

社交媒体在新能源汽车领域的影响力研究

——以Bilibili为例

桂远芬, 钱 颖

上海理工大学管理学院, 上海

收稿日期: 2025年4月27日; 录用日期: 2025年6月12日; 发布日期: 2025年6月19日

摘 要

数字化浪潮下, 社交媒体深刻重塑了消费者的信息获取与决策模式。而新能源汽车作为汽车行业的新兴力量, 视频形式的社交媒体对其产生的影响越来越大。本文以视频网站Bilibili和蔚来汽车为研究对象, 结合口碑传播理论和情绪传染理论探究视频网站及其平台信息对蔚来汽车销量的影响。本研究使用Python获取2018~2024年蔚来汽车相关视频数据以及视频下32万余条评论数据, 结合中国汽车工业协会的销量数据展开分析, 探讨社交媒体对汽车行业的影响。实证结果表明: 1) 视频累计曝光度(即播放量)、用户互动参与度(即评论量)和情感认同指数(即点赞量)对汽车销量均有显著的正向影响且均为销量的不同滞后期的格兰杰原因, 并且用户互动参与度的作用更为突出且持续时间更长; 2) 在线评论的正负面情绪值对销量存在显著影响, 正面情绪促进销量增长以及负面情绪抑制销量, 且正面情绪的促进作用更强; 3) 此外, 本文借助LDA主题聚类算法对评论内容继续分析, “电池”、“企业”、“技术”等主题受关注度较高且所有主题负面情绪占比较大, 其中“体验”、“电池”、“设计”主题的负面情绪值对销量产生了直接的负面影响, 其中“体验”主题的负面情绪影响最大。

关键词

社交媒体, 汽车销量, 向量自回归模型, 情感分析, LDA主题聚类

The Influence of Social Media in the New Energy Vehicle Sector

—A Case Study of Bilibili

Yuanfen Gui, Ying Qian

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

Received: Apr. 27th, 2025; accepted: Jun. 12th, 2025; published: Jun. 19th, 2025

Abstract

Under the digital transformation, social media has profoundly reshaped consumer information acquisition and decision-making patterns. As an emerging force in the automotive industry, new energy vehicles (NEVs) are increasingly influenced by video-based social media platforms. This study focuses on the video-sharing platform Bilibili and the Chinese electric vehicle manufacturer NIO, integrating word-of-mouth communication theory and emotional contagion theory to investigate how platform-driven content impacts NIO's sales. Utilizing Python, we collected data from NIO-related videos (2018~2024) and over 320,000 comment entries, combined with sales data from the China Association of Automobile Manufacturers, to analyze the role of social media in the automotive sector. Key empirical findings include: 1) Cumulative video exposure (play counts), user interaction engagement (comment volume), and sentiment affinity indices (like counts) exhibit significant positive effects on sales, with all variables acting as Granger causes of sales across different lag periods. User interaction engagement demonstrates a more pronounced and enduring impact. 2) Sentiment polarity in online comments significantly influences sales: positive emotions drive growth, while negative emotions suppress demand, with positive sentiment exerting a stronger promotional effect. 3) Through LDA topic clustering analysis, high-attention themes such as "battery," "corporate strategy," and "technology" were identified. Notably, negative sentiment in "user experience," "battery," and "design" topics directly reduced sales, with "user experience" negativity having the largest adverse effect.

Keywords

Social Media, Automotive Sales, Vector Autoregressive Model, Sentiment Analysis, LDA Topic Clustering

Copyright © 2025 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

随着数字化时代的快速发展，社交媒体平台已演变为关键性信息中介节点，改变了现代消费者的信息获取方式与决策过程[1]，对消费者购买决策的影响作用不断增加。中国互联网络信息中心(CNNIC)第55次统计报告显示，中国网民规模2024年达到11.08亿人，其中社交媒体用户人数达11.01亿，视频用户规模达10.70亿人，占整体网民的96.6%¹，庞大的用户群体为社交媒体发挥信息传播和消费引导作用奠定了坚实的基础。社交媒体信息传播范式的变革体现在三个维度：信息获取：76%的消费者会定期阅读社交媒体的在线评论[2]。例如旅游场景中，小红书上网络口碑的数量、可靠性以及负面口碑将显著影响消费者的出行意愿[3]；经验共享：大众点评等平台累积餐饮评价上亿条[4]，这些动态更新的评价信息帮助消费者在决策时做出更明智的选择；群体影响：社交媒体的网络效应使优秀的用户生成内容进行大量的传播，与传统媒体相比，这种即时信息发布更快速、更具互动性[5]。

同时，这种影响机制在耐用品消费领域尤为明显。根据中国汽车工业协会统计数据，全国机动车保有量达4.4亿¹。社交媒体平台汇集了海量来自不同领域、文化背景和消费层次的消费者，他们在社交媒体平台分享汽车驾驶体验，社交媒体使这些内容能够被快速且广泛地传播，同时，互动性提高了信息的

¹Sources: <https://www.cnnic.net.cn/n4/2024/0829/c88-11065.html> [last accessed: 29 August 2024].

可信度进而影响消费者的购买决策。其次, 消费者能够在短时间内从视频中获取全面且深入的产品信息, 满足快节奏生活。因此, 研究视频类社交媒体对汽车销量的影响具有实际意义。Bilibili 是国内拥有大量用户的视频网站, 它同时具有视频浏览和社交的特征, 用户可以作为视频观看者也可以成为视频创作者, 在视频下方进行评论和在播放页面发送弹幕, 对视频进行点赞、转发、收藏等操作。

本文的研究以 Bilibili 网站为研究对象, 探讨视频类社交媒体和该平台产生的信息对汽车销量的影响。本文研究基于电子口碑传播理论和情感传染理论, 收集并分析实证数据以检验所提出的假设。研究结果表明, 视频累计曝光度、用户互动参与度、情绪认同指数和评论的情绪值均显著影响了汽车的销量。本文的其余部分结构如下, 第二章概述了相关的先前文献, 总结先前研究的创新点和不足之处, 介绍本研究的理论基础并提出假设。在第三章中, 本文描述了实证分析的数据和方法。在第四章对结果进行了总结和讨论, 最后总结了研究局限性。

2. 文献综述

2.1. 社交媒体对汽车销量的影响研究

过往研究中, 社交媒体对汽车销量的影响主要体现在销量预测。经过文献梳理, 汽车销量预测有因果法[6]、统计时间序列法[7]、学习模型法[8][9]和数学模拟法[10]这四种基本方法, 大多研究多采用长时间框架预测汽车销量, 预测结果常用于战略性长期决策。

因果预测法通过识别因素关系来预测产品增长, Bass 扩散模型运用较多。Bass 模型将消费者分为创新者(主动尝试新产品)和模仿者(受他人影响才购买), 创新者借助社交媒体等外部渠道获取信息购买产品, 模仿者则通过口碑、交流和观察来决定是否购买[11]。此外, 因果预测法还涵盖灰色系统理论、增长模型[7]、多元线性回归[7][11]和系统动力学[7][12]。多元线性回归通过分析价格、充电时间、百公里耗能、电池类型等多个自变量与销量这一因变量的线性关系来预测汽车销量, 该方法涉及概率模型、Logit 模型、联合分析等多变量分析模型[11]。

此外, 统计时间序列预测使用按时间排列的历史数据来分析和预测未来的趋势, 常见的模型有自回归模型(VAR)、移动平均模型、自回归移动平均模型(ARIMA)[9]、季节性自回归移动平均模型(SARIMA)[13]等。一些研究在长期预测中使用了 ARIMA [14]模型, 短期研究则使用了三重指数平滑(TES)方法, VAR 模型也被用于 EV 预测[7]。另一种模型使用了机器学习技术, 如 Adaboost 回归[9]、反向传播神经网络(BPNN)[14]、双向长短期记忆(BiLSTM)[9][15]、卷积神经网络(CNN)等, 将因果预测方法和机器学习进行结合可以得到更好的预测效果。

在上述预测模型中, 加入社交媒体平台信息可以提高预测精度, 如加入文本的情感倾向和情感强度值, 文献[8]将在线评论、去噪算法和预测模型组合, 提高了预测精度。文献[11]用深度学习算法 LSTM 对模型进行优化, 提高了销量预测精度。文献[16]利用市场中所有可用的竞争搜索数据, 结果表明该模型比不利用完整竞争搜索数据的模型更适合预测销量。

2.2. 视频网站对消费者购买意愿的影响

文献[13]的研究发现通过自我提升和自我超越价值观构建的 SMIs 人设可促进消费者购买与转发意图且消费者自我认知起到中介作用, 同时发现该影响在女性消费者中存在显著差异。文献[8]关注视频营销对泰国千禧一代购买电动汽车意图的影响, 表明短视频广告的趣味性对态度和购买意图有重要影响。

除了品牌营销, 客户互动也是重要的营销环节, 文献[17]运用结构方程建模分析发现信息寻求和自我认同动机促进消费者与品牌互动, 进而影响消费者态度和购买意图。文献[1]收集汽车销售与社交媒体数据, 发现企业生成内容(FGC)和专业生成内容(PGC)存在替代关系, UGC 受价格和受欢迎程度调节影响销

售。文献[2]以小红书为例,发现短视频 UGC 比图片 UGC 更能激发购买意图,UGC 创作者类型和体验披露有调节作用。

上述研究从影响者人设、视频营销、社交媒体使用动机、不同类型社交媒体内容以及 UGC 类型等多角度深入探讨了社交媒体对消费者购买行为的影响,研究普遍采用实证方法,大多通过问卷调查、实验、数据分析等手段获取数据。然而,各项研究也存在如研究范围、样本代表性问题、变量选择不全面等局限性。

2.3. 相关理论与假设

“电子口碑”(eWOM)是指根据客户的真实消费体验和个人评价,对在线产品和服务的非正式和非商业评论[10]。客户经常在各种社交平台上分享多样化、有时甚至相互冲突的电子口碑(eWOM)信息,这给 eWOM 管理带来了新的挑战[18]。eWOM 作为消费者获取产品信息以促进其决策的主要方式,一直是再搜索者在决策过程中持续关注的问题[19]。

个人在日常基于社交媒体的微通信中产生的内容以及其中表达的情绪,可能会影响他人的情绪状态[20]。有研究发现,负面帖子发生在比基线多 4.34% 的负面内容过度曝光之后,而正面帖子发生在平均过度曝光 4.50% 的正面内容之后[20]。即假设有过于负面的内容曝光,之后将会产生更多负面的帖子,正面内容同理。这表明社交媒体上的评论内容的情绪倾向会影响到后续的在线内容的情绪倾向。

综上所述,本文提出如下假设:

H1: 社交媒体上视频累计曝光度对汽车销量有显著正向影响; H2: 社交媒体上情感认同指数对汽车销量有显著影响; H3: 社交媒体用户互动参与度对汽车销量有影响; H4: 社交媒体上产生的在线评论情绪值对汽车销量有显著影响; H5: 社交媒体上在线评论的不同主题情绪值对销量的影响程度不同。

3. 方法

3.1. 数据收集

Bilibili 网站是全国流行的社交媒体视频网站,为公众和消费者提供了一个表达意见的在线空间。本研究首先通过 Python 库检索关于蔚来汽车的视频链接,得到与蔚来汽车相关的视频链接,再使用 python 爬取视频下的评论内容。本文根据关键词爬取的视频链接数据,除 url 网页地址外,还包括该视频的播放数、点赞数、评论数、发布时间、发布作者、粉丝数量等信息。视频发布时间跨度 2018 年 9 月到 2024 年 9 月。评论数据包括评论内容、评论时间、评论者名称、评论者等级、评论点赞数、回复数和评论类型(一级评论/二级评论)。通过初步数据清洗,删除重复和无效内容后,最终获得蔚来汽车在线评论数量为 323,752 条。本文通过汽车的月销量数据来感知公司的市场表现,每月汽车销售数量数据来自中国汽车工业协会的月度汽车市场报告。

3.2. 关键变量描述

1) 蔚来汽车社交媒体热度指标。使用 python 抓取视频的播放量、点赞数量,转发量、评论数量等信息,研究每个变量与汽车销量的关系,其中播放量、点赞量、评论量很大程度体现出蔚来汽车的热度。

2) 在线评论的数量和内容。使用 python 程序根据视频链接进行评论文本信息的采集,按照评论发布时间进行按月汇总,得到的每月累计评论数量作为评论数量的指标。

3) 评论的正负面情绪值。通过调用百度 api 情感倾向分析接口获取在线评论的情绪值,情绪值范围从 0~1,越接近两端,情绪越强烈,越靠近 1 越正面,每条评论的正面情绪值与负面情绪值之和为 1。

4) 主题类别及其情绪值。将爬取的文本选取点赞量大于 5 的文本进行 LDA 主题分类,探讨不同主题及其情绪值对销量的影响。

3.3. 向量自回归模型 VAR

向量自回归模型简称 VAR 模型, 是 AR 模型推广后的模型, 在计量经济模型中很常用。向量自回归方法是联立方程模型, 模型具有多个变量, 但每个变量都通过其自身值与模型中其他变量的滞后项来解释[6]。VAR 模型常用于描述多变量时间序列之间的被动关系, 不需要理论作为基础, 是一种非结构化的模型。VAR 模型的一般表达式如下:

$$Y_t = \mu + \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \cdots + \Pi_k Y_{t-k} + u_t, u_t \sim \text{IID}(0, \Omega) \quad (1)$$

VAR 模型在进行分析时要先确定滞后的阶数 p , 目前用的较多的滞后阶数检验方法是通过分析准则, 选取满足准则条件数量最多的阶数作为模型的滞后阶数, 准则包括 AIC 准则、SC 准则、HQ 准则、LogL 准则和最终预测误差准则。本文采取该方法对 VAR 模型的滞后阶数进行估计和检验。

VAR 模型要求输入的时间序列数据是稳定的, 平稳是指该组数据的值与时间无关, 数据不会随着时间的变化而变化[6]。本文使用 ADF (Augmented Dickey-Fuller) 检验时间序列的平稳性, ADF 平稳性检验是一种广泛应用的时间序列平稳性检验方法, 其根据单位根存在性结果判断时间序列是否平稳。

3.4. 格兰杰因果检验

本文收集的销量、评论等数据均为时间序列样本, 本文主要探讨视频社交媒体特征对销量是否有显著影响, 因此选择格兰杰因果关系检验方法来判断哪些特征显著影响了汽车的销量。格兰杰因果关系检验(Granger Causality Test)是一种用于分析时间序列数据中变量之间因果关系的统计方法, 通过合并向 VAR 过程描述的变量信息来测量从一组时间序列到另一组时间序列的可预测性[1]。前提假设包含: (1) 时间序列的平稳性; (2) 线性: 假设真实的数据生成过程以及相应的变量相互之间的因果效应是线性的。(3) 已知滞后阶数[2]。以下是含有滞后项 X 的回归, Y_t 表示汽车月销量, Y_{t-1} 表示滞后项的月销量, X_{t-1} 表示滞后项的其他影响因素, 如月度总评论数量, 百度搜索指数, 月度视频总播放量, 月度平均情绪值等, s 为滞后阶数。

$$Y_t = \sum_{i=1}^s \alpha_i Y_{t-i} + \sum_{i=1}^s \beta_i X_{t-i} + u_{2t} \quad (2)$$

该等式假设当前月的 Y 与 Y 自身的过去值及 X 的过去值有关, 零假设为 $H_0: \beta_1 = \beta_2 = \cdots = \beta_s = 0$ 。明确一个变量是另一个变量变化的原因之后, 可借助脉冲响应深入探究影响的具体时间与幅度。对于平稳的时间序列而言, 初期序列会因受到外部冲击而处于变动状态, 但随着时间推移, 最终会回归至平稳状态。这是因为平稳时间序列具有均值回复特性, 外部冲击所带来的影响不会持续累积, 而是会逐渐被系统所吸收和调整。

4. 结果

4.1. 描述性分析

本文从 Bilibili 网站收集了 341,995 条与蔚来汽车相关的评论, 来自 1130 个视频, 时间跨度为 2018 年 9 月 1 日到 2024 年 10 月 28 日。对评论文本进行去重、移除 URL、HTML 标签、特殊字符、表情符号等无关信息, 最终得到 323752 条中文评论。表 1 显示了详细的描述性统计数据, 每月统计数据显示, 平均每月有 4916 条评论, 峰值为 35465 条评论, 最低为 5 条评论; 平均每月视频总播放量有 183 多万次, 峰值为 1124 万多, 播放量整体呈逐年上涨的趋势。

图 1 直观呈现了 2018 年 9 月至 2024 年 7 月期间, 销量(蓝色线)、视频曝光度(黄色线)、用户互动参与度(红色线)和情感认同指数(深蓝色线)的趋势变化轨迹。热度指标呈现出几乎相同的波动趋势, 均 2021

年初达到小峰值，随后在 2022 年末开始出现大幅度波动，最终与销量在 2024 年共同达到峰值。各变量在 2023 年后开始显著上升且波动幅度较大，这可能反映出市场需求的显著提升和公众对新能源汽车的关注度上升，同步变化的现象表明变量之间或许存在紧密联系。

Table 1. Descriptive analysis of NIO’s sales and video metrics (2018/9/1~2024/10/28)
表 1. 蔚来汽车的销量和视频指标描述性分析(2018/9/1~2024/10/28)

Months = 72	销量	Number of videos	Number of comments	View count	Like count_video	Favorite	Bullet comments count	Number of shares
Mean	8860	16	4916	1837582	46839	8751	10,379	7915
Std	5921.6	20.31	6831.92	2300210.31	79886.15	18,571	16,269	18851.43
Max	21,209	137	35,465	11,248,546	426,675	134,326	94,583	147,312
Min	707	1	5	4933	26	28	2	37
Median	8144	7	2429	959,000	16,715	2996	4483	2793
Skewness	0.57	3.27	2.63	2.01	2.98	4.94	2.98	5.98
Kurtosis	2.33	18.8	11.01	7.14	12.64	31.55	13.3	43.32

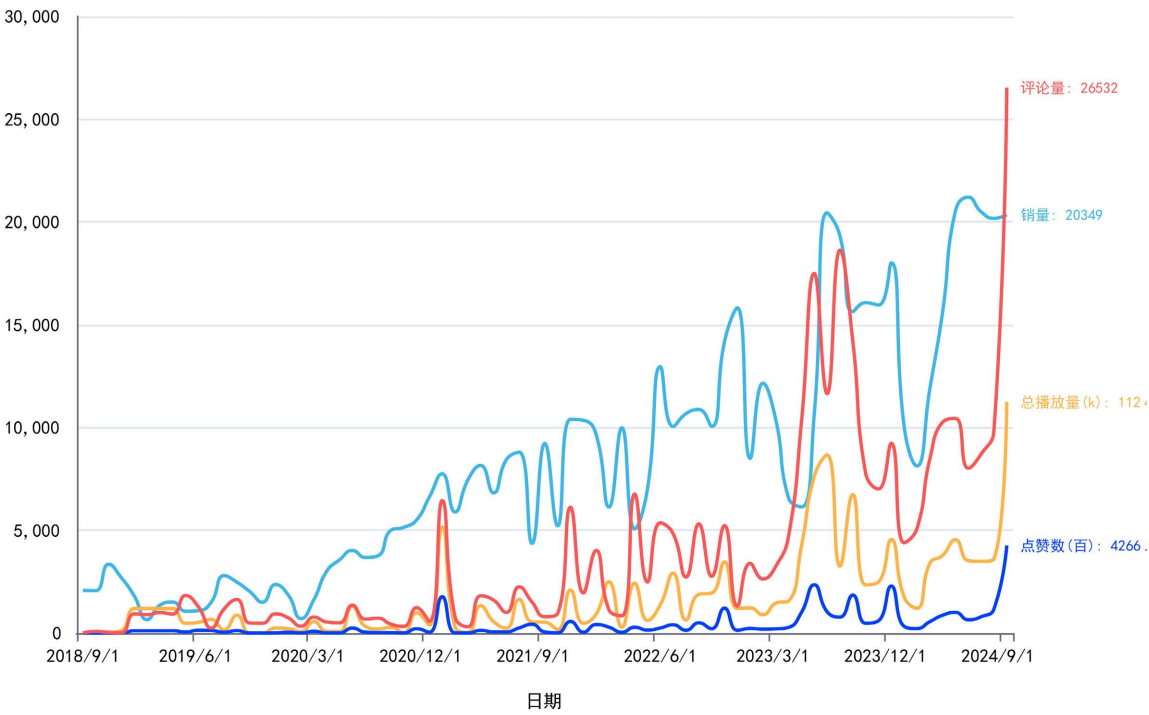


Figure 1. Sales volume and various indicators
图 1. 销量与各指标情况

4.2. 视频社交媒体热度对汽车销量的直接影响

本文选用 VAR 模型对蔚来汽车月销量与视频曝光量、用户互动参与度、情感认同指数之间的关系展开研究。运用差分法对时间序列进行处理，差分通过计算相邻两个观测值之间的差值，能够有效消除时间序列中的趋势成分和季节性成分，使数据更满足平稳性的要求。

初步探索不同变量与销量之间的关系时，分别将变量带入模型，能够清晰地看到每个变量自身的滞

后项如何影响产品销量，避免了多个变量之间可能存在的交互作用对结果解读的干扰。变量在进行一阶差分处理后均达到平稳状态，表 2 是 ADF 检验结果，根据 P 值判断，时间序列数据均呈现显著的平稳性。

Table 2. ADF test results
表 2. ADF 检验结果

变量	t	P	临界值		
			1%	5%	10%
销量	-5.035	0.000***	-3.541	-2.909	-2.592
视频曝光度	-7.553	0.000***	-3.526	-2.903	-2.589
用户互动参与度	-9.432	0.000***	-3.525	-2.903	-2.589
情感认同指数	-8.529	0.000***	-3.526	-2.903	-2.589

注：***、**、*分别代表 1%、5%、10%的显著性水平。

VAR 模型回归结果显示，视频累计曝光度、用户互动参与度和情感认同指数对销量均具有显著正向影响。销量与视频累计曝光度和情感认同指数的回归结果如表 3 所示，最佳滞后阶数均为 1，前一期的视频累计曝光度对当前销量有微小的正面影响，系数为 0.001，表明视觉化内容的高曝光度直接驱动消费者决策；同样，前一期的情感认同指数对当前的销量有显著的正面影响，情感认同指数对销量的直接促进效果大于视频曝光度带来的正面影响。

Table 3. The cumulative exposure of the video regressed to the emotional identity index
表 3. 视频累计曝光度与情感认同指数回归结果

参数	估计量	销量_cf	参数	估计量	销量_cf
销量(-1)	系数	-0.29	销量(-1)	系数	-0.179
	标准差	0.121		标准差	0.119
	t	-2.46		t	-1.512
视频累计曝光度(-1)	系数	0.001	情感认同指数(-1)	系数	0.007
	标准差	0		标准差	0.006
	t	2.438		t	1.27
常数	系数	343.013	常数	系数	287.1
	标准差	323.439		标准差	331.912
	t	1.061		t	0.865

$$\text{销量} = -0.299 * \text{销量}(-1) + 0.001 * \text{视频累计曝光度}(-1) + 343.013$$

$$\text{销量} = -0.179 * \text{销量}(-1) + 0.007 * \text{情感认同指数}(-1) + 287.1$$

为直观展示变量间动态关系，下图 2 呈现了视频累计曝光度对销量的脉冲响应分析结果。当在第 1 期曝光度变动后，销量在第 2 期迅速产生较为显著的正向响应且达到约 0.0007 的峰值，表明短期内视频曝光度的增加能够有力地带动销量上升，随后响应值开始波动下降，第四期后逐渐趋于平稳并向 0 收敛。

同理，情感认同指数对销量的脉冲响应分析如图 3 所示，销量受到冲击后显著上升，脉冲响应值达到约 0.018 的峰值，情感认同指数的正向带动效果大于播放量，随后响应值开始波动减弱，于第 7 期后逐渐消失。

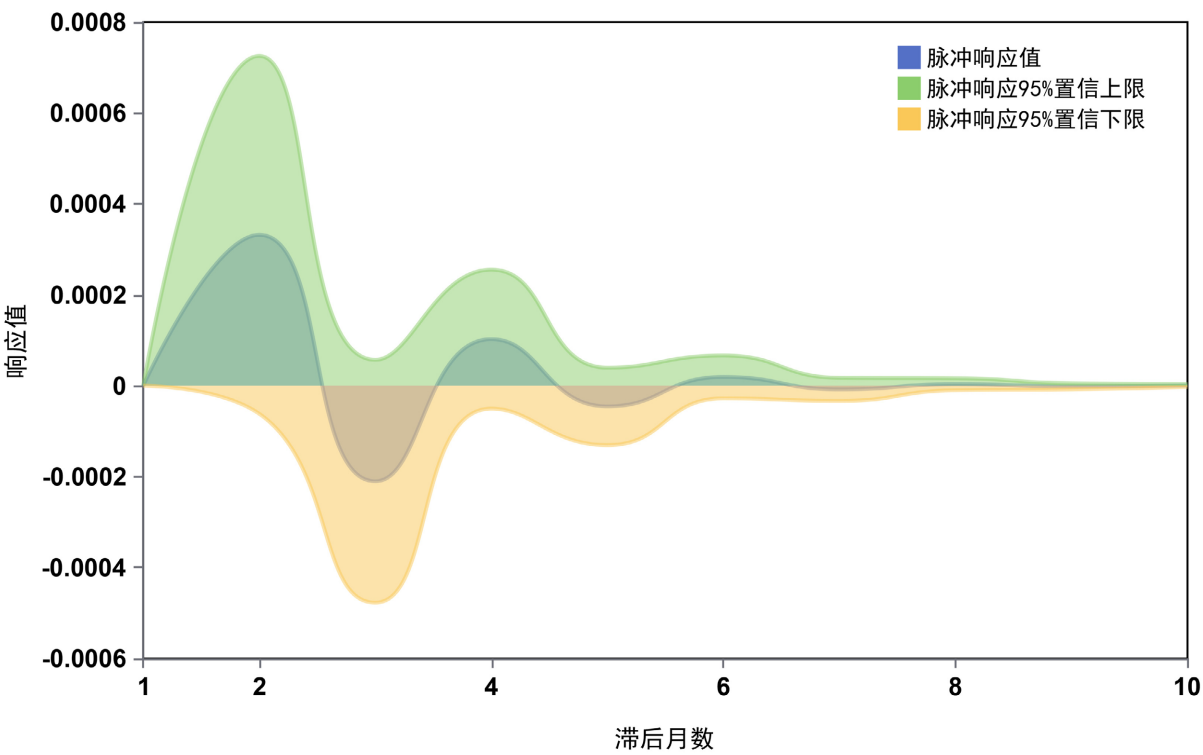


Figure 2. The impulse response of the cumulative exposure impact of the video to the sales volume
图 2. 视频累计曝光度冲击对销量的脉冲响应

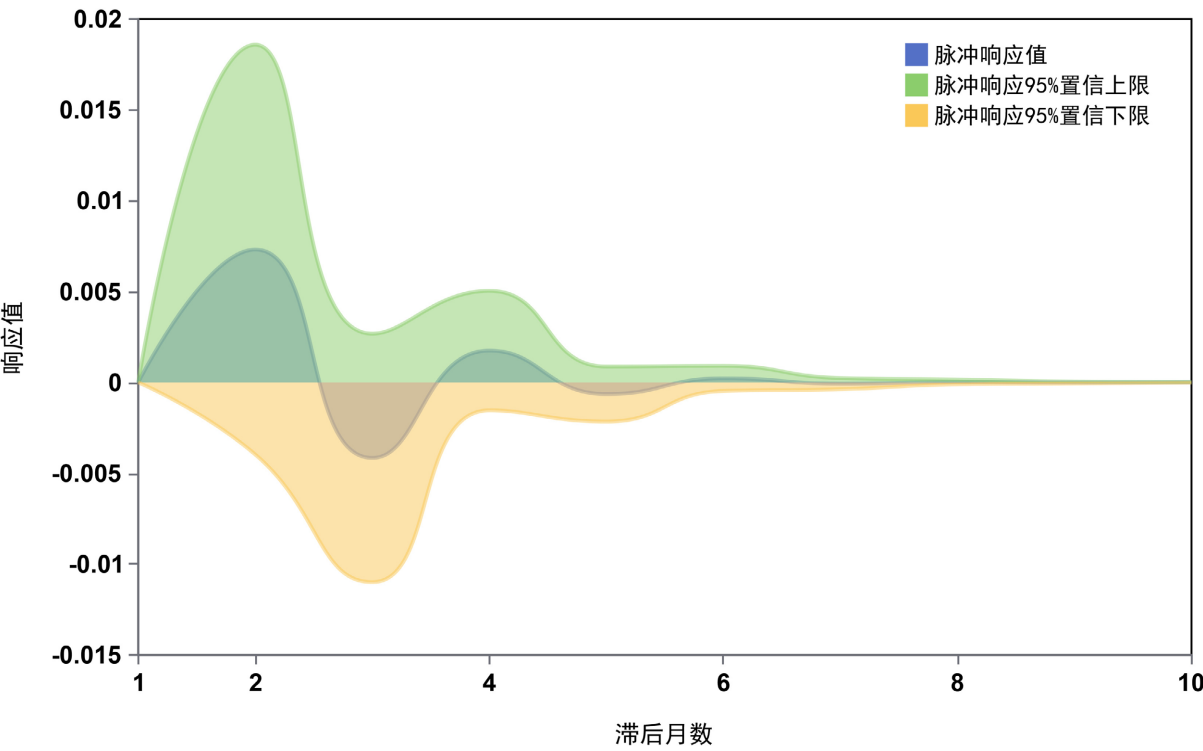


Figure 3. The impulse response of the emotional identity index to sales
图 3. 情感认同指数对销量的脉冲响应

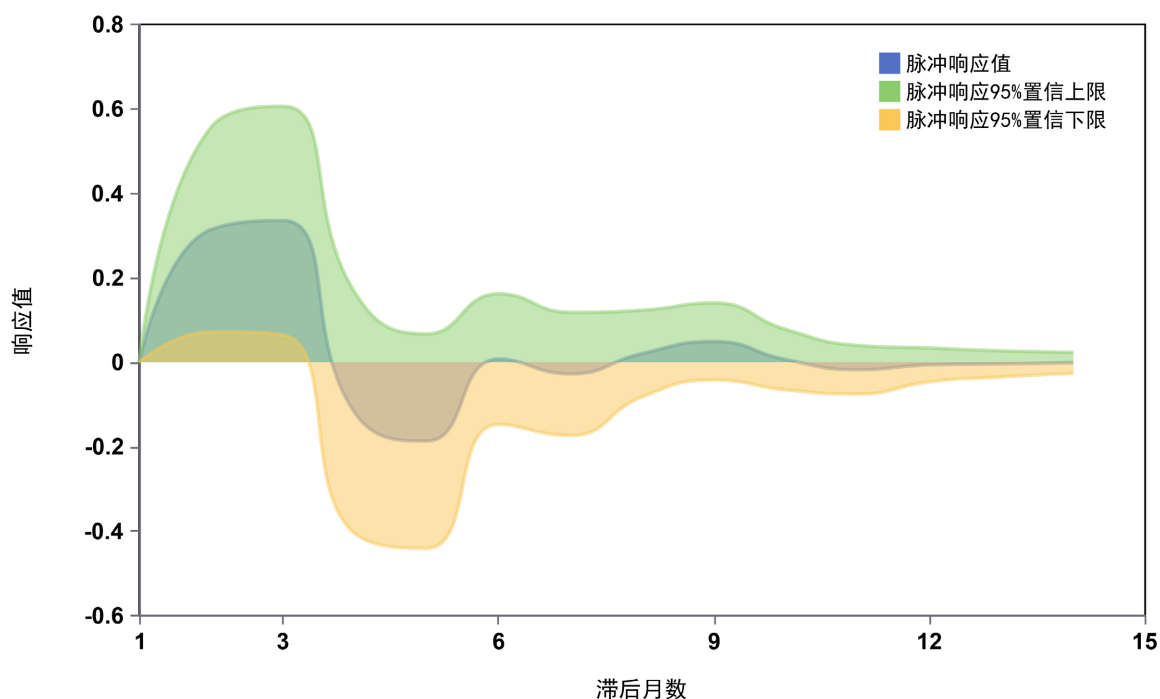


Figure 4. The impulse response of user engagement to sales
图 4. 用户互动参与度对销量的脉冲响应

此外, 将用户互动参与度与销量带入 VAR 模型, 最佳滞后阶数为 3, 过去三期的用户互动参与度对当期销量均有显著正面影响, 系数分别为 0.314、0.579、0.303, 影响效果远大于视频累计播放量和情感认同指数。进一步对比发现过去用户互动参与度所带来的正面影响大于销量本身带来的负面影响, 这表明用户互动参与度在推动销量增长方面具有重要作用。大量的互动和评论实际上起到了宣传和推广的作用, 众多的互动可以增加产品或服务的曝光度, 吸引更多潜在消费者的关注, 积极的评论内容能够提升产品或服务在消费者心中的口碑和形象, 增强消费者的购买意愿。

$$\begin{aligned} \text{销量} = & -0.401 * \text{销量}(-1) - 0.36 * \text{销量}(-2) - 0.207 * \text{销量}(-3) + 0.314 * \text{用户互动参与度}(-1) \\ & + 0.579 * \text{用户互动参与度}(-2) + 0.303 * \text{用户互动参与度}(-3) + 346.314 \end{aligned}$$

上图 4 体现了用户互动参与度对销量的脉冲响应情况。在冲击发生初期, 销量在第 2 期和第 3 期展现出显著的正向响应, 脉冲响应值分别达到约 0.5 和 0.45, 这表明短期内用户互动的增加能够有力地吸引消费者的注意力。然而影响的方向和程度从第 4 期开始急剧下降到负值(约-0.5), 用户互动参与度的增加也可能产生一定的负面效应, 例如大量的评论可能包含负面评价并引发消费者的担忧和疑虑从而抑制购买行为, 或者过多的评论信息使消费者产生信息过载, 降低了他们对产品的关注和兴趣。此后, 响应值在正负之间波动并于 12 期后逐渐趋于 0, 表明在长期内评论数量对销量的边际影响逐渐减小。

4.3. 格兰杰因果关系检验

本文使用格兰杰因果检验方法来检验视频网站相关指标对于汽车销量是否有预测作用, 检验结果如下表 4 所示, 视频累计曝光度是销量的第二个和第三个滞后期的格兰杰原因, 意味着视频累计曝光度会影响 2 个月和 3 个月后的销量, 反之销量对 3 个月后的视频曝光度也有显著影响; 同理, 用户互动参与度也对 2 个月和 3 个月后的销量产生影响, 反之无影响; 此外, 情感认同指数和视频曝光度之间相互影响, 相互促进。

Table 4. Results of Granger’s causality test
表 4. 格兰杰因果关系检验结果

关系	Lag = 2		Lag = 3	
	F-value	P-value	F-value	P-value
视频累计曝光度->销量	3.612	0.032**	3.52	0.020**
销量->视频累计曝光度	1.065	0.351	2.398	0.076*
用户互动参与度->销量	5.135	0.009***	4.641	0.006***
销量->用户互动参与度	1.165	0.319	2.005	0.124
	Lag = 1		Lag = 2	
	F-value	P-value	F-value	P-value
情感认同指数->曝光度	7.742	0.007***	4.134	0.020**
曝光度->情感认同指数	2.926	0.092*	0.588	0.559

4.4. 视频评论情绪分析

评论中所表达的情感倾向在变量之间起到一定的中介作用。例如，在消费者行为领域，产品质量与顾客满意度之间的关系可能受到评论情感倾向的影响，较高的产品质量可能会引发正面的评论情感倾向，进一步提升顾客的满意度从而提升销量。本文将情绪值大于 0.6 的评论归为正面评论，小于 0.4 的归为负面评论，其余为中立评论，由于评论数量较大，本文按照月份将情绪值进行汇总得到月度情绪值。

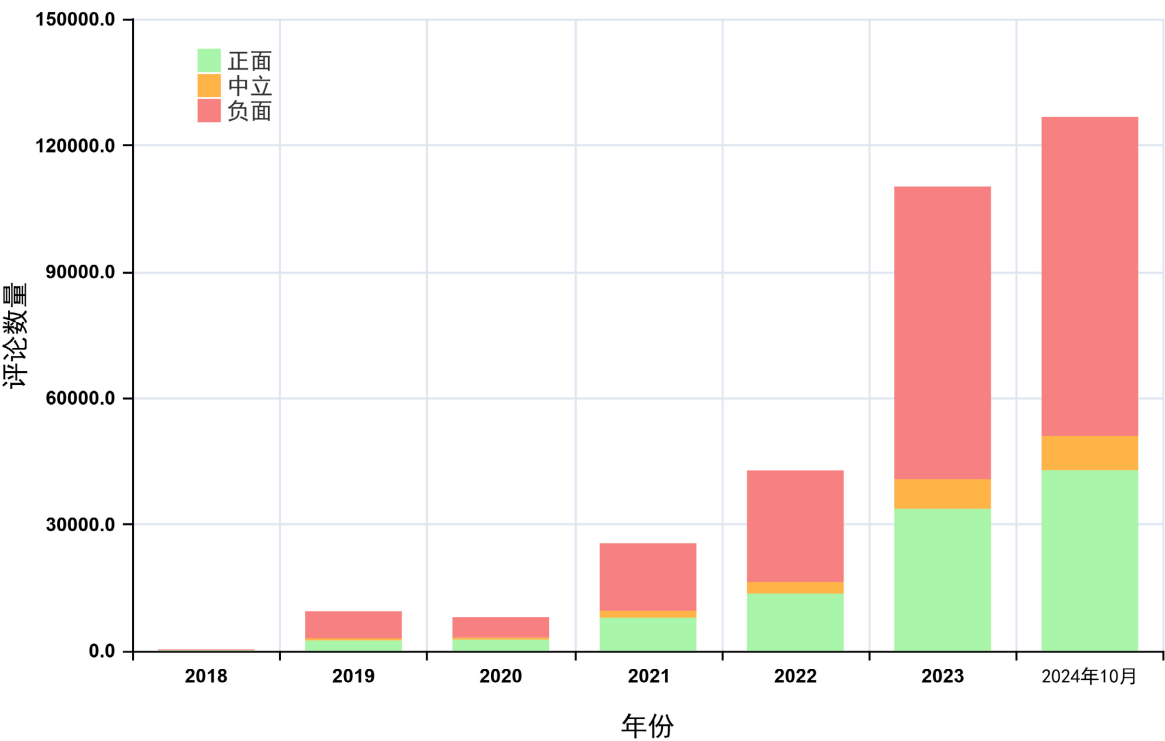


Figure 5. The number and proportion of positive and negative reviews in different years
图 5. 不同年份正负面评论数量及占比

上图 5 展示了从 2018 年到 2024 年每年的评论数量和正面、负面、中立评论的占比，由图可见，2020 年及之前评论数量较少，网民对蔚来汽车的关注度较少，但从 2021 年评论数量开始上升并在 2023 年急

剧增加, 大众对新能源汽车的关注度持续上升。此外, 负面评论的占比始终大于正面评论, 意味着网络中批评的声音较多。

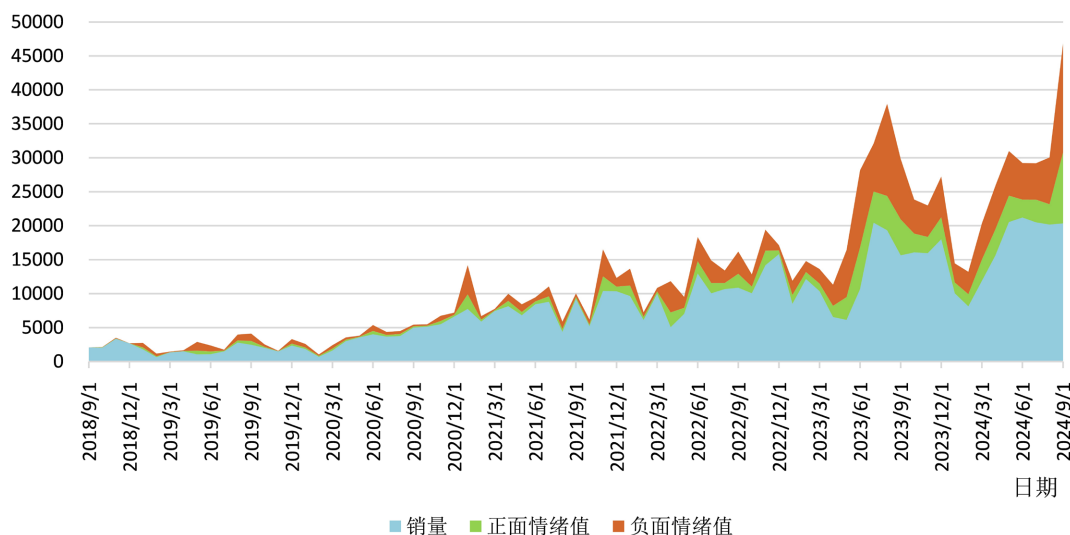


Figure 6. Sentiment and sales trends

图 6. 情绪值与销量变化趋势

图 6 是月度情绪值与销量的时间变化趋势, 图中蓝色部分表示销量的变化, 红色和绿色部分分别表示正、负面情绪值的变化。由图可见, 正、负面情感值都随着销量逐年增加, 情绪值在 2023 年 6 月出现了短时间内显著且急剧的增长, 表明当时评论情绪较为激烈, 情绪值波动较大并在 2024 年 10 月出现峰值。此外, 负面始终比正面情绪值要多一部分, 两者维持着相同的变化趋势, 但 2023 年后负面远高于正面情绪值, 这表明 2023 年后大多数评论较为负面。

将正、负面情绪值和销量分别进行一阶差分处理后变为平稳数据, 将其带入 VAR 向量自回归模型, 根据准则确定最佳滞后阶数为 3, 模型回归的方程如下:

$$\begin{aligned} \text{销量} = & -0.533 * \text{销量}(-1) - 0.44 * \text{销量}(-2) - 0.113 * \text{销量}(-3) - 0.667 * \text{销量}(-1) \\ & - 1.318 * \text{评论数量}(-2) - 1.546 * \text{评论数量}(-3) + 3.228 * \text{正面情绪值}(-1) \\ & + 6.066 * \text{正面情绪值}(-2) + 5.659 * \text{正面情绪值}(-3) + 328.121 \end{aligned}$$

根据回归结果, 正、负面情绪值对销量都有显著的影响。由于情绪值是按月进行加总, 会受到月度评论数量的影响, 评论数量越多一定程度上情绪值加总越多, 因此在回归模型中加入评论数量来消除评论数量对销量的影响。

正面情绪值与销量之间存在着紧密且显著的正向关联。销量这一变量受到过去 3 期正面情绪值的动态影响, 过去正面情绪值的系数分别为 3.228、6.066 和 5.695, 直观地表明了正面情绪值对销量具有正向的促进作用。当正面情绪值增多时, 销量会随之呈现出上涨的趋势。同理, 负面情绪值对销量的影响方程如下:

$$\begin{aligned} \text{销量} = & -0.485 * \text{销量}(-1) - 0.412 * \text{销量}(-2) - 0.158 * \text{销量}(-3) + 1.708 * \text{评论数量}(-1) \\ & + 2.983 * \text{评论数量}(-2) + 2.804 * \text{评论数量}(-3) - 2.036 * \text{负面情绪值}(-1) \\ & - 3.497 * \text{负面情绪值}(-2) - 3.685 * \text{负面情绪值}(-3) + 333.965 \end{aligned}$$

负面情绪值的增长对销量具有抑制作用, 过去三期的系数分别是 -2.036、-3.497 和 -3.685, 系数均为

负值，随着负面情绪值的增加，销量数值相应降低。此外，负面情绪的系数绝对值均小于正面情绪的系数值，这表明正面情绪值对销量的促进作用大于负面情绪对销量的抑制作用。

本研究聚焦于正、负面情绪值与销量之间的相互作用关系，并使用格兰杰检验方法来剖析它们之间潜在的因果联系。检验结果如下表 5 所示，正面情绪值、负面情绪值均是销量的格兰杰原因，从不同程度上影响了销量的趋势，滞后 2 阶和 3 阶的情绪值都对销量产生影响，反之，销量并不是情绪值的格兰杰原因。情绪值能够有效地预测销量的变化趋势，为后续的市场分析与销售策略制定提供了参考依据。

Table 5. Granger test results of sentiment value and sales volume
表 5. 情绪值与销量的格兰杰检验结果

relationship	Lag = 2		Lag = 3	
	F-value	P-value	F-value	P-value
正面情绪值->销量	7.069	0.002***	8.469	0.000***
销量->正面情绪值	0.659	0.521	1.707	0.175
负面情绪值->销量	5.355	0.007***	5.229	0.003***
销量->负面情绪值	1.002	0.372	2.114	0.107

注：***、**、*分别代表 1%、5%、10%的显著性水平。

4.5. 评论 LDA 主题分析

为进一步对评论内容进行分析，本文使用 LDA 算法对评论进行分类。LDA (Latent Dirichlet Allocation) 是一种基于概率图模型的无监督主题建模方法，属于生成式统计模型。其核心思想是假设文档由潜在主题的概率分布生成，每个主题对应特定词项的概率分布。该方法广泛应用于自然语言处理、信息检索和文本挖掘等领域，常用于文本分类、摘要生成和热点发现等任务。

4.5.1. 主题聚类

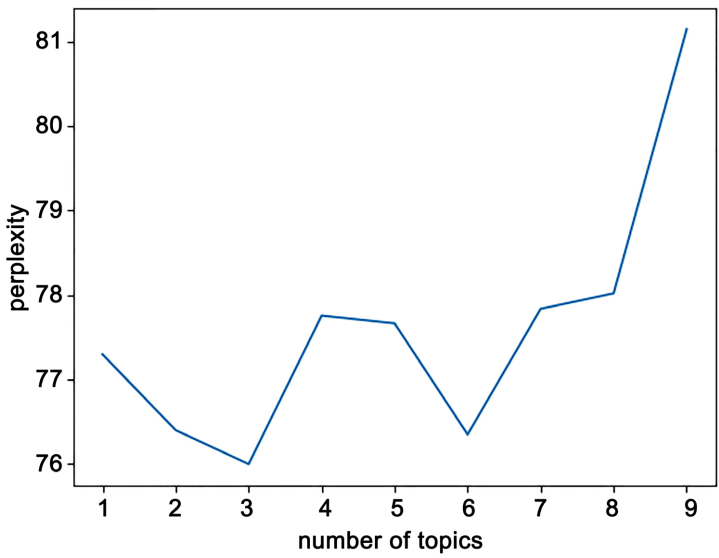


Figure 7. Number of topics and confusion
图 7. 主题数量与困惑度

在 LDA 主题模型中，主题困惑度是评估模型性能的关键指标，较低的困惑度意味着模型对数据有更

好的拟合与解释能力。上图 7 呈现了主题困惑度随主题数量变化的趋势。由图可见，主题数量为 3 和 6 时困惑度较低，结合新能源汽车行业的讨论通常集中于技术研发、市场竞争、政策支持、电池性能、智能化发展以及产业链协同等多个方面。因此选择 6 个主题，能够较为全面地覆盖新能源汽车行业的核心讨论维度，更有效地挖掘出评论中的潜在语义信息。

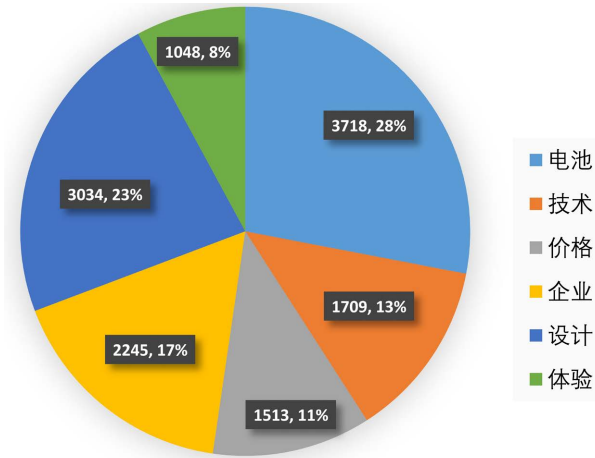


Figure 8. Number and proportion of each topic
图 8. 各主题数量与占比

由于评论数量极为庞大，共 32 万余条，其中包含大量无用信息，导致 LDA 主题分类效果不太理想，所以本文提取有用文本，删去点赞量小于 5 的评论信息，将剩下的 34,874 条评论带入 LDA 模型进行分类，最终得到 6 个主题分别是技术、价格、企业、体验、电池和设计。

以上饼状图 8 呈现了不同主题在评论数据中的分布情况。评论被归类为 6 个主题，从图中可知，“电池”主题以 3718 的数量成为主题中的焦点，电池相关话题的热度极高。作为新能源汽车的核心组件，电池的性能、成本、安全性以及续航能力等方面，始终是行业参与者、消费者以及研究者关注的重点；其次，“设计”主题的数量为 3034，“企业”主题的数量为 2245，均是消费者关注的话题。

4.5.2. 词云图

为了解每个主题谈论的内容，本文对各个主题的文本生成如下词云图。



词云图可以直观地展现每个主题的核心词汇。“电池”主题下，“电池”“电站”“充电”等词较为突出，该主题聚焦于新能源汽车电池技术、充电站覆盖程度、电池续航等相关内容；“设计”主题，“空间”“外观”“内饰”等词较为突出，表明从汽车的外观到内饰都是消费者重点关注的内容；“企业”主题下，“国产”“高端”“品牌”等词语较为突出，表明大众常常探讨企业的品牌定位、营销模式、研发能力和管理模式；“技术”主题下，“功能”“配置”“智能”等词较为突出，技术层面涉及到硬件和软件两方面；“价格”主题下，“性价比”“bba”等词较为突出，这表明消费者会将不同品牌、价位的汽车进行横向比较，对于蔚来汽车“高端”的定位，消费者容易将其与“bba”系列汽车进行对比，“bba”即奔驰、宝马和奥迪；“体验”主题下，“车主”“服务”“试驾”等词较为突出，从售前、售中到售后会给消费者带来不同的体验，品牌的服务是消费者体验好坏的重要指标。

4.5.3. 主题情绪对销量的直接影响

针对每个主题，本文详细探讨主题的情绪倾向对销量的影响。每个主题的负面情绪占比较大，因此着重探讨负面情绪对销量的影响。部分主题负面情绪对销量有显著的负面影响。将各主题的负面情绪值带入VAR模型，结果如表6所示。由各个主题负面情绪值的影响系数可知，“体验”主题对销量的负面影响最大，系数为-154.606，这表明关于体验产生的情绪会很大程度上对销量产生负面影响，其次，“电池”影响系数为-32.076，电池相关的舆论会给销量带来直接的负面影响；最后，“设计”主题系数为-19.255，虽然系数小于前两个主题，但对销量的直接影响也不容小觑。

Table 6. “Battery”, “Price”, “Design”, “Experience” regression results
表 6. “电池”、“价格”、“设计”、“体验”回归结果

参数	估计量	销量	参数	估计量	销量	参数	估计量	销量	参数	估计量	销量
	系数	-0.144		系数	-0.157		系数	-0.151		系数	-0.151
销量(-1)	标准差	0.121	销量(-1)	标准差	0.12	销量(-1)	标准差	0.123	销量(-1)	标准差	0.121
	t	-1.19		t	-1.312		t	-1.225		t	-1.25
电池负面 情绪(-1)	系数	-32.076	价格负面 情绪(-1)	系数	-0.347	设计情绪值 (-1)	系数	-19.255	体验情 绪值(-1)	系数	-154.606
	标准差	86.158		标准差	55.096		标准差	71.416		标准差	199.689
	t	0.372		t	-0.006		t	-0.27		t	-0.774
电池评论 数量(-1)	系数	-18.53	价格评论 数量(-1)	系数	-16.55	设计评论数量 (-1)	系数	10.755	体验评 论数量 (-1)	系数	100.789
	标准差	70.044		标准差	37.98		标准差	43.018		标准差	134.899
	t	-0.265		t	-0.436		t	0.25		t	0.747
	系数	215.03		系数	272.366		系数	237.337		系数	235.519
常数	标准差	337	常数	标准差	338.054	常数	标准差	339.214	常数	标准差	336.812
	t	0.638		t	0.806		t	0.7		t	0.699

进一步使用格兰杰来检验各主题负面情绪值对于销量是否有预测作用，结果如下表7所示。

由检验结果可知，“电池”和“价格”的情绪值是销量的第二个和第三个滞后期的格兰杰原因。此外，“设计”和“体验”是销量的第三个和第四个滞后期的格兰杰原因，销量也是“设计”“体验”话题情绪值的格兰杰原因，即该两个话题和销量之间互相影响；“技术”是销量的第四个滞后期的格兰杰原因，表明“技术”情绪值会在长期内对销量产生影响。

Table 7. Results of the Granger test of negative sentiment values and sales volume for each subject
表 7. 各主题负面情绪值与销量的格兰杰检验结果

关系	Lag = 2		Lag = 3	
	F-value	P-value	F-value	P-value
“电池”->销量	4.568	0.014**	2.85	0.044**
销量->“电池”	0.935	0.398	0.661	0.579
“价格”->销量	4.203	0.019**	3.787	0.015**
销量->“价格”	0.991	0.377	0.671	0.573
	Lag = 3		Lag = 4	
	F-value	P-value	F-value	P-value
“设计”->销量	2.774	0.048**	2.146	0.086*
销量->“设计”	3.212	0.029**	2.692	0.039**
“体验”->销量	2.976	0.038**	4.728	0.002***
销量->“体验”	2.788	0.048**	2.67	0.040**
“技术”->销量	1.328	0.273	3.394	0.014**
销量->“技术”	2.63	0.058*	2.628	0.043**

5. 结论与展望

本文以 Bilibili 平台的相关指标和在线评论以及蔚来汽车销售市场为研究对象，爬取在线评论，使用百度情感工具识别文本的情感倾向，将文本信息量化，然后运用 var 模型、格兰杰因果检验等方法，探究社交媒体与汽车销量之间的相关关系、因果关系以及评论内容对销量的影响。本文主要结论如下：

- (1) 视频累计曝光度(累计播放量)和情感认同指数(点赞量)在短期内对蔚来汽车销量有显著的正面影响。视频累计曝光度对 1 个月后的汽车销量影响系数只有 0.001，程度较为微小，而前一期的情感认同指数对当前的销量正面影响的系数为 0.007，情感认同指数对销量的直接促进效果大于视频曝光度带来的正面影响。
- (2) 过去三期的用户互动参与度(评论量)对当期销量均有显著正面影响，系数分别为 0.314、0.579、0.303，影响效果远大于视频累计曝光度和情感认同指数且影响的持续时间更长，用户互动参与度带来的正面影响大于销量本身带来的负面影响。
- (3) 评论文本的情绪值对销量有显著的影响。正面情绪值促进销量的上升，相反，负面情绪值抑制了销量的增长。
- (4) 经过格兰杰关系检验，视频累计曝光度、情感认同指数、用户互动参与度、情绪值均为销量的不同滞后期的格兰杰原因，即这些变量一定程度上可以对销量进行预测。
- (5) 对评论文本进行主题分类后，共聚类为 6 个主题，其中“电池”“技术”“企业”主题的占比较大，是消费者高度关注和讨论的话题，此外，“体验”“电池”“设计”的负面情绪值对汽车销量产生了显著的负面影响，“体验”的负面影响最大。

社交媒体信息作为一种实时、海量、可获取的资源，可以用来解释汽车销量、预测汽车销量和评估消费者购买行为。视频的累计曝光度体现了产品的热度，情绪认同指数和用户互动参与度中体现出大众对品牌和产品的认可和评价，这些指标均对销量产生影响。评论的情绪值和主题分类可以进一步了解消费者的立场和观点，从中挖掘详细的信息。本文使用了数据挖掘和实证分析的研究方法，探究社交媒体

上的信息对汽车销量产生的影响, 本文的研究也存在一定的不足, 未来可以从以下几个方面展开进一步研究:

首先, 本文采取了 Bilibili 网站作为研究对象, 尽管该网站具有一定的代表性, 但不同的平台在功能、评论方式等方面存在差异, 未来研究可考虑多平台抓取数据, 如抖音、快手、微博、小红书等, 进一步完善研究成果。其次, 本文只以蔚来作为研究对象, 研究结果是否适用于其他汽车品牌还有待进一步验证, 可进一步探讨其他品牌或消费品的社交媒体动态与其销量之间的关系, 以验证研究结果的可靠性。最后, 本文对评论内容进行了研究, 在未来可以对发表评论消费者的其他特征进行研究, 如评论者的专业程度、粉丝数量、活跃度等, 以丰富研究内容和成果。

参考文献

- [1] Chen, X., Wang, Y. and Wei, S. (2023) How Does Social Media Influence Car Sales: The Effects of Firm, User, and Professional-Generated Content. *Information Technology and Management*, **2023**, 1-18.
- [2] Qin, M., Qiu, S., Zhao, Y., Zhu, W. and Li, S. (2024) Graphic or Short Video? The Influence Mechanism of UGC Types on Consumers' Purchase Intention—Take Xiaohongshu as an Example. *Electronic Commerce Research and Applications*, **65**, Article 101402. <https://doi.org/10.1016/j.elerap.2024.101402>
- [3] Gao, J.J. and Ye, H.Y. (2023) The Influence of Internet Word-of-Mouth on Consumers' Choice of Tourist Destination: Take Consumers in East China on the Little Red Book as an Example. *Academic Journal of Business & Management*, **5**, 145-152.
- [4] Shu, Z., Carrasco, R.A., Sánchez-Montañés, M. and García-Miguel, J.P. (2024) A Multi-Criteria Decision Support Model for Restaurant Selection Based on Users' Demand Level: The Case of Dianping.com. *Information Processing & Management*, **61**, Article 103650. <https://doi.org/10.1016/j.ipm.2024.103650>
- [5] Zheng, Z., Yang, H., Fu, Y., Fu, D., Podobnik, B. and Stanley, H.E. (2018) Factors Influencing Message Dissemination through Social Media. *Physical Review E*, **97**, Article 062306. <https://doi.org/10.1103/physreve.97.062306>
- [6] Tseng, N., Hung, Y. and Nakano, J. (2023) Granger Causality Tests Based on Reduced Variable Information. *Journal of Time Series Analysis*, **45**, 444-462. <https://doi.org/10.1111/jtsa.12720>
- [7] Tenriawaru, A.N., Tawaha, A.R.A., Rukka, R., Ridwan, M., Chan, N.V. and Syam, S.H. (2023) Vector Autoregressive (VAR) Method in Analyzing the Effect of Inflation on Food Price Volatility (FPV) in Palopo City, Indonesia. *International Journal of Professional Business Review*, **8**, e01872. <https://doi.org/10.26668/businessreview/2023.v8i5.1872>
- [8] Bhovichitra, P. and Shrestha, A. (2022) The Impact of Video Marketing on Social Media Platforms on the Millennial's Purchasing Intention toward Electric Vehicles. *Journal of Economics, Finance and Management Studies*, **5**, 1-5.
- [9] Kumar, S., Singh, V. and Goel, R. (2024) Strategic Forecasting for Electric Vehicle Sales: A Cutting-Edge Holistic Model Leveraging Key Factors and Machine Learning Technique. *Transportation in Developing Economies*, **10**, Article No. 23. <https://doi.org/10.1007/s40890-024-00213-1>
- [10] Kemala, B.K.L., Surjandari, I. and Puspita, A.N.G. (2024) Forecasting Methods for the Electric Vehicle Ownership: A Literature Review. *Procedia Computer Science*, **234**, 87-95. <https://doi.org/10.1016/j.procs.2024.02.155>
- [11] Abu, N. and Ismail, Z. (2015) Forecasting Sales of New Vehicle with Limited Data Using Bass Diffusion Model and Grey Theory. *AIP Conference Proceedings*, **1643**, 467-475. <https://doi.org/10.1063/1.4907482>
- [12] Domarchi, C. and Cherchi, E. (2023) Electric Vehicle Forecasts: A Review of Models and Methods Including Diffusion and Substitution Effects. *Transport Reviews*, **43**, 1118-1143. <https://doi.org/10.1080/01441647.2023.2195687>
- [13] Shojaie, A. and Fox, E.B. (2022) Granger Causality: A Review and Recent Advances. *Annual Review of Statistics and Its Application*, **9**, 289-319. <https://doi.org/10.1146/annurev-statistics-040120-010930>
- [14] Liu, J., Chen, L., Luo, R. and Zhu, J. (2023) A Combination Model Based on Multi-Angle Feature Extraction and Sentiment Analysis: Application to EVS Sales Forecasting. *Expert Systems with Applications*, **224**, Article ID: 119986. <https://doi.org/10.1016/j.eswa.2023.119986>
- [15] Zhang, M., Xu, H., Ma, N. and Pan, X. (2022) Intelligent Vehicle Sales Prediction Based on Online Public Opinion and Online Search Index. *Sustainability*, **14**, Article 10344. <https://doi.org/10.3390/su141610344>
- [16] Liu, M., Zhao, Z., et al. (2022) A Novel Large-Scale Electric Vehicle Charging Load Forecasting Method and Its Application on Regional Power Distribution Networks. 2022 4th Asia Energy and Electrical Engineering Symposium (AEEES), Chengdu, 25-28 March 2022, 236-241.

- [17] Yakubu, H. and Kwong, C.K. (2021) Forecasting the Importance of Product Attributes Using Online Customer Reviews and Google Trends. *Technological Forecasting and Social Change*, **171**, Article ID: 120983. <https://doi.org/10.1016/j.techfore.2021.120983>
- [18] Liu, X., Ren, P., Lv, X. and Li, S. (2024) Service Experience and Customers' EWOM Behavior on Social Media Platforms: The Role of Platform Symmetry. *International Journal of Hospitality Management*, **119**, Article ID: 103735. <https://doi.org/10.1016/j.ijhm.2024.103735>
- [19] Song, M., Law, R. and Duan, Y. (2022) What Are the Obstacles in the Way to "Avoid Landmines"? Influence of Electronic Word-of-Mouth Dispersion on Order Decision from the Self-Construal Perspective. *International Journal of Hospitality Management*, **107**, Article ID: 103334. <https://doi.org/10.1016/j.ijhm.2022.103334>
- [20] Ferrara, E. and Yang, Z. (2015) Measuring Emotional Contagion in Social Media. *PLOS ONE*, **10**, e0142390. <https://doi.org/10.1371/journal.pone.0142390>