

文章编号: 2095-2163(2023)09-0059-08

中图分类号: F423;F251

文献标志码: A

基于 LightGBM 模型的离散制造业产品物料需求智能预测

李婷婷, 黄欣迪, 曹萌萌, 李剑锋

(中国计量大学 经济与管理学院, 杭州 310018)

摘要: 离散制造业产品的物料需求受多种因素影响,传统物料需求预测算法对数据要求高,企业需要进行大量运算,且预测提前期短、精准度低,不能及时满足企业的生产计划。为提高制造业物料需求预测精度,本研究采用美的集团离散型物料需求数据进行建模分析。首先进行数据预处理及特征工程,利用统计学中的统计量构建出滑动和滞后特征,然后构建并拟合 LightGBM 模型对物料需求量进行预测,并与传统时间序列 SARIMA 模型进行对比,引入平均绝对误差 MAE 评估模型的预测精准度,针对模型时间复杂度和预测精准度进行对比分析。结果表明,以月为时间粒度的情况下,LightGBM 机器学习模型对离散制造业物料需求预测的效率和准确率更高,更有利于提高离散制造企业的生产效率。

关键词: LightGBM 模型; 物料需求预测; 机器学习; SARIMA 模型; 对比分析

Intelligent forecast of material demand for discrete manufacturing products based on LightGBM model

LI Tingting, HUANG Xindi, CAO Mengmeng, LI Jianfeng

(School of Economics and Management, China Jiliang University, Hangzhou 310018, China)

[Abstract] The material demand of discrete manufacturing products is influenced by various factors. Traditional material demand prediction algorithms require high data requirements, and enterprises need to perform many calculations. Moreover, the prediction lead time is short, and the accuracy is low, which cannot meet the production plan of enterprises in a timely manner. To improve the accuracy of material demand prediction in the manufacturing industry, this study uses discrete material demand data from Midea Group for modeling and analysis. First, data preprocessing and feature engineering are carried out, and the statistics in statistics are used to build the sliding and lagging characteristics. Then, the LightGBM model is built and fitted to forecast the material demand and compared with the traditional time series SARIMA model. The prediction accuracy of the average absolute error MAE evaluation model is introduced, and the model time complexity and prediction accuracy are compared and analyzed. The results indicate that the LightGBM machine learning model has higher efficiency and accuracy in predicting material demand in discrete manufacturing industries with monthly time granularity, which is more conducive to improving the production efficiency of discrete manufacturing enterprises.

[Key words] LightGBM model; material demand forecast; machine learning; SARIMA model; comparative analysis

0 引言

大数据驱动制造产业智能化,人工智能助力制造业数字化转型升级,随着疫情环境下制造业现有供应链脆弱性的暴露,企业原有的预测算法已不能满足制造业企业对物料需求预测的准确性。为有效减少物料短缺与过度采购现象,助力企业降低库存成本,提高采购效率,缩短交货周期,提高企业抗风

险能力,迫切需要精准度更高的预测模型^[1]。

物料需求预测是指根据企业已有的销售或物料消耗数据,对企业未来一段时间内生产运营所需要的物料种类以及数量进行分析研究。预测方法大致分为定性预测法和定量预测法。其中,定性预测法根据预测者个人的知识、经验和主观判断,主观随意性较大,预测结果的准确度不能达到制造企业的实际需求。时间序列分析是较为常用的定量预测法,

基金项目: 国家级大学生创新创业训练计划项目(202210356031)。

作者简介: 李婷婷(2002-),女,本科生,主要研究方向:物流管理、大数据处理;黄欣迪(2002-),女,本科生,主要研究方向:机器学习、大数据处理;曹萌萌(2002-),女,本科生,主要研究方向:人工智能、大数据处理;李剑锋(1977-),男,博士,副教授,主要研究方向:商务数据分析、人工智能、大数据技术等。

通讯作者: 李剑锋 Email: ljfwinner@163.com

收稿日期: 2022-09-26

起源于英国统计学家 G.u.Yule 在 1927 年提出的自回归模型 (Auto Regressive, AR)^[2]。该模型与英国统计学家 G.T.Walker 在 1931 年提出的移动平均 (Moving Average, MA) 模型和 ARMA 模型, 构成了时间序列分析的基础, 至今仍被大量应用^[3]。这三个模型主要应用于单变量、同方差场合的平稳序列, 在解决线性时间序列的拟合问题上有着卓越表现。但现实中多为非线性复杂问题^[4], 相较于传统的时间序列预测方法, 引入机器学习算法^[5]可以更好地解决非线性时间序列的拟合问题。李福等人^[6]基于机器学习 XGBoost 算法, 对纽约市共享单车借车需求量进行了有效的预测。李卫星等^[7]使用 XGBoost 算法, 对 4 种工况下柴油机失火故障的平均预测准确率达 90% 以上。周宇阳^[8]提出了一种基于机器学习 SVM 算法的需求分析模型, 利用该模型对沿海地区 LNG 进行仿真预测。Ntakolia, C. 等^[9]认为传统方法未考虑历史数据信息, 因此对多种机器学习模型进行对比和校准, 得到了预测准确性较高的 LightGBM 算法, 从而对库存系统中的延期交货率进行了有效预测。Javad Feizabadi^[5]开发了机器学习混合需求预测方法, 将时间序列和解释因素都输入到开发模型中, 发现了具有统计学意义的供应链改善差异。

目前, 国内外需求量预测相关研究已积累了一些成果, 但较少考虑离散制造业产品种类多、定货数量多等特点对需求量预测的影响, 难以满足企业对预测准确性和实效性的要求。XGBoost 算法相比其它算法收敛慢、预测耗时大、超参数多且复杂, 人工调参具有较高难度, 各参数的取值大小直接影响模型的预测精度, 因此存在预测精度受限的问题^[10-11]。SVM 广泛应用于模式识别和回归估计等领域, 与其它算法相比在解决小样本、高维非线性决策问题时有很大的优势^[12], 但仍仍然存在计算复杂、

泛化能力差和过拟合等固有局限^[13], 在制造业需求预测问题上仍有一定的提升空间。针对上述问题, 相对于 XGBoost 和 SVM 算法, LightGBM 算法在样本规模较大、特征维度较高的预测任务中, 呈现出计算能力强且精准度高的优势。

本研究拟采用 LightGBM 机器学习集成算法对原始数据进行建模, 划分训练集和测试集。将时间序列下的物料需求量数据和库存数据输入模型进行训练, 利用训练好的 LightGBM 模型, 在测试集上预测未来时间段内的物料需求量, 同时将基于决策树的 LightGBM 物料需求预测模型与传统时间序列 SARIMA 模型的预测结果进行对比, 为离散制造业物料需求预测问题探寻更优的预测模型。

1 实验模型

1.1 LightGBM 算法

LightGBM 算法是一种并行学习的机器学习算法, 其中基于梯度的单边采样算法 (Gradient-based One-Side Sampling, GOSS) 和互斥特征捆绑算法 (Exclusive Feature Bundling, EFB) 解决了在大样本高维度数据情况下耗时长、占用内存大的问题。本文通过构建基于 Boosting 学习方式的 LightGBM 预测模型, 降低数据异常值对预测准确度的影响, 忽略异常值权重, 防止过拟合。

1.1.1 基于梯度单边采样的 GOSS 算法

梯度大的样本点会贡献更多的信息增益, 为了保持信息增益评估的精度, 对样本进行采样时保留梯度大的样本点, 对于梯度小的样本点按比例进行随机采样。如果一个样本点的梯度小, 则该样本点的训练误差就小, 直接的办法就是抛弃梯度小的样本点, 但该做法可能会改变数据的分布、损失学习的模型精度。而 GOSS 算法的提出, 则避免了上述问题。GOSS 算法描述详见表 1。

表 1 GOSS 算法描述

Tab. 1 GOSS algorithm description

过程	描述
输入	训练数据, 迭代步数 d , 大梯度数据的采样率 a , 小梯度数据的采样率 b , 损失函数和弱学习器的类型 (一般为决策树)。
采样过程	I. 将所有样本梯度取绝对值, 降序排序得到样本梯度索引全集; II. 选取前 $a * 100\%$ 的样本生成一个大梯度样本点的子集, 对剩下的样本集合 $(1 - a) * 100\%$ 的样本, 随机选取常数因子 $b * (1 - a) * 100\%$ 个样本点, 生成一个小梯度样本点的集合; III. 然后将大梯度样本和采样的小梯度样本合并赋给小样本一个权重系数, 采样过程结束; IV. 利用采样样本可以学习一个新的学习器。
迭代	重复上述步骤直到达到规定的迭代次数或者收敛为止。
输出	训练好的强学习器。

1.1.2 EFB 算法

LightGBM 算法不仅进行了数据采样,也进行了特征抽样,使得模型的训练速度进一步地提升,即将互斥特征绑定在一起,从而减少特征维度。其主要思想是,在实际应用中高纬度的数据往往都是稀疏数据,故而需要设计一种几乎无损的方法,来减少有效特征的数量。在稀疏特征空间中,许多特征都是互斥的,因此可将互斥特征绑定在一起,形成一个特征,从而减少特征维度。

EFB 算法将特征划分为更小的互斥绑定数量,是一个 NP-hard 问题,即在多项式时间内不可能找到准确的解决办法。因此,使用一种近似的解决办法,即特征之间允许存在少量的样本点并不是互斥的,允许小部分的冲突可以得到更小的特征绑定数量,更进一步地提高计算的有效性。

1.1.3 基于直方图的决策树算法

GBDT 算法由于要遍历每一个数据计算信息增益以获得最佳分裂点,时间复杂度高,而本研究样本数量非常大。为解决该问题,Ke 等^[14]提出了基于直方图的决策树算法,将连续的特征离散化为 k 个离散特征,也就是分桶 bins 思想,同时构造一个宽度为 k 的直方图用于统计信息(含有 k 个 bin)。利用直方图算法则无需遍历全部数据,只需要遍历 N/k 个数据即可找到最佳分裂点,极大地提高了训练效率,缩减了内存占用空间^[15]。

1.1.4 带深度限制的 leaf-wise 决策树生长策略

如图 1 所示,level-wise 生长策略,即数据可以同时分裂同一层的叶子,容易进行多线程优化,也可以较好控制模型复杂度,不容易过拟合。但实际上,level-wise 是一种低效的算法,很多分裂增益较低的叶子也会进行分裂和搜索,增加了计算量。

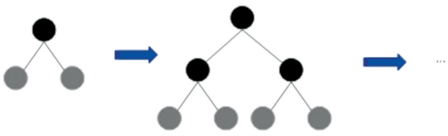


图 1 level-wise 生长策略

Fig. 1 level-wise growth strategy

为优化上述问题,提出 leaf-wise 生长策略,如图 2 所示。决策树每次从当前所有叶子中找到分裂增益最大(一般也是数据量最大)的一个叶子,然后分裂,如此循环。在分裂次数相同的情况下,和 level-wise 相比,leaf-wise 可以得到更高的精准度。但是,leaf-wise 可能会长出较深的决策树,产生过拟合。为解决该问题,LightGBM 在使用 leaf-wise 策略

时,对最大深度进行限制,在保证高效率、高精度的同时防止过拟合。

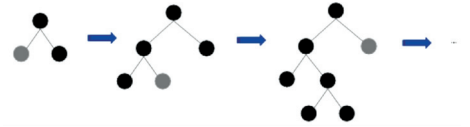


图 2 leaf-wise 生长策略

Fig. 2 leaf-wise growth strategy

使用了 GOSS 算法、Histogram 算法、EFB 算法的梯度提升树 (GBDT) 称之为 LightGBM,相较于 XGBoost 等的 GBDT 算法,LightGBM 的许多特性(如 leaf-wise 的决策树生长策略,类别特征值的最优分割策略,数据和特征的并行学习等)都使算法的性能迅速提升^[16]。

1.2 SARIMA 模型

为了验证 LightGBM 模型的预测准确率,本文采用基于 ARIMA 的季节性模型 SARIMA(p, d, q) \times (P, D, Q, s), 全称为季节性自回归移动平均模型进行对比,该模型是一种常见的时间序列分析模型^[17]。其中 AR 是“自回归”, p 为自回归项数; I 为差分, d 为使之成为平稳序列所做的差分次数(阶数); MA 为“滑动平均”, q 为滑动平均项数; 季节性序列的变化周期用 s 表示,对于月度序列 $s = 12$,对于季度序列 $s = 4$ 。

SARIMA 模型对时间序列的分析预测基本流程如图 3 所示:

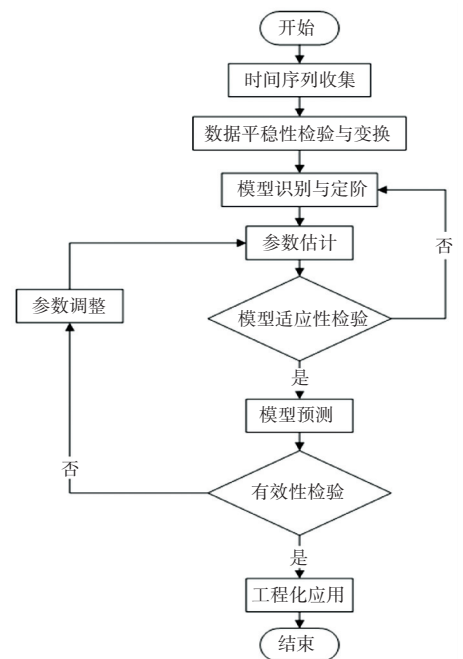


图 3 SARIMA 流程图

Fig. 3 SARIMA Flow Chart

2 实验过程与分析

本文研究的制造业原始数据集来源于美的集团,实验选取了该企业 2018~2020 年的历史数据,实验数据集共包含 1 183 594 条样本数据,其主要结构描述见表 2。

表 2 实验数据集

Tab. 2 Experimental data set

数据项	种类
过账日期(库存)	35
过账日期(需求)	1 050
工厂	16
物料	294
工厂+物料	962
物料品牌	12
物料类型	5
物料品类	24

2.1 LightGBM 模型实验

2.1.1 数据预处理

由于制造业原始物料需求数据存在类型不一,

残缺等问题,无法直接进行数据分析或建模。为提高数据挖掘的质量及预测精准度,需在数据分析之前进行数据预处理。

2.1.1.1 数据清洗

离散制造业物料原始数据大多来源于物流单据,可能存在时间空缺、错位等问题,真实且完整的数据更有利于提高模型的预测精准度。本研究通过数据清洗填充关键性空值数据,同时删除、更正错误或重复的数据。结合离散制造业实际业务,原始需求量数据部分特征不能应用于预测模型,故剔除了与预测结果无关的特征。

2.1.1.2 时间切片

数据集中含有 962 个“工厂编码_物料编码”组合,不能直接体现各个工厂物料具有的时间特征因素,因此选择以月为时间粒度,对现有数据进行时间切片。一条数据包括一个月内的物料种类、物料需求量与物料库存量等。切片后的数据从第一月开始生成对应 id,见表 3,作为数据的时间特征以及预测模型的影响因子之一参与模型优化。

表 3 经数据清洗与时间切片后样表

Tab. 3 Sample table after data cleaning and time slicing

	工厂编码_物料编码_id	工厂编码	物料编码	年	月	需求量	库存量
0	2000_1001	2000	1001	2018	1	12.53	2.71
1	2000_1001	2000	1001	2018	2	16.08	4.25
2	2000_1003	2000	1003	2018	3	72.67	15.47
3	2000_1003	2000	1003	2018	4	52.55	11.90
4	2000_1020	2000	1020	2019	5	2.00	0.44
5	2000_1020	2000	1020	2019	6	1.40	0.25
6	2001_1044	2001	1044	2019	7	20.52	4.69
7	2001_1044	2001	1044	2019	8	4.36	0.96
8	2003_1034	2003	1034	2020	9	9.72	1.47
9	2003_1034	2003	1034	2020	11	35.08	8.29

2.1.2 特征工程

特征工程是指用一系列工程化的方式,从原始数据中筛选出更好的数据特征,以提升模型的训练效果。特征工程通常包括数据预处理、特征选择、降维等环节。

表 3 所示的数据集包含了每个工厂对不同类型物料的月需求量,直接影响目标预测变量值的因素为过账日期。由于在构建模型时决定性特征数量较少,为使模型训练更准确,可解释性更强,研究利用

统计学中的统计量,构建出新的数据特征——滑动特征和滞后特征,见表 4。构建滑动窗口特征将时间序列在时间轴上划分窗口,是常用且有效的方法,其中包括滑动窗口和滚动窗口。窗口分析对平滑噪声或粗糙的数据非常有用,如移动平均法等。利用这种方式结合基础的统计方法,通过对同一特征在不同时间跨度下进行分析,得到整体数据更加一般化的变化趋势。

表 4 构建滞后和滑动特征

Tab. 4 Build hysteresis and slip characteristics

特征名称	特征含义
lag_1	window = 1 进行滞后
lag_2	window = 2 进行滞后
lag_3	window = 3 进行滞后
lag_4	window = 4 进行滞后
s_1_roll_2_min	在一阶滞后的基础上取 window = 2 窗口内的最小值
s_1_roll_2_max	在一阶滞后的基础上取 window = 2 窗口内的最大值
s_1_roll_2_median	在一阶滞后的基础上取 window = 2 窗口内的中位数
s_1_roll_2_std	在一阶滞后的基础上计算 window = 2 窗口内的标准差
s_1_roll_2_mean	在一阶滞后的基础上取 window = 2 窗口内的平均值
s_1_roll_3_min	在一阶滞后的基础上取 window = 3 窗口内的最小值
s_1_roll_3_max	在一阶滞后的基础上取 window = 3 窗口内的最大值
s_1_roll_3_median	在一阶滞后的基础上取 window = 3 窗口内的中位数
s_1_roll_3_std	在一阶滞后的基础上计算 window = 3 窗口内的标准差
s_1_roll_3_mean	在一阶滞后的基础上取 window = 3 窗口内的平均值

2.1.3 实验结果及分析

为探寻每年需求量的一般规律,本研究将每个工厂、每种物料的月需求量可视化。如图 4 所示,2018~2020 年的需求趋势在时间上并未出现显著的周期性规律;2019 年 3 月至 6 月数据呈现较大幅度的增长,同比往年数据,2019 年末的需求量也有一定幅度的增长,数据波动较大,可能是受线上大型电器促销活动和疫情等影响;2018 年和 2020 年数据增长趋势较为一致,均在 6、7 月份达到峰值,较为符合消费者对于美的电器的季节性需求规律。

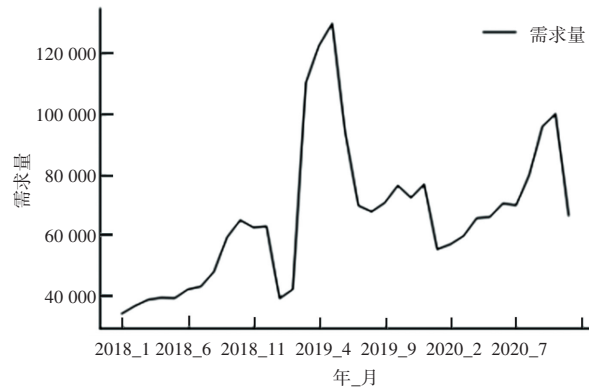


图 4 月需求量曲线(2018 年 1 月~2020 年 11 月)

Fig. 4 Monthly demand curve (January 2018 to November 2020)

基于上述分析,为预测 2020 年 12 月及 2021 年 1 月、2 月的需求量,本研究选取 2018 年 1 月至 2019 年 11 月的物料需求量数据作为训练集对模型进行拟合,将 2019 年 12 月及 2020 年 1 月、2 月的数据作为验证集,来验证模型的可行性和精准度。在此基础上对模型不断优化,最终得到拟合度较好的 LightGBM 模型。

通过预测分析,将 2019 年 12 月部分“工厂_物

料编码”组合的实际需求量(图 5)和预测需求量(图 6)可视化,并得出预测-实际对比需求量曲线(图 7)。可以看出,预测数据和实际数据基本一致,模型拟合度较好。

2.2 SARIMA 模型实验

2.2.1 数据预处理

SARIMA 模型适用于研究单变量在时间上的变化,而本研究需要解决多个不同工厂、不同物料随时间变化而引起的需求量变化问题,所以采用 LabelEncoder 方法构建类别特征见表 5。将工厂和物料编码结合起来,从而可以对任意一个 kind(类别)进行模型构建和预测。

表 5 数据预处理及类别编码

Tab. 5 Data preprocessing and category coding

特征	特征种类数
工厂编码	16
物料编码	294
年	3
月	12
需求量	11 034
工厂编码_物料编码_id	962
类别(kind)	962

2.2.2 数据平稳性检验与变换

为防止“伪回归”,选取工厂编码“kind = 30”的时间序列进行 ADF 平稳性检验,检验结果见表 6。可以看出,该工厂编码时间序列的 P 值为 0.460 481 744 342 412 24,大于 0.05,原始序列是非平稳的。进行一阶差分后的时间序列 P 值为 7.984 048 924 803 763e-05,小于 0.05,满足平稳性。因此,工厂编码“kind = 30”的时间序列是一阶单整序列,即 $ps \sim I(1)$ 。

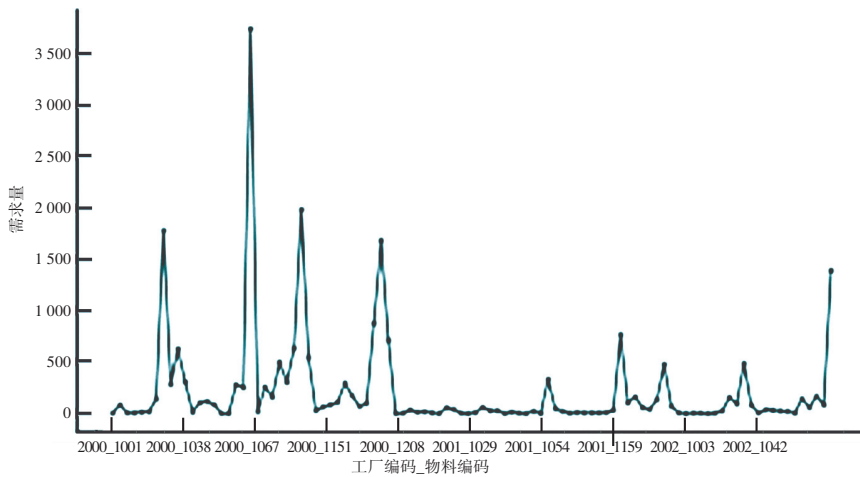


图5 实际需求量曲线

Fig. 5 Actual demand curve

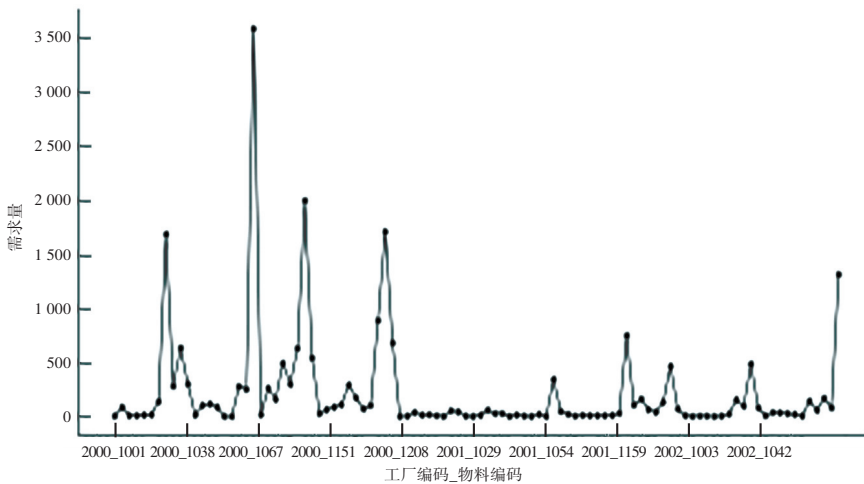


图6 预测需求量曲线

Fig. 6 Forecast demand curve

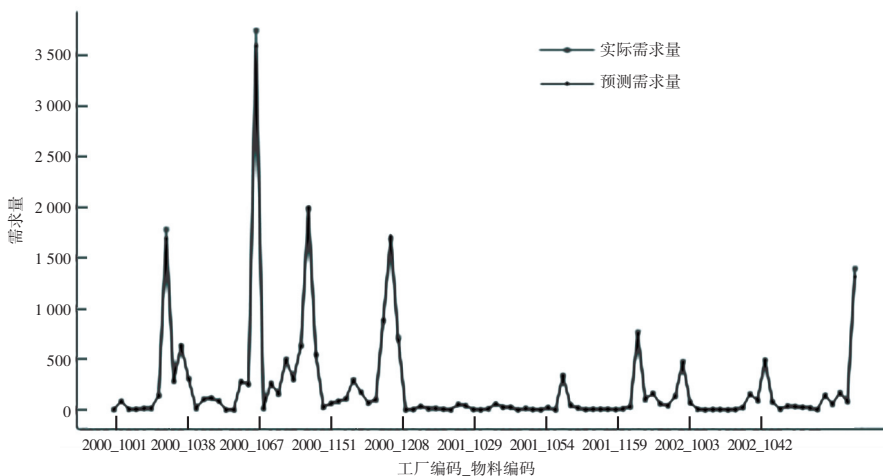


图7 预测-实际对比需求量曲线

Fig. 7 Forecast-actual demand curve

表6 序列平稳性检验

Tab. 6 Test of sequence stationarity

变量	ADF 统计量	P 值	检验结论
ps	-1.643 405 758 797 774	0.460 481 744 342 412 24	非平稳
	-4.712 147 611 368 79	7.984 048 924 803 763e-05	平稳

2.2.3 参数估计

从图 8(ACF)与图 9(PACF)中可以看出,一阶差分后的“工厂_物料”编码“kind=30”的时间序列自相关图和偏自相关图都有拖尾衰减特征,适合建立 ARMA 模型。结合表 7 的参数确认方法,确认出

$p = 1, q = 1$ 。综上分析,采用 ARIMA(1,1,1)模型进行拟合。

本实验对类别“kind = 30”的数据进行季节性分

解,得到原始数据的趋势、季节性因素和残差。图 10 显示出原始数据具有较明显的以年为周期的季节性特征,所以时间序列的周期 s 可以取 12。

表 7 参数确认方法

Tab. 7 Parameter confirmation method

模型	ACF	PACF
AR(p)	衰减趋于零(几何型或震荡型)	p 阶后截尾
MR(q)	q 阶后截尾	衰减趋于零(几何型或震荡型)
ARMA(p, q)	q 阶衰减趋于零(几何型或震荡型)	p 阶衰减趋于零(几何型或震荡型)

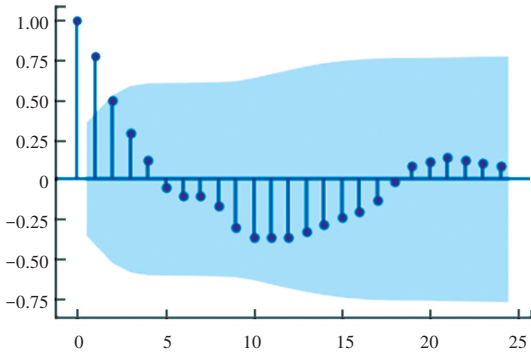


图 8 自相关图

Fig. 8 Autocorrelation graph

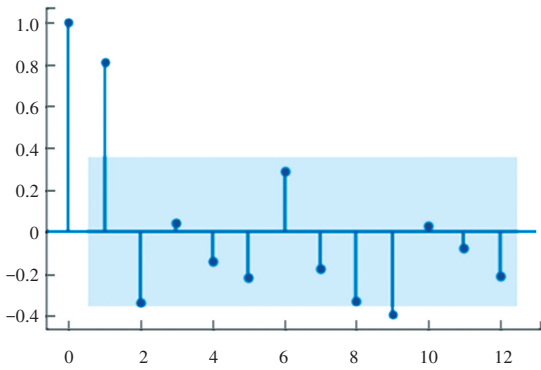


图 9 偏自相关图

Fig. 9 Partial autocorrelation graph

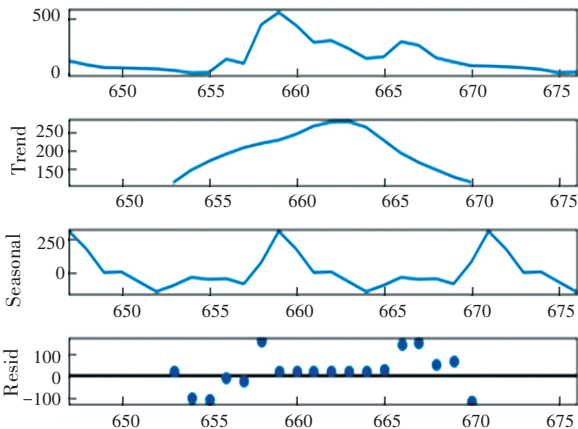


图 10 季节性分解结果图

Fig. 10 Seasonal decomposition results

对抽取出的季节性因素数据进行 ADF 平稳性

检验,得到 P 值为 0.0,说明原季节性因素数据满足平稳性,不需要进行差分,故 $D = 0$ 。

根据季节性因素的偏自相关图(图 11)和自相关图(图 12),结合网格搜索的 AIC 值,确定较优的模型参数 $P = 3, Q = 1$ 。最终得到较优的 SARIMA(1,1,1)×(3,0,1,1,2)模型。

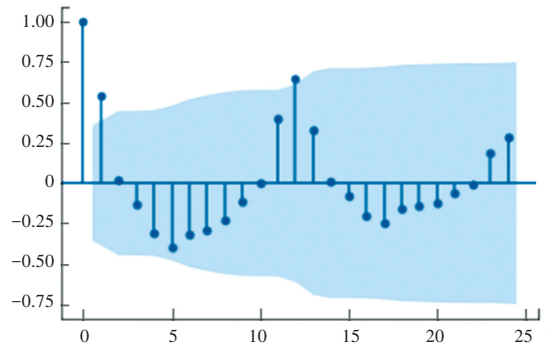


图 11 季节性因素偏自相关图

Fig. 11 Partial autocorrelation graph of seasonal factors

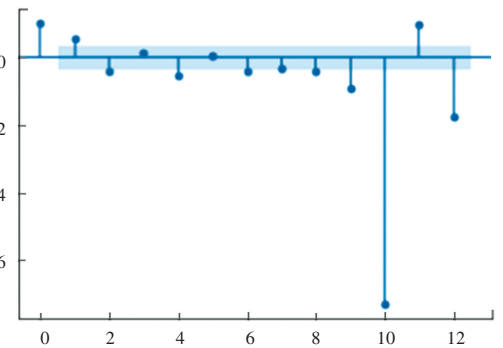


图 12 季节性因素自相关图

Fig. 12 Autocorrelation chart of seasonal factors

2.2.4 模型预测

利用上述确定的最优参数 SARIMA(1,1,1)×(3,0,1,1,2),来拟合有效模型进行动态预测。如图 13 所示,表明该模型对需求量拟合效果较好,但从长期来看,其拟合效果存在一定偏差。

2.3 模型对比分析

选取平均绝对误差 MAE 作为衡量两种模型预

测精准确度的指标,如式(1)所示。

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

其中, \hat{y}_i 指预测值; y_i 指真实值; n 为样本数量; MAE 的范围为 $[0, +\infty)$ 。当预测值与真实值完全吻合时等于 0,即完美模型,误差越大,该值越大。 MAE 的值越小,说明预测模型拥有更好的预测精度。

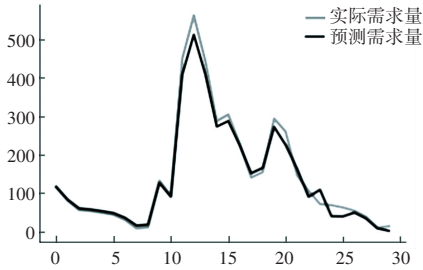


图 13 SARIMA 拟合预测-真实需求量对比图

Fig. 13 SARIMA fitting forecast-real demand comparison chart

表 8 MAE 值对比

Tab. 8 MAE value comparison

模型	MAE
LightGBM	4.427
SARIMA	14.627

由表 8 分析可知,与季节性的时间序列 SARIMA 模型相比,基于决策树的 LightGBM 机器学习模型平均绝对误差较小,拥有更高的预测精度。

3 结束语

本文以离散制造业企业美的为例,分析 2018 年至 2020 年的产品物料需求数据,利用基于决策树的 LightGBM 机器学习模型对企业物料需求量进行预测,为了研究该模型在实际应用中的优势,构建传统时间序列 SARIMA 对比模型。对比分析实验结果表明,SARIMA 需要针对每个“工厂编码_物料编码_id”,分别进行独立建模,模型构建过程复杂、耗时较长,计算量较大,且结果显示平均绝对误差 MAE 大、精度较低;基于决策树的 LightGBM 机器学习模型构建过程简单、平均绝对误差 MAE 小、预测精确度较高,在月粒度下的短期需求预测结果满足企业实际生产的准确性和实效性要求。

LightGBM 机器学习预测模型为物料需求预测提供了科学可行的解决方案,为企业智能化管理提

供了技术支持。然而,在构建特征工程时,并未考虑到影响需求量的其他外在因素(如疫情影响、节假日、线上优惠活动等),以期在后续研究中进一步分析完善。

参考文献

- [1] 赵明洲. 基于深度学习的制造企业物料需求预测研究[D]. 长春: 吉林大学, 2020.
- [2] 李洋莹, 陈智军, 张子豪, 等. 基于改进 Elman 神经网络的制糖企业原糖需求预测模型[J]. 计算机应用, 2021, 41(7): 2113-2120.
- [3] WU C L, CHAU K W. Data-driven models for monthly streamflow time series prediction[J]. Engineering Applications of Artificial Intelligence, 2010, 23(8): 1350-1367.
- [4] 张文华. ARIMA 模型用于时间序列预测的算法改进与仿真[J]. 信息与电脑(理论版), 2021, 33(5): 53-56.
- [5] Javad Feizabadi. Machine learning demand forecasting and supply chain performance, International Journal of Logistics Research and Applications, 2022, 25(2): 119-142.
- [6] 李福, 徐良杰, 朱然博, 等. 基于 XGBoost 算法的共享单车借车需求量预测[J]. 武汉理工大学学报(交通科学与工程版), 2021, 45(5): 880-884.
- [7] 李卫星, 陶建峰, 覃程锦, 等. 同步压缩小波与极限梯度提升树融合的柴油机失火故障诊断[J]. 西安交通大学学报, 2019, 53(2): 47-54, 169.
- [8] 周宇阳. 基于 SVM 的中国地区进口 LNG 需求量预测[J]. 石油化工设计, 2020, 37(1): 59-63, 66, 8-9.
- [9] NTAKOLIA C, KOKKOTIS C, KARLSSON P, et al. An explainable machine learning model for material backorder prediction in inventory management[J]. Sensors, 2021, 21(23): 7926.
- [10] 王献志, 曾四鸣, 周雪青, 等. 基于 XGBoost 联合模型的光伏发电功率预测[J]. 太阳能学报, 2022, 43(4): 236-242.
- [11] 孙斌, 储芳芳, 陈小惠. 基于贝叶斯优化 XGBoost 的无创血压预测方法[J]. 电子测量技术, 2022, 45(7): 68-74.
- [12] 石慧, 李芷萱, 彭壮壮. 基于高维多尺度核函数的模糊 SVM 剩余寿命预测[J]. 系统工程, 2022, 40(6): 148-155.
- [13] 代业明, 周琼. 基于改进 Bi-LSTM 和 XGBoost 的电力负荷组合预测方法[J]. 上海理工大学学报, 2022, 44(2): 138-147.
- [14] KE G, MENG Q, FINLEY T, et al. Lightgbm: a highly efficient gradient boosting decision tree [C]//Advances in Neural Information Processing Systems(NIPS), 2017: 3146-3154.
- [15] LIANG J, BU Y, TAN K, et al. Estimation of stellar atmospheric parameters with light gradient boosting machine algorithm and principal component analysis [J]. The Astronomical Journal, 2022, 163(4): 153.
- [16] 慕钢, 张宏烈, 党佳俊, 等. 基于 LightGBM 模型的二手房房价预测研究[J]. 高师理科学刊, 2020, 40(12): 27-31.
- [17] 咬登魁, 段功豪. 基于季节性 SARIMA 模型的武汉市长序列降雨趋势分析与预测[J]. 地下水, 2022, 44(2): 166-168.