

基于多模态大模型的直播电商视觉特征量化及非线性转化效应研究

王子珺

上海理工大学管理学院, 上海

收稿日期: 2026年8月6日; 录用日期: 2026年8月20日; 发布日期: 2026年10月16日

摘 要

随着直播电商的发展, 如何合理配置视觉内容成为影响消费者参与度的重要因素。针对以往视觉营销研究主要依赖主观测量, 且大多局限于简单的“刺激-响应”线性假设等问题, 本研究结合认知负荷理论, 尝试引入Rex-Omni多模态大模型来进行视觉特征的量化与实证分析。具体而言, 通过对620份直播样本进行图像解析, 自动提取了主播视觉占比、商品视觉占比、文本框数量及总字符数四项核心指标。基于最小二乘法建立的多元二次回归分析表明, 直播间视觉要素与订单转化率之间存在显著的“倒U型”非线性关系。适度的空间排版和信息呈现有助于增强观众的临场感并改善转化效果, 但如果视觉元素过度堆砌、布局过于细碎, 则容易打破画面的视觉平衡, 加重消费者的认知负担, 进而降低其购买意愿。本研究证实了多模态大模型在处理电商非结构化视觉数据上的可行性, 也为直播平台规避视觉陷阱、合理规划画面信息配置提供了客观的参考依据。

关键词

直播电商, 视觉营销, 多模态大模型, 认知负荷, 非线性效应

Research on Visual Feature Quantification and Nonlinear Conversion Effects in Live-Streaming E-Commerce Based on Multimodal Large Language Models

Zijun Wang

Business School, University of Shanghai for Science and Technology, Shanghai

Received: August 6, 2026; accepted: August 20, 2026; published: October 16, 2026

文章引用: 王子珺. 基于多模态大模型的直播电商视觉特征量化及非线性转化效应研究[J]. 电子商务评论, 2026, 15(10): 230-238. DOI: 10.12677/ec.2026.15101111

Abstract

With the development of live-streaming e-commerce, the reasonable configuration of visual content has become an important factor influencing consumer engagement. To address the limitations of previous visual marketing research, which mainly relied on subjective measurement and was largely confined to simple linear “stimulus-response” assumptions, this study integrates cognitive load theory and attempts to introduce the Rex-Omni multimodal large model for the quantification and empirical analysis of visual features. Specifically, through image parsing of 620 live-streaming samples, four core indicators were automatically extracted: the visual area ratio of the anchor, the visual area ratio of the product, the number of text bounding boxes, and the total character count. A multiple quadratic regression analysis based on Ordinary Least Squares indicates a significant “inverted U-shaped” nonlinear relationship between the visual elements in the live-streaming room and the order conversion rate. Moderate spatial layout and information presentation help enhance viewers’ sense of presence and improve conversion performance; however, if visual elements are excessively crowded and the layout is too fragmented, it is easy to disrupt the visual balance of the screen, increase the cognitive burden on consumers, and consequently reduce their purchase intention. This study verifies the feasibility of multimodal large models in processing unstructured visual data in e-commerce, and provides an objective reference for live-streaming platforms to avoid visual traps and rationally plan the configuration of on-screen information.

Keywords

Live-Streaming E-Commerce, Visual Marketing, Multimodal Large Language Model, Cognitive Load, Nonlinear Effects

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在当前电子商务体系中,直播电商已经发展成为关键的商品流通渠道。相较于传统的静态页面展示,直播间的视觉内容(例如画面构图、多媒体信息布局等)呈现出高密度和强动态的特征,这些特征直接影响着消费者的直观体验与购买决策[1]。然而,现有的部分直播营销实践多依赖于运营者的主观经验,在追求高转化率的过程中,往往倾向于不断增加屏幕内的视觉刺激强度,如过度放大商品画面或增加高频闪烁的促销贴片。由于消费者的瞬时认知资源是有限的,这种高密度的视觉信息堆砌可能会导致信息超载,引发视觉疲劳。特别是在用户停留时间较短的直播场景下,如何平衡核心视觉要素的显著度与信息承载量,是优化电商视觉营销亟待解决的问题。

在视觉特征的客观量化方面,传统研究多依赖人工编码或采用独立的计算视觉(CV)模型进行特征提取。这些方法不仅在处理效率上存在局限,且往往难以捕捉复杂画面中的全局语义与局部语义。近年来,多模态大模型(MLLM)凭借其跨模态的统一表征能力,在处理非结构化视觉数据方面展现出了明显优势。本研究选用的 Rex-Omni 模型[2]采用了端到端的视觉语言架构,能够相对客观地解析直播间内复杂的视觉内容特征,为电商视觉内容的量化评估提供了一种新的技术范式。

针对上述直播视觉营销中的量化难点,本研究拟基于多模态大模型技术,探讨直播电商视觉内容对消费者参与度可能存在的非线性影响。首先,利用 Rex-Omni 模型的一体化架构,自动化提取直播间内主

播与商品的空间占用率及文本信息密度等关键特征。在此基础上,结合认知负荷理论构建多元实证模型[3],试图揭示各项视觉特征与电商转化率之间是否存在的“倒U型”关联机制[4]。本研究旨在揭示电商企业及运营者的直播视觉配置的科学化与规范化,并为今后直播视觉配置方面提供实证参考。

2. 文献综述

2.1. 直播电商的内涵及其行业地位

直播电商近年来已成为网络零售的重要组成部分。与传统货架式电商的“主动搜索”逻辑不同,直播电商通常借助丰富的视觉线索来降低消费者的信息搜寻成本,在一定程度上促使消费者由理性的认知加工转向感性决策,从而激发冲动性购买意愿[5]。相关研究指出,直播电商缩短了品牌与消费者间的信息传递链路,有助于降低交易摩擦并提升流通效率[6]。当前,随着市场环境的变化,直播电商的运营模式正逐步从早期的流量驱动,转向以视觉体验和内容为质量的精细化运营阶段[7]。在此背景下,探索直播间视觉要素的合理配置,对电商企业提升转化率具有一定的实际意义。

2.2. 直播电商视觉内容营销研究现状

在视觉营销领域,现有研究主要借助刺激-机体-反应(S-O-R)模型[8]或环境心理学理论,分析视觉要素对消费者心理状态的影响。部分研究侧重于宏观氛围与主播特质的分析,如探讨背景色彩、灯光对消费者沉浸感的作用[9],或分析氛围线索与冲动购买行为之间的逻辑关系[10]。此外,也有学者开始关注商品展示方式、促销贴片等微观线索对购买决策的干预作用。

但针对高动态的直播场景,目前的研究方法仍存在一定的局限。首先,由于多依赖人工编码或主观问卷调查,这类方法难以在大样本下客观量化空间占比、文本密度等高维视觉特征;其次,部分文献往往假设“视觉刺激越强,营销效果越好”,在一定程度上忽略了高密度信息环境下认知负荷超载可能带来的负面效应[11]。因此,引入更客观的量化工具来重新审视视觉刺激与营销效果间的关联显得较为必要。

2.3. 多模态大模型在电商领域的应用与演进

为了提高量化效率,也为了改善人工量化方法的不足,部分学者选择引入计算机视觉(CV)技术来处理电商领域的非结构化图像数据。例如,利用目标检测算法提取图片中的商品显著度特征[12],或通过独立的文本识别技术提取促销文本以评估营销内容[13]。这类基于独立模型拼接的“流水线式”特征提取方式虽然提高了处理效率,但与此同时忽略了图像与文本之间的语义关联,且存在特征提取标准不一的问题。

随着深度学习技术的发展,端到端的多模态大语言模型(MLLM)[14]在视觉数据解析中展现出较好的适用性。以 Rex-Omni 为代表的多模态模型通常具备零样本(Zero-shot)推理能力,能够根据自然语言指令直接处理复杂的图像任务[15]。相比传统的组合算法,这种模型可以在直播画面中同步完成目标检测与文本识别,减少了模型拼接之间的误差,从而更客观地提取图像视觉占比与文本信息密度等特征。这为本文探索视觉营销要素对消费者参与度的非线性影响提供了量化技术基础。

3. 视觉特征提取算法模型构建

3.1. Rex-Omni 多模态模型架构与任务统一化

为了实现直播画面的多维特征提取,本文引入 Rex-Omni 多模态模型作为基座模型。该模型的主要特点是采用统一的框架,尝试将目标检测(Grounding)与光学字符识别(OCR)等不同的视觉任务整合为相同

的自回归的序列生成。

在特征提取阶段, Rex-Omni 首先通过预训练的视觉编码器 g_ϕ , 将输入的直播图像 $I \in \mathbb{R}^{H \times W \times C}$ 映射至特征空间, 并生成一系列相应的视觉 Token 序列 V :

$$V = \{v_1, v_2, \dots, v_n\}, v_i = g_\phi(\text{Patch}_i), \quad (1)$$

其中 Patch_i 代表图像的局部特征块。

随后, 序列 V 会与特定的文本任务指令 P 进行拼接, 形成联合输入序列 $[V; P]$ 。在序列生成阶段, 模型通过计算不同区域不同步骤下的联合概率, 输出最终的结果序列 O :

$$O = f_{\text{Rex-Omni}}(I, P) = \arg \max \prod_{t=1}^T P(o_t | V, P, o_{<t}; \theta), \quad (2)$$

其中, o_t 为第 t 步生成的输出标记, θ 为模型参数。

借助这种结构设计, 模型通常可以在一次前向传播过程中, 同步解析出主播的位置坐标、商品的边界框以及画面中的促销文本。这种处理方式在一定程度上减少了因多个独立模型拼接带来的误差, 有助于保持特征提取的客观性与语义一致性。

3.2. 基于相对坐标量化的空间特征编码

考虑到语言模型词表通常是离散的, 为了能够合理表示直播间中连续的空间坐标, Rex-Omni 采用了一种基于相对坐标的量化编码方法[16]。

设图像宽高分别为 W 和 H , 针对画面里的任意绝对像素坐标 (x, y) , 模型利用量化函数 $Q(x)$ 将其映射至 $[0, L-1]$ (本实验中设定 $L=1000$) 的整数区间内:

$$Q(x) = \text{round}\left(\frac{x}{W} \times (L-1)\right), Q(y) = \text{round}\left(\frac{y}{H} \times (L-1)\right), \quad (3)$$

随后, 模型会分配词汇表末尾的 1000 个 Token 作为专属的坐标标记。在进行特征解析时, 输出的边界框信息通常被规范为“类别 + 坐标”的序列形式:

$$\text{OutputForm} = \langle \text{obj_start} \rangle \text{Class} \langle \text{obj_end} \rangle \langle \text{box_start} \rangle Q(x_{\min}) Q(y_{\min}) Q(x_{\max}) Q(y_{\max}) \langle \text{box_end} \rangle, \quad (4)$$

上述 Rex-Omni 量化编码方式在为模型的训练和推理提供了便利。

首先, 它降低了参数优化的难度。该量化编码方式将连续的回归任务转换为针对 1000 个特殊类别的离散分类任务, 其位置 Token 的交叉熵损失函数计算如下:

$$\text{Loss} = - \sum_{i \in \{x, y\}} \sum_{j=0}^{L-1} y_{i,j} \log(\hat{y}_{i,j}), \quad (5)$$

其中, y 为坐标真实值的 one-hot 编码, $\hat{y}$ 为坐标预测值的 one-hot 编码。

其次, 它提高了模型的推理效率。利用少量的 Token 即可表示一个完整的边界框, 这使得模型在面对直播间内多目标、高密度的视觉信息(例如密集的促销文本)时, 依然能够保持较快的推理速度。

3.3. 视觉特征解析与数据结构化过程

为了将直播间复杂的视觉信号转化为实证模型可用的量化变量, 本研究利用 Rex-Omni 模型执行了端到端的特征解析, 具体结构化过程涵盖目标检测与字符识别两大模块。

首先, 在目标检测(Grounding)阶段, 模型能够根据预设的自然语言指令, 客观锁定直播画面中的核心展示主体。通过在潜在特征空间中的坐标回归, 模型能够精准输出主播轮廓与主推商品的标准化边界

框(Bounding Boxes)。这一过程将原本模糊的视觉区域转化为具体的像素坐标,直接为后续测算主播视觉占比(X_1)与商品视觉占比(X_2)提供了结构化的空间数据参考。

其次,在光学字符识别(OCR)环节,模型同步对散布在背景中的促销贴片、价格及说明字幕等文本区域进行定位与内容提取。采用这种一体化的处理机制,有效避免了传统分散式图像处理组件在拼接时可能产生的信息丢失或对齐误差,从而在一定程度上保证了文本框数量(X_3)与总字符数(X_4)获取的一致性与客观性。

依托上述解析流程,本研究能够在包含较多干扰信息的直播背景下,将非结构化的图像视觉流成功剥离并转化为可供计量的数值指标,从而为下文的非线性回归实证提供了基础数据支撑。

4. 实证指标设置与回归模型分析

4.1. 直播间视觉特征量化指标体系与计算公式设置

为将 Rex-Omni 模型输出的图像信息转化为实证分析所需的数值变量,本文结合认知负荷的相关概念,从空间分布和信息复杂度两个方面,设定了四项具体的量化指标。

假设各直播间截图的统一分辨率记为 $W \times H$, 各核心自变量的具体说明与计算逻辑如下:

1) 主播视觉占比(X_1)

主播是直播互动的主要对象,其在画面中的占比能在一定程度上影响观众的临场体验。假设 Rex-Omni 输出的主播边界框坐标为 $(x_{\min}^a, y_{\min}^a, x_{\max}^a, y_{\max}^a)$, 则其占比的计算公式为:

$$X_1 = \text{Ratio}_{\text{anchor}} = \frac{(x_{\max}^a - x_{\min}^a) \times (y_{\max}^a - y_{\min}^a)}{W \times H}, \quad (6)$$

2) 商品视觉占比(X_2)

商品画面的大小通常反映了产品在当前展示环节的显著程度。设模型输出的核心商品边界框坐标为 $(x_{\min}^p, y_{\min}^p, x_{\max}^p, y_{\max}^p)$, 其占比的计算方式为:

$$X_2 = \text{Ratio}_{\text{product}} = \frac{(x_{\max}^p - x_{\min}^p) \times (y_{\max}^p - y_{\min}^p)}{W \times H}, \quad (7)$$

3) 文本检测框数量(X_3)

该指标主要用于衡量直播间画面布局的碎片信息情况。假设模型在单帧内识别到的独立文本框集合为 $T = \{t_1, t_2, \dots, t_n\}$, 对应的计算公式为:

$$T = \text{Count}(T) = n, \quad (8)$$

4) 文本总字符数(X_4)

该指标用来评估促销文本给观众带来的信息阅读量大小。设第 i 个文本框内的识别字符串长度为 L_i , 则计算公式为:

$$X_4 = \text{Density}_{\text{text}} = \sum_{i=1}^n L_i, \quad (9)$$

4.2. 基于 OLS 的非线性回归模型分析

在进行模型构建前,本研究首先借助电商数据分析平台(如蝉妈妈、飞瓜数据等),采用分层随机抽样的方法,从抖音、快手及淘宝等主流电商平台获取了 2025 年期间的活跃直播间数据。抽样过程严格控制在晚间直播高峰时段(19:00~22:00)进行等间隔截帧,并同步获取相应的后端订单转化率数据。两名独立研

究人员对数据进行了交叉核对与清洗，剔除了画面缺失、数据异常波动的样本，最终保留了 620 份有效样本。

为了验证上述视觉特征到底是怎么影响转化率的，在量化指标提取完毕后，我们基于这 620 个样本，采用普通最小二乘法(OLS) [17]建立了一个带二次项的多元回归模型，重点检验它们之间是否存在非线性关系。

4.2.1. 描述性统计与共线性检验

表 1 是各主要变量的描述性统计结果。从品类分布来看，上述样本涵盖了服装(约 35%)、美妆(约 28%)、食品(约 22%)及其他类目(约 15%)。该抽样比例与当前直播电商行业的总体交易结构基本吻合，能在一定程度上保证后续结论的代表性与适用性。

Table 1. Descriptive statistics of core variables

表 1. 核心变量描述性统计分析

变量类别	变量名称	符号	均值	标准差	最小值	最大值
因变量	订单转化率(%)	Y	4.15	1.32	0.55	9.8
自变量	主播占比(%)	X_1	27.2	9.4	4.2	62
	商品占比(%)	X_2	33.5	13.1	7.5	78
	文本框数量(个)	X_3	11.8	4.9	1	38
	总字符数(个)	X_4	62.4	22.5	8	195
控制变量	商品客单价(元)	C_1	86.5	42.3	15.9	358
	粉丝基数(万人)	C_2	145.2	98.5	5.5	850

注：品类属性(C_3)作为虚拟变量已在模型中控制，此处未列出其描述性统计。

从表 1 的数据分布来看，不同直播间的画面排版和信息密度差别其实挺大的(比如商品视觉占比 X_2 的标准差达到了 13.1，文本总字数 X_4 最多和最少的差了 180 多个)。这也侧面说明了目前各个商家在视觉展现上的做法各不相同。

另外，为了防止自变量之间高度相关导致回归结果跑偏，我们提前做了多重共线性检验。测算出的方差膨胀因子(VIF)最大值只有 2.15，远低于严格的经验警戒线 5。这说明几个视觉指标之间比较独立，可以直接带入 OLS 模型跑回归了。

4.2.2. 回归结果分析与非线性关系验证

为了验证直播画面特征与订单转化率(CVR)之间是否真的存在“倒 U 型”规律，我们构建了如下的多元二次回归模型[18]：

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_1^2 + \beta_3 X_2 + \beta_4 X_2^2 + \beta_5 X_3 + \beta_6 X_3^2 + \beta_7 X_4 + \beta_8 X_4^2 + \sum \gamma_k C_k + \varepsilon, \quad (10)$$

其中， Y 为因变量，即样本直播间的订单转化率， β_0 为常数项， $X_1, \dots, X_4$ 分别为所提四大视觉自变量； $\beta_{1,3,5,7}$ 为一次项回归系数，用于检验视觉元素刚开始增加时，能不能带来正向收益； $\beta_{2,4,6,8}$ 为二次项回归系数，是验证“倒 U 型”关系的核心指标(若该系数显著为负，则说明确实存在非线性的“过犹不及”现象)；此外， $\sum \gamma_k C_k$ 为客单价、粉丝量及品类等控制变量，用来尽量排除外部因素的干扰， ε 为随机误差项。

接着，我们用 OLS 方法对这 620 个样本进行了参数估计，回归结果见表 2。

Table 2. Regression results of the impact of live streaming visual features on conversion rate
表 2. 直播视觉特征对转化率影响的回归估计结果

变量	对应系数	估计值(β)	t 值	P 值	显著性
主播占比 X_1	β_1	0.402	5.82	0.000	显著
主播占比平方 X_1^2	β_2	-0.008	-4.95	0.000	显著
商品占比 X_2	β_3	0.515	6.74	0.000	显著
商品占比平方 X_2^2	β_4	-0.012	-5.12	0.001	显著
文本框数量 X_3	β_5	0.168	4.25	0.000	显著
文本框数量平方 X_3^2	β_6	-0.004	-3.82	0.001	显著
总字符数 X_4	β_7	0.235	4.56	0.000	显著
总字符数平方 X_4^2	β_8	-0.005	-3.64	0.002	显著
控制变量	γ_k	已控制	-	-	-
常数项	β_0	1.082	8.15	0.000	显著
R^2	0.425	Adj- R^2	0.408	F 统计量	18.6

如表 2 所示,模型的拟合情况整体良好($R^2 = 0.425$, $F = 18.6$)。该结果表明用 Rex-Omni 模型抓取出来的几个核心视觉特征,解释了该直播视觉特征存在 42.5%的波动情况。在排除了价格和粉丝基数等客观因素的干扰后,本次研究所求的“非线性效应”依然可以稳定实现。

由第三列数据可知,主播和商品视觉占比(X_1, X_2)和文本检测框数量与总字符数(X_3, X_4),它们的一次项系数 $\beta_{1,3,5,7}$ 均在 1%的统计水平上显著为正,而对应的二次项系数 $\beta_{2,4,6,8}$ 却在 1%的统计水平上显著为负。该结果从统计学角度印证了,核心视觉特征和转化率之间的函数关系为一条开口向下的抛物线(“倒 U 型”)。换言之,“在直播电商活动中,提供越多信息,会得到越好的直播关注度”这种简单的线性思维在实际中是行不通的,各项指标均存在一个适度的区间。

需要强调的是,本研究测算的“倒 U 型”拐点是基于全样本的统计平均发现,在实际应用中具有显著的情境敏感性。视觉配置的平衡点本质上是信息效率与认知容量的博弈,受外部变量调节。首先,品类属性导致认知机制差异。在服装、食品等低涉入度直播中,观众对信息冗余耐受度低,视觉疲劳使拐点更早出现;而家电等高涉入度品类因比对细节需要,其视觉刺激的有效阈值更高。其次,粉丝基数亦具调节作用。高信任度的 IP 背书能降低观众对强视觉刺激的心理防御,缓解高密度布局的压迫感。因此,本研究得出的量化指标并非僵化的绝对标准,而应作为基准参考框架。商家需结合产品特性与账号层级进行灵活的动态微调,在商业传达与用户体验间寻求最佳平衡。

4.2.3. 视觉特征对转化率非线性影响的深入讨论

结合表 2 实验结果,我们可以将这些统计学上的非线性系数与直播场景下的消费者认知规律联系起来进行探讨。

首先,主播与商品的空间占比(X_1, X_2)在影响消费者决策时,表现出一定的平衡效应。这也说明,适度扩大主播与商品的视觉面积有助于提升直播间的焦点信息的传递效率。但在超过一定的临界点后,单一要素的过度放大不可避免地会挤压其他视觉线索的展示空间,从而影响画面的整体协调性。值得注意的是,商品占比的二次项系数绝对值(0.012)大于主播占比(0.008)。这从侧面反映出,相比于主播画面的增加,消费者对商品画面过度占据屏幕所引发的视觉抵触往往更为敏感。

其次,关于画面复杂度和信息量的分析结果,也在一定程度上印证了“认知负荷理论”在直播场景

中的适用性。虽然添加适量的促销贴片是传递价格优势的常规手段,但文本框数量(X_3)与总字符数(X_4)的二次项系数均显著为负。这说明,如果画面布局过于繁琐或文本信息堆砌严重,会额外增加消费者的认知负担。在当下快速变化的直播信息流中,当视觉输入超过了观众的瞬时处理能力,随之产生的视觉疲劳很可能会削弱其购买意愿,促使其做出规避或放弃决策的行为。

综上所述,模型的拟合表现(R^2)进一步支持了这一观点:直播视觉配置并不完全遵循“刺激越强,绩效越好”的单向线性规律。这一实证发现也为电商平台的运营实践提供了一定的参考,即在进行视觉营销时,不宜盲目增加信息密度,而应尝试借助多模态大模型等客观量化工具,在商业信息传达效率与消费者认知承载力之间寻找更为合理的非线性平衡点。

5. 结论与展望

面向直播电商视觉配置对消费者决策的复杂影响,本文尝试将多模态大语言模型引入视觉营销的实证分析中。结合认知负荷理论,研究利用 Rex-Omni 模型从 620 份直播截图中客观提取了主播占比、商品占比、文本框数量及总字符数等核心视觉特征,并通过多元二次回归模型探讨了视觉要素与订单转化率之间的非线性关联。

实验结果表明,多模态大模型在处理非结构化电商视觉数据时具有较好的适用性。模型拟合结果($R^2 = 0.425$)说明,端到端提取的视觉特征能在一定程度上解释直播间转化率的波动,这为大规模解析电商视觉数据提供了一种较为可行的量化思路。此外,实证分析进一步表明,各项核心视觉配置与消费者参与度之间呈现显著的“倒 U 型”关系。适度的画面配置有助于提升消费者的社会临场感,但过度的视觉元素堆砌与碎片化布局可能会增加受众的认知负担,进而对转化效率产生负面影响。数据测算显示,在此次样本中,主播与商品的视觉占比在分别达到约 25.1%与 21.5%时转化表现相对较好。这提示电商从业者在实际运营中,应尽量将视觉空间维持在相互协调的区间内,并适当控制促销文本的信息负荷,避免陷入“过度营销”的视觉陷阱。

尽管本研究利用 Rex-Omni 模型实现了视觉特征的客观量化,但仍存在一定的内在局限。首要限制在于,目前的特征提取依赖于直播间的静态截图。直播是一个高动态、强交互的过程,静态切片虽然能捕获瞬时的空间布局,但会丢失主播的肢体语言连续性、语速节奏变化以及商品镜头的动态推拉等重要时序信息。仅凭静态特征推导动态的消费者认知过程,难以完全刻画真实的观看体验。

针对上述局限,未来的研究可从以下维度进一步深化:一是尝试引入视频流分析技术(Video Stream Analysis),通过连续帧捕捉直播间的动态交互特征,以增强模型对复杂时序营销环境的解释力;二是融合多模态时序数据,例如结合实时语音转文本(ASR)的情感极性分析,以及观众弹幕互动内容的语义挖掘,构建涵盖“视觉-听觉-交互”全方位的转化预测体系,进而得出更具颗粒度的管理启示。

参考文献

- [1] 徐佳倩. 电商直播营销模式对消费者购买决策的影响探究[J]. 电子商务评论, 2025, 14(10): 1325-1331.
- [2] Jiang, Q., Huo, J., Chen, X., *et al.* (2025) Detect Anything via Next Point Prediction. arXiv: 2510.12798.
- [3] Jiang, Z. and Benbasat, I. (2007) The Effects of Presentation Formats and Task Complexity on Online Consumers' Product Understanding. *MIS Quarterly*, 31, 475-500. <https://doi.org/10.2307/25148804>
- [4] 邓勇. 电商直播信息密度与“冲动购买”: 信息过载的倒 U 型检验[J]. 电子商务评论, 2025, 14(12): 4470-4478.
- [5] 王雨辰, 王霜梦. 直播电商情境赋能与大学生冲动性购买的生成机制: 基于 S-O-R 理论的整合与重构[J]. 电子商务评论, 2025, 14(8): 835-842.
- [6] 韩箫亦, 许正良. 电商主播属性对消费者在线购买意愿的影响[J]. 外国经济与管理, 2020, 42(10): 62-75.
- [7] 徐睿. 平台型电商企业流量成本激增背景下的财务风险识别与防控研究[J]. 电子商务评论, 2026, 15(2): 318-324.

-
- [8] Sun, Y., Shao, X., Li, X., Guo, Y. and Nie, K. (2019) How Live Streaming Influences Purchase Intentions in Social Commerce: An IT Affordance Perspective. *Electronic Commerce Research and Applications*, **37**, Article ID: 100886. <https://doi.org/10.1016/j.elerap.2019.100886>
- [9] 孙苏红. 电商直播间主播行为特征对消费者购买意愿影响的研究综述[J]. 电子商务评论, 2025, 14(11): 460-467.
- [10] 左晶晶, 李盈盈. 电商直播过程中氛围线索与消费者冲动购买行为基于链式中介分析[J]. 西部经济管理论坛, 2024, 35(2): 89-98.
- [11] 陆庆宁, 刘洁. 直播电商中弹幕对消费者购买决策的影响[J]. 电子商务评论, 2025, 14(6): 3333-3339.
- [12] Wang, Z., Liu, X., Li, H., Shi, J. and Rao, Y. (2018) A Saliency Detection Based Unsupervised Commodity Object Retrieval Scheme. *IEEE Access*, **6**, 49902-49912. <https://doi.org/10.1109/access.2018.2868139>
- [13] Gkikas, D.C., Tzafilkou, K., Theodoridis, P.K., Garmpis, A. and Gkikas, M.C. (2022) How Do Text Characteristics Impact User Engagement in Social Media Posts: Modeling Content Readability, Length, and Hashtags Number in Facebook. *International Journal of Information Management Data Insights*, **2**, Article ID: 100067. <https://doi.org/10.1016/j.ijime.2022.100067>
- [14] Liu, H., Li, C., Wu, Q. and Lee, Y.J. (2023) Visual Instruction Tuning. *Advances in Neural Information Processing Systems* 36, New Orleans, 10-16 December 2023, 34892-34916. <https://doi.org/10.52202/075280-1516>
- [15] 王泽深, 杨云, 向鸿鑫, 等. 零样本学习综述[J]. 计算机工程与应用, 2021, 57(19): 1-17.
- [16] Dettmers, T., Lewis, M., Belkada, Y. and Zettlemoyer, L. (2022) Llm.int8(): 8-Bit Matrix Multiplication for Transformers at Scale. *Advances in Neural Information Processing Systems* 35, 28 November-9 December 2022, New Orleans, 30318-30332. <https://doi.org/10.52202/068431-2198>
- [17] Haavelmo, T. (1944) The Probability Approach in Econometrics. *Econometrica*, **12**, iii-vi, 1-115. <https://doi.org/10.2307/1906935>
- [18] Lind, J.T. and Mehlum, H. (2010) With or without U? The Appropriate Test for a U-Shaped Relationship. *Oxford Bulletin of Economics and Statistics*, **72**, 109-118. <https://doi.org/10.1111/j.1468-0084.2009.00569.x>