和刘璇, 2025)。平台经营者从自身利益出发滥用规则制定权和执行权并影响平台内商家行为,应当受到反不正当竞争的法律规制(殷继国, 2025)。目前,电商平台垄断行为

^① 资料来源:《政府工作报告——二〇二六年三月五日在第十四届全国人民代表大会第四次会议上》,《人民日报》2026年3月14日第1版。

受到较多关注,而对其中同样可能导致市场机制失效的不正当竞争问题的研究仍不充分。不难理解,如果线上市场规则不够完善,缺乏约束的不正当竞争行为演变为商家的集体选择,则完全有可能成为行业性内卷的原因。

线上市场竞争的维度和方式复杂多样,有待拓展的研究空间非常广阔。从竞争的维度来看,除电商平台间的寡头垄断行为外,平台内不同商家间的竞争同样需要关注(Jin *et al.*, 2021);从竞争方式看,对价格竞争的研究已经包括价格水平本身和策略性调价(张昊和冯永晟,2022;刘重阳等,2025),同时,“二选一”限制交易(蔡祖国和李世杰,2022)、产品差异化(甄艺凯和崔凯南,2025)、好评返现(Li *et al.*, 2014;魏瑾瑞和王金伟,2022)等非价格竞争行为也受到关注。在此基础上,完善信誉评价、排序规则等方面的平台规则设计成为线上市场竞争行为治理的重要途径(王宇等,2019;刘小鲁和鲍仁杰,2020)。必须看到,受研究方法和数据等因素的限制,线上市场中仍有大量不同于传统市场的竞争形式未得到充分研究,其中不乏不正当竞争行为。2025年修订的《中华人民共和国反不正当竞争法》第21条做出了“平台经营者应当在平台服务协议和交易规则中明确平台内公平竞争规则”的授权性规定。考察现有制度规则下商家之间是否存在不正当竞争行为,对于依托竞争政策防范治理内卷式竞争具有重要现实意义。

值得注意的是,日本对“过度竞争”问题的讨论也常常与不正当竞争行为相关联。类比内卷式竞争,日本将低集中度行业中,低或负利润率长期持续,而生产要素无法顺利退出的状态称为过度竞争;与之相对的概念是“过当竞争”,后者还包含不希望或不应当出现这种竞争状态的价值判断(小宫隆太郎等,1984)。过当竞争常常表现为同质企业同时生产超出需求的商品并大力推销,导致商品价格水平下降,企业利润微薄,不正当交易行为滋生(飯田経夫,1961),并朝破坏市场秩序的方向演化(两角良彦,1962)。这也说明,深入考察内卷式竞争背后可能存在的不正当竞争行为,对于认识市场机制失效的原因十分必要。

综上所述,本文将线上消费品市场出现的标题党行为,即商品名称信息模糊化作为一种吸引消费者的非价格竞争手段,考察并分析了其导致内卷式竞争后果的原因。一般而言,商品名称中会包括品牌、品种以及附加性描述信息等内容,对于吸引消费者至关重要(Hopkins,1923)。在线上市场中商家还会通过选择商品标题中包含的关键词影响搜索及相关性排序,对购买行为具有更深入的影响。标题党则是指商品名称中包含与实际商品有一定关联但不完全相符的内容。消费者在看到这样的商品名称后会形成模糊认知,进而影响购买选择。例如,某棉服商品的名称“新款××棉衣中长款加厚羽绒服棉服中老年男装大毛领男冬装棉袄外套”中,羽绒服与商品实际并不

相符。在标题中直接出现羽绒服和棉服两个存在矛盾互斥关系的名词,会使消费者产生模糊的商品意象^①。又如,在某太阳镜商品的标题“xx太阳镜男女眼镜潮人炫彩蛤蟆镜复古反光墨镜男太阳眼镜时尚新款太阳镜”中,关键词太阳镜出现了两次,同时还包括若干个同义词或近义词。商家采用这种尽可能多列近义词的命名方式,使自身的产品更有可能出现在用户检索结果中,而面对其中混杂信息的消费者则往往难以抓住重点特征。再如,在某毛线帽商品的名称“xx秋冬帽子女毛线帽棒球帽女冬季韩版大头围羊羔毛茸茸洋气护耳鸭舌帽冬天帽子女款”中,不仅同时出现了毛线帽、棒球帽、鸭舌帽3种具体的帽子款式,还提到了护耳和鸭舌两种在帽子上可能有的配饰。一般而言,棒球帽有鸭舌而没有护耳,商家采取这样的命名方式主要是为了涵盖相近产品的检索关键词,但顾客则会面临信息模糊带来的困扰^②。

消费者注意力是有限的(Clippel *et al.*, 2014),通过商品名称命名影响消费者注意,从而争夺流量的做法也因此具有了非价格竞争行为的属性。本文将以理论分析和经验检验相结合的方式证明,在对不合理商品命名缺乏惩罚机制的情况下,这种非价格竞争将会被异化:大部分商家都会在博弈均衡状态下选择信息模糊化策略,由此导致竞争同质化并带来逆向选择,而产品的特质或创新将无法得到消费者的足够关注。其实质是竞争规则不健全导致了竞争行为扭曲,即本身具有不正当竞争行为特点的商品名称模糊化做法演变成集体行动,进而导致商品总体价格水平下降且销量无法增加、投入耗费无效的内卷式竞争后果。

本文可能的创新之处在于:一是聚焦微观层面的企业竞争行为,从诸多经营者共同陷入囚徒困境并导致竞争同质化的角度对内卷式竞争的成因机制进行探讨,说明治理内卷式竞争应充分关注市场规则不完善导致的竞争机制扭曲。二是将研究重点放在消费品领域,重点关注竞争本身较为激烈的线上市场,从非价格竞争行为视角认识企业在经营活动中为获取消费者注意、争夺流量而采取的做法。这将内卷式竞争的形式由价格延伸到了非价格方面,说明其同样可能导致整体价格水平下降的结果。三是在对商品名称进行文本分析并构建测度指标的过程中,将相对成熟的词典匹配方法与基于生成式人工智能大语言模型相结合,并充分发挥后者的强大能力,对涉及自然语言处理的研究具有工具意义。

^① 中国羽绒工业协会在2024年11月发布《羽绒制品消费提示》指出,羽绒制品的产品名称必须是羽绒服、羽绒被,其他如羽绒棉、羽丝棉等都不是羽绒产品。

^② 《淘宝网商品发布规范》(2025年7月修订)中还提及了家电、化妆品、食品等其他品类的名称命名不规范问题。受数据条件等因素限制,后文经验分析将主要针对服装品类展开。

二 理论分析

下面对线上市场商品名称信息模糊化引发内卷式竞争的过程进行理论分析。其重点是构建一个相对直观的博弈模型,分析线上市场经营者如何在缺乏惩罚成本的条件下陷入囚徒困境,说明竞争激化、内卷加深背后的机制。以此为基础,进一步考察标题信息模糊导致竞争同质化与逆向选择,从而对价格、销量等市场表现产生的影响。

(一) 竞争博弈行为分析

首先,假设在一个线上平台中有两个商家A、B进行完美信息同时博弈,两者的产品具有自身特点,且均拥有一定数量的目标顾客群体。若双方确定商品名称时均采用准确简练表示商品信息的精准策略(J),使两者的目标顾客分别找到偏好商品,则双方都可获得由各自顾客群体带来的基准净收益 α 。为简化分析,假定商家成本一致且总能被收益覆盖($\alpha > 0$)。

其次,考虑一方采用以矛盾或近义关键词表示商品信息的模糊策略(M),而另一方仍采用精准策略(J)的情况。此时,采取模糊策略的一方同时将A、B的特征词或其同义词、近义词加入标题名称,使自身商品能够出现在偏好另一方商品的消费者的检索结果中,由此获得更多的流量并转化为销量,带来增幅为 $g(0 < g < 1)$ 的额外净收益。这一做法可能会受到因竞争对手或消费者举报等带来的平台惩罚,以及消费者差评导致的信誉损失等,由此损失销量导致的净收益降幅为 $d(0 < d < 1)$ 。采取精准策略(J)的一方仍然只采用自身商品的特征词,其目标顾客的检索结果中包含竞争对手商品,用户注意力将被分散,假定因流量被对手攫取而损失销量导致的净收益损失比率为 $f(0 < f < 1)$ 。

最后,考虑双方都采取模糊策略的情况。此时,消费者以各方特征词检索商品时都会得到混杂的结果,流量转移效应消失,两个商家都无法通过标题模糊策略获得额外流量并增加收益,但都面临潜在的惩罚。综上,双方的支付矩阵可由图1表示。

		商家B	
		精准 (J)	模糊 (M)
商家A	精准 (J)	α, α	$\alpha(1-f), \alpha(1+g)(1-d)$
	模糊 (M)	$\alpha(1+g)(1-d), \alpha(1-f)$	$\alpha(1-d), \alpha(1-d)$

图1 标题信息决策行为基本模型的支付矩阵

说明:支付矩阵中双方所得以“A、B”列示。

由支付矩阵可知,当满足(1)式条件时,模糊策略为双方占优策略:

$$\left. \begin{array}{l} \alpha(1+g)(1-d) > \alpha \\ \alpha(1-d) > \alpha(1-f) \end{array} \right\} \Rightarrow 1-d > \max\left(\frac{1}{1+g}, 1-f\right) \quad (1)$$

也就是说,在双主体博弈情境下,模糊策略带来的预期损失越小(d 越小),一方以模糊策略引诱对方流量的总体效果越好(g 和 f 越大),则(1)式成立的可能性越大。这一结果意味着可能出现一个囚徒困境,即在 d 不太大的情况下,无论对方如何选择,采取模糊策略总是更优的:若对方采取精准策略,则可以通过攫取对方的流量获益;若对方采取模糊策略,则可以避免自身的流量单方面被对方获取。此时,双方虽然获得的净收益 $\alpha(1-d)$ 不如都采取精准策略的净收益 α ,但自发选择必然会带来这样的结果。现实中,一些数字平台虽然给出了商品名称不应包含无关或不一致信息的规则及监管承诺,但多采取由消费者先向平台投诉,平台再进行处理的方式监管,赔偿金额也被限定为商品实际成交金额的较小比例;加之消费者在举证、主张过程中需要付出“烦扰成本(hassle cost)”,实际发生的索赔事件会进一步减少(Hviid and Shaffer, 1999)。这些情况决定了商家面临的 d 处于较低的水平,而 $1/(1+g) < 1$ 、 $1-f < 1$,所以(1)式得以成立,模糊策略成为占优策略。并且,由于占优策略均衡一定是纳什均衡,所以双方都没有意愿主动改变这一选择,即竞争参与者无法自我摆脱囚徒困境。

下面我们将仅有两个主体的基本博弈模型拓展至存在多个主体的情形。考虑其中的一个代表性主体A,市场中其他差异化经营者总数为 $N=N_J+N_M > 1$, N_J 和 N_M 分别为采取精准策略和模糊策略的经营者数量。 g 和 f 转变为与 N_J 和 N_M 取值相关的函数 $g(N_J, N_M)$ 和 $f(N_J, N_M)$,两者的取值范围仍在 $(0, 1)$ 区间。此时,A选择模糊策略的条件为:

$$(1-d)[1+g(N_J, N_M)] > 1-f(N_J, N_M) \quad (2)$$

其中,不等号的左侧和右侧分别为选择模糊策略和精准策略的收益。(2)式可转化为:

$$1-d > [1-f(N_J, N_M)]/[1+g(N_J, N_M)] \quad (3)$$

由于 $1-f(N_J, N_M) < 1$ 、 $[1+g(N_J, N_M)] > 1$,故(3)式右侧必然小于1。当 d 较小时,(3)式依然成立。这说明,其他商家数量由1变为 N 时,代表性商家A仍然会将模糊策略作为最优选择。并且,当采取模糊策略的商家数量 N_M 较多时,即使 d 稍大、 $g(N_J, N_M)$ 较小,商家A选择模糊策略的可能性仍会很大,因为此时由市场中商品信息模糊造成的消费者注意力分散已经较为严重,采取精准策略面临的流量损失比率 $f(N_J, N_M)$ 将接近于1。

从经济含义上看,上述分析刻画了一个多主体囚徒困境:在缺乏足够有效惩罚的背景下,采取模糊策略的商家可以先获得几乎整个市场的流量,然后在彼此间进行分配;采取精准策略的商家则因为流量被采取模糊策略的商家掠走而面临销售损失。因此,模糊策略成为理性选择。如果竞争主体具有完全的对称性,则均衡状态下所有商家都将选择模糊策略,市场总收益会比都采取精准策略时更低。现实中,不同商家决策参数有所差异(如一些商家可能更担心消费者被模糊的商品名称诱导后给出差评,其面临的 d 更大),所以市场中仍会有采用精准策略的情况存在。但是,当大量竞争对手都采用信息模糊的商品名称获取流量时,采取精准策略的商家也将因为无利可图而更多地转向模糊策略,采取模糊策略的商家则会使商品信息变得更为模糊以提升流量获取能力。由此形成具有自我强化特征的多主体囚徒困境,这也是内卷式竞争不断激化的一个重要原因。据此,我们归纳出本文假说1。

假说1:商家在商品命名时选择模糊策略的做法会形成不断强化的多主体囚徒困境。

值得注意的是,(1)和(3)式的条件中均不包含参数 α 。也就是说,在上述静态分析情境下,只要市场中存在正的经营利润,在惩罚成本较小的条件下通过模糊策略诱导消费者购买行为而攫取更多利润的做法就成为最优策略。

(二)市场影响效应分析

在线上市场中,大量商品的名称信息模糊化且程度不断加深,会对市场竞争格局及消费者选择行为产生影响,使代表商品差异化的特征信息无法充分、有效地传递给消费者,从而导致竞争同质化。商品名称中同时堆叠大量近义甚至矛盾关键词,会在很大程度上分散消费者有限的注意力,使品质、功能等产品识别信息难以引起消费者足够的关注。同时,以特定功能关键词进行检索的结果中也将混入大量针对性不强、使用场景匹配不当的相关产品,导致检索质量下降。加之综合排序结果还受到销量、评价、价格等其他诸多因素影响,真正符合消费者特定需求的商品很可能难以获得靠前排序。一些商家在未对自身产品改进的情况下,将竞争对手体现产品创新特质的关键词稍加调整处理(如将“不起球”改为“不易起球”,“免烫防皱”改为“不易皱”),甚至直接加入自身商品名称中,使不同产品呈现给消费者的信息感知几乎没有差别,最终导致通过功能设计等方面创新获取产品溢价的差异化竞争策略难以收效。

进一步地,消费者购买行为将体现出逆向选择的特点。在商品名称无法充分、有效传递信息的情况下,消费者与商家之间的信息不对称程度自然提升,与商品价格有关的核心品质感知将不可避免地受到影响。典型的,当多个商品的名称中均同时包

含“丝绵”“蚕丝”“天丝”及“丝绸”等一系列近义关键词时,消费者很难分辨其中哪个产品真正使用了纯天然面料。加上线上商品价格除受原料成本影响外,还与渠道费用、品牌溢价、促销活动等诸多因素相关(Jin *et al.*, 2021;张昊和冯永晟,2022),当其价格存在差异时,消费者也无法判断是否因品质所致。经典的逆向选择理论认为,理性的消费者只会按其中的最低品质支付可以接受的价格(Akerlof, 1970)。现实中,虽然不排除有商家成功通过搭便车获得不合理溢价,但消费者因此付出的额外价款常被称为“智商税”,最终会有越来越多的高品质商品因为无法被消费者相信而被迫降价。随着时间的推移,这种逆向选择将导致高品质商品供给逐渐减少,相应需求无法被满足,市场陷入低水平价格竞争,造成总体社会福利的损失(Rothschild and Stiglitz, 1976;Einav *et al.*, 2010)。正因为如此,以各种信息模糊的标题党商品名称吸引消费者眼球的做法在演变成集体行动后,就具有了内卷式竞争的特点。

上述分析说明了商品名称信息模糊化带来的负面影响,其结果将突出表现为商品总体价格水平因同质化竞争与消费者逆向选择而下降。在前文理论模型的设定中,这一点将体现为商家的基准收益(α)不断减小。由此可以看出,商品名称信息模糊化除了可能给商家带来直接损失(d)外,还会产生中长期的收益减损。同时也不难发现,上述结果不会改变商家之间选择同时采取模糊策略的内卷式囚徒困境,因为 α 并不作为决策参数出现在支付矩阵中。由此,我们提出本文假说2。

假说2:商品名称信息模糊化会导致市场总体价格水平下降。

需要强调的是,这种降价并非生产效率提升带来的,更不意味着社会福利的改善,而是企业为了抵消逆向选择负面影响采取的被迫之举,且还伴随着消费者搜寻成本和认知负担的上升。同时,这种由商品信息模糊导致逆向选择带来的价格下降并不能扩大市场总量,而只能使销量由精准策略商家向模糊策略商家转移;对于非必需品而言,由于信息模糊化导致消费者搜寻成本上升,市场总销量甚至还有可能减少。也正因如此,无论商家围绕产品命名的非价格竞争手段如何复杂细微,甚至包括采用基于销量与转化率大数据分析的关键词推荐,其结果都无法推动产品创新与消费升级,这进一步体现了内卷式竞争的核心特征,不利于消费品市场的长期发展。

三 数据来源与预处理

本文使用国内某知名电商平台的数据开展经验分析,该平台兼有自营和入驻两种模式,且拥有较为完善快捷的配送体系。原始数据包括存货单位(SKU)层面的商品名

称、编号、分类、链接、价格、销量、店铺、评价等信息,能够支持本文的经验研究。综合考虑样本代表性及处理可行性等方面的因素,本文选择服装品类数据进行分析。服装是居民消费中的重要品类,且线上市场已较为成熟,由于其标准化程度较低,在无法直接感触商品的情况下,商品信息对于购买决策具有较大影响。服装产品品种众多,除提供基础使用功能外,还在品牌价值、顾客对象、穿着场景等方面存在差异化,消费者在购买过程中先通过关键词检索进行筛选的必要性较高,服装面料、抗皱功能等信息也需要从文字内容中获取。服装品类下的商品被进一步分为男装、女装、内衣、配饰和其他等5类子类。除其他类外,每1个子类又包括数十个细类。数据提供方剔除了本身无法打开(如跳转到平台首页)而得不到有效记录的失效链接。需要说明的是,服装购买受图片展示的影响也较大,但从本文数据看,商品名称和展示图片两者间的变更并无明显关联,消费者考虑展示图片并不会影响本文对商品名称模糊化的分析结果^①。

数据采集自2023年4月开始,11月初结束。考虑采集初期的程序调试以及“双十一”前夕平台维护对数据完整性及可靠性产生的影响,本文采用4月中旬至10月中旬的半年期数据进行经验分析,并以周度作为时间单位。数据主要进行了以下清洗:一是剔除其中的错链。二是对价格异常商品进行处理。这两者是同时进行的,本文根据商品价格排序结果对可疑情况进行甄别,错链和价格奇高的现象主要出现在积分兑换商品、区域结算链接之中。三是检查调整部分明显的错误分类,如归入女装大类下的男士服装等。经过处理以后,本文所用数据集包含近20万个服装单品,时间跨度为26周。

四 商品名称模糊程度指标构建

下面本文对标题党现象,即商品名称模糊的程度进行测度。我们将人工智能生成式大语言模型与基于词典匹配的传统文本分析相结合,构建定量指标完成对商品名称信息模糊化程度的测算。

(一)导致语义信息模糊的原因维度

标题党的实质是通过在商品名称中加入各种关键词造成信息模糊,所以第一步要

^① 本文按照商品的名称和展示主图是否变化分别构建日度虚拟变量 *namechg* 和 *imgchg*,两者全样本期相关系数为0.0409,名称变化前后5日内子样本相关系数为0.0388,在展示主图发生变更的样本中仅有1.64%同时出现了名称变更。

归纳造成这种语义信息模糊的原因,为定量分析指标确定具体维度。本文随机选取了20个商品名称,采取前述语义解读的方式进行分析。总的来看,造成信息模糊的原因可以归为两类:一是用词矛盾,其中既包括细分品种(如羽绒服与棉服)、用料(如纯棉与羊绒)、特征属性(如吊带与短袖)等方面的直接矛盾,又包括词语引申含义体现出的间接矛盾(如一字领与长裤、条纹与纯色)。二是语义重复,其中包括将同一词语反复使用的直接重复,以及大量罗列近义词的间接重复两种情形。接着,本文又随机抽取了50个商品名称,检查其中的语义模糊是否可以由矛盾和重复涵盖,结果表明这一维度划分具有可行性。表1列举了两个典型商品名称的分解示例。当然,一个商品名称可能同时存在这两种问题,并且两个维度也无法绝对区分,后文将在处理中予以具体考虑和说明。

表1 语义信息模糊的维度

商品名称:xx夹克男装可拆卸连帽外套春秋冬季新款防晒夏季百搭休闲户外男装上衣 2230 深灰【春夏款】2XL(136-150斤)	
矛盾	(季节矛盾)冬季、夏季
重复	(近义词)夏季、夏款;夹克、外套;(相同词语反复)男装
商品名称:xx长裤男士夏季春秋款休闲裤装宽松运动潮流男学生束脚抽绳工装长裤 622 黑色(松紧口) 2XL(建议150-170斤)	
矛盾	(季节矛盾)夏季、春秋款;(款式矛盾)潮流、工装
重复	(近义词)抽绳、松紧;男士、男学生;(相同词语反复)长裤

(二)主要使用生成式大语言模型的指标构建

下面构建标题内容矛盾与重复的具体测度指标。本文先考虑主要使用人工智能方法的指标构建。现有大语言模型在文本分析中的应用主要集中在辅助词典构建、文本贴签或归类等方面,较少直接用于指标评分。近年来,生成式大语言模型的出现,为完成这一任务提供了机会。经过学习训练的模型能精确理解用户输入提示语(prompt)中的文本含义,然后根据要求以自然语言输出所需结果(completion)。并且,生成式大语言模型能高效完成大批量数据处理,这也是人工难以实现的。本文尝试使用国产AI模型Deepseek-R1完成标题内容矛盾与重复的人工智能评分。该模型在逻辑推理方面具有良好表现,能够进行深度语义分析并具备一定的自主推理能力,因在一系列应用场景中表现出色而备受关注。

但是,将基于生成式大语言模型的语义分析应用于科研的做法也面临着诸多挑战。典型的,模型“黑箱”式的处理方式难以让人获得直观理解,两次人工智能评分生成的结果可能存在差异,以及出现AI幻觉造成偏差等。对此,本文考虑采取3种方式

应对:(1)使用模型对每个商品名称进行两次打分,以平均值作为最终得分;(2)设置模型温度(temperature)参数,使之以最为精确而非艺术的方式输出结果内容,并使用固定的随机数种子;(3)要求在输出结果中列出被判定为矛盾或重复的词语,提高评分结果的直观度和可理解性。

具体而言,本文搭建了Python 3.9+PyCharm环境调用Deepseek API,以每次评分50条完成共405 192个商品名称的两轮评分^①。调用过程中选择“deepseek-reasoner”模型,设置温度参数值为0,固定随机数种子为1234。进行矛盾维度的评分时使用的提示语模板为:“给下面50个商品名称按内容是否存在细节上的矛盾打分(每行是1个独立名称,不考虑名称间的矛盾),用0到1之间的小数(保留小数点后2位,越接近于1矛盾程度越高),然后列出存在矛盾的词语(1行以内)。按商品名称|分数|矛盾的词语(如有)给出结果,解析说明等等都不需要。”进行重复维度的评分时,将上述提示语模板中的“矛盾”替换为“重复”,其余不变。所有结果写入文本文件中,以便进行后续整理。从输出结果看,尽管出现了一些错误^②,有96%左右的标题都按提示语中预设的格式要求完成了评分。Deepseek在精确模式下的评分是较为稳定的,对重复、矛盾的两次得分结果差异绝对值大于0.3的情况占比分别为1.1%和1.7%。同时,绝大部分自动给出的评分依据也是较为合理的。本文对输出结果进行整理、补缺后,得到了基于Deepseek的矛盾与重复程度评分结果,分别记作“ DS_{score} _{矛盾}”和“ DS_{score} _{重复}”。

(三)主要基于人工词典开发方法的指标构建

由于基于生成式大语言模型的指标构建方式仍处于探索阶段,本文还主要基于词典开发方法构建衡量指标,其分析结果可以与基于生成式大语言模型构建的结果形成印证。这种方法通常需要完成较大工作量的人工标注、判断等任务,但可应用生成式人工智能辅助提升效率。在矛盾与重复这两个造成语义模糊的成因维度中,矛盾产生的负面影响更为突出,但实际出现的频次相对较少。因此,本文考虑构建专门的矛盾配对词典,以文本匹配与统计计数的方法进行测度^③。

^① 在去掉重复后的商品名称列表中,仍有诸多商品名称几乎是一样的,只在末尾涉及尺寸(S、M、L、XL、XXL)、颜色(如“黑”“灰”“藏青”)等部分有变化。为提高人工智能评分效率,本文将这样的标题归并为1组,只进行1次打分。实际每轮由Deepseek评分的标题数量为204 277个。

^② 发生错误的情况主要包括:(1)商品名称文本中含有换行符等异常情况;(2)将不包含矛盾或重复的多个商品名称以“得分均为0”等形式合并给出,而非逐条列出;(3)服务器超时未返回结果等。

^③ 为保证基于人工词典的指标构建具有独立性,这里没有使用构建 DS_{score} _{矛盾}的打分过程中作为评分理由自动给出的矛盾内容进行二次加工。

在构建矛盾配对词之前,首先要建立基本的电商商品名称关键词词表。相比通用语境,电商商品名称中包含较多的特殊组合或用词。例如,在“袜子女”中,“子女”不再组合为词语,正确的分词应该是“袜子|女”;又如,一些商家为了同时覆盖更多检索关键词而造词,出现了“青年轻”等通用语境下没有的词。对此,本文采取逆向思维,先利用大语言模型的深度思考能力辅助识别特殊搭配或用词,即由 Deepseek 先进行智能分词,再提取汇总关键词。本文从数据集包含的电商名称中随机抽取了 10%,使用以下提示语模板进行人工智能分词:“请对这些商品名称进行分词,使用‘|’分割,列出结果即可。不要修改商品名称,不要遗漏。”将分词结果输出到文本文件后,结合原词出处进行以下分析整理。一是剔除词频仅为 1 的关键词。二是剔除仅包含 1 个汉字的单字分词结果,因为在商品名称中通常难以根据单字间矛盾关系得到可靠的判断。例如,“黑”和“白”为矛盾词,但可以同时出现在条纹图案中。三是剔除大部分包含汉字数量超过 5 的分词结果,除“七天无理由退货”“收藏店铺优先发货”等固定搭配或名词外,大多可以进一步拆分。由此筛选得到 9931 个词语或短语,包括 3198 个难以再拆分的双音节词,同时仍然包含大量可进一步拆分的合成词(组)。直接拆解其中的合成词会影响矛盾关系判断的准确性。例如,“内衣”在表示一个大类时可以包括“棉毛裤”等贴身穿着的服装,因而不应判为矛盾;而合成词“无钢圈内衣”等应作狭义理解,与“棉毛裤”矛盾。因此,对于这些合成词,本文后续根据现实应用中的意义再做处理。

在建立词语间矛盾配对关系时,本文采用了先归类、再匹配整理的做法以提高效率。归类环节依然由 Deepseek 人工智能大语言模型辅助完成。本文先随机抽取 2000 个关键词,分 10 次进行不限定类别的归类,然后对其中的类别进行分析归纳,最终确定材质、场景、服装(细分品类)、功能、款式、配饰、人群、时间(含季节)、颜色等维度,另外增加其他类以处理疑难或分词错误产生的关键词,形成共 10 个分类维度。然后,由 Deepseek 完成所有关键词的归类,使用的提示语模板为:“对下面 200 个电商服装商品的关键词,材质、场景、服装(细分品类)、功能、款式、配饰、人群、时间(含季节)、颜色和其他进行分类,不要遗漏、不要新增,不要修改原词。”其余程序设定与之前相同。

矛盾词匹配整理环节共 4 轮,主要由人工完成。第 1 轮先抽取一半关键词进行处理,并考虑两种矛盾情况:一是同类别内的矛盾配对,如款式中的“吊带”与“短袖”,人群中的“学生”与“中老年”等;二是跨类别的矛盾配对,如功能中的“蓄热”与时间中的“夏季”等。第 2 轮是对结果进行删减和提炼:一方面,人为设想列出的矛盾词组合在

现实中可能并未出现;另一方面,如前所述,大量合成词需要精简为引起矛盾的核心词,如“长袖”与“短袖衬衫”“短袖衬衣”“短袖裙”等矛盾,其中矛盾的核心用词是“长袖”与“短袖”。本文将核心用词(“短袖”)与合成词(“短袖衬衫”)共同作为比对词(“长袖”)的矛盾词列表,并根据矛盾词列表中词语间是否存在完整的相互包含关系判断和剔除非核心用词。第3轮主要是增补。将得到的核心词与第一轮未处理的另一半关键词进行比对,主要关注合成词,若其包含之前整理中涉及矛盾关系较多的核心词,则重点进行语义判别和处理。第4轮是复核。以两两配对列出并逐一确认的方式,重点处理合成词涉及的矛盾层次关系。例如,一方面,单字词“帽”与“衬衫”“裤子”等存在矛盾关系,因而列出这些配对组合后,就不需要再将“毛线帽”“雷锋帽”等与“衬衫”“裤子”再建立矛盾配对关系;另一方面,由于存在带耳捂的帽子,故“帽”不能与“耳捂”构成矛盾关系,但通常“棒球帽”不含耳捂,故应当在合成词“棒球帽”的层次确定与“耳捂”构成矛盾关系。

在完成上述配对整理后,本文得到了包含3342个矛盾词配对的词表。由于处理过程中已经考虑了合成词的层级包含关系,故可以直接使用该词表(而不需要先进行分词)计算数据集中商品名称包含的矛盾词配对个数。同时,如果商品名称中有两两间矛盾的关键词,如“帽子”“面巾”“袖套”彼此互斥,则矛盾词个数也会因计算配对组合数而相应增长,即形成自然的“惩罚效应”。本文将矛盾词配对个数进行0-1标准化处理后,得到这一部分的最终测算结果,记作“ $HMscore_{矛盾}$ ”。

对于语义重复这一维度,基于词典方法的测度主要需考虑两种情形。一是罗列近义词或同义词,如前述商品名称中的“太阳镜”与“墨镜”“蛤蟆镜”“太阳眼镜”等。在文本分析中,“余弦相似度”被大量应用于近义词分析中,该指标能通过两个文本单元在语义空间的夹角余弦值衡量其相似性。本文需要计算电商语境下词语间的余弦相似度,完成准确的分词及选择合适的模型对提高结果准确度至关重要。就分词而言,本文在矛盾词的人工配对整理环节已经对基于人工智能分词结果构建的关键词词表进行了细致梳理,形成了包含商品名称专有名词及特定含义核心词的词表,这里进一步通过实际应用与人工检查对其进行完善。将其作为用户自定义词表,每轮随机抽取1000个商品名称,使用“哈工大文本分析工具包”pyltp 0.4.0和ltp 4.2.13进行分词后,由人工进行词表调整修改,最终得到包括1184个词语(或词组)的电商名称用户自定义词典。就模型选择而言,本文采用“PAI BERT Base Chinese”模型,该模型使用阿里巴巴AI平台(Platform for AI, PAI)团队托管的包括电商文本在内的大规模语料完成预训练,并提供多种自然语言工具(Wang *et al.*, 2022)。同时,该模型作为

基于 transformers 架构的动态模型,在余弦相似度计算过程中会考虑一个词语出现时不同的上下文语义,这与 word2vec 等静态词向量模型的分析是不同的。本文采用 Python 调用 transformers 库,将前述分词结果与 BERT 模型标记的子词对齐(subword alignment),然后提取每个词的向量,最后计算得到余弦相似度。二是同一关键词多次出现,如前述商品名称中两次出现的“太阳镜”。由于本文在计算词语余弦相似度的过程中使用了考虑上下文语义的动态模型,重复词配对的余弦相似度计算结果会因其出现的位置不同而有所差异。为消除这一影响,本文统一将完全相同的词语间余弦相似度赋值为 1,这一做法还能对关键词本身重复产生惩罚效应^①。

综上所述,本文先对数据集中的每 1 个商品名称进行分词处理,然后两两配对计算词语间余弦相似度,并以 1 替换相同词语配对的计算结果。接下来,若直接计算所有词语配对的均值,则可能出现因商品名称中含有描述性词语较多而拉低测算结果的问题。这通常也出现在商品名称较长的情况下,是需要避免的结果。为此,本文仅保留配对余弦相似度计算结果位于前三(含替换后的 1)的值,并求取平均数。在将该均值进行 0-1 标准化处理后,得到这一部分的最终测算结果,记作“ $HMscore_{\text{重复}}$ ”。

还需要说明的是,尽管在基于人工词典开发的语义信息模糊中分别考虑了矛盾、重复两个维度,但事实上两者并不能实现绝对的区分,并且从本文的研究目的看也没有必要这样做。就 $HMscore_{\text{矛盾}}$ 而言,商品名称中若存在矛盾词配对,那么配对中一个词的近义词在多数情况下也会被认定为与配对中的另一个词矛盾,所以测算结果也会因近义词语义重复而被拉高。就 $HMscore_{\text{重复}}$ 而言,余弦相似度是根据词语在语境中出现的前后文语义计算空间向量关系的,因而在原理上完全可能出现矛盾词之间余弦相似度也较高的情况。就此而言,本文计算的 $HMscore_{\text{矛盾}}$ 或 $HMscore_{\text{重复}}$ 应被理解为主要基于人工词典开发的“侧重”矛盾或重复程度的评分结果。

(四)综合指标合成

综合指标能简化计量分析结果的呈现形式,但计量分析需要确保结果的可靠性与稳健性,并且本文还需要进一步对采用不同方法得出的分维度测算结果进行比较。因此,本文一方面通过设定权重合成综合指标,另一方面在经验分析中分别考察分项指标的回归结果以排除权重变化对计量结果可能产生的影响。考虑到矛盾与重复是两种商品名称模糊化的具体形式,本文主要围绕两者构建合成指标 $score_{\text{矛盾}}$ 和 $score_{\text{重复}}$ 。同时,由于两者本身就可作为侧重反映语义模糊一个方面的互替指标,故这

^① 采用 BERT 模型计算的重复词语间余弦相似度通常在 0.8 左右。

里不再进一步构建反映整个商品名称模糊程度的综合指标,后文的经验分析也将在区分矛盾、重复两种模糊方式的情况下进行。

对于计算 $score_{\text{矛盾}}$ 和 $score_{\text{重复}}$ 两个综合指标时,主要使用生成式大语言模型和主要基于人工开发词典的测算结果应当具有的权重,考虑人工开发词典方法较为成熟、应用较广泛,而基于生成式大语言模型评价的指标测算仍处于尝试阶段,本文采取以人工开发词典方法为主的综合指标计算原则,确定采用(4)式中的权重计算综合指标(X =矛盾或重复)。

$$score_X = 0.7 \times HMscore_X + 0.3 \times DSscore_X \quad (4)$$

(五)测度结果分析

我们对使用上述测度方法得到的指标进行汇总分析,描述性统计见表2。从矛盾维度的测算结果看,主要使用生成式大语言模型的 $DSscore_{\text{矛盾}}$ 的75分位数为0。更具体地,结果中有77.9%的值为0。换句话说,Deepseek认为在分析的商品名称中有超过20%存在矛盾。而主要基于人工词典开发方法的 $HMscore_{\text{矛盾}}$ 的中位数为0,75分位数不为0。更具体地,其中有36.1%的结果值不为0,即在商品名称中发现存在配对矛盾词。两者间存在的差异,既可能是因为基于人工词典开发方法的矛盾关系判别相对使用生成式大语言模型时更为严格,又可能是因为前者较为机械而可能存在误判所致。但若以综合指标 $score_{\text{矛盾}}$ 判断(相当于并集),则用词存在矛盾的商品名称约占49.9%。该比例同时高于 $DSscore_{\text{矛盾}}$ 和 $HMscore_{\text{矛盾}}$ 的非零值占比,说明 $DSscore_{\text{矛盾}}$ 和 $HMscore_{\text{矛盾}}$ 可能都存在漏判。这也表明,由两者共同构建合成指标具有一定的必要性与合理性。同时本文还统计出, $DSscore_{\text{矛盾}}$ 和 $HMscore_{\text{矛盾}}$ 均不为0的商品名称(相当于交集)占比为9.2%,即有较为充分的理由认为,接近10%的商品名称中存在矛盾用词。

表2 指标测算结果描述统计

指标	观测数	平均值	最小值	25分位数	中位数	75分位数	最大值	标准差
$DSscore_{\text{矛盾}}$	404 811	0.1074	0	0	0	0	1	0.2361
$DSscore_{\text{重复}}$	404 811	0.2175	0	0	0.1500	0.3000	1	0.2548
$HMscore_{\text{矛盾}}$	404 811	0.0226	0	0	0.0000	0.0370	1	0.0396
$HMscore_{\text{重复}}$	404 811	0.7610	0	0.6766	0.7431	0.8319	1	0.1154
$score_{\text{矛盾}}$	404 811	0.0480	0	0	0	0.0519	0.7000	0.0779
$score_{\text{重复}}$	404 811	0.5979	0	0.4990	0.5739	0.6825	1	0.1275

从重复维度的指标结果看, $DSscore_{\text{重复}}$ 的中位数非 0, 具体而言有 64.5% 的商品名称被 Deepseek 判断为存在语义重复。这一比例明显高于矛盾维度的判断, 符合语义矛盾出现频次少于语义重复的一般感受。 $HMscore_{\text{重复}}$ 作为余弦相似度进行 0-1 标准化后的结果, 25 分位数已经明显大于 0.5。综合指标计算结果 $score_{\text{重复}}$ 的 25 分位数也接近 0.5, 说明观测值多集中在取值较大的区间中。我们对过程数据进行具体分析, 结果发现商品名称中至少有 1 组、2 组和 3 组完全相同的关键词重复出现的占比分别为 40.5%、15.2% 和 7.2%^①。这也反映出关键词重复是商品名称中更为常见的现象。

本文还计算了不同子类测算结果的均值(见表 3)。若不考虑观测数明显偏少的“其他”类别, 则女装的 4 个分项指数中有 3 个均为最低($DSscore_{\text{矛盾}}$ 略大于男装), 2 个综合指数都低于其他类别。女装子类中商品名称的模糊化程度最低, 与女性顾客群体在挑选商品时对服装材质、款式等方面的了解更为深入, 商家能够通过名称模糊策略获得的额外收益(g)和采取精准策略时被误导的销量(f)均较小有关。同时, 除 $HMscore_{\text{矛盾}}$ 指标以外, 内衣类和配饰类的测算结果均高于男装类和女装类, 并且重复类指标的差别还较为明显。这两类商品本身的差异化区分维度与消费分层空间都相对较小, 加之单品价格较低, 因此商家通过模糊策略能获得的额外收益(g)和采取精准策略时被误导的销量(f)都相对较大。

表 3 分子类指标测算结果均值比较

	观测数	$DSscore_{\text{矛盾}}$	$DSscore_{\text{重复}}$	$HMscore_{\text{矛盾}}$	$HMscore_{\text{重复}}$	$score_{\text{矛盾}}$	$score_{\text{重复}}$
男装	123 808	0.0950	0.2011	0.0265	0.7571	0.0470	0.5903
女装	125 607	0.1057	0.1801	0.0187	0.7276	0.0448	0.5634
内衣	88 660	0.1258	0.2437	0.0216	0.7859	0.0529	0.6232
配饰	66 685	0.1090	0.2835	0.0239	0.7980	0.0494	0.6436
其他	51	0.0373	0.1565	0.0102	0.7385	0.0183	0.5639

五 经验分析策略与结果分析

(一) 对商品名称模糊化的计量分析

1. 回归模型构建。本文假说 1 描述了商家在竞争中采取模糊化商品名称的策略并形成多主体囚徒困境的情境, 其核心机制在于竞争主体之间的相互影响, 即代表性

① 由于采取配对计算, 3 组完全相同的关键词重复出现也包括 1 个关键词同时出现 3 次的情况。

商家因为竞争对手模糊化商品名称而采取相同的策略或加深自身商品名称的模糊化程度。这种现象在很大程度上与个体行为会受其所在群体中相似成员影响的“同群效应”类似,这为本文设计经验分析策略提供了参照(Bisztray *et al.*, 2018; 张川川和朱涵宇, 2021; 杜勇等, 2023)。本文在商品(SKU)层面构建以名称信息模糊程度($vague_i$)为被解释变量,以前1期中其他同群竞品的平均信息模糊程度($gvague_{-1}$)为核心解释变量的计量回归模型:

$$vague_{it} = \beta_0 + \beta_1 gvague_{-i,t-1} + \beta_c controls_{it} + \eta + \varepsilon_{it} \quad (5)$$

其中, $vague$ 由前述 $score_x$ 衡量,包括矛盾、重复两个维度。计算 $gvague_{-1}$ 时需要先确定同群竞品的范围。同属于一个细类的商品具有较为相近的材质、功能与目标顾客,相互之间存在竞争关系。但这样的划分仍然过于宽泛,对于消费者而言,价格是选购商品时的重要决策因素,价格接近的同一细类商品间具有更加明显和直接的竞争关系。因此,本文以满足属于同一产品细类,且价格较为接近这两个要求,并剔除同店铺商品的方式来确定同群竞争者。在基准回归中,价格接近的标准以当期价差绝对值小于所属于子类中非同一店铺销售产品两两间价差绝对值的30%分位数确定。使用子类层次的价差绝对值分位数,一方面是为了更好地体现细类、价格双重要求的作用,因为同属一个子类、但属于不同细类的产品(如不同材质但功能接近的服装)之间也可能存在一定的替代关系;另一方面是由于部分细类产品的数量较少,如果其价格彼此接近,则可能全部属于一个竞品群体。使用滞后一期既是为了在最大程度上避免反向因果问题,也是考虑到商家因竞争对手的做法影响自身商品销售而做出的反应存在时滞。 $Controls$ 是一系列控制变量,在产品层次加入商品的历史评分($goodcmt$)、店铺的历史评分($shopcmt$)以控制商家维护信誉对采取模糊策略的影响,加入是否有折扣券促销($coupon$)以控制采取其他促销措施的影响;在竞争群体层次加入涉及其他店铺的数量($numshop$)及其销售单品总数($numsku$)以控制竞争强度的影响;在细类层次加入全细类平均名称信息模糊程度($meanvague$)以控制其对所关注商品及其竞品的共同影响。 η 是一组固定效应,控制时间(周)和个体(单品)两个维度^①。基准回归采用聚类到单品的稳健标准误。本文对其他店铺数量及其销售单品总数($numshop$ 、 $numsku$)加1后做对数化处理; $goodcmt$ 和 $shopcmt$ 的原始满分分别为100分和5分,本文将其满分标准化为1。

2. 基准回归结果与稳健性检验。我们采用(5)式进行回归,结果见表4。其中第

^① 限于篇幅,描述性统计结果未报告,详见本刊网站(www.jweonline.cn)本文附录。

表4 假说1计量分析结果(1)

	(1)	(2)	(3)	(4)	(5)
<i>L.gvague</i> ₋₁ 重复	0.0121*** (0.0018)	0.0053*** (0.0019)	0.0066*** (0.0019)	0.0053* (0.0028)	0.0035** (0.0015)
<i>meanvague</i> 重复		0.1168*** (0.0087)	0.0781*** (0.0107)	0.1168*** (0.0159)	0.1207*** (0.0089)
<i>goodcmt</i>		-0.0069* (0.0039)	-0.0048 (0.0041)	-0.0069*** (0.0023)	-0.0073* (0.0039)
<i>shopcmt</i>		-0.0011 (0.0016)	-0.0006 (0.0026)	-0.0011 (0.0019)	-0.0013 (0.0016)
<i>coupon</i>		-0.0004** (0.0002)	-0.0001 (0.0002)	-0.0004 (0.0003)	-0.0003* (0.0002)
<i>numshop</i>		-0.0004** (0.0002)	-0.0008*** (0.0002)	-0.0004 (0.0003)	-0.0005** (0.0002)
<i>numsku</i>		0.0000 (0.0002)	-0.0003 (0.0003)	0.0000 (0.0002)	-0.0000 (0.0002)
周固定效应	控制	控制	控制	控制	控制
店铺×周固定效应	未控制	未控制	控制	未控制	未控制
F值	45.3099***	30.3125***	11.9388***	6.96e+07***	30.0404***
调整后的R ²	0.9718	0.9715	0.9747		0.9716
观测数	1165 100	1109 023	1094 734	1147 713	1078 304
	(6)	(7)	(8)	(9)	(10)
<i>L.gvague</i> ₋₁ 矛盾	0.0117*** (0.0018)	0.0065*** (0.0018)	0.0072*** (0.0016)	0.0065** (0.0026)	0.0050** (0.0023)
<i>meanvague</i> 矛盾		0.1483*** (0.0116)	0.1318*** (0.0140)	0.1483*** (0.0268)	0.1473*** (0.0112)
<i>goodcmt</i>		-0.0037 (0.0022)	-0.0024 (0.0025)	-0.0037** (0.0017)	-0.0033 (0.0022)
<i>shopcmt</i>		0.0018* (0.0011)	0.0011 (0.0016)	0.0018 (0.0011)	0.0022** (0.0011)
<i>coupon</i>		0.0000 (0.0001)	-0.0001 (0.0001)	0.0000 (0.0002)	0.0000 (0.0001)
<i>numshop</i>		0.0003* (0.0002)	0.0002 (0.0002)	0.0003** (0.0001)	0.0002 (0.0002)
<i>numsku</i>		0.0001 (0.0002)	-0.0002 (0.0002)	0.0001 (0.0003)	-0.0001 (0.0002)
周固定效应	控制	控制	未控制	控制	控制
店铺×周固定效应	未控制	未控制	控制	未控制	未控制
F值	42.3060***	25.8525***	15.2351***	20 291.6808***	26.4855***
调整后的R ²	0.9527	0.9519	0.9604		0.9516
观测数	1119 112	1067 056	1052 935	1105 570	1114 336

说明:第(4)、(9)列中括号里的数字为D-K稳健标准误,其余为聚类到单品的稳健标准误。所有回归都控制了单品固定效应。*、**、***分别表示在10%、5%和1%水平下显著,后表同。

上栏和下栏分别采用重复、矛盾两个维度的商品名称模糊测度指标,第(1)、(6)列没有加入控制变量,第(2)、(7)列加入了控制变量。可以看到, $L.gvague_{-1}$ 的回归系数均显著为正,表明特定商品的名称模糊程度确实会受同一竞争集合内其他商品的影响,支持了本文假说1的判断。

除使用重复、矛盾两个替换指标外,本文还从以下方面考量回归结果的稳健性。一是改变固定效应控制方式。商家是商品命名的主体,在线上市场中以店铺形式呈现,并且一个店铺可以经营不同细类的商品。同时,一家店铺的商品命名策略也会随不同时点的促销要求而变化。因此,本文考察同时控制店铺 \times 周固定效应、单品固定效应的情况。从表4第(3)列结果可知,在改变固定效应控制方式后, $L.gvague_{-1}$ 的回归系数仍显著为正。

二是使用更为稳健的标准误。表4第(4)列使用了对面板组间相关、组内序列相关和异方差等更一般残差形式稳健的D-K标准误(Driscoll and Kraay, 1998), $L.gvague_{-1}$ 的估计结果依然显著为正。

三是改变确定同群竞品时采用的价格分位数临界值,由原先的30%放宽为35%。从表4第(5)列结果可知,在加入控制变量后, $L.gvague_{-1}$ 的回归系数仍显著为正但值略有减小,符合略微放宽同群竞品确定范围后的结果特征。

四是以不同测度方法细分指标 $DSscore_x$ 、 $HMScore_x$ 计算被解释变量和核心解释变量,以说明改变综合指标计算权重的稳健性。表5展示了对应基准回归设定的结果, $L.gvague_{-1}$ 的回归系数均显著为正^①。这说明,虽然 $DSscore_x$ 、 $HMScore_x$ 在评分尺度及程度衡量存在差异,但这并没有影响本文的计量分析结果。

表5 假说1计量分析结果(2)

	(1)	(2)	(3)	(4)
$L.gvague_{-1}$	0.0060*** (0.0019)	0.0048*** (0.0016)	0.0021** (0.0010)	0.0070*** (0.0019)
F值	12.7804***	30.3631***	30.9161***	23.5396***
调整后的R ²	0.9726	0.9641	0.9750	0.9470
观测数	1127 438	1113 918	1077 761	1115 715
指标计算基础	$HMScore_{重复}$	$DSscore_{重复}$	$HMScore_{矛盾}$	$DSscore_{矛盾}$

说明:所有回归都加入了控制变量,控制了周固定效应和单品固定效应。

① 本文同时使用细分指标进行了改变固定效应、改变稳健标准误计算方式、改变同群竞品计算分位数的交叉稳健性检验,结果详见本刊网站本文附录。

五是安慰剂检验。商品名称模糊化程度的同群效应表现可能由其他竞争集合内商品受到难以直接观测的共同因素影响导致,本文接下来进行安慰剂检验。先选取同属一个细类的其他商品,这些商品更有可能受到共同因素的影响;然后筛选出非同店铺销售的,且价差绝对值处于子类价差绝对值60%~90%区间内的其他商品,计算其商品名称模糊程度的均值 $ngvague$,并取最后1期作为核心被解释变量进行回归。选择90%的上限可以规避价差离群值带来的影响,60%的下限则可以尽可能排除与关注商品属于同一竞争集合的可能性。从回归结果可知,无论采用重复还是矛盾的测算指标作为计算基础, $L.ngvague_{-1}$ 的回归系数均未出现显著为正的结果^①。

六是采用工具变量法。为进一步排除遗漏变量、反向因果等带来的内生性问题,本文使用竞争集合中其他商家进行决策时可能依赖的个体特征均值作为工具变量(Duflo *et al.*, 2002; 张川川和朱涵宇, 2021)。具体而言,本文计算了同群竞品所在店铺的评价得分均值,以其滞后项($L.gmshopcmt_{-1}$)作为工具变量。店铺评分由过往诸多产品销售与客户服务决定,通常是不会迅速发生明显变化的前定变量。同时,以模糊化商品名称信息误导消费者是有可能对店铺评分造成负向影响的投机行为,商家需要考虑当前评分水平进行决策。但是,该工具变量对关注商品的商家策略选择没有直接影响。使用滞后项是为了使工具变量与以之为基础决策的核心解释变量处于同期。采用两步最小二乘法进行回归,第一阶段的F统计量都明显大于10,且模型通过了识别不足和弱工具变量检验。同时,无论以 $score_{重复}$ 还是 $score_{矛盾}$ 为计算基础, $L.gvague_{-1}$ 的回归系数均显著为正^②。

七是使用动态面板模型。动态面板模型引入了被解释变量的滞后项,能更加有效地控制随时间缓慢变化的因素。估计动态面板模型使用的工具变量可以包括内部与外部两类。本文采用系统广义矩估计(GMM)动态面板模型,该模型将差分方程和水平方程作为一个系统进行联合估计,并分别使用水平滞后项和差分滞后项作为内部工具变量。根据工具变量构建要求,需要采用Arellano-Bond自回归(AR)检验,通过考察原始模型误差项的自相关情况判断内部工具变量与其对应方程的误差项是否存在相关性,若结果显示存在1至*l*阶自相关而不存在*l*+1阶自相关(*l*不应过大),则应在回归过程中从*l*+1阶的滞后项开始构建工具变量。本文根据Roodman(2009)的建议,加入时间固定效应以提高动态面板自回归检验结果的有效性。为减少工具变

① 限于篇幅未列出,详见本刊网站本文附录。

② 限于篇幅未列出,详见本刊网站本文附录。

量数量过多造成的不良影响,我们一方面限定各内生变量作为工具变量的滞后期数,并结合工具变量有效性检验情况以折叠(collapse)形式引入全部或部分内部工具变量,另一方面采用主成分分析法对内生形式的工具变量组进行降维处理(Kapetanios and Marcellino, 2010)。就外部工具变量而言,本文继续采用前述滞后1期的同群竞品所在店铺评价得分均值($L.gmshopcmt_{-1}$)。在过度识别检验方面,本文主要考察存在异方差时更加稳健的Hansen检验结果。根据AR检验结果,内部工具变量均从滞后3阶开始。回归结果显示,核心被解释变量 $L.gvague_{-1}$ 仍然显著为正^①。

(二)对市场影响效应的计量分析

本文假说2描述了商品名称信息模糊化对市场价格水平的影响,下面分别从细类和个体两个层次进行检验。

1. 细类层次的检验。将每一个产品细类(j)看作一个行业,则可以在细类层面构建以平均价格($mprice$)为被解释变量,商品名称模糊度均值($mvague$)为核心解释变量的回归模型。其中,商品名称模糊度均值可以在前述各竞争群组商品名称平均信息模糊程度的基础上进一步计算均值得到。在一个细类市场中,一部分产品可能与许多产品具有竞争关系,而另一部分产品则仅与少量产品具有竞争关系,在按各个单品计算其同群竞品的名称信息模糊度时就体现出了这一点。在此基础上进一步计算均值,得到的结果就考虑了各个商品名称信息模糊程度对整个细分市场影响大小的权重。但直接估计由此构建的回归模型将面临内生性问题。进入门槛低、商家数量多的领域本身更容易发生同质化竞争,其中低价、低质商品无法在功能创新上进行竞争,而更倾向使用信息模糊策略;相反,能够形成差异化优势的商家则可能倾向在制定较高价格的同时使用信息精准策略。也就是说,价格与信息模糊度之间可能存在反向因果且影响方向同样为负。为此,本文采用(6)式的动态面板模型进行回归。

$$mprice_{jt} = \beta_0 + \beta_1 mprice_{j,t-1} + \beta_2 mvague_{jt} + \eta_t + \varepsilon_{it} \quad (6)$$

如前所述,动态面板模型通过引入被解释变量的滞后项控制中长期因素,而细类产品的平均价格及其决定因素刚好符合这一特点,由此得到的回归结果可以更好地反映解释变量在当期产生的直接影响。同时,动态面板模型采用工具变量估计,能够在较大程度上缓解内生性问题。根据Roodman(2009)的建议,本文采取不同滞后期数、主成分分析降维数量等设定,在各项检验结果满足要求的情况下观察结果的变化,以此作为稳健性检验的一个途径。表6给出了采用上述动态面板模型的回归结

^① 限于篇幅未列出,详见本刊网站本文附录。

果,第(1)–(4)列和(5)–(8)列分别采用重复、矛盾两个维度的商品名称模糊测度指标,其中第(1)和(2)列为1组,是改变动态面板工具变量设定得到的可彼此验证的结果。从中可知,*mvague*的回归系数均显著为负,印证了本文假说2关于商品名称信息模糊化对产品细类均价具有负向影响的推断。

表6 假说2计量分析结果:细类层面

因变量: <i>mprice</i>	(1)	(2)	(3)	(4)
<i>L.mprice</i>	0.8370*** (0.0958)	0.8278*** (0.0887)	0.8844*** (0.1220)	0.9686*** (0.3528)
<i>mvague</i> 重复	-2.7120** (1.3458)	-2.5194*** (0.9614)	-1.8985** (0.8797)	-7.4953** (3.7087)
χ^2	2.05e05***	3.04e05***	2.46e05***	7.71e04***
AR(1)检验 p 值	0.0056	0.0044	0.0084	0.0105
AR(2)检验 p 值	0.9458	0.9229	0.9827	0.5622
Hansen 检验 p 值	0.1094	0.6564	0.1951	0.6991
KMO 值	0.8786	0.9003	0.8767	0.8918
工具变量数量	34	32	35	31
观测数	2321	2321	2259	2916
	(5)	(6)	(7)	(8)
<i>L.mprice</i>	0.9638*** (0.2474)	0.9680*** (0.2750)	0.7522*** (0.1476)	0.7202*** (0.2677)
<i>mvague</i> 矛盾	-14.5716** (6.1543)	-16.5412** (6.8307)	-9.1374** (4.3876)	-7.2046** (3.6473)
χ^2	9.45e04***	6.13e04***	1.26e05***	7.96e04***
AR(1)检验 p 值	0.0072	0.0115	0.0037	0.0325
AR(2)检验 p 值	0.1165	0.1057	0.2363	0.2905
Hansen 检验 p 值	0.5132	0.8262	0.4041	0.4825
KMO 值	0.9089	0.9154	0.9144	0.8600
工具变量数量	32	31	35	34
观测数	2270	2270	2415	2916

除使用不同基础计算指标和改变工具变量设定外,本文还采取以下方式检验结果的稳健性。一是调整确定同群竞品时采用的临界值,改为35%,由此得到的结果见表6第(3)和(7)列。二是改变核心解释变量的计算方式,由基于各竞争群组商品名称的平均信息模糊程度计算改为直接使用各单品的名称信息模糊程度计算算术均值,由此得到的结果见表6的第(4)和(8)列。三是以细分指标 $DScore_{\chi}$ 和 $HMScore_{\chi}$ 计

算核心解释变量,以排除综合指标构建时所用权重的影响。在结果中 *mvague* 的回归系数仍显著为负,说明细类层次的回归结果具有一定稳健性^①。

2. 单品层次的检验。在单品层次的分析也需要构建以价格为被解释变量,名称模糊程度为解释变量的回归模型。若直接以单品的上述指标构建回归模型,则同样面临前述价格水平与商品名称信息模糊化之间存在反向因果的内生性问题。为此,本文采取以下回归模型构建方式。

被解释变量方面,本文使用单品各期价格与基期价格的比值($pstd_{it}=p_{it}/p_{i0}$)作为被解释变量,其中的基期($t=0$)为样本期间的首期或首次观测到商品进入市场的当期。经过上述方式计算的被解释变量可以理解为只包含自身各期内的价格变化信息,而与价格水平高低无关的相对价格。考虑到在半年的样本期内,行业进入门槛、生产经营成本、商家经营定位等决定商品价格水平、市场竞争策略的因素不会发生明显变化,这样的处理能在较大程度上排除产品间价格水平与命名策略存在反向因果带来的影响。为避免极端异常值的影响,本文对 $pstd_{it}$ 做双侧各 0.5% 的删尾处理。

核心解释变量方面,本文采用各单品所属竞品群体中全部商品名称模糊程度的均值以更好地刻画多主体共同行为,并取滞后 1 期项(L.*gm vague*)。这样的设定能在较大程度上减少产品内部生命周期阶段变化带来的定价调整与命名策略之间的反向因果问题,因为一方面单个产品的命名策略变化不会给同群竞品名称模糊度平均水平带来较大影响,即便考虑存在相互模仿行为,这种影响也不会迅速显现;另一方面,计算同群竞品名称平均模糊度的产品范围是由滞后 1 期的价格水平确定的,也可以排除定价水平调整引起竞品集合变化造成的影响。

控制变量方面,继续在产品层次控制商品的历史评分(*goodcmt*)、所在店铺的历史评分(*shopcmt*)、是否有折扣券促销(*coupon*),在竞争群体层次控制其他店铺数量(*numshop*)和销售单品总数(*numsku*),这些因素也会对产品相对价格产生影响^②。同时,仍然控制产品及时间固定效应。由此得到计量回归模型:

$$pstd_{it} = \beta_0 + \beta_1 gm vague_{i,t-1} + \beta_c control_{it} + \eta + \varepsilon_{it} \quad (7)$$

按(7)式进行回归,表7第(1)和(7)为加入控制变量的回归结果,第(2)和(8)列为不加入控制变量的回归结果,其中 L.*gm vague* 的系数均显著为负,表明竞品群体的

① 结果详见本刊网站本文附录。

② 这里不再控制细类层面的商品名称模糊程度均值(*meanvague*),因为该变量与作为核心解释变量的竞争群体商品名称模糊程度均值(*gm vague*)高度相关,而与作为被解释变量的关注商品相对价格(*pstd*)几乎无关,会对统计推断造成不利影响(张子尧和黄炜,2025)。

名称信息平均模糊程度提高会使关注商品价格水平下降,这是市场价格效应假说2在单品层面的体现。表7同时给出了改变固定效应控制方式、改变标准误计算方法、改变同群划分临界值的稳健性检验结果,结果见第(3)-(5)和(9)-(11)列,其中L.gmvague的回归系数依然显著为负。同时,本文继续使用细分测算指标进行回归以排除综合指标构建权重的影响^①。

除此之外,本文还通过以下方式进行稳健性检验^②。一是使用被解释变量未删尾的全部样本进行分析,结果见表7第(6)和(12)列。可以看到,L.gmvague的回归系数仍显著为负。此外,全样本回归得到的回归系数明显大于使用删尾后样本回归的结果,说明本文以删尾样本作为基准回归是较为审慎的选择。

表7 假说2计量分析结果:单品层面

	(1)	(2)	(3)	(4)	(5)	(6)
L.gmvague _{重复}	-0.0979*** (0.0077)	-0.1036*** (0.0071)	-0.0967*** (0.0074)	-0.1036*** (0.0130)	-0.0741*** (0.0054)	-0.1946*** (0.0207)
控制变量	未控制	控制	控制	控制	控制	控制
周固定效应	控制	控制	控制	控制	控制	控制
店铺×周固定效应	未控制	未控制	控制	未控制	未控制	未控制
单品固定效应	控制	控制	未控制	控制	控制	控制
F值	163.0492***	1435.1799***	1329.8501***	12049.4685***	1391.8046***	520.9648***
调整后的R ²	0.5890	0.6395	0.7155		0.6389	0.7176
观测数	1150 144	1094 238	1079 910	1133 383	1063 909	1108 944
	(7)	(8)	(9)	(10)	(11)	(12)
L.gmvague _{矛盾}	-0.0515*** (0.0113)	-0.0271** (0.0108)	-0.0388*** (0.0103)	-0.0271* (0.0138)	-0.0712*** (0.0146)	-0.1137*** (0.0253)
控制变量	未控制	控制	控制	控制	控制	控制
周固定效应	控制	控制	未控制	控制	控制	控制
店铺×周固定效应	未控制	未控制	控制	未控制	未控制	未控制
单品固定效应	控制	控制	未控制	控制	控制	控制
F值	20.6118***	1335.7074***	1243.0559***	2223.5179***	1403.8323***	482.8076***
调整后的R ²	0.5881	0.6380	0.7152		0.6390	0.7488
观测数	1104 865	1052 930	1038 769	1091 893	1099 533	1066 978

说明:第(4)和(10)列括号里的值为D-K稳健标准误,其余为聚类到单品的稳健标准误。第(5)和(11)列使用35%作为确定同群竞品是采用的价格分位数临界值,其余各列为30%。

① 限于篇幅未列出,结果详见本刊网站本文附录。

② 限于篇幅第二至四项稳健性检验结果未列出,详见本刊网站本文附录。

二是使用自身商品命名模糊程度未发生改变的样本。如前所述,之所以本文使用同群竞品的名称模糊程度作为核心解释变量能减少内生性问题带来的影响,是因为同一竞争集合中的其他竞品虽然与关注商品的命名行为存在相互模仿现象,但不会迅速、明显地显现。为考察这一前提假设与结论稳健性的关系,本文筛选出 $score_x$ 的当期值不高于前期值、不小于后期值的样本。这些样本在三周内均未提升自身名称的模糊程度,故可以进一步排除商家改变经营策略而采取模糊策略并降低销售价格的解释,同时该商品不会通过同群效应提升竞品的名称模糊程度。采用(7)式进行回归的结果显示, $L.gmvague$ 的回归系数仍然显著为负。

三是使用严格筛选的未采用模糊策略的样本。根据本文的 $DSscore_x$ 评分规则,得分为0表示人工智能系统未发现商品名称中存在矛盾或重复现象。仅保留当期 $DSscore_{重复}$ 和 $DSscore_{矛盾}$ 为0样本时的样本,或仅保留当期及前后期 $DSscore_{重复}$ 和 $DSscore_{矛盾}$ 均为0的样本,结果中 $L.gmvague$ 的回归系数仍显著为负。由于这些商品本身未采用任何(可被人工智能察觉到的)名称模糊策略,这一结果表明竞争集合中其他商品的信息模糊策略对未采取模糊策略的商品相对价格也具有负向影响,假说2具有稳健性。

四是使用工具变量回归。为进一步排除内生性问题,本文采用前述同群竞品所在店铺评价得分均值的滞后项($L.gmshopcmt_{-1}$)作为工具变量进行两步最小二乘(2SLS)回归。该变量不会直接影响所关注商品的相对价格,仍然符合工具变量构建条件。结果显示,第一阶段回归排除IV的F值大于10,且 $L.gmvague$ 的回归系数仍保持显著为负。

六 机制检验

在前文的理论分析中,采取商品名称信息模糊策略的商家能从采取精准策略的商家那里掠走销量,是形成多主体博弈囚徒困境的核心原因。由此带来逆向选择问题,是造成市场总体价格水平下降的关键机制。

本文仍然关注由各个商品构建的竞争集合,先根据 $DSscore_x$ 的得分,将各个竞争集合中的商品分为采取精准策略(J)和模糊策略(M)的两组,然后参考 Zhuo(2017)的做法,分别计算两组商品的平均价格(gmp^J 、 gmp^M),并求取其差值($dgmp = gmp^J - gmp^M$)。这一做法是将采取模糊策略的商品作为参照,因为该差值反映的是 gmp^J 、 gmp^M 的相对变化方向。以此为被解释变量,竞争集合中商品名称平均信息模糊程度的滞后项($L.gmvague$)作为核心解释变量构建回归模型。如果 $L.gmvague$ 回归

系数显著为负,则表明 gmp^l 、 gmp^m 受到商品名称信息模糊化的影响是不同的,采取精准策略的商品价格发生了相对下降。模型中的控制变量及固定效应设定与前文保持一致。由此得到:

$$dgmp_{it} = \beta_0 + \beta_1 gm vague_{i,t-1} + \beta_c controls_{it} + \eta + \varepsilon_{it} \quad (8)$$

这里还要同时考察商品销量的变化。一方面,如前文理论分析部分所述,这样的价格下降无法带来销量的增长,本文需要加以验证;另一方面,采取精准策略的商品销量会被掠走是多主体博弈囚徒困境的核心成因,若能同时排除其价格相对发生上涨,则可以同时验证这一机制。因此,本文按照与价格相一致的方式计算两组商品平均销量变化的差值变量($dgmq$),并构建如(9)式所示的回归模型:

$$dgmq_{it} = \beta_0 + \beta_1 gm vague_{i,t-1} + \beta_c controls_{it} + \eta + \varepsilon_{it} \quad (9)$$

表8给出了对(8)和(9)式分别进行回归的一系列结果($dgmp$ 、 $dgmq$ 均做对数化处理),包括采用不同的评分计算基础指标(该指标与作为商品分组标准的指标相对应)、不加入或加入控制变量、使用聚类标准误或D-K标准误的情形。可以看到,在以 $dgmp$ 为被解释变量的回归结果中, $L.gm vague$ 的回归系数显著为负或不显著;而在以 $dgmq$ 为被解释变量的回归结果中, $L.gm vague$ 的回归系数均显著为负。

表8 机制分析结果

指标计算基础	score 重复			score 矛盾		
被解释变量: $dgmp$	(1)	(2)	(3)	(4)	(5)	(6)
$L.gm vague$	-0.0311*** (0.0046)	-0.0248*** (0.0043)	-0.0248** (0.0094)	0.0084 (0.0055)	0.0025 (0.0052)	0.0025 (0.0143)
F值	44.8679***	52.5027***	13867.18***	2.3247	5.7454***	6.85e05***
调整后的R ²	0.5781	0.5680		0.7605	0.7547	
观测数	1148 874	1094 623	1133 045	1116 909	1065 119	1103 516
被解释变量: $dgmq$	(7)	(8)	(9)	(10)	(11)	(12)
$L.gm vague$	-0.6669*** (0.0750)	-0.6121*** (0.0767)	-0.6121** (0.2198)	-2.0034*** (0.0821)	-1.8718*** (0.0842)	-1.8718*** (0.2820)
F值	79.1240***	92.2499***	1.23e05***	594.7588***	95.2114***	55211.11***
调整后的R ²	0.5499	0.5480		0.5766	0.5730	
观测数	1139 856	1086 320	1124 697	1113 948	1062 450	1100 753
控制变量	未控制	控制	控制	未控制	控制	控制

说明:第(3)、(6)、(9)、(12)列括号中的值为D-K稳健标准误,其余为聚类到单品的稳健标准误。

综上所述,竞争集合中商品名称信息平均模糊程度的提高对采取精准策略的商品存在明显不利,其相对销量下降而相对价格未上涨甚至同时下跌,或者相对价格下跌而相对销量还同时下降。这些结果体现了本文关于商品名称信息模糊化导致消费者购买选择变化,进而产生逆向选择的核心机制。

七 结论与启示

内卷式竞争以经营主体无谓的精细化、复杂化作为主要行为特征,无法实现资源优化配置并推动迈向高水平的发展。本文围绕线上服装消费品市场,聚焦于商品名称标题党现象展开分析,发现这种做法的实质是模糊化商品信息,具有不正当竞争行为的特点,并在线上规则不完善的情况下演变为内卷式的集体行动,由此带来的竞争同质化与逆向选择会导致商品总体价格水平下降的内卷式竞争后果。进一步地,本文在对标题党商品名称实例进行分析的基础上,将导致信息模糊的原因解构为用词矛盾和语义重复两个维度,然后综合使用生成式大语言模型“深度求索(Deepseek)”智能评分与人工构建关键词词典匹配两种方法,对2023年4月到11月共计26周内某大型电商购物平台服装装饰物品类中涉及约20万个单品的40余万个名称进行分析,发现这些商品名称中有接近10%存在用词矛盾,有大约40%存在至少1组关键词重复。在此基础上构建商品名称模糊程度的综合测度指标,采用同群效应的分析思路进行计量分析并进行多种稳健性检验,结果验证了商品名称信息模糊化存在共同行为特征,且会导致行业性的内卷式竞争后果。上述分析结论对于认识内卷式竞争,尤其把握线上市场消费品领域内卷式竞争的成因机制并讨论相应对策具有重要现实意义。本文将带来的政策启示归纳为以下4个方面。

第一,总体上,应统筹运用竞争政策与产业政策对内卷式竞争进行防范与治理。竞争政策应侧重于建立健全可监督、可执行的市场竞争规则,对各类不正当竞争行为形成有效约束;产业政策则应注重通过制度引导与机制设计,推动行业开展以创新升级为导向的自觉协调。内卷式竞争常伴随不正当竞争行为,例如商品名称标题党等现象中存在的虚假宣传倾向。如果缺乏有效的惩戒与纠正机制,此类行为可能持续蔓延并固化为行业惯性,导致竞争重心偏离正轨、价格信号失真,最终表现为无序低价竞争。因此,尽管内卷式竞争成因复杂,但强化竞争政策、维护市场秩序,对减少其对资源配置效率的损害具有基础性作用。同时,产业政策所推动的行业自律与协调,并非针对低价竞争表象而构建价格或产量联盟,而是要通过建立行业规范、树立良性

竞争导向,推动形成以创新和品质为核心的健康竞争文化。

第二,建立更加完善的消费品线上市场经营规则体系,重点规范、治理包括商品命名等在内的信息展示与宣传促销中的信息模糊化行为。传统观点认为,线上市场搜寻成本较低、信息更为透明,而这一点恰恰为企业在竞争中采取信息模糊化手段提供了动机,而经营规则的漏洞则为之提供了条件。在2024年底国家市场监督管理总局督办制售假冒伪劣羽绒制品问题的同时,中国羽绒工业协会发布《羽绒制品消费提示》,明确提到要看清产品名称细节,其中列举的问题在前文实例分析中也得到了体现。线上平台中的商品名称、信息标注等作为消费者感知商品特性的重要信息来源,应考虑适用与线下实体商品相同的质量规则,包括要求完整清晰展示成分表、水洗标等。对于经营者使用与商品实际不符的关键词进行标注的做法,应考虑适用虚假宣传或销售假冒伪劣商品的法律规则。电商平台经营者应当充分利用基于大数据、人工智能等技术的文本分析手段,强化对平台内包括入驻商家销售商品在内的名称命名及宣介内容的把关。目前的人工智能技术已可以用来筛选出违规可能性较高的信息内容,大大降低人工检查的成本投入。而第三方推广机构在根据流量或销量向线上商家提供命名关键词推荐等服务时,应当对可能承担的法律风险予以关注。

第三,要拓展对线上市场竞争维度的认识,建立与线上市场快速发展相适应的不正当竞争监管机制。线上市场价格竞争激烈,低于成本价销售商品的不正当竞争行为一直是监管部门的关注重点。但随着线上市场竞争焦点逐渐转向流量与注意力,加之商品呈现、顾客沟通方式与传统线下市场相比有诸多不同,以此延伸出的新型竞争行为变得纷繁复杂,甚至低价本身也成为流量竞争的手段^①。这正是数字经济时代市场监管面临更大挑战的重要原因。本研究表明,除以低价吸引流量外,包括商品名称标题党行为在内的信息呈现与沟通途径同样可能成为诱导流量的手段,并由此引发不正当竞争。对此,需要建立更加动态、灵活的市场监管机制,基于具体场景与数据监测识别竞争行为,根据行业特征与行为危害程度采用差异化监管工具。

第四,要明确防范内卷式竞争的行业协调重点,以推动消费升级、支撑行业长期发展为导向,优化电商平台及入驻商家经营行为。在生产环节,可通过设立专项资金,对开展新材料、新工艺、个性化设计的企业给予支持。同时,应鼓励企业基于消费需求分层,开发适应不同需求的产品线,避免陷入单一价格战。对消费者存在刚需且关注性价比的商品,则可推动建立行业质量分级与标识制度,通过标准公开与认证引

^① 《“内卷”之因》,《经济日报》2025年4月1日第1版。

导有序竞争。在流通环节,应推动电商平台优化流量分配机制,通过设立“新品孵化专区”“品质认证频道”等方式,使流量向创新型、品质型产品倾斜。

参考文献:

- 蔡祖国、李世杰(2022):《互联网平台“二选一”策略性行为的垄断机理研究》,《世界经济》第12期。
- 杜勇、娄靖、胡红燕(2023):《供应链共同股权网络下企业数字化转型同群效应研究》,《中国工业经济》第4期。
- 李晓华(2025):《中国“新三样”产业发展面临的挑战与机遇——“内卷式”竞争如何“破圈突围”》,《人民论坛》第1期。
- 刘彤(2025):《向“内卷式”竞争亮剑:食品行业需要一场质量革命》,《中国食品报》2025年4月2日第1版。
- 刘小鲁、鲍仁杰(2020):《组合排序规则、产品质量与平台生态》,《经济研究》第6期。
- 刘志彪、王兵(2024):《中国制造业“内卷式”恶性竞争的发生机制与破解路径》,《财经问题研究》第12期。
- 刘重阳、曲创、王夕琛(2025):《互联网平台中的信息不对称与卖方策略性定价行为——来自中国电商平台的证据》,《财经问题研究》第3期。
- 欧阳日辉、刘璇(2025):《数字平台“内卷式”竞争的发生机制、潜在风险与破解对策》,《财经问题研究》第3期。
- 王海燕、张占斌(2025):《“内卷式”竞争:表现、成因及治理》,《改革》第3期。
- 王宇、王梅、黄广映(2019):《平台可以做到大而美吗——不同排序机制下的厂商质量选择研究》,《中国工业经济》第4期。
- 魏瑾瑞、王金伟(2022):《在线评论回报的动态声誉机制研究》,《中国管理科学》第1期。
- 魏银科(2025):《价格内卷风暴下,线下女装店淡季靠啥赚钱?》,《齐鲁晚报·齐鲁壹点》2025年7月16日, <https://www.qlwb.com.cn/detail/26367230.html>。
- 殷继国(2025):《平台滥用规则行为的反不正当竞争法规制》,《比较法研究》第5期。
- 张川川、朱涵宇(2021):《新型农村社会养老保险参与决策中的同群效应》,《金融研究》第9期。
- 张昊、冯永晟(2022):《线上价格为何频繁大幅波动:引导购买行为的策略性定价》,《世界经济》第3期。
- 张杰、任元明(2025):《中国“内卷式”竞争的主要特征、形成机制及其破解之道》,《北京交通大学学报(社会科学版)》第2期。
- 张子尧、黄炜(2025):《实证研究中的控制变量选择:原理与原则》,《管理世界》第10期。
- 钟春平、翟乃森(2022):《产能利用率具有阈值和评判价值吗——“产能过剩”误区的再考察》,《财贸经济》第11期。
- 甄艺凯、崔凯南(2025):《双寡头电商平台竞争研究:基于纵向产品差异化视角》,《经济理论与经济管理》第1期。
- Akerlof, G. A. “The Market for Lemons: Quality Uncertainty and the Market Mechanism.” *Quarterly Journal of Economics*, 1970, 84(3), pp. 488–500.
- Bisztray, M.; Koren, M. and Szeidl, A. “Learning to Import from Your Peers.” *Journal of International Economics*, 2018, 115, pp. 242–258.
- de Clippel, G.; Eliaz, K. and Rozen, K. “Competing for Consumer Inattention.” *Journal of Political Economy*, 2014, 122(6), pp.1203–1234.

- Driscoll, J. C. and Kraay, A. C. “Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data.” *Review of Economics and Statistics*. 1998, 80(4), pp. 549–560.
- Duflo, E. and Emmanuel S. “Participation and Investment Decisions in a Retirement Plan: The Influence of Colleagues’ Choices.” *Journal of public Economics*, 2002, 85(1), pp. 121–148.
- Einav, L.; Finkelstein, A. and Cullen, M. R. “Estimating Welfare in Insurance Markets using Variation in Prices.” *Quarterly Journal of Economics*. 2010, 125(3), pp. 877–921.
- Geertz, C. *Agricultural Involvement: The Processes of Ecological Change in Indonesia*. California: University of California Press, 1969.
- Goldenweiser, A. “Loose Ends of Theory on the Individual Pattern, and Involution in Primitive Society.” *Essays in Anthropology*, 1936, 99, pp.104.
- Hopkins, C. C. *Scientific Advertising*. Chicago/New York: Lord & Thomas, 1923.
- Hviid, M. and Shaffer, G. “Hassle Costs: The Achilles’ Heel of Price-Matching Guarantees.” *Journal of Economics and Management Strategy*, 1999, 8(4), pp. 489–521.
- Jin, G. Z.; Lu, Z.; Zhou, X. and Fang, L. “Flagship Entry in Online Marketplaces.” *NBER Working Paper*, No. 29239, 2021.
- Li, L. and Xiao, E. “Money Talks: Rebate Mechanisms in Reputation System Design.” *Management Science*, 2014, 60(8), pp. 2054–2072.
- Kapetanios, G. and Marcellino, M. “Factor-GMM Estimation with Large Sets of Possibly Weak Instruments.” *Computational Statistics and Data Analysis*, 2010, 54(11), pp. 2655–2675.
- Roodman, David. “How to do xtabond2: An Introduction to Difference and System GMM in Stata.” *The Stata Journal*, 2009, 9(1), pp. 86–136.
- Rothschild, M. and Stiglitz, J. E. “Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information.” *Quarterly Journal of Economics*. 1976(90), pp. 630–649.
- Wang, C.; Qiu, M.; Shi, C.; Zhang, T.; Liu, T.; Li, L.; Wang, J.; Wang M.; Huang, J. and Lin, W. “EasyNLP: A Comprehensive and Easy-to-Use Toolkit for Natural Language Processing.” *arXiv preprint*, arXiv: 2205. 00258, 2022.
- Zhuo, R. “Do Low-Price Guarantees Guarantee Low Prices? Evidence from Competition between Amazon and Big-Box Stores.” *The Journal of Industrial Economics*. 2017, 65(4), pp. 719–738.
- 飯田経夫 “過当競争と二重構造”, 季刊理論経済学, 1961, 11(3–4), pp. 23–31.
- 两角良彦 “産業協調体制論”, 通商産業研究, 1962, 10(2), pp. 38–50.
- 小宮隆太郎、奥野政寛、鈴木興太郎 “日本の産業政策”, 東京: 東京大学出版会, 1984, pp. 13–15.

Online Sellers’ Clickbait Tactics and Involutionary Competition

Zhang Hao; Yan Yushan; Wang Xinyi; Wang Weilin

Abstract: In online marketplaces, a “clickbait” phenomenon has emerged whereby sellers deploy product titles containing specific keywords to attract consumers’ attention and, in so doing, influence

their purchasing decisions. This paper examines this phenomenon within the online apparel consumer market. The analysis finds that the practice centres on the obfuscation of product information – a characteristic feature of unfair competition. Against a backdrop of inadequate online regulation, such behaviour evolves into a form of collective action that precipitates involution. The resulting homogenisation of competition and adverse selection ultimately drive an involutory outcome – namely, a decline in the aggregate price level of goods. Based on an examination of actual clickbait product titles, the study deconstructs the sources of information obfuscation into two distinct dimensions: contradictory wording and semantic repetition. A mixed-method approach is then employed, combining intelligent scoring generated by the generative large language model “DeepSeek” with manual keyword dictionary matching. This approach is applied to approximately 400,000 product titles – associated with 200,000 individual items in the clothing and accessories category of a major e-commerce platform – over a 26-week period from April to November 2023. The results reveal that nearly one-tenth of these titles contain contradictory wording, while approximately 40% contain at least one set of repeated keywords. On this basis, a comprehensive metric for measuring the degree of obfuscation in product titles is constructed. An empirical analysis is subsequently conducted using a peer-effects framework, supported by multiple robustness tests. The findings confirm that information obfuscation in product titles exhibits herd behaviour characteristics and gives rise to industry-wide involutory consequences.

The contributions of this paper are threefold. First, by focusing on micro-level competitive behaviour among firms, it explores the mechanism underlying involutory competition from the perspective of multiple operators collectively trapped in a prisoner’s dilemma – a situation that drives the homogenisation of competition. This analysis underscores that addressing involutory competition requires due attention to the distortions in competitive mechanisms arising from incomplete market rules. Second, the research concentrates on the consumer goods sector, with particular emphasis on the highly competitive online market. It examines the non-price competitive strategies that firms adopt to capture consumer attention and compete for web traffic – strategies that are central to understanding rivalry in digital marketplaces. Third, in the process of conducting textual analysis on product titles and constructing measurement indicators, the study makes full use of generative artificial intelligence based on large language models. This methodological choice carries instrumental significance for research involving natural language processing, offering a replicable framework for future empirical work.

The conclusions of this paper imply that preventing and addressing involutory competition requires the coordinated application of competition policy and industrial policy. Competition policy should focus on establishing and refining enforceable and monitorable market competition rules – thereby imposing effective constraints on all forms of unfair competition. Industrial policy, in contrast, should emphasise institutional guidance and mechanism design, with the aim of promoting industry-wide self-coordination oriented towards innovation and upgrading.

Key words: online marketplaces, product titles, information obfuscation, multiple operators trapped in a prisoner’s dilemma, involutory competition

JEL codes: D23, D83, L13, L41

(截稿:2026年3月 责任编辑:王 徽)