

基于机器学习算法的促销销量预测研究

吕艳琳

重庆理工大学

DOI:10.12238/acair.v3i1.11920

[摘要] 本文研究了基于机器学习算法的促销销量预测方法,针对电商企业面临的促销销量不稳定问题,提出了一种融合促销敏感度分类与机器学习模型的综合预测方法。首先,根据历史数据计算各商品的促销敏感度,并据此将商品划分为不同的类别。然后,针对每一类别,利用该类别的促销数据样本对机器学习模型进行训练。在进行销量预测时,先预测商品的促销敏感度,确定其所属类别,再选择对应的机器学习模型进行预测。实验结果表明,该方法能够有效提升促销销量的预测精度,为电商企业的运营决策提供了有力支持。

[关键词] 销量预测; 随机森林; LightGBM; 梯度提升树

中图分类号: F272.1 **文献标识码:** A

Research on Promotional Sales Prediction Based on Machine Learning Algorithm

Yanlin Lv

Chongqing University of Technology

[Abstract] This paper investigates the promotional sales prediction method based on a machine learning algorithm. It proposes a comprehensive prediction method integrating promotional sensitivity classification and machine learning models for the unstable promotional sales problem e-commerce enterprises face. First, the promotion sensitivity of each product is calculated based on historical data, and the products are classified into different categories accordingly. Then, for each category, a machine learning model is trained using promotional data samples from that category. When making sales predictions, the promotion sensitivity of the products is predicted first to determine the category they belong to, and then the corresponding machine learning model is selected for prediction. The experimental results show that the method can effectively improve the prediction accuracy of promotional sales, which provides powerful support for the operational decision-making of e-commerce enterprises.

[Key words] sales volume prediction; random forest; LightGBM; gradient boosting tree

引言

精确预测销量对于入驻京东、亚马逊等电商平台的企业而言,是制定高效运营规划和实施有效供应链管理的关键。预测促销销量的准确性直接关系到电商企业在采购、库存管理、生产调度及人力资源配置等多个环节的决策质量。过于乐观或悲观的预测都可能引发不必要的经营成本与收益损失。例如,在促销期间,企业通常会根据预测结果,提前将产品从区域配送中心调配至更接近客户的前置物流中心以备发货。若预测销量低于实际需求,前置物流中心的库存将难以满足订单需求,需由区域配送中心进行二次配送,这不仅延长了交货时间,降低了客户满意度,还可能导致客户流失,或者因为库存不足导致促销活动提前结束,导致企业利润受损。相反,若预测销量过高,则会导致库存积压,增加仓储成本及运营风险,同时造成人力资源和物资的双

重浪费。

因此,提升促销产品销量的预测精度,已成为电商企业提升核心竞争力的当务之急,对于优化资源配置、降低运营成本、提高客户满意度及保障企业稳健发展具有重要意义。本文通过针对商品类别构建预测模型,解决了电商企业面临的诸多挑战,包括商品种类繁多且特性各异以及促销活动多样化导致促销销量不稳定等问题。具体而言,本文提出了一种基于促销敏感度的商品分类方法,弥补了传统商品分类在促销情境下难以有效区分相似商品的不足。

王新锋的研究聚焦于深度学习在电商销量预测中的应用。该研究基于消费者购买决策过程理论,提炼出影响电商销量的各类要素,并构建了融合卷积神经网络(CNN)和XGBoost的预测模型。实验结果显示,非线性模型在预测准确率上远超线性模型,

而基于深度学习的CNNXG融合模型在特征集较少时表现更佳。这表明特征学习能有效减少特征工程工作,提升预测精度,为电商企业的销量预测提供了新视角^[1]。白云则设计了一个基于神经网络和集成学习的商品销量管理系统。该系统初步采用BP网络构建预测模型,但因原始数据维度过高导致效果不佳。因此,提出PCA-BP模型,通过主成分分析(PCA)降维,提高了模型效率和精度。同时,研究还基于XGBoost算法设计了销量预测模型,并通过特征工程优化了性能。实验表明,经过特征工程的XGBoost模型预测结果更优。最终,提出的PCA-BP与特征工程优化后的XGBoost组合模型,预测效果超越单一模型,为商品销量管理提供了有效工具^[2]。邢添威针对零售药房数据,研究了流感类药品销量和流动客户数量对运营效率的影响。使用GBDT+Ridge融合模型预测药品销量,并利用XGBoost模型分析流动客户数量的影响因素。实验证明,所用模型在销量预测上有效,且给出了流动客户数量影响因素的重要度排名,为零售药房制定经营决策提供了科学依据^[3]。

Trapero J R等探究了在促销活动中,基于专家经验的销量调整方法的预测效果。结果显示,专家经验有助于减小基础销量预测偏差,尤其在历史数据稀缺时仍适用,体现了人工判断在商业预测中的价值^[4]。应舟丹基于电商平台活动数据,融合随机森林、XGBoost及Light GBM算法,构建了特征融合的机器学习预测模型。为提升模型可解释性,引入SHAP值法分析商品特征与销量的关系,对营销商品进行分层聚类并重新训练模型。研究揭示,该模型预测准确度高,且识别出影响销量的关键因素,为营销商品选择策略提供了科学支持^[5]。

在预测医药零售及跨境电商的销量时,选择并优化模型对于提升预测精度极为关键。前人研究发现了这样一个事实:针对不同行业和应用场景,可能需要采用不同的模型与优化策略,此外在促销销量预测领域,前人的研究主要聚焦于两种主要方法:人工调整与模型预测。

1 数据预处理

数据预处理的具体过程如下:因为本文的研究对象为促销销量预测,折扣为0说明该产品并未参与促销,销量为0的数据,说明参与活动也未对其销售量产生影响,对于研究问题亦无用,故删除销量、销售额、折扣为0的数据。实验数据集中主要包含促销活动特征和商品特征两大类影响因素,具体涵盖活动类型、活动名称、活动时间、SKU(库存持有单位)、机型、折扣等。然而,实际上影响消费者购买决策的因素远不止于此。为了提升预测的准确性,本文基于大量的数据分析、文献回顾以及企业调研所得出的结论,进一步构造了一系列对促销销量具有潜在影响的新特征。这些新特征的具体内容如下所述:

(1) 活动持续天数:计算同一活动名称下活动的持续天数,将“促销活动持续天数”加入到数据集中,以便进一步探究其对促销效果的具体作用。

(2) 促销活动数量:将商品同时参与的促销活动数目作为一个新的特征变量,并将其命名为“促销活动数量”。

(3) 促销价格、原价格与价格下降幅度:根据促销销售额和销售量得到该产品的促销价格,再根据折扣信息得到该产品的原价格,根据求出的促销价格与原价格得到产品促销期间价格的下降幅度。

(4) 年、月:将活动时间中的年、月分离出来,考虑活动的时间因素

2 评价指标选取

在评估本研究提出的预测方法及其基准预测方法的性能时,我们采用了三个核心的评价指标:平均绝对误差(Mean Absolute Error, MAE)、均方根误差(Root Mean Square Error, RMSE)以及平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)。这些指标因其计算简便且不受样本规模影响的特性,而被广泛采纳为衡量回归预测模型精度的标准。三个评价指标的定义如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.1)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (2.2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.3)$$

3 促销销量预测方法的构建

本文提出了一种针对电商促销销量预测的综合方法,该方法融合了促销敏感度分类与机器学习模型的应用。具体而言,我们首先依据历史数据计算各商品的促销敏感度,并据此将其划分为不同的类别。随后,针对每一类别,我们利用该类别的促销数据样本对机器学习模型进行训练。在进行销量预测时,我们预测商品的促销敏感度,利用预先训练的模型确定其所属类别,进而选择对应的机器学习模型对该商品的促销销量进行预测。

3.1 促销敏感度

本文提出了一种基于促销敏感度的商品分类方法,该方法的核心在于将展现出相似促销响应特性的商品归入同一类别,旨在提升促销预测的精确度。促销敏感度作为本文新定义的一种商品属性特征,能够有效表征商品在促销活动中的表现特性。鉴于不同商品对促销活动的响应程度存在差异,其促销敏感度亦呈现多样性。本文提出促销敏感度1及促销敏感度2,具体定义如下:

$$\text{商品的促销敏感度1} = \frac{\sum_{j=1}^n \frac{y_{ij}}{P_{ij}(1-d_{ij})}}{n} \quad (3.1)$$

$$\text{商品的促销敏感度2} = \frac{y_{ij}}{p_{ij}(1-d_{ij})} \quad (3.2)$$

n 代表商品 i 参与促销活动的总次数; y_{ij} 是商品 i 在促销

活动 j 中的销售量; d_{ij} 表示商品在促销活动 j 中的折扣,

p_{ij} 表示商品 i 的原始价格。由公式可知, 促销敏感度1表示每下降1美元所带来的平均促销销量, 促销敏感度2表示每下降1美元所带来的促销销量。

3.2 促销敏感度预测

对于曾经参加过促销活动的产品来说, 促销敏感度1是已知的, 鉴于促销敏感度2无法直接获取, 且其对于预测促销销售量的影响至关重要, 我们首先需要采取一种间接的方法来预测这一敏感度。我们分别采用了LightGBM、随机森林、Xgboost、梯度提升树、神经网络模型对敏感度2进行预测, 通过对比促销敏感度2测试集上不同预测方法的误差(加入促销敏感度1与否), 即从MAE、MAPE和RMSE来看, 加入敏感度1后的随机森林模型对于敏感度2的预测相对准确, 所以选取随机森林模型对于敏感度2进行预测。

3.3 销量预测模型

3.3.1 模型的建立

对于促销销量预测, 主要有以下步骤:

(1) 敏感度分类: 在获取促销敏感度的预测值基础上, 依据业务实际需求及数据分析的严谨结果, 我们将敏感度科学划分为多个类别, 例如高敏感度、中敏感度及低敏感度等。此分类步骤有助于深化理解各类促销活动对消费者吸引力的差异, 并为后续的促销销量预测提供更具指向性的数据支持。

(2) 分级别促销销量预测模型: 基于促销敏感度分类的结果, 我们为每个类别量身定制相应的促销销量预测模型。这些模型在构建时充分考虑了各类别特有的促销敏感度特征, 旨在实现针对不同类别促销活动下销售量的精确预测, 从而提升预测的针对性和准确性。

3.3.2 模型结果及分析

(1) LightGBM.

表1 LightGBM预测结果

预测方法	分类指标	自变量	MAE	MAPE	RMSE
LightGBM			70.2252	139.0986	117.5931
LightGBM	敏感度1		26.1331	39.3051	44.0566
	敏感度2		23.9242	34.1495	40.5035

(2) 随机森林.

表2 随机森林预测结果

预测方法	分类指标	自变量	MAE	MAPE	RMSE
随机森林			23.8291	46.8178	56.195
随机森林	敏感度1		17.2834	30.5958	39.4141
	敏感度2		16.4661	29.4151	36.5877

根据对LightGBM、随机森林算法的实验结果分析, 我们发现了一个有趣的现象: 当商品销量按照敏感度进行分类预测时, 模型的预测效果优于未分类结果。此外, 按照敏感度2的分类预测比敏感度1的效果更好, 说明直接利用具体的敏感度值可能比使用它们的平均值更能提供有用的信息, 从而帮助模型做出更准确的预测。

4 结论

本文通过研究基于机器学习算法的促销销量预测方法来解决电商企业面临的促销销量不稳定问题。实验结果表明, 本文提出的预测方法能够有效提升预测精度, 为电商企业的运营决策提供了科学依据。未来, 将继续深化研究, 探索更多影响促销销量的因素, 并优化预测模型, 以进一步提高预测效率和准确性, 为电商企业的稳健发展提供有力支持。

[基金]

重庆理工大学研究生创新项目资助。

[参考文献]

- [1]王新锋.基于深度学习的电商销量预测方法研究[D].南京大学,2019.
- [2]白云.面向商品销量预测模型的研究与实现[D].西安电子科技大学,2020.
- [3]邢添威.零售药房的总销量预测与流动客户数量影响因素分析[D].吉林财经大学,2020.
- [4]Trapero J R,Kourentzes N,Fildes R.On the identification of sales forecasting models in the presence of promotions [J].Journal of the operational Research Society, 2015, 66(2): 299-307.
- [5]应丹丹.营销促销场景下电商平台销量预测研究[D].南京大学,2021.

作者简介:

吕艳琳(2000--),女,汉族,四川广安人,研究方向: 大数据可视化。