文章编号：1003-1421(2026)08-0207-10
引用格式：赵海军，孙浩，吴林鸿，等. 基于信息融合的铁路货运场站运量预测方法[J]. 铁道运输与经济，2026，48(8)：207-216.
ZHAO Haijun，SUN Hao，WU Linhong，et al. Prediction Method for Freight Volume of Railway Freight Stations Based on Information Fusion[J]. Railway Transport and Economy，2026，48(8)：207-216.

基于信息融合的铁路货运场站运量预测方法

赵海军[1]，孙　浩[1]，吴林鸿[2]，薛　锋[2]

(1.国能朔黄铁路发展有限责任公司 物流分公司，河北 肃宁　062350；2.西南交通大学 交通运输与物流学院，四川 成都　611756)

摘　要：铁路货运场站的运量预测精准性直接影响着运输资源的合理配置与作业效率。受运营环境复杂性、多源数据异构性等多重因素影响，传统的铁路货运场站运量预测方法往往难以满足实际需求，为此提出了基于多源信息融合的货运场站运量预测框架。系统梳理了影响货运场站运量的内外部关键因素，通过信息融合策略对数据结构和特性进行处理，并采用特征筛选优化输入变量，构建了偏最小二乘回归、岭回归以及先进梯度提升树模型广泛谱系等多元预测模型，同时结合海铁联运港站进行了验证分析。结果表明，轻量级梯度提升机模型(LightGBM)拟合系数R^2高达0.959，在全局预测精度上表现最优，同时说明基于信息融合构建的货运场站运量预测方法可以有效提升模型对运营环境的适应性，为铁路货运场站的精细化管理提供数据支撑。

关键词：铁路运输；货运场站；信息融合；运量预测
中图分类号：U292.16　　**文献标识码**：A　　DOI：10.16668/j.cnki.issn.1003-1421.20251223001

Prediction Method for Freight Volume of Railway Freight Stations Based on Information Fusion

ZHAO Haijun[1]，SUN Hao[1]，WU Linhong[2]，XUE Feng[2]

(1. Logistics Branch, Guoneng Shuohuang Railway Development Co., Ltd., Suning 062350, Hebei, China; 2. School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, Sichuan, China)

Abstract: The prediction accuracy of freight volume in railway freight stations directly affects the rational allocation of transportation resources and operational efficiency. Affected by multiple factors such as the complexity of the operating environment and the heterogeneity of multi-source data, traditional prediction methods for freight volume of railway freight stations often fail to meet actual needs. Therefore, a prediction framework for freight volume of freight stations based on multi-source information fusion was proposed. The internal and external key factors affecting the freight volume of freight stations were systematically sorted out. The data structures and characteristics were processed through information fusion strategies, and feature screening was adopted to optimize input variables. Multiple prediction models, including partial least squares regression, ridge regression, and a broad spectrum of advanced gradient boosting tree models, were constructed. Meanwhile, verification and analysis were conducted in combination with sea-rail intermodal port

收稿日期：2025-12-23　　**修订日期**：2026-03-04
通信作者：薛锋(1981—)，男，山东邹城人，西南交通大学交通运输与物流学院副教授。
基金项目：国能朔黄铁路发展有限责任公司科技创新项目(SHTL-22-11)

stations. The results show that the lightweight gradient boosting machine (LightGBM) model has a fitting coefficient R^2 of up to 0.959, performing the best in global prediction accuracy. At the same time, it indicates that the prediction method for freight volume of freight stations constructed based on information fusion can effectively improve the adaptability of the model to the operating environment, providing data support for the refined management of railway freight stations.
Keywords: Railway Transportation; Freight Station; Information Fusion; Freight Volume Prediction

0　引言

准确的货运场站运量预测是优化铁路运输资源配置、提升物流效率的前提和关键环节。随着铁路货运需求的持续增长，提升货运场站运量预测的精度对于提高运输计划兑现率、降低运营成本、保障运输秩序的稳定性具有重要意义。货运场站的运量受多方面因素影响，如场站内部作业效率、外部经济环境、国家政策调控等，这些因素相互交织，使得传统预测方法难以全面捕捉其复杂性和关联性。因此，构建一种能够融合多源数据，且能适应需求动态变化的运量预测模型，对提升铁路货运场站管理的智能化水平具有重要作用。

近年来，铁路货运量预测研究在建模、方法创新和实践方面都取得了比较多的成果，整体研究方向从传统统计方法向XGBoost、深度学习等机器学习技术演进。Khan等[1]系统评估了经济整合模型等传统时间序列方法在巴基斯坦铁路货运预测中的表现，研究发现经济活动及运费费率是其国家铁路货运需求的关键因素；Patil等[2]提出了一种结合加性回归和时间序列技术的预测模型，将孟买港进出港货运量的预测误差减小到了5%。国内学者在铁路货运量预测方面也取得了诸多进展，通过理论分析、数值计算、综合研究等建立了不同环境下的预测模型[3-4]，包括基于经济周期改进的动态网络预测模型[5]、加入混沌特性识别的时间序列分析模型[6]、基于数据挖掘的时序模型[7]、结合误差修正和权重调整的深度学习模型[8-9]等。

随着铁路信息化的快速发展，信息融合技术逐渐应用到了货运量预测领域。史樱[10]通过大数据和互联网平台在物流业的实际应用中发现了现代信息技术在铁路货运中的重要作用；郭颖霞[11]结合数据库，运用CELL报表以及J2EE技术，设计了铁路货运信息管理系统；王静[12]则在设计各个铁路系统数据仓库的基础上建立决策系统，提升了数据利用效率。

尽管既有研究在建模、算法优化方面取得一定突破，但针对铁路货运场站特定场景的研究，缺乏考虑运营数据与外部经济因素的交互影响。针对上述问题，提出一种融合多源数据的运量预测框架，通过信息融合策略将关键因素进行量化，提取有用的数据特征。首先构建货运场站的多维特征层，包括作业动态、经济指标、环境因素等，利用信息融合策略对部分特征层进行处理，克服数据异构问题，并在此方法上选择偏最小二乘回归、岭回归以及先进梯度提升树模型的广泛谱系进行验证，从传统线性模型到机器学习算法展开对比讨论。

1　基于信息融合策略的影响因素分析

1.1　多源信息融合策略

影响铁路货运场站运量的因素众多，既包括场站内部的装卸作业效率、设备运行状态和货类结构特征等运营参数，也涉及外部的经济环境、政策调控和市场供需等宏观条件[13]。这些因素通过复杂的非线性机制相互作用，共同决定了货运场站的作业能力和效率水平[14]。为准确预测货运场站运量，拟建立多源信息融合的分析框架，系统考察各因素的综合影响效应。

1.1.1　不同数据结构融合

在多源信息融合过程中通常包含多种类型的数据。一是结构化数据，如运量数据、设备状态分类，这类数据可以直接用表格处理；二是文本类数据，如政策文件、调度日志，需要通过自然语言处理提取关键词(NLP处理)；三是图像视频数据，需要计算机视觉技术识别物体和动作(CV处理)，此类数据提取较为困难。不同数据结构融合流程图如图1所示。

1.1.2　不同时空特性融合

除要对不同数据结构进行融合处理外，还需处理多时空尺度的数据特征。不同时空特征融合流程图如图2所示。在时间维度上，需分析不同尺度数据，如经济指标月数据、运量统计日数据等，可以

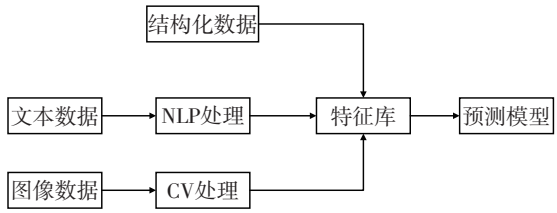

图 1　不同数据结构融合流程图

Fig.1　Flowchart of fusion process for different data structures

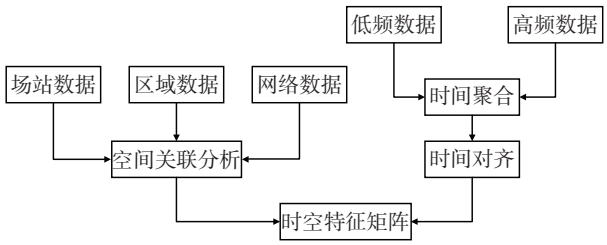

图 2　不同时空特征融合流程图

Fig.2　Flowchart of fusion process for different spatio-temporal features

利用固定窗口聚合的形式实现时间序列聚合，通过对数据定义一个固定大小的窗口，并在窗口内进行统计计算来实现；在空间维度上，需整合场站运营数据、区域宏观经济指标以及物流网络节点状态数据，构建包含空间相关性和距离衰减效应的特征层。

1.1.3　信息融合方式

（1）数据层融合。数据层融合直接对原始数据进行合并处理，适用于同构数据源的实时监测场景，其优势在于理论上能保留最完整的信息，但由于不同数据源的采样频率和维度差异，在实际应用中存在一定的限制，且对计算资源要求较高，在较为复杂的系统中采用较少。

（2）特征层融合。特征层融合是比较常用的方法，主要通过分别提取各类数据源的特征后拼接形成高维特征向量。这种方法能有效保留各数据源的关键信息，并允许模型学习跨模态的交互关系，但在应用中需要专业的特征工程处理，经常面临维度问题。

（3）决策层融合。决策层融合通过训练多个独立子模型后组合预测结果，对异构数据的兼容性最佳，该方法允许为不同数据类型选择最优模型。该方法计算成本较高，且有可能丢失底层特征相关性，但是当数据源差异比较显著时，这种融合方式往往有更好的鲁棒性。

1.2　铁路货运场站运量影响因素分析

1.2.1　内部运营因素

铁路货运场站的内部运营因素主要包括货类结构、装卸作业和设备运行三大类，这些因素通过相互作用直接影响货运场站运量[15]。在货类结构中，煤炭与矿石的占比决定了基础运输规模，同时煤炭品质、矿石特性等会影响装卸工艺选择和效率；在装卸作业方面，装卸速度、车辆停留时间是关键指标，前者反映单位时间的货物处理能力，后者体现站点的运作效率，二者共同决定装卸效率，在一定程度上反映货运周转速度；设备运行状态则通过故障率等指标影响作业连续性，特别是一些关键设备的可靠性直接影响装卸能力。这些因素相互关联，共同影响铁路货运场站的运量变化。综上，铁路货运场站的内部运营因素可以作如下量化。

（1）货类结构影响模型。

$$Q_{\text{base}} = \alpha \cdot C + \beta \cdot M \tag{1}$$

式中：Q_{base}为特定基准条件下，由煤炭和矿石2类核心货物的占比所共同决定的基准运量，万t；C为煤炭占比，%；M为矿石占比，%；α和β为工作经验系数。

（2）装卸效率指数η。

$$\eta = \frac{Q_a}{Q_t} = \frac{\sum_{i=1}^{n} P_i \cdot T_i}{24 \cdot N} \tag{2}$$

式中：P_i为单机作业能力，t/h；T_i为有效作业时间，h；N为作业天数，d。

（3）设备可靠性模型。

$$\lambda(t) = \lambda_0 \cdot e^{0.002\,3t} \tag{3}$$

式中：$\lambda(t)$为设备运行时间t下的故障率，次/h；λ_0为设备初始故障率，次/h；t为运行时间，h；0.002 3为货运场站设备故障率的指数增长率。

1.2.2　外部关联因素

外部环境对铁路货运场站运量的影响呈现多层次动态特征。从需求侧来说，区域经济发展水平[16]与铁路货运量之间存在显著的正相关关系，这不仅反映了区域经济增长与交通基础设施建设的相互促进作用，也体现了国家产业布局与运输网络规划的紧密联系[17]。特别是在能源资源富集区域，区域经济结构对特定大宗商品运输需求的塑造作用表现得尤为突出。政策调控措施对货运结构的影响则呈现鲜明的结构性特征，如“公转铁”政策[18]通过行政引导与市场力量的互动，推动了特定区域和货类运输方式的结构性转移。这种政策效应在重点区域表

现出梯度扩散特征，由核心区域向周边辐射，带动整体运输网络的优化调整；自然条件因素则显示了运输系统固有的脆弱性，极端天气是导致运量波动的重要因素，其影响程度说明了系统在应对突发事件时弹性恢复不足，而且这种影响还具有季节性差异，不同季节的极端天气对运输网络的干扰存在明显的不对称性。

上述外部因素并非孤立作用于铁路货运系统，它们与系统内部运营因素之间存在复杂的交互与耦合。为了更加全面地分析铁路货运场站运量受到的诸多影响，综合选择区域生产总值、煤炭价格指数[19]、工业用电量、极端天气天数、政策影响系数[20]作为外部关联影响因素。同时，利用信息融合策略对部分因素进行特征层融合。

（1）工业用电量传导模型。

$$Q_t = \alpha + \sum_{k=3}^{6} \beta_k E_{t-k} + \gamma \boldsymbol{X}_t + \varepsilon_t \tag{4}$$

式中：α为货运量基准水平；Q_t为t时期货运场站的运量；β_k为运量增加系数；E_{t-k}为某一段时间内的工业用电量；$\boldsymbol{\gamma}$为控制变量的系数向量；$\boldsymbol{X}_t$为控制变量矩阵(矩阵元素为货运场站运量的其他影响因素)；ε_t为随机误差项；k为滞后阶数，基于国际能源署对中国电煤消费变化的数据分析，此处定义$k \in [3, 6]$。

（2）政策影响系数。

$$P_t = \sum_{i=1}^{n} w_i \cdot I_{i,t} \cdot \mathrm{e}^{-\lambda(t-t_i)} \tag{5}$$

式中：$I_{i,t} \in (0, 1]$，为第i项政策在t时期的实施强度；w_i为政策权重；$\lambda = 0.2$，为计算后所取的政策效应衰减率平衡值。

1.3　关键特征筛选

通过多源信息融合方法整合得到铁路货运场站的内部运营因素与外部环境因素，关键特征表如表1所示。

表1　关键特征表

Tab.1　Key features

自变量	因变量	对应关键特征
铁路货运场站运量	X_1	煤炭占比
	X_2	矿石占比
	X_3	装卸效率指数
	X_4	设备故障率
	X_5	区域生产总值
	X_6	煤炭价格指数(2009=100，以2009年数据为基准换算)
	X_7	工业用电量
	X_8	极端天气天数
	X_9	政策影响系数

2　铁路货运场站运量预测方法

考虑到货运量数据兼具时序相关性和多因素耦合特征，现选取：用于高维数据降维的偏最小二乘回归模型，擅于处理特征共线性问题的岭回归以及能捕捉复杂非线性关系的先进梯度提升树模型的广泛谱系，从不同角度量化各因素对运量的贡献度。

2.1　偏最小二乘回归模型(PLSR)

偏最小二乘回归[21]采用降维思想，通过提取与因变量协方差最大的潜变量来建立预测模型。这种方法适用于样本量少而特征多的场景，能够有效处理高度相关的特征。PLSR通过迭代提取成分，每个成分都最大程度地解释因变量的变异，但模型的可解释性相对较差。虽然其计算过程相对复杂，但能很好地解决传统回归方法在小样本高维数据上的局限性，其核心计算步骤[22]为

$$w = \frac{\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}}{\left\| \boldsymbol{X}^{\mathrm{T}}\boldsymbol{y} \right\|} \qquad \boldsymbol{t} = \boldsymbol{X}\boldsymbol{w} \tag{6}$$

式中：$\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y}$为特征矩阵与目标变量的协方差向量；$\| \ \|$为欧几里得范数；$\boldsymbol{t}$为投影得分向量；$\boldsymbol{w}$为投影方向向量。

2.2　岭回归模型(Ridge)

岭回归[23]常作为针对线性回归在共线性数据上表现不佳的改进方案，通过在目标函数中引入L2正则化项来约束系数大小。这种方法通过牺牲一定的无偏性来换取更高的稳定性，使得在特征相关性较强时仍能获得可靠的系数估计。模型性能受正则化参数λ影响较大，过大的λ会导致模型欠拟合，而过小的λ则无法有效解决共线性问题。其优化目标[24]及拟合公式为

$$\begin{cases} y_i = \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i \\ \min\limits_{\boldsymbol{\beta}} \left\{ \sum\limits_{i=1}^{n} (y - \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{X}_i)^2 + \lambda \sum\limits_{j=1}^{p} \beta_j^2 \right\} \end{cases} \tag{7}$$

式中：$\boldsymbol{X}_i$表示第i个样本的特征行向量；$\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_p]^{\mathrm{T}}$，代表回归系数向量；$\lambda$为控制系数收缩程度的正则化强度参数；$\sum\limits_{j=1}^{p} \beta_j^2$为L2正则化项。

2.3　先进梯度提升树模型的广泛谱系

梯度提升树模型作为一种集成学习算法，常通过迭代的方式将多个预测结果进行加权结合，同时通过拟合损失函数的负梯度来减少上一轮模型的偏差，提升整体的预测性能。其中经典梯度提升树模型(GBDT)的基础原理公式如下[25]。

$$\min \sum_{i=1}^{n} L(y_i,\ F_{t-1}(x_i)+f_t(x_i)) \tag{8}$$

$$r_{ti}=-\left[\frac{\partial L(y_i,\ F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{t-1}(x)} \tag{9}$$

式中：F_{t-1}是前$t-1$棵树的集成模型；f_t为当前树的输出模型；r_{ti}为样本i的负梯度公式。

XGBoost[26]则是GBDT的高效扩展，其优化目标建立在结构风险最小化原则上，通过二阶泰勒展开优化损失函数，同时引入正则化项数控制模型的复杂度，具体公式如下。

$$Obj^{(t)}=\sum_{i=1}^{n} L(y_i,\ \hat{y}_i^{(t)})+\sum_{k=1}^{K}\Omega(f_k) \tag{10}$$

$$Obj^{(t)}\approx\sum_{i=1}^{n}\left[g_i f_t(x_i)+\frac{1}{2}h_i f_t^2(x_i)\right]+\Omega(f_k) \tag{11}$$

式中：$L(\cdot)$为损失函数；$\Omega(f_k)=\gamma H+\frac{1}{2}\theta\|\phi\|^2$为针对第$k$棵树的复杂度惩罚项($H$为叶子数，$\phi$为叶子权重，$\theta$为正则化系数)；$g_i$和$h_i$分别为一阶梯度和二阶梯度。

LightGBM[27]则遵循GBDT的加法模型与前向分步算法，为更高效地找出最优分裂点，引入直方图算法，其分裂增益公式如下。

$$Gain=\frac{S_L^2}{n_L}+\frac{S_R^2}{n_R}-\frac{S_P^2}{n_P} \tag{12}$$

式中：$Gain$为分裂增益，即一次节点分裂中损失函数的减少量；S_L为一阶梯度之和；n_L为样本数；S_P为父节点梯度总和；n_P为父节点样本数。

3　铁路货运场站运量预测分析

以某铁路海铁联运站2010—2024年货运月度数据为基础，进行验证分析，货运量及关键特征数据(部分)如表2所示。

3.1　货运场站运量预测分析

在完成数据的融合处理后，为量化各因素与运量之间的关联程度，采用Spearman相关性分析法对关键特征进行筛选，为后续预测模型提供特征选择依据，关键特征相关性分析如图3所示。

根据相关性分析结果可知，货运场站运量与区域生产总值、装卸效率指数、工业用电量和政策影响系数呈显著正相关，表明经济增长、物流效率提升、能源需求增加和政策支持是推动货运量增长的关键因素。同时，货运量与煤炭占比的相关性系数为0.943，与矿石占比相关性系数为−0.940，说明煤炭运输仍是主要货类，而矿石运输可能受替代效应影响；设备故障率与货运量呈显著负相关，凸显设备稳定性对运输效率的重要性；极端天气影响较弱，表明极端天气对能源市场会产生一定扰动，但实际相关性不大。

3.1.1　偏最小二乘回归模型

分析偏最小二乘回归模型的累积投射重要性(VIP)，累积投射重要性如图4所示，显示各变量对模型的解释贡献程度，其中VIP值大于1的变量通常被视为对预测目标具有显著重要性。区域生产总值和工业用电量在因子1至因子5中VIP值均高于1.08，表明两者对模型的解释力最强，可能是核心驱动因素；装卸效率指数、设备故障率及煤炭与矿石占比的VIP值始终高于1.0，说明这些变量对预测具有稳定贡献。极端天气天数的VIP值显著低于1，说明其对模型的边际贡献度最低，可能需结合其他分析验证其必要性。各变量VIP值在因子间变化幅度有限，反映模型结构稳定性比较高，但区域生产总值与极端天气天数的动态差异显示二者对潜在潜变量的敏感性不同。成分矩阵表如表3所示，研究变量可归纳为4个核心驱动因素：因子1代表经济规模与能源需求，由区域生产总值、工业用电量等正向驱动；因子2和因子3共同反映了政策与价格对经济活动的调控与抑制效应；因子4关联能源成本；因子5代表物流运输维度。上述分析同时揭示了变量间存在协同与拮抗作用，极端天气的影响具有多面性。

3.1.2　岭回归模型

岭迹图作为岭回归分析中的核心诊断工具，其作用在于通过可视化不同正则化强度下回归系数的变化趋势，帮助模型确定最优的正则化参数。岭迹图如图5所示，模型根据方差扩大因子法最终确定λ=0.097。

岭回归分析结果如表4所示，模型F统计量为201.554(P<0.01)，表明模型具有显著解释力。在λ=

表 2　货运量及关键特征数据(部分)

Tab.2　Data of freight volume and key features (partial)

月度	货运量/万 t	煤炭占比/%	矿石占比/%	装卸效率指数(基数=100)	设备故障率(单位为 1)	区域生产总值/万元	煤炭价格指数(基数=100)	工业用电量/(万 kW·h)	极端天气天数/d	政策影响系数(单位为 1)
2022.1	587.8	80.22	13.75	126.5	0.78	25 850	159.0	25.1	0	0.961
2022.2	578.4	80.18	13.84	127.1	0.81	25 710	158.7	25.0	2	0.962
2022.3	673.7	80.65	13.42	128.8	0.64	27 330	160.8	26.1	1	0.963
2022.4	689.1	80.58	13.51	129.4	0.61	27 550	161.1	26.4	0	0.964
2022.5	738.8	80.95	13.12	131.0	0.54	28 270	162.0	27.2	0	0.965
2022.6	758.0	80.88	13.21	131.6	0.51	28 490	162.4	27.5	0	0.966
2022.7	795.6	81.32	12.75	133.2	0.44	29 050	163.4	28.3	1	0.967
2022.8	814.7	81.25	12.84	133.8	0.41	29 270	163.8	28.6	0	0.968
2022.9	768.8	80.82	13.25	131.2	0.54	28 210	162.7	27.0	0	0.969
2022.10	731.9	80.75	13.34	130.6	0.57	27 990	162.3	26.7	0	0.970
2022.11	665.7	80.38	13.73	128.0	0.74	27 240	161.4	25.9	1	0.971
2022.12	626.1	80.32	13.82	127.4	0.77	27 020	161.1	25.6	0	0.972
2023.1	637.8	81.35	12.62	130.2	0.48	27 680	160.0	26.8	0	0.961
2023.2	627.5	81.31	12.71	130.8	0.51	27 540	159.7	26.7	0	0.962
2023.3	730.8	81.78	12.29	132.5	0.34	29 160	161.8	27.8	0	0.963
2023.4	747.5	81.71	12.38	133.1	0.31	29 380	162.1	28.1	0	0.964
2023.5	801.3	82.08	11.99	134.7	0.24	30 100	163.0	28.9	0	0.965
2023.6	822.1	82.01	12.08	135.3	0.21	30 320	163.4	29.2	0	0.966
2023.7	863.0	82.45	11.62	136.9	0.14	30 880	164.4	30.0	1	0.967
2023.8	883.8	82.38	11.71	137.5	0.11	31 100	164.8	30.3	1	0.968
2023.9	834.0	81.95	12.12	134.9	0.24	30 040	163.7	28.7	0	0.969
2023.10	794.1	81.88	12.21	134.3	0.27	29 820	163.3	28.4	0	0.970
2023.11	722.3	81.51	12.60	131.7	0.44	29 070	162.4	27.6	0	0.971
2023.12	679.4	81.45	12.69	131.1	0.47	28 850	162.1	27.3	0	0.972
2024.1	691.5	82.48	11.49	133.9	0.18	29 510	157.0	28.5	1	0.851
2024.2	680.4	82.44	11.58	134.5	0.21	29 370	156.7	28.4	0	0.852
2024.3	792.5	82.91	11.16	136.2	0.04	30 990	158.8	29.5	0	0.853
2024.4	810.7	82.84	11.25	136.8	0.01	31 210	159.1	29.8	0	0.854
2024.5	869.2	83.21	10.86	138.4	0.06	31 930	160.0	30.6	0	0.855
2024.6	891.7	83.14	10.95	139.0	0.09	32 150	160.4	30.9	0	0.856
2024.7	935.9	83.58	10.49	140.6	0.06	32 710	161.4	31.7	0	0.857
2024.8	958.3	83.51	10.58	141.2	0.09	32 930	161.8	32.0	1	0.858
2024.9	904.5	83.08	10.99	138.6	0.06	31 870	160.7	30.4	0	0.859
2024.10	861.2	83.01	11.08	138.0	0.03	31 650	160.3	30.1	0	0.860
2024.11	783.4	82.64	11.47	135.4	0.04	30 900	159.4	29.3	0	0.861
2024.12	736.8	82.58	11.56	134.8	0.07	30 680	159.1	29.0	1	0.862

	货运量	煤炭占比	矿石占比	装卸效率指数	设备故障率	区域生产总值	煤炭价格指数	工业用电量	极端天气天数	政策影响系数
货运量	1	0.943	−0.94	0.966	−0.955	0.977	0.879	0.978	0.024	0.91
煤炭占比	0.943	1	−1	0.976	−0.998	0.989	0.921	0.986	−0.003	0.989
矿石占比	−0.94	−1	1	−0.975	0.998	−0.988	−0.92	−0.985	0.006	−0.988
装卸效率指数	0.966	0.976	−0.975	1	−0.981	0.982	0.903	0.982	0.015	0.96
设备故障率	−0.955	−0.998	0.998	−0.981	1	−0.993	−0.923	−0.991	−0.002	−0.984
区域生产总值	0.977	0.989	−0.988	0.982	−0.993	1	0.915	0.998	0.013	0.969
煤炭价格指数	0.879	0.921	−0.92	0.903	−0.923	0.915	1	0.911	−0.003	0.914
工业用电量	0.978	0.986	−0.985	0.982	−0.991	0.998	0.911	1	0.024	0.963
极端天气天数	0.024	−0.003	0.006	0.015	−0.002	0.013	0.003	0.024	1	−0.025
政策影响系数	0.91	0.989	−0.988	0.96	−0.984	0.969	0.914	0.963	−0.025	1

图 3　关键特征相关性分析

Fig.3　Correlation analysis of key features

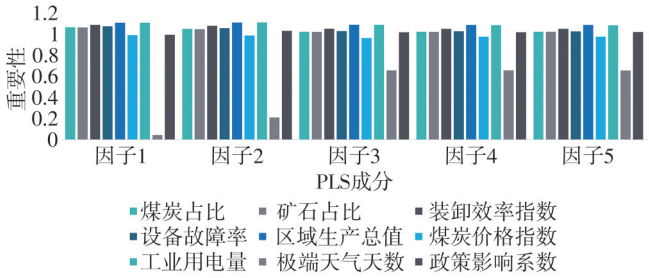

图 4　累积投射重要性

Fig.4　Cumulative projection importance

表 3　成分矩阵表

Tab.3　Component matrix

变量	因子1	因子2	因子3	因子4	因子5
X_1	0.354	0.108	0.213	−0.302	−0.496
X_2	−0.354	−0.122	−0.239	0.338	0.620
X_3	0.362	−0.275	−0.490	0.044	0.139
X_4	−0.358	0.002	0.002	0.114	−0.167
X_5	0.369	−0.400	−0.746	0.244	0.409
X_6	0.330	0.289	0.467	0.793	−0.014
X_7	0.369	−0.408	−0.732	0.106	−0.033
X_8	0.014	−0.380	0.390	−0.002	0.003
X_9	0.331	0.586	1.078	−0.397	0.461
货运量	0.423	−0.422	−0.497	0.264	0.756

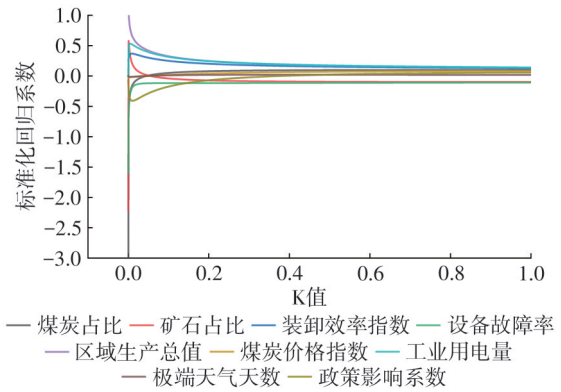

图 5　岭迹图

Fig.5　Ridge trace plot

表 4　岭回归分析结果

Tab.4　Results of ridge regression analysis

	非标准化系数		标准化系数	P
	B	标准误差	β	
常数	−53.068	22.178	—	0.019**
X_1	1.469	0.351	0.049	0.000***
X_2	−1.191	0.408	−0.039	0.004***
X_3	2.285	0.278	0.252	0.000***
X_4	−12.571	1.724	−0.118	0.000***
X_5	0.680	0.044	0.329	0.000***
X_6	0.088	0.187	0.016	0.637
X_7	6.540	0.403	0.319	0.000***
X_8	1.385	2.705	0.011	0.610
X_9	−122.148	20.047	−0.180	0.000***

注：***，**，*分别代表1%，5%，10%的显著性水平。

0.097的正则化参数下，各变量的非标准化系数及其显著性水平说明：设备故障率(β=−0.118)和政策影响系数(β=−0.180)，对货运量产生显著负向影响且边际效应较强，反映运营效率损失和政策调控对货运增长存在一定的抑制作用。区域生产总值(β=0.329)和工业用电量(β=0.319)则均拥有较大标准化系数，呈现显著正向关联。煤炭和矿石占比虽通过显著性检验但标准化系数较低，表明其实际影响有限。煤炭价格指数和极端天气天数未达显著水平，表明短期价格波动和气候因素对货运量影响有限。模型通过岭回归有效缓解了潜在多重共线性问题，保留变量并未出现方差膨胀因子异常现象。

3.1.3　机器学习模型

基于梯度提升树模型的原理，本研究选择经典的GBDT，引入正则化与二阶导数的XGBoost以及针对较大规模数据深度优化的LightGBM这3种模型展开分析。其中，GBDT模型擅长利用梯度提升与残差拟合处理中小规模数据，XGBoost模型则在正则化与泛化性能上更具优势，而LightGBM模型侧重于训练效率与大规模数据的处理。为保证分析的严谨性，统一确定模型训练集与测试集的划分比例为7∶3，并通过数据洗牌的方式提高模型的综合学习性能，降低模型过拟合的风险。机器学习模型的特征重要性如表5所示，从3种模型的特征重要性结果比对可以看出，不同机器学习模型在训练过程中各特征对目标变量的解释力存在差异。在GBDT模型中装卸效率指数和区域生产总值构成核心驱动因素，两者累计贡献率达83.8%；XGBoost模型则主要依赖于经济活动和物流效率指标，工业用电量、装卸效率指数和区域生产总值是关键驱动因

表 5　机器学习模型的特征重要性　%

Tab.5　Feature importance of machine learning models

特征名称	特征重要性		
	GBTD	XGBoost	LightGBM
X_1	0.20	0.50	4.90
X_2	0.20	0.20	4.20
X_3	42.80	29.90	24.90
X_4	0.20	0.10	1.50
X_5	41.00	28.80	21.20
X_6	0.20	0.30	7.70
X_7	4.00	38.10	17.10
X_8	0.40	0.50	5.00
X_9	1.10	1.60	15.10

素，贡献率之和高达96.8%；相较于前2种模型，LightGBM模型中各因素的特征分布率稍均衡且贡献率较高的指标集中在经济活动、物流效率以及政策影响类指标，累计贡献率为78.3%。

3.2　预测结果分析

基于各模型回归结果的分析，选取预测结果中的30个样本对5种模型的预测性能展开进一步的对比讨论，同时对3种梯度提升树模型加入数据洗牌这一预处理方式，以提高模型的泛化能力。偏最小二乘模型拟合结果如图6所示，岭回归模型拟合结果如图7所示，GBDT模型拟合结果如图8所示，XGBoost模型拟合结果如图9所示，LightGBM模型拟合结果如图10所示，模型拟合结果比对如表6所示。

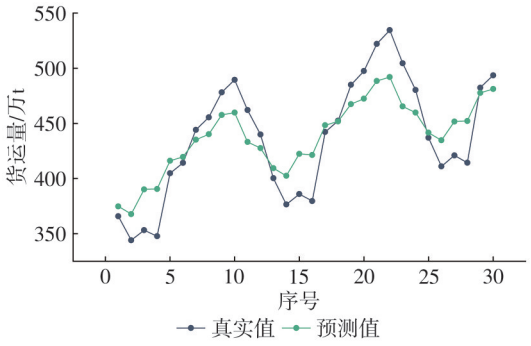

图6　偏最小二乘模型拟合结果

Fig.6　Fitting results of partial least squares model

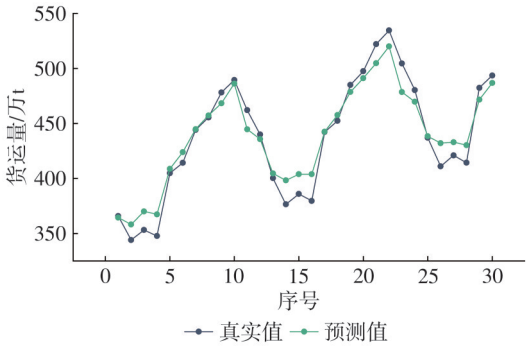

图7　岭回归模型拟合结果

Fig.7　Fitting results of ridge regression model

结合表6中的数据，对5种模型的拟合效果与预测性能进行综合分析：传统线性模型中岭回归模型拥有更优的多重共线性问题解决能力，预测曲线虽能跟随真实值的基本趋势，但序列波动细节与极值点处拟合相对平滑，易产生偏差。与此同时，梯度提升树模型更能在波动剧烈区域保持较高的预测精度，体现了其捕捉复杂的非线性关系的能力。从

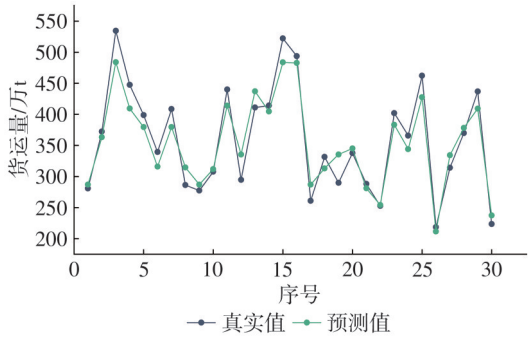

图8　GBDT模型拟合结果

Fig.8　Fitting results of GBDT model

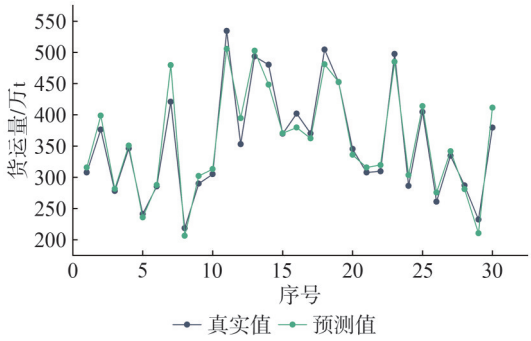

图9　XGBoost模型拟合结果

Fig.9　Fitting results of XGBoost model

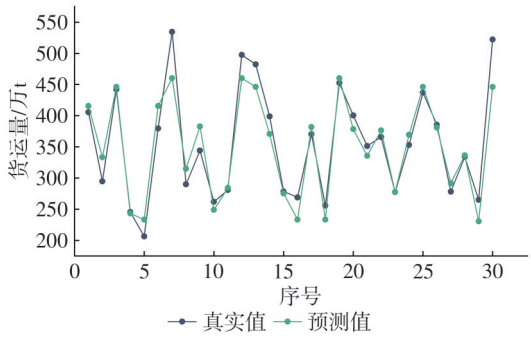

图10　LightGBM模型拟合结果

Fig.10　Fitting results of LightGBM model

表6　模型拟合结果比对

Tab.6　Comparison of model fitting results

模型	PLSR	Ridge	GBDT		XGBoost		LightGBM	
			训练集	测试集	训练集	测试集	训练集	测试集
R^2	0.917	0.948	0.999	0.916	0.969	0.930	0.955	0.959
MSE	647.923	394.333	842.131		606.450		174.524	
RMSE	25.454	19.858	29.019		24.626		13.211	

拟合系数上看，LightGBM拟合效果最优，R^2达到0.959，且在提升精度的同时有效控制了过拟合。综上，本研究构建的多源信息融合预测框架与梯度提升树模型相结合，能够更准确、稳健地实现铁路货运场站的运量预测，为运营决策提供了可靠依据。

4　结论

基于多源信息融合的思想，构建了涵盖运营效率、外部关联等多维因素的运量预测模型，并验证分析了不同模型的预测性能。

（1）综合考虑货运场站内部运营数据和外部关联因素，提出了多源信息融合的运量预测框架，提高了预测的全面性和适应性。

（2）对比分析了传统回归模型以及机器学习算法的预测效果，验证了信息融合策略在铁路货运运量预测中的可运用性。结果表明，LightGBM模型在全局预测精度上表现最优(R^2=0.959)，能够较好地拟合铁路货运场站的运量变化趋势。

（3）研究发现，在使用信息融合法对铁路货运场站运量进行预测研究的前提下，岭回归模型能够有效解决模型在分析过程中极易产生的多重共线性问题，LightGBM模型能够更全面地撷取关键因素的数据特征。基于此，未来可以此模型展开运量监测的优化，为场站精准管控运营效率提供可靠依据。

虽然通过研究发现基于多源信息融合思想构建的LightGBM预测模型能够较好地拟合铁路货运场站的运量变化趋势，但在个别样本上仍存在一定的偏差，且目前所用数据时间维度较大。未来可考虑引入动态权重调整或结合非线性模型以进一步提升模型的鲁棒性和适应能力。

参考文献：

[1] KHAN M Z，KHAN F N. Estimating the Demand for Rail Freight Transport in Pakistan：A Time Series Analysis[J]. Journal of Rail Transport Planning & Management，2020，14：100176.

[2] PATIL G R，SAHU P K. Estimation of Freight Demand at Mumbai Port Using Regression and Time Series Models[J]. KSCE Journal of Civil Engineering，2016，20(5)：2022-2032.

[3] 孟建军，陈鹏芳，李德仓，等. 铁路货运量预测研究综述[J]. 铁道标准设计，2022，66(10)：18-26.
MENG Jianjun，CHEN Pengfang，LI Decang，et al. Summary of Research on Railway Freight Volume Forecast[J]. Railway Standard Design，2022，66(10)：18-26.

[4] 江路路，高天泽. 基于组合模型的黄大铁路货运量预测研究[J]. 铁路物流，2025，43(9)：58-64.
JIANG Lulu，GAO Tianze. Research on Freight Volume Forecasting of Huanghua-Dajiawa Railway Based on a Combined Model[J]. Railway Logistics，2025，43(9)：58-64.

[5] 郭玉华，陈治亚，冯芬玲，等. 基于经济周期的铁路货运量神经网络预测研究[J]. 铁道学报，2010，32(5)：1-6.
GUO Yuhua，CHEN Zhiya，FENG Fenling，et al. Railway Freight Volume Forecasting of Neural Network Based on Economic Cycles[J]. Journal of the China Railway Society，2010，32(5)：1-6.

[6] 吴华稳. 混沌时间序列分析及在铁路货运量预测中的应用研究[D]. 北京：中国铁道科学研究院，2014.

[7] 张　璇. 基于数据挖掘的铁路货运量预测应用研究[D]. 成都：西南交通大学，2016.

[8] 蒋泽华. 基于乘积季节模型和LSTM神经网络的铁路货运量预测研究[D]. 南昌：华东交通大学，2019.

[9] 徐　莉，薛　锋. 基于GM(1，1)残差二次修正的铁路货运量预测[J]. 交通运输工程与信息学报，2019，17(2)：44-50.
XU Li，XUE Feng. Prediction of Rail Freight Volume of China Based on Secondary Residual Error Modification of GM (1，1) Model[J]. Journal of Transportation Engineering and Information，2019，17(2)：44-50.

[10] 史　樱. 现代信息技术与铁路货运的融合应用[J]. 铁路采购与物流，2019，14(2)：59-61.
SHI Ying. Fusion Application of Modern Information Technology and Railway Freight Transportation[J]. Railway Purchasing and Logistics，2019，14(2)：59-61.

[11] 郭颖霞. 铁路货运工作信息管理系统的设计及实现[D]. 北京：北京工业大学，2012.

[12] 王　静. 铁路货运信息数据挖掘系统的设计与实现[D]. 成都：电子科技大学，2014.

[13] 颜保凡，郭垂江，李夏苗. 铁路货运量的内部影响因素及其敏感度分析[J]. 铁道科学与工程学报，2018，15(5)：1341-1346.
YAN Baofan，GUO Chuijiang，LI Xiamiao. Internal Influence Factors of Railway Freight Volume and Its Sensitivity Analysis[J]. Journal of Railway Science and Engineering，2018，15(5)：1341-1346.

[14] 邓文青. 铁路现代物流中心货运量预测及布局规划[D]. 长春：吉林大学，2018.

[15] 刘晓彤. 基于遗传算法优化BP神经网络的铁路货运量预测及影响因素探究[D]. 北京：北京交通大学，2020.

[16] 王相平. 基于协整理论的铁路运输周转量与国民经济关系

实证分析[J]. 铁道运输与经济，2014，36(5)：59-63.
WANG Xiangping. Example Analysis on Relationship between Railway Transport Turnover and National Economy Based on Co-Integration Theory[J]. Railway Transport and Economy，2014，36(5)：59-63.

[17] 王　宇. 铁路货运区域联系及发展研究[J]. 铁路物流，2025，43(6)：1-5.
WANG Yu. Research on Regional Connection and Development of Railway Freight[J]. Railway Logistics，2025，43(6)：1-5.

[18] 晁攀迪. “公转铁”运输结构调整博弈机制与治理策略研究[D]. 北京：中国铁道科学研究院，2024.

[19] 张　岘. 铁路货运量预测及影响因素研究[D]. 北京：北京交通大学，2016.

[20] 王德胜. 河北省铁路货运影响因素分析及运量预测研究[D]. 石家庄：石家庄铁道大学，2024.

[21] 赵　威，王　伟. 偏最小二乘回归在响应面法可靠度分析中的应用[J]. 工程力学，2013(2)：272-277.
ZHAO Wei，WANG Wei. Application of Partial Least Squares Regression in Response Surface for Analysis of Structural Reliability[J]. Engineering Mechanics，2013(2)：272-277.

[22] 申　艳，刘次华. 偏最小二乘回归的应用效果分析[J]. 应用数学，2004，17(S1)：140-143.
SHEN Yan，LIU Cihua. Analysis of Applied Effect of Partial Least Squares Regression[J]. Mathematica Applicata，2004，17(S1)：140-143.

[23] 杨　洋，朱芳阳. 基于岭回归模型的广西物流需求预测研究[J]. 物流科技，2025，48(9)：4-7，37.
YANG Yang，ZHU Fangyang. Research on Logistics Demand Forecasting in Guangxi Based on Ridge Regression Model[J]. Logistics Sci-Tech，2025，48(9)：4-7，37.

[24] 朱青山. 科技创新对物流业效率的影响实证分析：基于岭回归模型[J]. 商业经济研究，2021(9)：91-93.

[25] 吕　佳. 梯度提升回归树算法研究及改进[D]. 上海：上海交通大学，2017.

[26] 李占山，刘兆赓. 基于XGBoost的特征选择算法[J]. 通信学报，2019，40(10)：101-108.
LI Zhanshan，LIU Zhaogeng. Feature Selection Algorithm Based on XGBoost[J]. Journal on Communications，2019，40(10)：101-108.

[27] 谢　勇，项　薇，季孟忠，等. 基于XGBoost和LightGBM算法预测住房月租金的应用分析[J]. 计算机应用与软件，2019，36(9)：151-155，191.
XIE Yong，XIANG Wei，JI Mengzhong，et al. An Application and Analysis of Forecast Housing Rental Based on XGBoost and LightGBM Algorithms[J]. Computer Applications and Software，2019，36(9)：151-155，191.

(责任编辑　魏　然)

(上接第179页)

ZHENG Yang. Quality Improving and Upgrading Strategy of China Railway Container Company's Domestic Multimodal Transport Trains[J]. Railway Freight Transport，2024(7)：18-23，30.

[16] QIU X，XU G Y. Consolidated Transport Service Pricing in Supply Hub in Industrial Park (SHIP) with Heterogeneous Manufacturers[J]. IEEE Transactions on Systems，Man，and Cybernetics：Systems，2023，53(3)：1370-1383.

[17] LU Y，XU X F，YIN C Z，et al. Network Optimization of Railway Cold Chain Logistics Based on Freight Subsidy[J]. Transportation Research Record：Journal of the Transportation Research Board，2021，2675(10)：590-603.

[18] DABIA S，LAI D，VIGO D. An Exact Algorithm for a Rich Vehicle Routing Problem with Private Fleet and Common Carrier[J]. Transportation Science，2019，53(4)：986-1000.

[19] 王能民，史玮璇，崔　巍，等. 考虑分散协作及数量折扣的双目标电动车辆路径优化[J]. 工程管理科技前沿，2024，43(4)：27-36.
WANG Nengmin，SHI Weixuan，CUI Wei，et al. Dual-Objective Electric Vehicle Routing Optimization Considering Decentralized Collaboration and Quantity Discount[J]. Frontiers of Science and Technology of Engineering Management，2024，43(4)：27-36.

[20] 王思睿，王　林，王志刚，等. 考虑数量折扣的多中心仓库联合采购-配送集成优化研究[J]. 管理工程学报，2022，36(2)：238-249.
WANG Sirui，WANG Lin，WANG Zhigang，et al. Optimizing Integrated Multi-Warehouse Joint Replenishment and Delivery Problem with Quantity Discount[J]. Journal of Industrial Engineering and Engineering Management，2022，36(2)：238-249.

(责任编辑　魏　然)