

基于消费者偏好的A公司国产新能源汽车区域订单预测

延安然, 李 芳

上海理工大学管理学院, 上海

收稿日期: 2025年2月22日; 录用日期: 2025年3月26日; 发布日期: 2025年4月3日

摘 要

随着新能源汽车行业竞争的加剧与消费者偏好的快速变化, 传统的基于历史数据的订单预测方法已经无法满足企业精确决策的需求。现有研究主要忽视了消费者情感的动态变化, 导致预测准确性有偏差。如何结合消费者情感的实时变化进行订单预测, 已成为提高企业竞争力的重要课题。本文提出了一种基于消费者情感的订单预测模型, 以A公司为例, 通过情感分析量化消费者评论中的情感特征, 并将其与历史销量结合, 构建考虑消费者情感的预测模型。该模型通过生成消费者偏好指数(Behavioral Index, BI), 充分反映消费者情感在订单预测中的重要作用, 并根据这一信息优化区域订单预测的精度。实证研究结果表明, 考虑消费者偏好后的订单预测模型提高了预测准确性。特别是西南和华南地区的预测精度达到93%以上, 其他大区的预测精度也均超过90%。这一结果表明, 消费者情感分析在订单预测中具有重要价值, 能够为企业的生产和供应链管理提供更科学的决策支持。

关键词

新能源汽车, 订单预测, 情感分析, 在线评论, 区域销量预测

Order Prediction for Company A's Domestic New Energy Vehicles Based on Consumer Preferences

Anran Yan, Fang Li

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

Received: Feb. 22nd, 2025; accepted: Mar. 26th, 2025; published: Apr. 3rd, 2025

Abstract

With the intensifying competition in the new energy vehicle industry and the rapid changes in

文章引用: 延安然, 李芳. 基于消费者偏好的A公司国产新能源汽车区域订单预测[J]. 运筹与模糊学, 2025, 15(2): 237-251. DOI: 10.12677/orf.2025.152080

consumer preferences, traditional order prediction methods based on historical data can no longer meet the precise decision-making needs of enterprises. Existing research mainly overlooks the dynamic changes in consumer sentiment, leading to deviations in prediction accuracy. How to integrate real-time changes in consumer sentiment for order prediction has become a crucial issue for improving enterprise competitiveness. This paper proposes an order prediction model based on consumer sentiment, taking Company A as an example. It quantifies the emotional characteristics in consumer reviews through sentiment analysis and combines these with historical sales data to construct a prediction model that considers consumer sentiment. The model generates a Consumer Behavioral Index (BI), which fully reflects the significant role of consumer sentiment in order prediction and optimizes the accuracy of regional order forecasts based on this information. Empirical research results show that the order prediction model considering consumer preferences improves prediction accuracy. Specifically, the prediction accuracy for the Southwest and South China regions reaches over 93%, with other major regions also exceeding 90%. These results indicate that sentiment analysis of consumers plays a crucial role in order prediction and provides more scientific decision-making support for enterprise production and supply chain management.

Keywords

New Energy Vehicles, Order Prediction, Sentiment Analysis, Online Reviews, Regional Sales Prediction

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

近年来, 新能源汽车产业在全球范围内蓬勃发展。国内新能源汽车的市场规模与技术能力都实现了显著提升。根据中国汽车工业协会的数据, 2023 年全年我国新能源汽车的产销量均居全球首位, 市场渗透率高达 31.6%。然而, 市场竞争的加剧与消费者偏好的快速变化对企业提出了更高的要求。

消费者偏好的多样化不仅影响着产品设计与营销策略, 还直接决定了企业的生产计划与供应链布局。现有的新能源汽车订单预测大多依赖于历史数据, 这种方法忽略了消费者行为的动态变化, 难以准确反映市场需求。因此, 如何将消费者偏好的实时变化纳入预测模型, 成为优化企业决策的重要研究方向。

本文以 A 公司为例, 提出了一种考虑消费者情感偏好的订单预测方法。通过情感分析, 模型能够较为精准地提取消费者评论中的情感特征, 并结合历史订单数据, 为区域订单预测提供更为科学的依据。本研究不仅从理论层面探索了消费者偏好在订单预测中的应用, 还通过实证分析, 为企业的生产和供应链管理提供了切实可行的优化建议。

2. 文献回顾

新能源汽车产业的快速发展引发了学术界对消费者偏好与销量预测的广泛关注。相关研究主要集中于消费者偏好的研究、销量预测方法的发展以及二者的结合应用。

2.1. 消费者偏好的研究

消费者偏好是影响企业决策的重要因素。现有研究表明, 消费者偏好的变化直接影响产品设计、营销策略及供应链管理。周柏旭等[1]提出, 消费者的低碳偏好与碳税政策相结合, 能够优化供应链决策,

提高市场协调效率; 段永瑞等[2]从消费者行为出发, 研究了后悔厌恶对企业定价策略的影响, 为企业制定动态定价提供了有益参考。

随着互联网的普及, 用户生成内容(UGC)成为分析消费者偏好的新兴数据来源。研究表明, UGC 中的在线评论、评分及情感表达能够真实反映消费者的体验与需求。例如, 蔺秀媛[3]通过文本挖掘技术分析实木家具的评论数据, 优化产品设计, 提高消费者满意度。

2.2. 销量预测方法的发展

销量预测是企业制定生产和营销计划的关键环节。传统的销量预测方法多依赖于历史数据, 主要包括时间序列模型、计量经济模型及混合模型等。早期研究中, ARMA 和向量误差修正模型被广泛应用于分析销量数据的趋势和经济变量的动态关系; 而近年来, 随着机器学习技术的发展, 基于深度学习的预测模型逐渐成为主流。

如卢志平等[4]提出了一种融合变分模态分解(VMD)和长短期记忆网络(LSTM)的销量预测方法, 有效解决了销量数据的非线性与非平稳性问题; 刘吉华等[5]通过百度指数结合历史销量数据, 利用主成分分析法提取关键词特征, 构建了精准的销量预测模型, 预测精度高达 96%。这些研究表明, 网络搜索数据和深度学习技术在销量预测中的潜力巨大。

2.3. 消费者偏好与销量预测的结合应用

将消费者偏好纳入销量预测模型是近年来的研究热点。基于消费者行为的动态变化, 研究者逐渐将文本情感分析技术与销量预测相结合。蔡卫星[6]通过分析消费者在线投诉与公司业绩的相关性, 证明了消费者反馈数据对企业预测能力的重要性; 段文杰[7]利用 LSTM 模型结合消费者信心指数, 显著提升了经济指标预测的准确性。

此外, 黄婕等[8]利用网络舆情数据分析新冠疫情期间的公众关注趋势, 揭示了网络评论和搜索行为对消费决策的深远影响; 陈婷[9]则研究了直播电商弹幕内容对消费者购买行为的影响, 表明在线互动中的情感表达对销量预测具有重要价值。

2.4. 研究现状总结与研究方向

现有研究为本文的研究提供了丰富的理论与技术支持。然而, 在新能源汽车(NEV)领域, 将消费者偏好的动态变化与订单预测相结合的研究仍然不足。尤其在通过深度学习模型挖掘消费者评论中的情感特征, 并将其与历史销量数据融合, 提升订单预测的准确性方面, 仍是一个亟待解决的问题。

因此, 本文创新性地提出了一种结合 CNN-Attention-Bi-LSTM 模型的预测框架, 通过对网络评论的情感分析, 捕捉消费者情感, 并基于此优化订单预测模型, 为 NEV 企业的市场决策提供理论支持和实践指导。

3. 考虑消费者偏好的订单预测模型构建

随着互联网普及, 网络评论成消费者表偏好的重要渠道。本文基于网络评论情感分析与历史订单数据, 构建了融合消费者偏好的订单预测模型。本章先分析情感特征与订单数据关联性, 再介绍订单预测模型构建过程与关键步骤。

3.1. 情感特征与订单数据关联分析

网络评论反映了消费者对 NEV 产品的偏好、体验与需求, 其情感倾向对订单量有一定的指示作用。消费者在决定购买 NEV 前, 会通过网络获取相关产品的信息, 并在评论中表达对品牌、功能或性价比的

看法。这些评论中的情感特征直接影响潜在消费者的购买决策, 进而反映在订单数据中。

本文通过收集网络评论数据, 并结合情感分析技术, 量化消费者的情感特征, 参考股票市场中的股票看涨指数, 改编为消费者偏好指数(Behavioral Index, BI)。BI 值的变化趋势与订单数据之间具有显著的相关性, 情感特征作为订单预测的重要补充信息, 能够反映出历史数据难以捕捉的消费者行为变化。

3.2. 模型设计与构建

本文提出的订单预测模型结合了网络评论情感分析与历史订单数据, 主要包括以下三个模块:

第一, 数据预处理模块: 这个模块是对网络评论数据进行清洗、分词和向量化处理; 对历史订单数据进行归一化处理。

第二, 情感分析模块: 将处理好的文本信息输入到 CNN-Attention-Bi-LSTM 模型中, 提取消费者评论中的情感特征, 生成消费者偏好指数(BI 值)。

第三, 订单预测模块: 将 BI 值与历史订单数据结合后, 输入双向长短期记忆网络(Bi-LSTM), 完成订单预测。

3.3. 情感分析模型

本文提出的情感分析模型基于 CNN-Attention-Bi-LSTM 的架构, 旨在通过 Attention 机制对输入文本的特征进行加权处理, 筛选出与情感表达更相关的信息, 再通过 Bi-LSTM 提取文本的序列特征, 从而生成消费者偏好指数(BI 值)。具体步骤如下:

首先, 引入卷积神经网络(CNN), CNN 由卷积层、输入层、激活函数、池化层等构成。输入层是用来接收文本数据的, 通常会处理为词向量的形式。CNN 的核心组成部分是卷积层, 用于提取文本的特征, 它通过使用不同大小的卷积核(如 2-gram、3-gram 和 4-gram)在输入的数据上滑动来提取局部特征, 可以提取像词汇、短语和句子类的信息。在卷积操作后, 使用 ReLU 激活函数来增加模型的非线性表达能力, 最后数据经由池化层来降低特征维度, 以此保留重要的特征。

CNN 数学表达式如下:

$$h_{i,j} = f(w \cdot x_{i:i+h-1} + b) \quad (1)$$

其中, w 为卷积核, x 为输入词向量, h 为窗口大小。

其次, 加入 Attention 机制, 用于对卷积层提取的特征进行加权, 突出对情感表达具有关键影响的特征。该机制首先要通过比较目标分量之间的相似度得到初步的权值, 然后使用 softmax 函数对初步的权值进行归一化处理, 使权值的和为 1。最后将归一化后的值与对应的特征进行加权求和, 得到最终的注意力特征。

CNN 提取的局部特征输入到 Attention 机制, 计算各特征的权重的数学公式如下:

$$a_t = \frac{\exp(s_t)}{\sum_t \exp(s_t)} \quad (2)$$

其中, s_t 为特征的重要性得分, a_t 为归一化后的权重。

根据权重对特征加权, 生成加权特征表示:

$$c_t = \sum_t a_t \cdot h_t \quad (3)$$

再次, 融入双向长短期记忆网络(Bi-LSTM), 将 Attention 加权后的特征输入 Bi-LSTM, 以捕捉文本的上下文序列关系:

$$\overrightarrow{h_t} = \text{LSTM}(c_t, \overrightarrow{h_{t-1}}) \quad (4)$$

$$\overleftarrow{h_t} = \text{LSTM}(c_t, \overleftarrow{h_{t-1}}) \quad (5)$$

最后, 生成消费者偏好指数(BI 值), BI 值(Behavioral Index)原本应用于金融领域, 是股票情绪看涨指数, 衡量的是股民情绪与股票价格之间的关系, 现在引入 NEV 领域。原始评论经由 CNN-Attention-Bi-LSTM 模型, 已经生成每条评论所对应的主题词。先用知网情感词典对预处理(去停用词)的文本进行初步情感得分计算, 之后结合每条评论的用户评分, 给出权重, 得到每条评论最终的文本情感因子 E_t 。将每条评论的最终情感得分赋值给其对应的主题词。每个大区有多条评论, 分别对应不同的主题词, 整合每个主题词的平均最终情感得分, 可以得到各大区消费者关注主题的排名。

通过 Attention 和 Bi-LSTM 处理后的文本特征, 生成消费者偏好指数(BI 值)。BI 值的计算公式为:

$$\text{BI}_t = \alpha \cdot E_t + \beta \cdot S_t \quad (6)$$

其中, E_t 是文本情感因子, 这是从情感分析模型中提取的, S_t 是历史销量数据。 α 和 β 都是权重参数, 通过实验优化来确定。

3.4. 订单预测模型

订单预测模型通过整合 BI 值与历史订单数据, 基于双向长短期记忆网络(Bi-LSTM)构建。

首先, 情感分与历史数据相融合。

情感分析模块生成的 BI 值与历史订单数据结合, 形成综合特征矩阵:

$$X = [\text{BI} \quad H] \quad (7)$$

其中, BI 是消费者偏好指数, H 是历史订单数据。

其次, 进行时间序列建模。

综合特征矩阵 X 输入到 Bi-LSTM 模型中, 捕捉时间序列中的长期依赖关系, 预测未来的订单量:

$$y_t = \sigma(W_y \cdot h_t + b_y) \quad (8)$$

h_t 是隐藏层状态, 表示当前时间步的序列特征, w_y, b_y 都是模型参数。

最后, 模型输出。

Bi-LSTM 输出的预测结果为未来某一段时间内的区域订单量。通过对预测结果进行误差分析, 验证模型的准确性。

4. 实证研究及分析结果

为验证本文提出的订单预测模型的有效性, 本文以 A 公司 NEV 区域订单数据及网络评论为基础, 开展实证研究。首先, 对数据来源及处理方式进行说明; 其次, 对模型的实验结果进行分析, 最后通过对比实验验证模型的性能优势。

4.1. 数据来源与处理

本文实证研究数据包括历史订单数据和网络评论数据。历史订单数据来自 A 公司 2022 年 1 月至 2024 年 6 月期间在各大区的实际订单量, 为确保与网络评论统一完整, 进行了缺失值填补和归一化处理, 并按时间和空间维度划分。时间维度上, 2022 年 1 月至 2024 年 3 月数据作为训练集和验证集, 2024 年 4~7 月数据作为测试集; 空间维度上, 将十余万条数据按购车地属性归类为七个大区。网络评论数据来自主流汽车论坛, 通过 Python 爬虫技术抓取, 包含 A 公司销量最高的五款车型的评论共 107085 条。经文本

预处理后按购车地点分类为七个大区文档，再对评论数据进行去重、去噪、分词及停用词过滤等清洗操作，最后用预训练的 Word2Vec 模型将分词结果转化为词向量，以便输入情感分析模型。

4.2. 模型训练与评估

4.2.1. 情感分析实验步骤

如前文所述，将处理好的数据输入到情感模型后可以输出各大区每条评论的两个主题词，这两个主题词是该条评论的中心凝练，汇总评论主题词，可以得到各大区购车者的关注主题排名，见表 1。

Table 1. Ranking of the most popular topics in each region
表 1. 各大区最受关注主题排名

大区	受关注主题排名第一	受关注主题排名第二	受关注主题排名第三
东北	品牌与售后	空间舒适性	外型设计
华北	价格与性价比	安全性能	科技配置
华东	价格与性价比	科技配置	内饰与装饰
华南	价格与性价比	动力性能	驾乘舒适性
华中	驾乘舒适性	价格与性价比	品牌与售后
西北	驾乘舒适性	内饰与装饰	外型设计
西南	动力性能	空间舒适性	外型设计

根据知网情感词典遍历每条评论，得到每条评论的情感评分，结合每条评论中的网友评分，得到每条评论最终的文本情感因子，示例见表 2。

Table 2. Example of the final sentiment score
表 2. 文本情感因子得分示例

购车地区	评论时间	评分	评论	情感分	文本情感因子	关键词 1	关键词 2
华北	2023-8	3.86	买这个车主要就是图便宜。车只值这个价钱，并不代表物超所值。品控的工艺一般……前提你得能等。以上是个人 的感受，不喜勿喷。	0.75	0.7566	价格与性价比	科技配置

4.2.2. 订单预测实验步骤

为了证明模型有效性，设置考虑消费者偏好的订单预测与只考虑历史销量的订单预测进行对比。

1) A 公司只考虑历史订单量的订单预测

本文采用 Bi-LSTM 模型对订单进行预测，为了有高质量的数据，先对 A 公司 2022 至 2024 年 7 月的历史销量数据进行全面的清理和预处理。预处理过程具体包括：缺失值填补、异常值检测与修正、数据标准化处理等。

在模型测试阶段，为了全面评估 Bi-LSTM 模型的预测能力，本文将模型的预测值与 2022 年至 2024 年 7 月的实际销量数据进行了对比。测试过程中，模型使用 2022 年和 2023 年的数据进行训练，然后对 2024 年的销量进行预测。图中蓝色实线表示实际销量，橙色虚线表示预测销量。图 1~图 7 是七个大区的测试和验证结果通过绘制 2024 年预测销量与实际销量的对比图，对模型的预测效果进行直观分析。对比显示，Bi-LSTM 模型的预测结果与实际销量趋势基本一致，能够较好地反映销量的周期性波动和整体增长趋势。

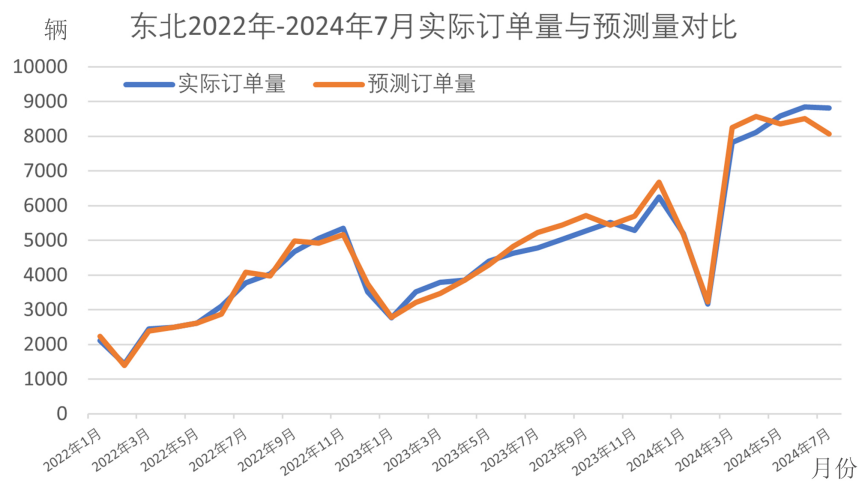


Figure 1. Daily sales forecast vs. actual sales comparison chart for the northeast region
图 1. 东北地区每日销售预测与实际销量对比图

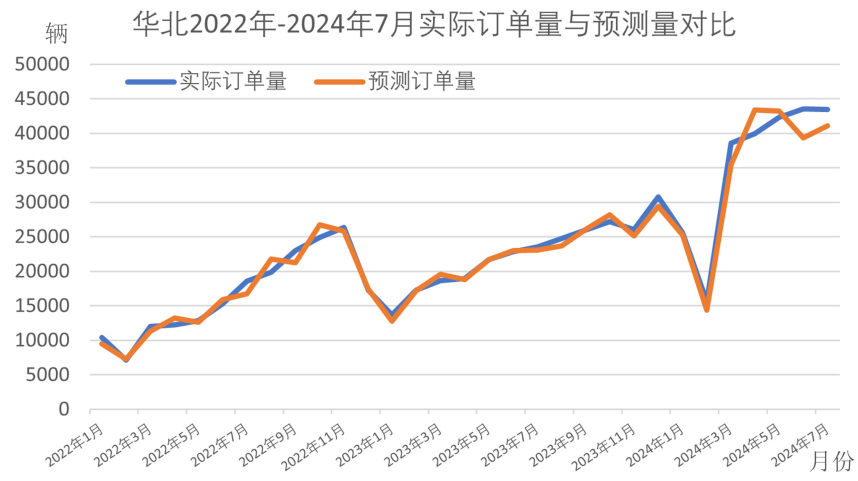


Figure 2. Daily sales forecast vs. actual sales comparison chart for the north China region
图 2. 华北地区每日销售预测与实际销量对比图

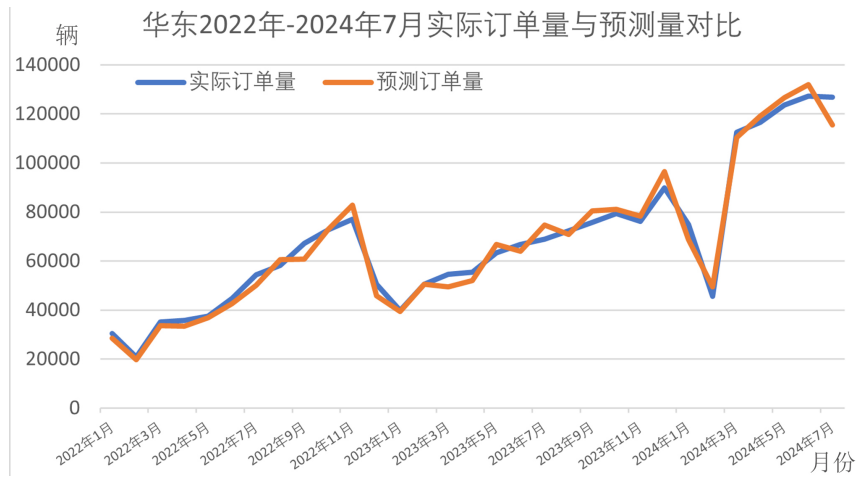


Figure 3. Daily sales forecast vs. actual sales comparison chart for the east China region
图 3. 华东地区每日销售预测与实际销量对比图

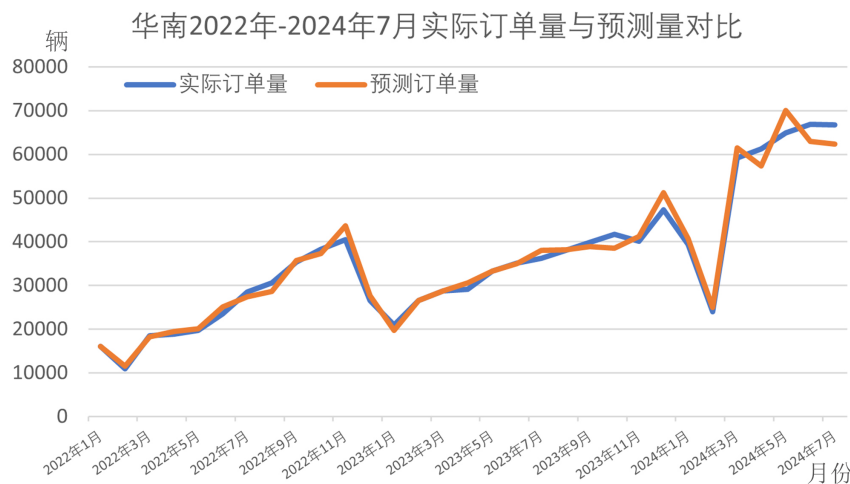


Figure 4. Daily sales forecast vs. actual sales comparison chart for the south China region

图 4. 华南地区每日销售预测与实际销量对比图

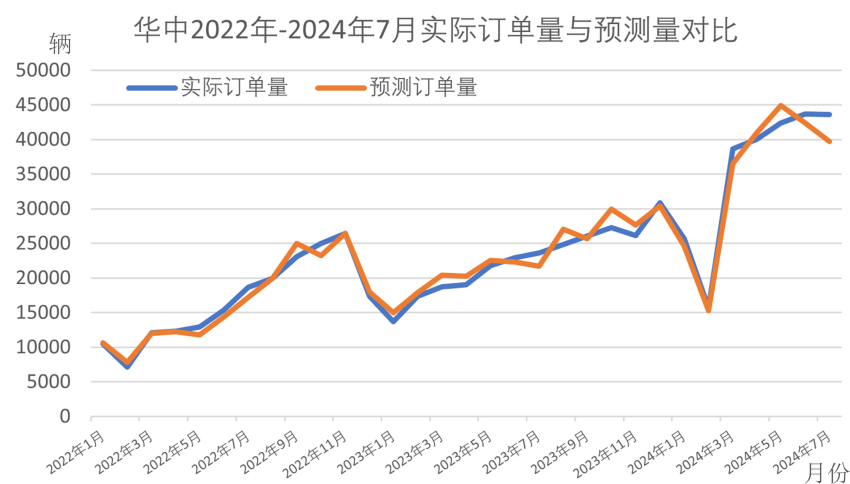


Figure 5. Comparison of daily sales forecast and actual sales in the central China region

图 5. 华中地区每日销售预测与实际销量对比图

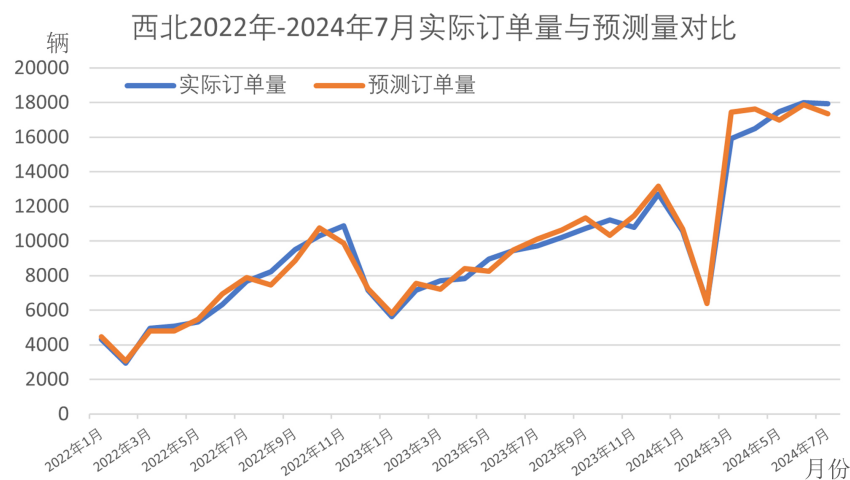


Figure 6. Daily sales forecast vs. actual sales comparison chart for the northwest China region

图 6. 西北地区每日销售预测与实际销量对比图

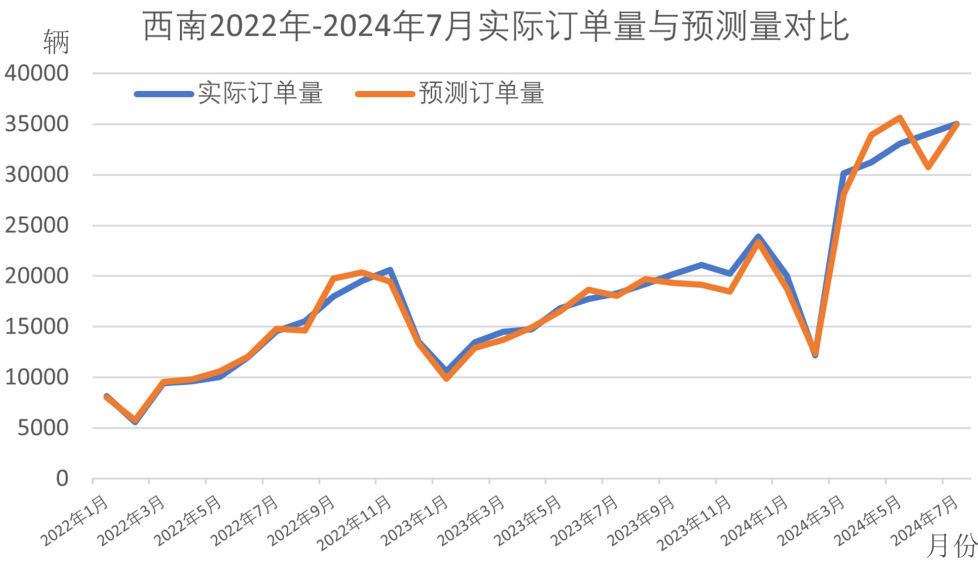


Figure 7. Daily sales forecast vs. actual sales comparison chart for the southwest China region
图 7. 西南地区每日销售预测与实际销量对比图

表 3 为对此基础模型的效用评估。

Table 3. Evaluation metrics for the model considering only historical sales volumes
表 3. 只考虑历史销量的模型效用评估指标

大区	SSE	MSE	RMSE	R ²	Dev
东北	12838267.89	414137.68	706.90	0.8814	915.57
华北	143102016.47	4616194.12	2269.68	0.9007	4511.92
华东	2323516145.54	74952134.17	9687.17	0.8926	13171.99
华南	484363067.75	15624615.09	4023.78	0.8827	6921.39
西北	46189509.46	1489984.17	1497.01	0.8931	1862.43
西南	108087412.09	3592913.74	1947.54	0.9116	3443.64
华中	152044292.00	4904654.58	2367.08	0.9081	4524.09

2) A 公司考虑消费者偏好的订单预测

为了考虑到消费者的情感因素，本节引入 BI 看涨指数模型。BI 指数原本用于金融领域的股票趋势预测，通过整合情感因子和市场交易数据来衡量市场的看涨情绪。本文将 BI 模型扩展到销量预测场景中，通过融合区域消费者的情感评价分数与历史订单数据，生成一个综合性特征，用以优化该 Bi-LSTM 的预测能力。

BI 指数由区域情感因子和时间加权订单数据构成，数学表达式为：

$$BI = w_s \cdot S_s + w_o \cdot H_w \tag{9}$$

其中， w_s 、 w_o 分别为情感因子权重与订单量权重； S_s 是基于情感分析计算的综合得分； H_w 为做了归一化处理的历史订单量。

情感因子通过提取该区域消费者评价中的主题词频次与最终情感分所得出的。

$$S_s = \frac{\sum_{i=1}^N F_i \cdot SA_i}{\sum_{i=1}^N F_i} \quad (10)$$

其中, F_i 表示该主题在所有消费者评价中的出现次数; SA_i 是平均情感分, 代表消费者对该主题的整体情感强度; N 是主题数, 在 A 公司的具体案例中, N 为 9。表 4 为各大区对应的 S_s 。

Table 4. The overall S_s in each region
表 4. 各地区总体 S_s

大区	东北	华北	华东	华南	华中	西北	西南
S_s	0.8577	0.4780	0.4543	0.5088	0.5890	0.5851	0.5030

情感得分体现了各大区消费者对主题的评价强度, 从最终的数据中可以看出, 东北区域的情感得分最高, 表明东北消费者对其购车体验有更为积极的反馈。

历史订单数据通过时间衰减权重计算得出。以区域 2022-2024 年订单数据为例, 权重设置为 2022 年为 0.32, 2023 年为 0.32, 2024 年为 0.34。

BI 指数是通过结合销量数据与情感因子数据加权计算得到, 用以反映市场趋势与情感的综合表现能力的值。因为销量数据与情感因子数值差别太大, 为了保证二者具有可比性, 本文对原始数据进行了标准化处理, 并采用权重网格优化法寻找参数 w_s 和 w_o 的最优值, 公式如下:

$$R_{total} = \text{corr}(BI, S_s) + \text{corr}(BI, H_w) \quad (11)$$

最终确定情感因子权重 w_s 为 0.443; 销量数据权重 w_o 为 0.557。

接下来构建 BI 值与历史销量数据按时间序列组合所形成双特征矩阵:

$$X = \begin{bmatrix} S_1 & B_1 \\ \vdots & \vdots \\ S_n & B_n \end{bmatrix} \quad (12)$$

基于 Bi-LSTM 模型。可以得到考虑到消费者偏好的订单量预测结果, 见图 8~图 14。

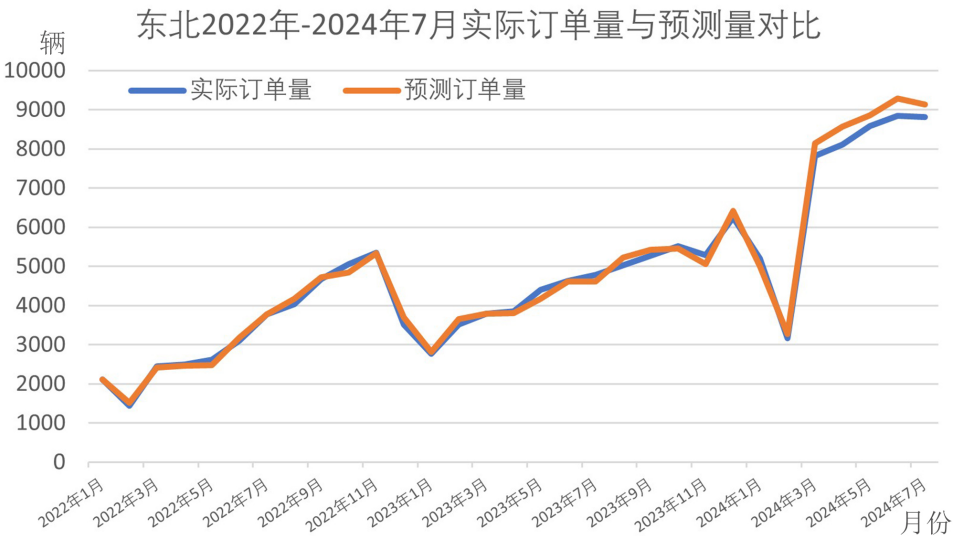


Figure 8. Daily sales forecast vs. actual sales comparison chart for the northeast region after integrating the BI value
图 8. 融合 BI 值后东北地区每日销售预测与实际销量对比图

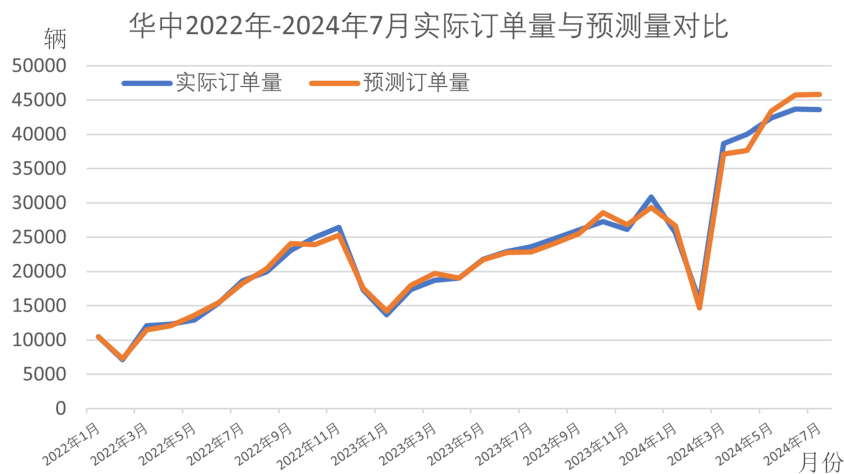


Figure 9. Daily sales forecast vs. actual sales comparison chart for the central China region after integrating the BI value
图 9. 融合 BI 值后华中地区每日销售预测与实际销量对比图

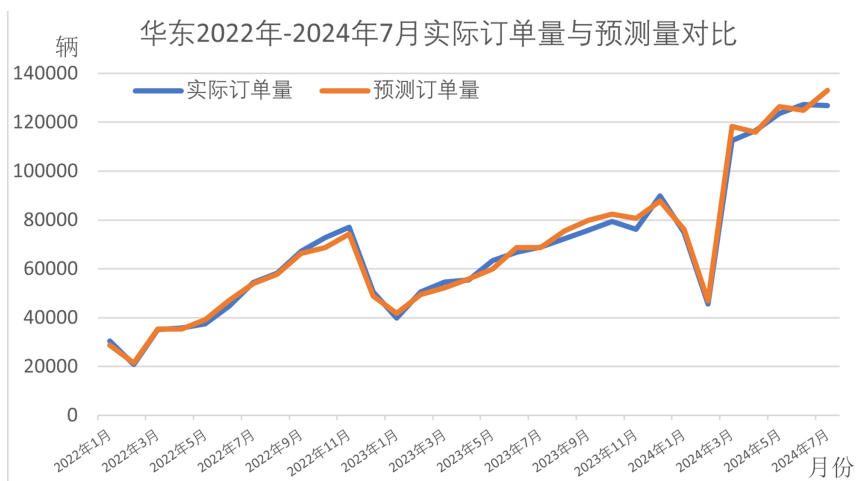


Figure 10. Daily sales forecast vs. actual sales comparison chart for the east China region after integrating the BI value
图 10. 融合 BI 值后华东地区每日销售预测与实际销量对比

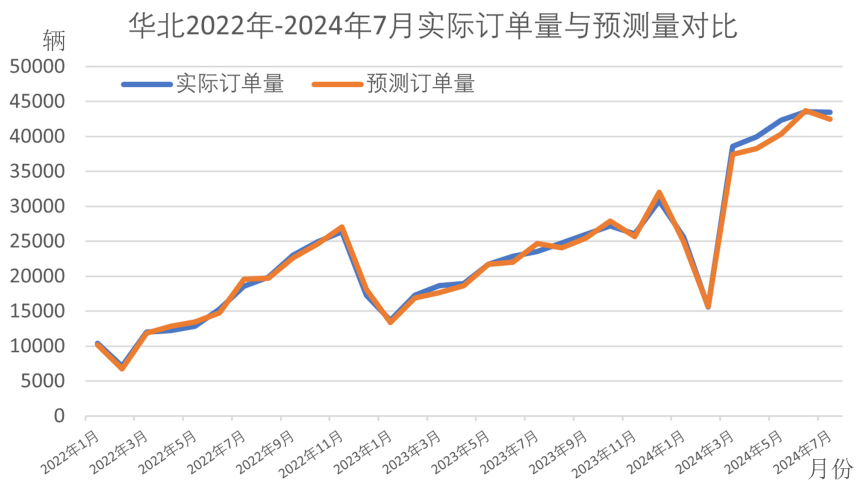


Figure 11. Daily sales forecast vs. actual sales comparison chart for the north China region after integrating the BI value
图 11. 融合 BI 值后华北地区每日销量预测与实际销量对比

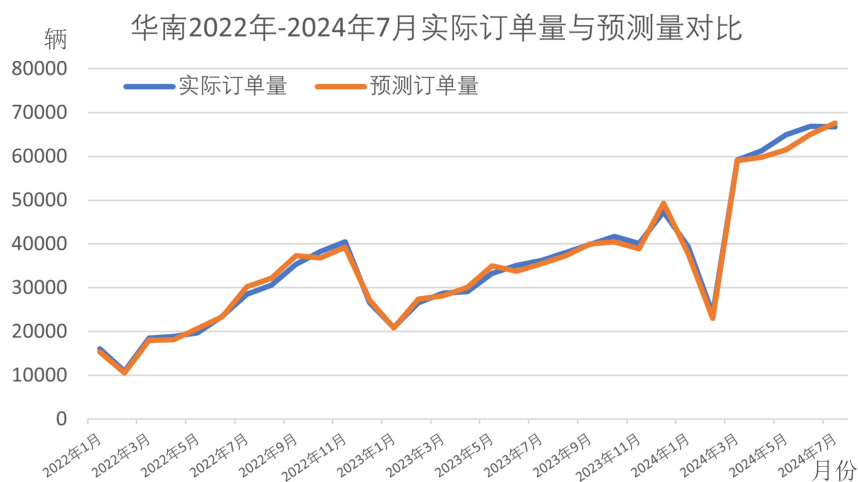


Figure 12. Daily sales forecast vs. actual sales comparison chart for the south China region after integrating the BI value
图 12. 融合 BI 值后华南地区每日销量预测与实际销量对比

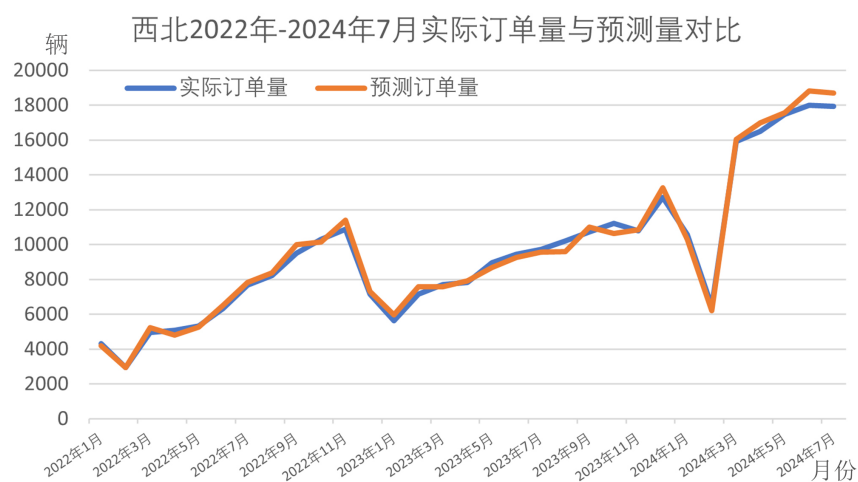


Figure 13. Daily sales forecast vs. actual sales comparison chart for northwest China region after integrating the BI value
图 13. 融合 BI 值后西北地区每日销量预测与实际销量对比

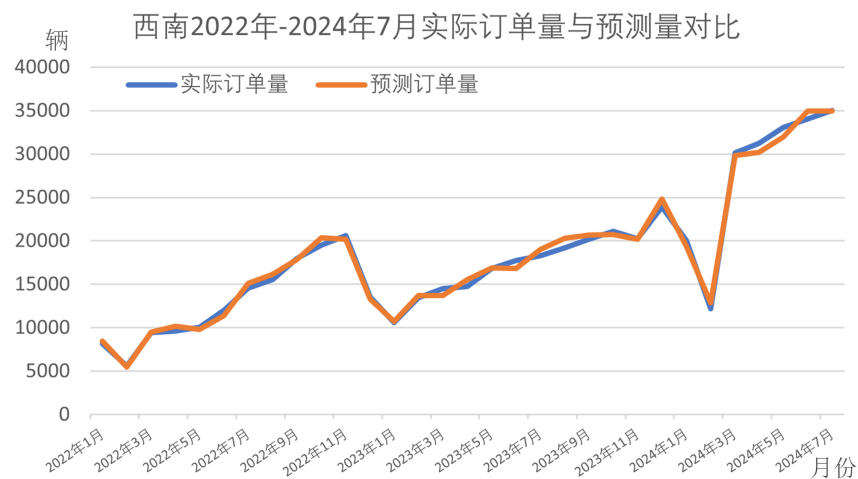


Figure 14. Daily sales forecast vs. actual sales comparison chart for southwest China region after integrating the BI value
图 14. 融合 BI 值后西南地区日销量预测与实际销量对比

(3) 效用评估

Table 5. Model utility evaluation metrics
表 5. 模型效用评估指标

大区	SSE	MSE	RMSE	R ²	Dev
东北	11163711.21	360119.72	644.10	0.9078	834.53
华北	124436536.93	4014081.84	2068.52	0.9277	4111.29
华东	2020448822.21	65175768.84	8835.15	0.9194	12008.32
华南	421185276.3	13586621.82	3669.00	0.9092	6311.29
西北	40164790.83	1295638.41	1365.26	0.9199	1698.18
西南	93989053.99	3132968.47	1776.02	0.9389	3140.24
华中	132212427.83	4264917.03	2158.17	0.9353	4124.48

表 5 与表 4 对比可知，两个模型的效用分析都使用了 SSE (残差平方和)、MSE (均方误差)、RMSE (均方根误差)、R² (决定系数)和 Dev (偏差)五个指标。优化后的模型在各项指标上均优于未优化版本，尤其是在 R² 和 RMSE 上表现出显著提升。

R² 是衡量模型解释力的重要指标，加入情感分析的模型的 R² 在各个大区平均提高了约 3%。

RMSE 反映预测值与真实值之间的偏差，优化后的 RMSE 显著降低。华南地区的 RMSE 从 4023.78 降低到 3669.00，降幅达 8.8%，表明加入情感分析后的模型预测精度更高。

SSE 和 MSE 是衡量模型误差总量的指标，优化后这两项指标平均降低了 15%左右。西北地区的 SSE 从 46189509.46 降低到 40164790.83，表明优化后的模型整体误差减少。

Dev 是衡量模型预测偏差的指标，优化后的 Dev 在各个大区均有所降低。华中和华北地区的 Dev 降幅分别达到 9.7%和 9.6%，表明优化后的模型在预测时更加稳定。

从具体的各区域表现来看，华东地区的数据量最大，本文使用的模型在 R² 和 RMSE 上的提升尤为显著，说明模型在大量数据下的表现更稳定。西北和东北地区这两个地区的数据量较小，但优化后的模型仍然较之前有改进效果(R² 分别提升 2.1%和 2.3%，RMSE 分别降低 8.1%和 8.4%)，说明加入情感分析的模型在数据量较小的场景下依然有效。

Table 6. Benchmark model evaluation
表 6. 基准模型评估

大区	ARIMA			随机森林			本模型		
	RSME	Dev	R ²	RSME	Dev	R ²	RSME	Dev	R ²
东北	678.30	876.41	0.87641	732.50	924.76	0.8503	644.10	834.53	0.9078
华北	2171.95	4316.85	0.8813	2245.90	4489.20	0.8721	2068.52	4111.29	0.9277
华东	9276.91	12608.74	0.8734	9492.80	12975.20	0.8651	8835.15	12008.32	0.9194
华南	3852.45	6626.86	0.8637	3980.90	6845.20	0.8534	3669.00	6311.29	0.9092
西北	1433.52	1783.09	0.8739	1478.60	1835.20	0.8642	1365.26	1698.18	0.9199
西南	1864.82	3297.25	0.8912	1923.50	3400.90	0.8810	1776.02	3140.24	0.9389
华中	2266.08	4330.70	0.8881	2340.80	4475.20	0.8785	2158.17	4124.48	0.9353

表 6 是本模型与 ARIMA 和随机森林的模型效用分析对比表，整体上看，本模型仍然优于这两个模

型。本模型整体 RSME 表现良好, ARIMA 的 RSME 平均约高出约 4.8%, 表明预测误差相对较大。随机森林的平均 RSME 更是高出约 8%~13%不等, 误差范围进一步扩大。本模型 Dev 值整体较低, 表明模型预测结果的稳定性较高 ARIMA 的波动性较大, 随机森林的 Dev 值高出约 8%~10%, 稳定性更低, 不太适用于此案例。三个模型 R^2 也遵从上述趋势, 由本模型、ARIMA 和随机森林依次降低。

因此, 无论是与基础模型相对比, 还是与未加入情感分析的模型相对比, 本文所使用的模型都具有较好的预测能力。

4.3. 预测结果差异分析

通过对比不同区域的预测精度(如表 5 所示), 西南和华中地区的预测准确率显著高于其他区域。可能与数据质量、消费者情感的一致性和市场特性三个因素有关。

从数据量大小与质量来看, 评论 IP 为华中地区的数据高达 30472 条, 占据数据总量的 28.46%, 且消费者评论所关注的重点更集中在“价格与性价比”和“动力性能”方面, 情感分析的稳定性更高, 与此相反的是西北和东北的数据, 两者数据的和占总数居比例 27.31%, 由此可能导致模型训练不足。

从消费者情感一致性来看, 西南地区消费者对“动力性能”和“空间舒适性”的情感表达较为一致(平均情感分达 0.503), 正向情感占比高, 降低了模型噪声; 而华北地区消费者对“价格与性价比”的负面评论较多, 情感波动增加了预测难度。

从市场特性来看, NEV 市场受政策因素影响大, 华东地区是国内 NEV 市场竞争较为激烈的区域, 发行促销活动等因素未被纳入模型, 可能导致偏差, 西南和华中地区相对华东地区来说, 市场较为稳定。

5. 结论与展望

通过网络评论提取信息构建消费者偏好的情感因素与历史销量相结合的预测模型, 本文提出的模型可以为 A 公司的具体预测提供相对可靠的解决方案, 对于西南、华中地区, 预测精准度可以达到 93%, 其余大区的预测精准度也都达到 90%以上; 除此之外, 还可以揭示各大区消费者对于 A 公司产品属性的不同关注度。

本文用户价值可以在预测及策略优化两方面体现。

从订单预测角度来说, 通过融合消费者情感与历史数据, 企业可动态调整区域产能。较为精准的订单预测可以倒逼供应链上的各个环节, 降低不必要的时间与财力浪费。

从策略优化方面来说, 针对各大区消费者关注的产品特性不同, A 公司可以调整每个地区的销售运营侧重点。比如, 华北地区“安全性能”情感分较低(0.478), A 公司针对性改进 ADAS 系统并加强宣传, 可以再占领一部分华北地区的市场份额, 达到利润最大化的目的。

在本文的研究过程中发现, 政策对于 A 公司销量有很大影响, 但目前的预测方式只能观察到政策出台后民众对 A 公司产品的情绪变化, 无法推测政策何时出台以及造成的正负面影响。之后的研究重点可以放在政策与销量的关系上。

参考文献

- [1] 周柏旭, 吴金卓, 王卉. 考虑消费者低碳偏好和碳税政策的三级纸制品供应链协调机制研究[J]. 西安理工大学学报, 2024, 40(4): 461-470.
- [2] 段永瑞, 高一分, 喻博峰, 等. 预售环境下考虑消费者后悔厌恶的定价问题[J]. 工业工程与管理, 2024, 29(3): 200-206.
- [3] 蔺秀媛, 许柏鸣. 实木床产品消费者满意度文本特征挖掘[J]. 商场现代化, 2023(20): 31-33.
- [4] 卢志平, 玉晓晶, 陆成裕. 融合 VMD 和 LSTM 模型的新能源汽车销量预测方法[J]. 武汉理工大学学报(信息与

管理工程版), 2023, 45(4): 546-551.

- [5] 刘吉华, 张梦迪. 基于百度指数的大众汽车销量预测研究[J]. 统计与管理, 2020, 35(10): 23-31.
- [6] 蔡卫星, 蒲雨琦, 李浩民. 顾客至上: 消费者在线投诉的基本面预测能力研究[J]. 管理世界, 2024, 40(5): 139-154.
- [7] 段文杰. 基于文本特征挖掘和 LSTM 组合模型的消费者信心指数预测研究[D]: [硕士学位论文]. 大连: 东北财经大学, 2023.
- [8] 黄婕, 刘慧, 刘维斯, 等. 基于百度指数的我国新型冠状病毒肺炎网络舆情分析[J]. 预防医学情报杂志, 2022, 38(2): 172-177.
- [9] 陈婷, 丁宏昊, 周昊宇, 等. 直播电商弹幕内容与行为特征对消费者购买行为的影响研究——主播与产品关联度的调节效应[J/OL]. 数据分析与知识发现, 1-19.
<http://kns.cnki.net/kcms/detail/10.1478.G2.20240524.1107.004.html>, 2025-02-11.